UNIVERSITÄT ZU LÜBECK
INSTITUTE OF MEDICAL INFORMATICS

From the Institute of Medical Informatics
of the University of Lübeck
Director: Prof. Dr. rer. nat. habil. Heinz Handels

Analyzing the Progression of Pathologies in Medical Images:

Deep Learning for Joint Image Registration and Unsupervised Pathology Segmentation of Longitudinal Medical Image Data

Dissertation
for
Fulfillment of Requirements for the Doctoral Degree
of the University of Lübeck

from the Department of Computer Sciences and Technical Engineering

Submitted by
Julia Andresen
from Münster

Lübeck, 2025

First referee: Prof. Dr. rer. nat. habil. Heinz Handels
Second referee: Prof. Dr. rer. nat. Jan Lellmann

Date of oral examination: 20 November 2025

Approved for printing, Lübeck, 24 November 2025

Abstract

Medical image analysis is a key component of modern healthcare, required not only for diagnosis, but also for treatment planning and disease monitoring. The number of medical images acquired every day is constantly increasing and with it the need for automated tools to process, segment and interpret these images efficiently and reliably. Over the past decade, deep learning-based approaches, especially convolutional neural networks, have revolutionized the field providing unprecedented performances for almost all medical image analysis tasks, including semantic segmentation and image registration. However, the training of deep neural networks needs vast amounts of data, whereas most annotated medical datasets are small. The manual delineation of anatomical and pathological structures needs expert knowledge, and is both time-consuming and error-prone. These problems are even more severe in the analysis of disease progression, where not just one image but several have to be analyzed together. Furthermore, pathologies exhibit higher variability than anatomical structures and occupy comparatively small image areas, further increasing the data demands for training.

This dissertation aims to develop deep learning-based algorithms for the automatic analysis of medical time series image data, focusing on pathological progression over time, such as retinal fluid in optical coherence tomography and brain lesions in magnetic resonance imaging. The main goal is to segment pathologies across all time points in order to monitor disease progression. Expert segmentations are typically unavailable for extensive time series data, requiring weakly supervised or fully unsupervised methods. Therefore, longitudinal registration of medical images is investigated as a tool for pathology tracking and unsupervised segmentation.

To achieve the goals described, the present work follows three complementary research directions. First, unsupervised clustering is used to segment individual images. Second, registration-based approaches are developed for the joint analysis of longitudinal data with simultaneous segmentation of non-correspondences that reflect evolving or disappearing pathologies. Third, registration approaches inspired by metamorphosis models are used to model the formation of new pathologies. To improve the plausibility of the resulting deformations, these models are designed to separate displacements of anatomical structures from volumetric changes of the pathologies.

The methods presented in this thesis enable the unsupervised segmentation of pathological structures, without relying on manually generated pathology segmentations. By leveraging weak supervision through anatomical labels and exploiting temporal information in longitudinal data, the proposed approaches can identify disease-related changes in an unsupervised manner. Overall, this work provides novel, annotation-efficient strategies for the automated analysis of medical image time series data, with the potential to support clinical workflows in the assessment of disease progression.

Kurzfassung

Die medizinische Bildverarbeitung ist eine zentrale Komponente der modernen Gesundheitsversorgung, die nicht nur zur Diagnosestellung, sondern auch zur Behandlungsplanung und Krankheitsüberwachung benötigt wird. Die täglich aufgenommene Menge medizinischer Bilder steigt stetig und damit wird auch der Bedarf an Algorithmen größer, die diese Bilder automatisiert, effizient und zuverlässig segmentieren und interpretieren können. In den letzten zehn Jahren hat das Deep Learning, insbesondere in Form von Faltungsnetzwerken, die medizinische Bildverarbeitung revolutioniert, indem es bis dahin unerreichte Leistungen in Aufgaben wie der semantischen Segmentierung und der Bildregistrierung erzielt hat. Das Training tiefer neuronaler Netze erfordert jedoch sehr große Datenmengen, während medizinische Datensätze häufig nur in begrenztem Umfang und mit wenigen Annotationen verfügbar sind. Anatomische oder pathologische Strukturen manuell zu segmentieren, ist eine zeitaufwändige und fehleranfällige Aufgabe, die zudem Expertenwissen erfordert. Diese Probleme sind bei der Analyse von Krankheitsverläufen noch verstärkt, da nicht nur ein Bild, sondern mehrere gemeinsam ausgewertet werden müssen. Zudem zeigen Pathologien eine größere Variabilität als anatomische Strukturen und nehmen dabei gleichzeitig meist nur kleine Bildbereiche ein, was den Datenbedarf für das Netzwerktraining zusätzlich erhöht.

In der vorliegenden Dissertation werden Deep-Learning-Algorithmen für die automatisierte Analyse medizinscher Bildzeitreihen entwickelt, wobei der Fokus auf der Untersuchung von Pathologieveränderungen über die Zeit liegt. Dabei werden insbesondere retinale Fluide in Aufnahmen der optischen Kohärenztomografie untersucht sowie Gehirnläsionen in Magnetresonanztomografiebildern. Das Hauptziel der Arbeit ist die Segmentierung von Pathologien für alle Zeitpunkte, um Krankheitsverläufe zu überwachen. Für umfangreiche Zeitreihendaten liegen typischerweise keine manuellen Segmentierungen von medizinischen Experten vor, sodass Methoden benötigt werden, die schwach überwacht oder vollständig unüberwacht trainiert werden können. Ein zentraler Fokus der Dissertation liegt daher auf der zeitlichen Registrierung medizinischer Bilder, wobei die Registrierung als Werkzeug für die Pathologieverfolgung und unüberwachte Segmentierung dient.

Um die beschriebenen Ziele zu erreichen, nutzt diese Arbeit drei sich ergänzende Forschungsansätze. Erstens wird unüberwachtes Clustering verwendet, um einzelne Bilder zu segmentieren. Zweitens werden Registrierungsansätze für die gemeinsame Auswertung von Bildern verschiedener Zeitpunkte entwickelt. Diese Methoden sind in der Lage, gleichzeitig mit der Registrierung nicht-korrespondierende Bereiche, welche neu auftretende oder verschwindende Pathologien widerspiegeln, zu segmentieren. Drittens werden von Metamorphosemodellen inspirierte Registrierungsalgorithmen eingesetzt, um die Entstehung neuer Pathologien zu modellieren. Zur Verbesserung der Realitätsnähe der resultierenden Deformationsfelder werden die Modelle so konzipiert,

dass sie Volumenveränderungen der Pathologien ermöglichen, während die umgebende Anatomie durch die Pathologien verschoben wird.

Die in dieser Arbeit vorgestellten Methoden ermöglichen die unüberwachte Segmentierung pathologischer Strukturen, ohne auf manuell erstellte Pathologiesegmentierungen angewiesen zu sein. Durch die Nutzung einer schwachen Überwachung mittels anatomischer Annotationen sowie die Auswertung zeitlicher Informationen in longitudinalen Bilddaten können die vorgeschlagenen Ansätze krankheitsbedingte Veränderungen auf unüberwachte Weise identifizieren. Insgesamt liefert diese Arbeit neuartige, annotationseffiziente Strategien für die automatisierte Analyse medizinischer Bildzeitreihendaten, die das Potenzial haben, klinische Arbeitsabläufe bei der Beurteilung von Krankheitsverläufen zu unterstützen.

Contents

Nomenclature

AMD	Age-related macular degeneration
ANCR-Net	Appearance adaptation, non-correspondence segmentation and image registration network
ANN	Artificial neural network
ANTs SyN	Advanced normalization tools – Symmetric image normalization
ASSD	Average symmetric surface distance
BCE	Binary cross entropy
BM	Bruch's membrane
BN	Batch normalization
BraTS	Brain tumor segmentation challenge
CNN	Convolutional neural network
CNS	Central nervous system
CSCR	Chronic central serous chorioretinopathy
CT	Computed tomography
DL	Deep learning
DRAMMS	Deformable registration via attribute matching and mutual-saliency weighting
DSC	Dice similarity coefficient
ELM	External limiting membrane
EZ	Ellipsoid zone
f-AnoGAN	Fast anomaly detection generative adversarial network
FC	Fully-connected
FCNN	Fully-connected neural network
FluidRegNet	Fluid registration network (FRN)
FN	False negative
FP	False positive
FPR	False positive rate
FRN	Fluid registration network (FluidRegNet)
FRN-Ano	Anomaly detection with FluidRegNet
GAN	Generative adversarial network
Grad-CAM	Gradient-weighted class activation mapping
HD	Hausdorff distance
IDIR	Implicit deformable image registration
IIC	Invariant information clustering

ILM	Inner limiting membrane
ILSVRC	ImageNet large-scale visual recognition challenge
INR	Implicit neural representation
IRF	Intraretinal fluid
ISLES	Ischemic stroke lesion segmentation dataset
IZ	Interdigitation zone
LNCC	Local normalized cross-correlation
LP	Locality preservation
LP loss	Locality-preserving loss
LPBA40	LONI probabilistic brain atlas dataset
LPIPS	Learned perceptual image patch similarity
MAE	Mean absolute error
MI	Mutual information
MLP	Multi-layer perceptron
MOOD	Medical out-of-distribution analysis challenge
MR	Magnetic resonance
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
MSE	Mean squared error
MSSEG-2	Multiple sclerosis new lesions segmentation challenge
NCC	Normalized cross correlation
NCR-Net	Non-correspondence segmentation and image registration network
NSA	Natural synthetic anomalies
OCT	Optical coherence tomography
OCTA	Optical coherence tomography angiography
ORA	Outer retinal atrophy
PED	Pigment epithelial detachment
PEZ	Photoreceptor ellipsoid zone
ProbReg	Probabilistic registration
PSNR	Peak signal-to-noise ratio
ReLU	Rectified linear unit
RETOUCH	Retinal OCT fluid challenge
R-Net	Registration network
RNN	Recurrent neural network
RORA	Retinal pigment epithelium and outer retinal atrophy
RPE	Retinal pigment epithelium
SINR	Spline-enhanced implicit neural representation
SIREN	Sinusoidal representation network
SLO	Scanning laser ophthalmoscopy
SRF	Subretinal fluid

SSIM	Structural similarity index measure
SSL	Self-supervised learning
SSM	Statistical shape model
STAPLE	Simultaneous truth and performance level estimation
STN	Spatial transformer network
TN	True negative
TP	True positive
TPR	True positive rate
UKSH	Universitätsklinikum Schleswig-Holstein
VarReg	Variational registration toolbox
ViT	Vision transformer
VXM	VoxelMorph
WIRE	Wavelet implicit neural representation

Chapter 1

Introduction

1.1 Motivation

Medical imaging visualizes the interior of the body and, thus, enables the diagnosis, monitoring, treatment planning and follow-up of diseases. Nowadays, imaging procedures are widely used in almost all medical disciplines and the number of images taken every day is constantly increasing. In Germany alone, more than 13 million computed tomography (CT) and also more than 13 million magnetic resonance (MR) images were taken in 2022, compared to around 11 million CT scans and 10 million MR scans in 2013 [Eurostat, 2024]. As the number of images increases, so does the need to examine and evaluate them. Semantic segmentation describes the process of partitioning an image into several regions, that correspond to anatomical or pathological structures, by assigning a class label to each pixel in the image. Segmentation masks are the basis for many subsequent medical image analysis and interpretation tasks, such as size measurements of organs or pathologies, 3D rendering, implant design and surgery planning [Handels, 2000; Semmlow, 2008]. However, manual segmentation of medical image data is labor-intensive and error-prone, especially for 3D volumes, where the structures of interest need to be delineated in each 2D slice separately. In addition, for many medical segmentation tasks, expert knowledge is required, which makes the manual annotation process extremely expensive and ties up physicians' time. These problems are even more severe when analyzing time series data, where not only one but several images need to be analyzed jointly to assess disease progression based on differences between the time points.

During the last decade, deep learning-based (DL-based) approaches, especially convolutional neural networks (CNNs), have set new benchmarks for almost all medical image processing tasks, including semantic segmentation. Since its development in 2015, the U-Net [Ronneberger et al., 2015] has been the predominant architecture, whose extensions won well-known medical image processing challenges BraTS, e.g., [Isensee et al., 2021b; Ferreira et al., 2024], MOOD [Baugh et al., 2023] and many more. Compared to classical image processing algorithms, CNNs allow the beforehand crucial task of feature engineering to be skipped, since they automatically learn to extract relevant image features during a data-driven training process. CNNs can, thus, be applied more flexibly, being directly adaptable to a specific use case by retraining with a custom dataset. As they are also very fast and deliver reproducible results, automated DL solutions for medical image segmentation have already substantially improved the clinical workflow and will continue to do so in the future.

However, one of the major obstacles for the training of CNNs in the medical field is data scarcity. CNN training requires large amounts of data, which is particularly important for medical images to cover the wide variety of patient anatomy and the even greater differences in pathologies. Despite the large number of images taken every day, most medical image datasets are small due to privacy restrictions and the rarity of certain diseases. In addition, the training of supervised segmentation networks does not only require the image data but also the corresponding ground truth annotations, which, as described above, are difficult to get. Therefore, annotated medical image data is even more seldom, oftentimes preventing the application of DL-based methods. Many approaches have been proposed to ease the data problem, data augmentation being the most popular choice. Here, the dataset is artificially enlarged by applying intensity or spatial transformations to the given images. Another approach to expand or even replace datasets is the generation of synthetic images, e.g. through generative adversarial networks or diffusion models [Goodfellow et al., 2020; Ho et al., 2020]. However, these methods themselves require large, yet annotation-free, datasets.

A different approach frequently pursued is not to enlarge the training data, but to use transfer learning. Here, CNNs are first pre-trained on separate, large-scale datasets, so that they have a good starting point for adapting to small medical datasets, and then fine-tuned for the respective tasks. Thereby, either the entire network is fine-tuned or only the last layers of the network are replaced and trained. Self-supervised learning (SSL) uses labels that are artificially produced from the data to pre-train the networks rather than manual labels. Among others, SSL tasks may include predicting transformations applied to the input images, superresolution, inpainting, denoising or, in the case of time series data or videos, frame ordering and forecasting. Contrastive SSL, similar to clustering methods, aims to align similar input pairs while pushing apart dissimilar inputs. Similar or positive pairs are generated by randomly augmenting the same input twice, while dissimilar or negative pairs are built from different images. The network is then trained to produce the same features for positive and dissimilar features for negative pairs [Chen et al., 2020]. Such pre-training methods can greatly improve the supervised fine-tuning of CNNs on small datasets, and have the potential to be used as stand-alone unsupervised techniques for medical image analysis.

Anomaly detection methods tackle the data scarcity problem from a different perspective. Rather than relying on datasets featuring the pathologies or diseases of interest, the aim is to detect everything that deviates from the "normal" appearance of images. For this purpose, networks are trained on healthy images and applied to pathological data during inference. In autoencoders, deviations from the healthy distribution can then be found based on the reconstruction error or by analyzing the feature distribution in the latent space [Zhou and Paffenroth, 2017; Uzunova et al., 2019]. Self-supervised anomaly detection, in turn, uses artificially introduced anomalies reflecting the appearance of real pathologies to train anomaly segmentation networks in a supervised manner [Baugh et al., 2023]. Such methods allow increasingly precise delineations of pathologies, but need a careful selection of pretext tasks, i.e. a sophisticated creation of artificial anomalies, to ensure that the vast variety of real pathologies is covered. Furthermore,

methods relying on no or on synthetic labels usually still lag behind solutions that use manual annotations.

While semantic segmentation allows to interpret individual images, the pixel-precise comparison between images from different time points, patients or modalities is only possible if corresponding structures in the images are exactly aligned. To achieve such an alignment is the goal of image registration, which describes the process of transforming images into a common coordinate system. Typically, image registration is done in an unsupervised manner by using image similarity and deformation regularity as optimization objective for either iterative algorithms [Beg et al., 2005; Avants et al., 2008; Klein et al., 2009] or DL approaches [Jaderberg et al., 2015; de Vos et al., 2019; Balakrishnan et al., 2019; Mok and Chung, 2020]. Here, the pathologies often observed in medical images pose difficulties since they differ greatly in their size, appearance and location among patients. This means that for the pixels affected by the pathology in one image, no corresponding pixels can be found in the other image. The intensity differences introduced by such non-correspondences are disparate from the differences that occur in other areas of the images. In many registration frameworks, this leads to unrealistic deformation fields that overcompensate for the intensity differences, making them unsuitable for subsequent volume change or displacement analyses [Chen et al., 2015]. Similar problems arise when analyzing time series image data: New and disappearing pathologies or pathologies that have changed in size lead to non-correspondences between time points, as they are either not visible for one time point or occlude and displace the surrounding anatomy.

One approach to achieve more realistic deformation fields in the presence of pathologies is to use segmentation masks to regularize the deformation in non-correspondent areas. This, however, requires that the segmentations are available beforehand. In addition, the segmentations of individual images for time series data do not directly contain information on which pathologies do not correspond to the other time points. Due to the as yet unestablished spatial alignment, it is unclear which pathology has newly appeared, has changed in size, or has disappeared. Finally, even with given segmentations, structures that are only visible at one time point cannot be compensated for by deformation-based approaches alone. Metamorphosis approaches, therefore, estimate both shape and appearance differences to match the images [Niethammer et al., 2011; François et al., 2022; Tian et al., 2023]. Oftentimes, these methods suffer from an entanglement of form and appearance, i.e. differences in form are compensated for in appearance and vice versa, which again hinders the plausibility of the generated deformation fields.

Given the high necessity for automated tools to analyze medical images and the developments and problems described above, the goal of this work is to develop algorithms for the interpretation of medical time series image data. More specifically, the goals are to segment pathologies and analyze disease progression over time, with the main focus on retinal and brain diseases on optical coherence tomography (OCT) and MR images. Owing to the success of DL methods, the algorithms presented in this thesis use CNNs and multi-layer perceptrons. Since medical images are available in larger quantities than their corresponding labels, a further methodological focus of the work is to reduce or

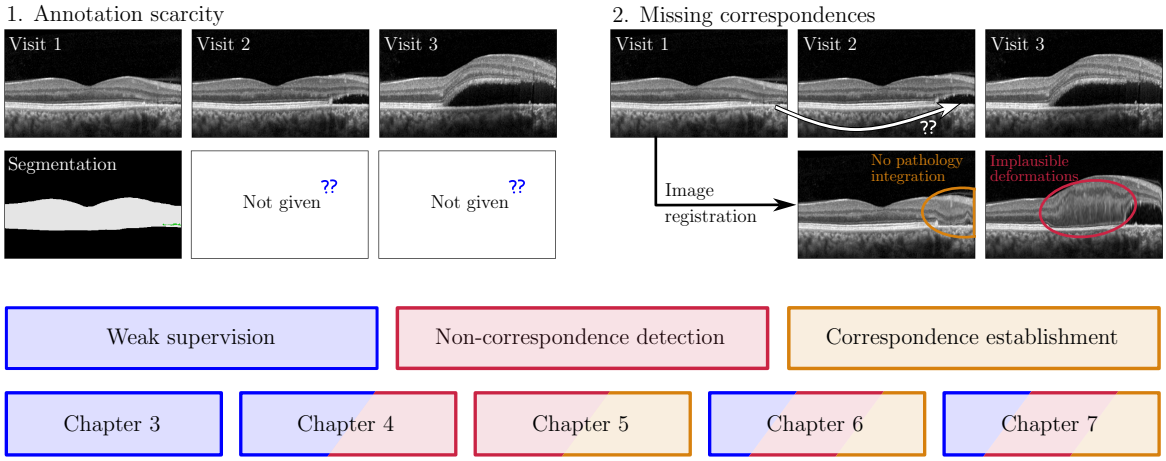


Figure 1.1: Medical time series image analysis faces common challenges: manual annotations are often missing, and analyzing disease progression requires comparing multiple time points despite evolving pathologies causing non-correspondent image areas. In image registration, these non-correspondences result in unrealistic deformation fields (red ellipse) that, additionally, cannot introduce evolving pathologies (marked in orange). For this example, registration was performed using ANTs SyN [Avants et al., 2008], mapping visit 1 and 2 to their subsequent follow-up. Below, the methodological chapters of this thesis are listed and color-coded by the approaches used to address data scarcity and registration challenges. Weakly, partially, and completely unsupervised training methods are explored. The proposed image registration methods improve deformation plausibility by detecting and masking non-correspondent areas and explicitly incorporating new pathologies into the registration process.

even eliminate the need for manual segmentations by exploiting the information hidden in the images. Time series data bares the potential to identify disease progression as variations between images from different visits. Many of the considered methods, therefore, use image registration to enable the comparison and detection of changes between images. However, image registration suffers from the evolving diseases depicted in the images, as they may contain new or disappearing pathologies. Hence, another goal of this thesis is the integration of such evolving processes into image registration networks to overcome the problem of unrealistic deformations caused by non-correspondences. The objectives of this work can be summarized as follows:

Develop algorithms that

- analyze medical time series data
- minimize the manual annotation need
- are pathology-aware.

More specifically, the goal is to develop deep learning-based algorithms for the automatized segmentation of medical time series image data, while minimizing the need for man-

ual annotations to train such algorithms. A particular focus of this work lies in the implementation of image registration solutions that enable unsupervised non-correspondence detection and result in realistic, i.e. pathology-induced, deformations.

These goals are pursued in three ways throughout this work: First, images are analyzed individually using a clustering approach. Second, image registration with simultaneous non-correspondence segmentation is used to jointly analyze images from subsequent time points. The third class of approaches is also based on longitudinal image registration, whereby the focus here is on metamorphosis-inspired approaches to integrate new pathologies in the deformation process. Fig. 1.1 provides a brief overview of the problems addressed in this thesis and also indicates the chapters in which they are dealt with. A more detailed description of the structure of this thesis and the proposed methodologies is given in the following section.

1.2 Organization and Contributions

This work tackles the DL-based analysis of longitudinal medical image data, with applications focusing on the analysis of OCT images of the skin and eye and MR images of the brain. Thereby, the overall goal to reduce or even completely eliminate the need for difficult-to-get manual annotations is pursued in three ways. First, images are analyzed individually using a contrastive-learning approach to separate healthy and pathological image regions. This approach is presented in Chapter 3 and trained partially supervised to segment retinal atrophy and fully unsupervised to segment wounds in OCT images. Second, images from subsequent time points are analyzed jointly using image registration approaches that detect non-correspondences between images automatically. Deformation fields and non-correspondence segmentations are generated within one network to regularize the registration task, leading to more plausible deformations, as shown in Chapter 4. In Chapter 5, this approach is extended for the detection, segmentation and modeling of new multiple sclerosis lesions. Third, metamorphic registration approaches for the decoupling of common and discriminative image features between time points are presented. The separation of the (pathological) changes from the constant anatomy helps to create more realistic, pathology-aware deformations. In Chapter 6, a convolutional neural network is used for the longitudinal registration of OCT images with new pathologies, whereas in Chapter 7 implicit neural representations are used for the same purpose.

In more detail, this thesis is structured as follows:

- **Chapter 2** gives the theoretical background necessary for a deeper understanding of the methodologies described in Chapters 3 to 7. First, the basic concepts and a brief overview of the existing methods for automatic image segmentation and registration are given. Second, DL is introduced by explaining the functionality of artificial and convolutional neural networks. Finally, the imaging modalities and datasets analyzed throughout this thesis are introduced.

- **Chapter 3** uses invariant information clustering (IIC) in a contrastive learning-like setup to segment pathologies in OCT images. IIC encourages network predictions for differently augmented versions of the same image patch to be the same. Here, it is proposed to combine IIC with a locality-preserving loss that favors spatially close patches to be assigned to the same class, whereas distant patches are classified independently of each other. This new combination of loss functions allows the unsupervised segmentation of wounds and the partially supervised segmentation of atrophic eye regions. For both applications, pathology extension in the en-face plane is the major biomarker of interest, which is why a sophisticated network architecture is used that classifies single columns (A-scans) of the 3D image volumes. The proposed method for wound segmentation was published in:

  [Andresen et al., 2022b] J. Andresen, T. Kepp, M. Wang-Evers, J. Ehrhardt, D. Manstein and H. Handels: "Unsupervised Segmentation of Wounds in Optical Coherence Tomography Images Using Invariant Information Clustering" in *Bildverarbeitung für die Medizin (BVM) 2022* (editors: K. Maier-Hein, T. Deserno, H. Handels, A. Maier, C. Palm, T. Tolxdorff), Heidelberg, Germany, 2022

- **Chapter 4** presents the joint non-correspondence segmentation and image registration network (NCR-Net), at the time one of the first and few CNNs capable of registering pathological images while simultaneously segmenting areas of missing correspondences between the images. Network training uses a combination of segmentation sparsity, deformation field regularization and an image distance measure that is masked with the segmentation output of the network. This masking encourages the segmentation of voxels with a large appearance difference, while ensuring smooth deformations inside these regions, thus leading to more plausible transformations. No pathology annotations are required for training and yet the resulting segmentations can be used for the unsupervised delineation of pathologies, as demonstrated for retinal OCT and brain MR images and published in:

  [Andresen et al., 2022a] J. Andresen, T. Kepp, J. Ehrhardt, C. von der Burchard, J. Roider and H. Handels: "Deep Learning-Based Simultaneous Registration and Unsupervised Non-Correspondence Segmentation of Medical Images with Pathologies" *International Journal of Computer Assisted Radiology and Surgery* 17, 2022

- **Chapter 5** proposes application-driven extensions of NCR-Net to enable the detection and segmentation of new multiple sclerosis (MS) lesions. Most importantly, ANCR-Net is introduced, where the "A" stands for appearance adaptation. Instead of only segmenting the new lesions, these are modeled with appearance offsets, inspired by metamorphosis approaches. Thus, the registration, segmentation and modeling of new lesions is enabled within a single CNN. Furthermore, the potential of the modeled lesions to reveal different lesion areas is shown. Unsupervised approaches proved unsuitable for the difficult task of delineating the very small and often subtle MS lesions, but for comparison to ANCR-Net, NCR-Net is

trained again, and its outputs serve as prior knowledge to improve the performance of a segmentation U-Net. These methods were originally proposed in:

[Andresen et al., 2022c] J. Andresen, H. Uzunova, J. Ehrhardt, T. Kepp and H. Handels: "Image Registration and Appearance Adaptation in Non-Correspondent Image Regions for New MS Lesions Detection" *Frontiers in Neuroscience* 16, 2022

[Andresen et al., 2021] J. Andresen, H. Uzunova, J. Ehrhardt and H. Handels: "New Multiple Sclerosis Lesion Detection with Convolutional Neural Registration Networks" in *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure. MICCAI 2021 - 24th International Conference on Medical Image Computing and Computer Assisted Intervention* (editors: O. Commowick, F. Cervenansky, F. Cotton and M. Dojat), Strasbourg, France, 2021

- **Chapter 6** focuses on the registration of OCT images with dynamically changing pathologies. A fluid registration network (FluidRegNet) is proposed that models the onset of fluid formation by estimating a sparse appearance "seed" which is deformed to match the pathology in the reference image. In longitudinal registration, this allows to not only align existing fluids but to integrate new pathologies in the registration process. The resulting deformation fields concentrate volume changes in fluid regions and can, thus, be used to visualize areas of disease progression. Appearance offsets and deformation fields are used to segment new fluids in an unsupervised manner. In the second part of the chapter, FluidRegNet is trained to register healthy onto pathological images, allowing to model and segment all pathologies observed in the images, i.e. to use FluidRegNet for anomaly detection. The first part of the chapter is based on:

  [Andresen et al., 2024] J. Andresen, J. Ehrhardt, C. von der Burchard, A. Tatli, J. Roider, H. Handels and H. Uzunova: "FluidRegNet: Longitudinal Registration of Retinal OCT Images with New Pathological Fluids" in *Proceedings of The 7th International Conference on Medical Imaging with Deep Learning (MIDL)* volume 250 of *Proceedings of Machine Learning Research* (editors: N. Burgos, C. Petitjean, M. Vakalopoulou, S. Christodoulidis, P. Coupe, H. Delingette, C. Lartizien and D. Mateus), Paris, France, 2024

- **Chapter 7** presents methods based on implicit neural representations (INRs) for two different applications. First, a generalizable INR is proposed for the improved shape interpolation of sparsely sampled retinal OCT images. Additional context information from another imaging modality is integrated to enable the reconstruction and correct localization of blood vessels. Second, INRs are used for the registration of pathological images. Two INRs are used for each registration problem, which allows to not only learn the deformation field, but also to integrate developing pathologies into the registration process. Similar to FluidRegNet, the proposed INR delivers deformation fields with larger volume changes in pathologic areas than in the surrounding healthy tissue, and the generated outputs are

used for unsupervised fluid segmentation. The interpolation method was originally published in:

[Kepp et al., 2025] T. Kepp, J. Andresen, and H. Handels: "Bridging Gaps in Retinal Imaging - Fusing OCT and SLO Information with Implicit Neural Representations for Improved Interpolation and Segmentation" in *Bildverarbeitung für die Medizin (BVM) 2025* (editors: C. Palm, K. Breininger, T. Deserno, H. Handels, A. Maier, K.H. Maier-Hein, and T.M. Tolxdorff), Regensburg, Germany, 2025

Each of the methodological Chapters 3 to 7 first motivates the presented approaches and gives an overview of the relevant literature before describing the method itself, presenting and discussing the results achieved. The concluding Chapter 8 summarizes the results of the methodological chapters and looks at them from an overarching perspective in order to clarify whether the intended objectives were achieved and what further research questions arise.

For the publications [Andresen et al., 2021, 2022a,b,c, 2024] listed above, the methodological development, implementation and evaluation were carried out by the author of this thesis, who also took the lead in writing the manuscripts. Methodological development and the writing of the manuscript of the publication [Kepp et al., 2025] were done in equal parts by T. Kepp and the author of this thesis, who share the first authorship. The data acquisition and annotation of the images used in the publications were carried out by M. Wang-Evers, C. von der Burchard, A. Scharf and A. Tatli. J. Ehrhardt contributed comparison algorithms for publications [Andresen et al., 2022a], [Andresen et al., 2024] and [Kepp et al., 2025]. H. Uzunova also contributed a comparison algorithm to [Andresen et al., 2024]. H. Handels and H. Uzunova supervised the work. All co-authors provided critical feedback and helped shape the research, analyses and manuscripts.

Chapter 2

Background

The purpose of this chapter is to provide the theoretical foundations for the following methodological chapters. Sec. 2.1 introduces the two main applications of the algorithms presented in this thesis: semantic segmentation and spatial registration of medical images. All considered algorithms are based on deep learning, the theoretical background of which is described in Sec. 2.2. Subsequently, Sec. 2.3 gives a description of the metrics used in this thesis to evaluate the proposed algorithms. Finally, in Sec. 2.4, the datasets used throughout this thesis are briefly introduced.

2.1 Medical Image Analysis Tasks

The algorithms presented in this thesis serve two main medical image analysis tasks: semantic segmentation and spatial registration, both used to monitor diseases over time. The basic concepts of these two tasks will be presented in the following.

2.1.1 Semantic Segmentation

Semantic segmentation describes the process of assigning each pixel or voxel in an image a class label to section the image into regions (segments) that correspond with the displayed content. The most prominent applications of semantic segmentation in the medical field are the quantification of pathologies such as tumors or fluids, organs or other anatomical structures and their temporal monitoring. Furthermore, segmentation is an important pre-processing step for other image analysis tasks like image registration, three-dimensional rendering of the considered structures and computer-assisted surgery planning. Formally, the image $I : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$ is divided into segments by the mapping $S : \Omega \rightarrow \mathcal{L}$ with $\mathcal{L}$ being the set of possible labels $l \in \mathcal{L} = \{0, 1, ..., L-1\}$. The spatial dimension d typically is $d = 2$ or $d = 3$ for 2D or 3D medical images.

In clinical practice, semantic segmentations most often are generated manually by physicians, which is a time-consuming and error-prone process. Automated segmentation methods can, therefore, significantly improve the clinical routine by saving expensive off-patient expert time and harmonizing segmentation results. Existing algorithms for automatic image segmentation can be divided into point-, region-, edge-, model-, clustering- and learning-based approaches [Handels, 2000; Ramesh et al., 2021]. The simplest, point-based approaches rely on pixel intensities only to achieve a pixel-wise classification. One or several thresholds are used to assign each pixel a class label. Ob-

viously, these approaches are highly susceptible to noise and intensity variations in the images. Region-based approaches, such as the region growing algorithm [Adams and Bischof, 1994], consider local image regions rather than single pixels. Starting from initial seed points, neighboring pixels are iteratively assigned to different segments based on a pre-defined homogeneity criterion. Typically, methods based on region growing are only semi-automatic, as they require manual seed point selection. Despite being less prone to noise than point-based segmentation methods, region-based methods still suffer from low contrast, local intensity variations and fuzzy edges.

Methods based on clustering follow a similar idea, grouping homogeneous pixels into the same segment and separating dissimilar pixels into different segments. In contrast to region growing, clustering methods are fully unsupervised. The most popular clustering algorithm is k-means clustering [Lloyd, 1982]. Starting from k initial cluster centers, each voxel is assigned the cluster whose cluster center has the minimum distance to this voxel. The distance typically depends on voxel intensities, texture, location or a combination of these. In the next step, the cluster centers are recalculated via averaging of all voxels in the cluster. These steps are repeated until convergence is reached, or a stopping criterion is met.

Edge-based methods seek to find strong local alterations of intensities or other image features. A popular half-automatic edge-based segmentation method is the live wire segmentation [Mortensen et al., 1992; Barrett and Mortensen, 1996, 1997]. The user marks successive contour points, and, based on image edge information, the most cost-effective path between these points is found using graph-theoretic methods. Additionally, the cost function can also take into account external information, such as the smoothness of the contour and the consistency of the new contour section with the partial contour already found. This enables segmentation in the case of blurred edges or several closely spaced edges.

In contrast to point- and region-based methods, model-based approaches can consider the entire image to create meaningful segmentations, even in the case of poor-quality images or partial occlusions. Statistical shape models (SSMs) are a prominent model-based approach for semantic segmentation [Cootes et al., 1992; Cootes and Taylor, 1992]. Starting from a collective of segmented images, the mean shape of the considered structure is extracted and deformed in an iterative optimization process to fit new data. SSMs are only applicable to structures that exhibit a typical shape throughout different patients, e.g. organs, bones or other anatomical structures. Pathologies, however, are characterized by large variations in their size and shape so that they are difficult or impossible to represent using SSMs. Similar problems arise for another popular class of segmentation algorithms: Atlas-based segmentation transfers the given segmentation of one patient to a similar image of another patient. Pathologies are normally not transferable from one patient to another, limiting the applicability of atlas-based segmentation to anatomical features.

The most recent and currently most widely used approaches for semantic segmentation are based on DL. Starting with the U-Net [Ronneberger et al., 2015], which surpassed many standards of the time, a whole range of segmentation methods based on convolu-

tional network architectures (cf. Sec. 2.2.4) has emerged. U-Net-like architectures are characterized by the fact that they combine low- and high-level features of the images to generate segmentations. These properties are learned automatically in a training phase, during which a set of images, often annotated, is used to iteratively adjust the network parameters.

2.1.2 Image Registration

The goal of image registration is to geometrically align images from the same or at least similar scenes. That is, given two images taken, for example, at different times, perspectives or from different devices, the goal is to find a plausible transformation, such that a deformed version of the first image is similar to the second one [Modersitzki, 2009]. Registered images can be analyzed simultaneously, as corresponding structures also match spatially, allowing direct comparison between images. The applications of image registration in the medical field are manifold, including treatment planning, disease monitoring, fusion of different modalities and motion correction. Recent surveys of medical image registration approaches can be found in [Abbasi et al., 2022; Pan and Chen, 2023; Chen et al., 2025].

Given a fixed reference image $F : \Omega \subset \mathbb{R}^d \to \mathbb{R}$ and a moving image $M : \Omega \to \mathbb{R}$, image registration involves finding a transformation $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ that minimizes the image distance $\mathrm{D}[F(\boldsymbol{x}), M(\varphi(\boldsymbol{x}))]$ for the fixed and warped moving images. The transformed image is denoted by $M[\varphi] = M[\varphi](\boldsymbol{x}) = M(\varphi(\boldsymbol{x})) = (M \circ \varphi)(\boldsymbol{x})$. Finding the optimal φ is an ill-posed problem that can only be solved with additional constraints. Smooth transformations are often required in the medical field because the anatomy of the body cannot deform freely. This constraint on the deformation field can be enforced by joint optimization of the image distance and a regularizer R on φ:

$$\mathrm{D}\left[F(\boldsymbol{x}), M(\varphi(\boldsymbol{x}))\right] + \alpha \mathrm{R}\left[\varphi(\boldsymbol{x})\right] \xrightarrow{\varphi} \min . \tag{2.1}$$

The weighting parameter α serves to balance between image similarity and smoothness of the deformation field.

Image registration algorithms can be classified by (1) the spatial transformation model used (rigid, affine, deformable), (2) the measure of image similarity used (feature- or intensity-based), and (3) the optimization techniques used to solve Eq. (2.1). More recently, the widespread use of DL for solving image processing tasks has led to a further distinction between traditional, i.e. non-learning-based, machine learning-based and deep learning-based registration methods [Cao et al., 2020]. While traditional methods solve registration problems individually, learning-based methods use an entire dataset of registration problems to move the optimization to an offline training phase. This enables faster run times of the learning-based methods once training is complete. Regardless of the method used, the main idea is to use a gradient descent procedure to iteratively fit a transformation model so that Eq. (2.1) is minimized.

A typical medical image registration pipeline includes two registration steps. First, a linear registration is performed to globally align the images. This global transformation

may include rotation, translation, scaling and shearing. Second, a deformable registration is performed to establish voxel-to-voxel correspondences between moving and fixed images [Cao et al., 2020]. Many image registration approaches assume correspondence between moving and fixed images for the entire image domain. That is, it is assumed that for every structure in the fixed image, there is a matching structure in the moving image. Problems with this assumption arise, for example, in the registration of infant brains, which look very different at different points in time, or in the registration of healthy and pathological images. Pathologies are not present in healthy images, so that no correspondence can be established for the diseased tissue. Various approaches to address missing correspondences include down-weighing the image distance in non-correspondent regions based on previously created manual segmentations [Brett et al., 2001; Schaffland and Schöning, 2024], or segmentations estimated during registration [Ou et al., 2011; Chen et al., 2015; Krüger et al., 2019], and metamorphoses models that jointly estimate deformation fields and appearance differences to make moving and fixed images look the same [Niethammer et al., 2011; François et al., 2022; Uzunova et al., 2023].

2.2 Deep Learning

Deep learning (DL), a subfield of machine learning, includes algorithms that can learn from data using artificial neural networks (ANNs). The functionality of ANNs is inspired by the information processing and storage mechanisms in the biological brain. Similar to the visual cortex, information is processed and passed on by small units, the neurons, which are arranged in several layers one behind the other. Connected via so-called weights, these neurons form a network-like structure that processes input data, extracts information and finally draws conclusions from the data. DL is used wherever there are huge amounts of data from which patterns or trends can be derived. Popular applications are speech and face recognition, autonomous driving and personalized advertisements. Many medical applications of DL are based on convolutional neural networks (CNNs). These are especially suitable for image analysis tasks like image segmentation or classification. Other medical application fields are electronic health records, genomics and drug development [Yang et al., 2021].

The weighting parameters that are decisive for the output of ANNs are learned problem-specifically and data-driven in a training process. This is done by specifying a loss function that determines the error between desired and actual network output. Based on this error, the training data is used to iteratively adjust the network weights using the backpropagation algorithm so that the network outputs converge to the desired outputs. Once training is complete, the learned weights extract features that are critical to the task from the data. Thus, no manual specification of features is required, as was the case with other machine learning algorithms such as random forest classifiers, e.g. [Lin et al., 2020], or support vector machines, e.g. [Nanni et al., 2017]. Therefore, and because the large number of parameters makes it impossible to understand the net-

work's decision-making, ANNs are often considered black box systems. Currently, much research is being done on the interpretability of neural networks, which is particularly important in medical applications to assure correct and safe decisions. Counterfactual explanations are generated by perturbing the network's input to analyze which changes cause deviating outputs. For image analysis, approaches such as Grad-CAM [Selvaraju et al., 2017] or saliency maps [Simonyan et al., 2014] serve to determine those image regions that have contributed the most to the network's output. Different from these post-hoc methods, case-based reasoning compares the input to a particular instance (e.g. training instances or learned prototypes) to explain the prediction, allowing the user to evaluate the network's decision by inspecting the most similar instances [Narayanan and Bergen, 2024].

2.2.1 Artificial Neural Networks

Biological neurons consist of a cell body, multiple dendrites, and a single axon as displayed in Fig. 2.1. Dendrites are highly branched cell processes that receive excitatory or inhibitory signals from other neurons. The axon is another cell extension that carries signals to other cells at the synapses. The effect of incoming signals is cumulative if they occur in quick succession or are received by multiple synapses. When the accumulated excitatory signals reach the threshold potential of the neuron's cell membrane, an action potential is triggered. This action potential follows the all-or-nothing principle: it either occurs in full size or not at all [Campbell et al., 2015].

Artificial neurons mimic this behavior of accumulating incoming signals, weighting them and only passing them on when an internal threshold is reached (cf. Fig. 2.1). For this purpose, each artificial neuron has a certain number of inputs that receive signals (usually real numbers between 0 and 1 or -1 and 1) from previous neurons. These input signals are multiplied by weights and then summed up. The all-or-nothing principle in a biological neuron is technically realized with an activation function applied to the resulting summed input signal. The obvious choice for performing this thresholding would be a simple step function. However, as will be explained in the following section, any operation performed in the ANN must be differentiable, which is not the case for step functions. Commonly used activation functions include the sigmoid function, the hyperbolic tangent, and the rectified linear unit (ReLU), each of which is a differentiable approximation to the threshold function. A comparison between biological and artificial neurons, also called perceptrons, is shown in Fig. 2.1.

The output y of a perceptron is calculated via

$$y = g\left(\sum_{i=1}^{n} x_i w_i + b\right), \tag{2.2}$$

where g is the activation function, x_i for $i = 1, 2, ..., n$ are the n inputs of the perceptron and w_i are its weights. The weight b is the bias weight of the perceptron and realizes the internal threshold potential of biological neurons. A single perceptron can already per-

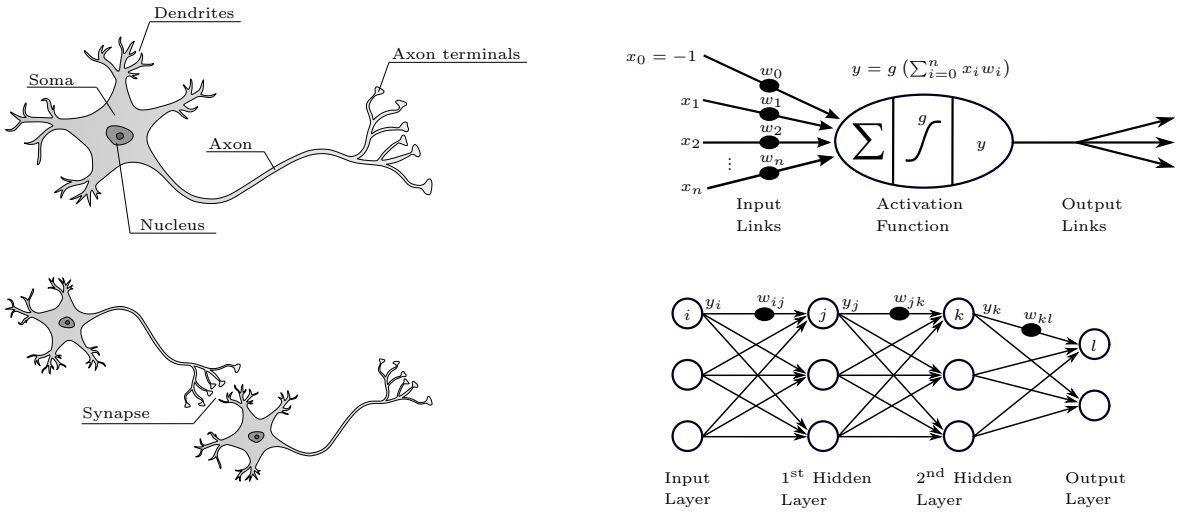


Figure 2.1: Analogy between biological and artificial neurons. Own illustration inspired by [Vadapalli, 2021; Burkhard, 2006; Meng et al., 2020]

form linearly separable classification tasks. However, the true power of neural networks lies in the combination of many such perceptrons, same as the brain's computational power results from the complex interconnection of individual cells. The "intelligence", of ANNs results from the weights in the network, that copy the connections between biological neurons, as they determine which signals are passed on or attenuated [Otte et al., 2020].

Like biological neurons, perceptrons are arranged in layers one behind the other. Three types of layers are distinguished from each other. The first layer of an ANN is called the input layer, while the last layer is the output layer. Each layer in between is called hidden layer. In current ANNs many of these hidden layers are being used, which has led to the term Deep Learning. In typical fully-connected neural networks (FCNNs), the perceptrons of one layer are connected to all perceptrons of the subsequent layer, as shown in Fig. 2.1, and the data flows through the layers sequentially. For very deep networks, it has proven to be beneficial to use skip-connections. These connections represent a shortcut in the network, offering an alternative path for the data that is directly passed to later layers of the network. The outputs from both paths are combined via concatenation, or, in the case of so-called residual connections, via element-wise addition. Skip-connections stabilize and speed up the learning process of ANNs and are used very frequently.

2.2.2 The Backpropagation Algorithm

The goal of training ANNs is to use examples to adjust the weights in the network so that the network outputs match the given desired outputs. For this purpose, an objective

or loss function $\mathcal{L}$ is defined in advance, which is then minimized during training over all given training data. Normally, training datasets are too large to be sent completely through the ANN in a single run. Minimization is, therefore, performed using stochastic gradient descent. A fixed amount of data points (mini batch) is drawn randomly and passed through the network. Subsequently, the network weights are adjusted based on the error made on this partial dataset. This step is repeated until the entire dataset has been passed through the network. One pass through the dataset is called an epoch. The training of ANNs requires many epochs to converge.

The algorithm used to adapt the network's weights and biases is called the backpropagation algorithm and is a generalization of the delta learning rule. For example, let the loss function be defined by the quadratic error $\mathcal{L} = \frac{1}{2}\sum_{j=1}^{N}\left(y_j - y_j^{\mathrm{GT}}\right)^2$ between network outputs y_j and ground truth values y_j^{GT}, whereby N is the number of output neurons of the network. The connecting weight w_{ij} of neurons i and j is updated based on the gradient of the loss function with respect to w_{ij}. The gradient descent direction for w_{ij} is given by

$$w_{ij}^{\mathrm{new}} = w_{ij}^{\mathrm{old}} - \eta \cdot y_i \cdot \delta_j \tag{2.3}$$

where y_i is the output of neuron i, η is the learning rate and δ_j is calculated via

$$\delta_j = \begin{cases} g'\left(\sum_{i=1}^{n} x_i w_{ij}\right)\left(y_j - y_j^{\mathrm{GT}}\right) & \text{if neuron } j \text{ is output neuron} \\ g'\left(\sum_{i=1}^{n} x_i w_{ij}\right)\sum_k \delta_k w_{jk} & \text{if neuron } j \text{ is hidden neuron.} \end{cases} \tag{2.4}$$

The index k runs over all neurons located in the layer following the layer in which neuron j is located. Whereas the parameters of output neurons can be updated directly by comparing them with the desired outputs, the parameters of hidden neurons are updated by propagating the error from the end of the network to the input of the network.

The learning rate η in Eq. (2.3) defines how strongly the parameters are changed in each iteration. This parameter may be chosen fixed throughout the entire training process, but typically it is changed with each epoch. The Adam algorithm [Kingma and Ba, 2015] computes individual adaptive learning rates for each parameter based on estimations of the first and second moments of the gradients. Like this, updates do not solely depend on the current gradient but also on gradients of previous iterations. Adam can help overcome problems arising from very steep or very shallow gradients, and is a popular choice for the training of ANNs.

2.2.3 Implicit Neural Representations

Recently, a new way of using FCNNs or multi-layer perceptrons (MLPs) has emerged, where the data is not used as input to the network, but is represented by the network itself. For example, an image is encoded implicitly in an MLP by passing the spatial coordinates of the pixels as input to the network and training the network to predict the intensity or color value of the pixels. Such an MLP is called implicit neural representation (INR), where the idea is to store the image not as a discrete matrix of pixel

values, but as a continuous function from which the pixel values are sampled. Once the network has been adapted to an image, pixel values can also be predicted at intermediate locations outside the original pixel grid. This allows a memory-efficient and resolution-independent storage of images or other signals.

However, INRs that use widespread activation functions, such as ReLU, show a low-frequency bias, i.e. they fail to represent the high-frequency components of the signals. To mitigate the tendency to generate smooth, piece-wise linear functions of ReLU-based networks, two complementary strategies have emerged: First, replacing the activation functions with periodic ones and second, transforming the input coordinates to higher-dimensional features. Sitzmann et al. proposed sinusoidal representation networks (SIRENs), that use the periodic sine function for activation [Sitzmann et al., 2020]. The activation function g in Eq. (2.2) becomes

$$g_{\text{SIREN}}(\boldsymbol{x}) = \sin(\omega \boldsymbol{x}), \tag{2.5}$$

where ω is used to match the frequency spectrum of the signal. In practice, $\omega = 30$ has proven successful for many applications [Sitzmann et al., 2020; Byra et al., 2023a]. Since then, several works proposed alternative activation functions that, like SIREN, also improve the ability of the INR to capture fine details, e.g. [Ramasinghe and Lucey, 2022; Saragadam et al., 2023; Liu et al., 2024]. Wavelet INRs (WIRE) use complex Gabor wavelet activation functions

$$g_{\text{WIRE}}(\boldsymbol{x}) = \exp(j\omega \boldsymbol{x}) \cdot \exp(-|s\boldsymbol{x}|^2), \tag{2.6}$$

improving the representation quality over SIREN for both natural and medical images [Saragadam et al., 2023].

Alternatively or additionally, the input coordinates can be transformed to frequency-rich, higher dimensional features. For example, positional encodings map input coordinates $\boldsymbol{v} \in [0, 1)^d$ to a set of sinusoids before passing them through the network. The mapping is defined by

$$\gamma(\boldsymbol{v}) = \left[\sin(2^0 \pi \boldsymbol{v}), \cos(2^0 \pi \boldsymbol{v}), \sin(2^1 \pi \boldsymbol{v}), ..., \cos(2^{K-1} \pi \boldsymbol{v})\right], \tag{2.7}$$

with a pre-defined number of frequencies K [Mildenhall et al., 2021]. Following, it was proposed to use a random projection

$$\gamma(\boldsymbol{v}) = \left[\sin(2\pi \boldsymbol{B}\boldsymbol{v}), \cos(2\pi \boldsymbol{B}\boldsymbol{v})\right]^\top, \tag{2.8}$$

where each element in the matrix $\boldsymbol{B}$ is randomly drawn from a standard normal distribution [Rahimi and Recht, 2007; Tancik et al., 2020]. There are several alternative encodings, that can be used, e.g. multi-resolution grids [Barron et al., 2021] or hash encodings [Müller et al., 2022]. Input encodings are often used with ReLU activation functions, but are also increasingly being combined with periodic activation functions [Byra et al., 2023b]. An overview of the different ways to implement an INR is given in Fig. 2.2, using the example of an OCT image of a skin wound. In this work, SIREN and WIRE will be used in conjunction with positional encodings.

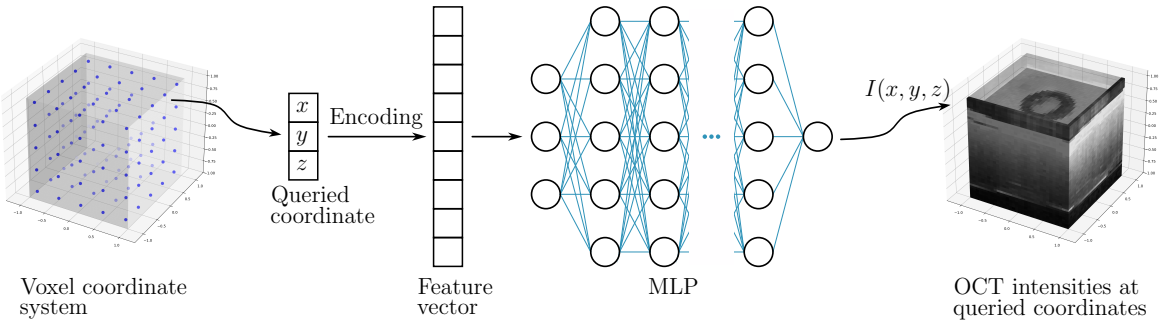


Figure 2.2: Overview of a general INR framework for OCT image representation. Input are the voxel coordinates, that are optionally transformed to higher dimensional feature vectors. The network used is a simple MLP with dedicated activation functions. Output of the INR are the OCT intensity values at the respective voxel locations. The example image shown is taken from the wounds dataset (cf. Sec. 2.4).

2.2.4 Convolutional Neural Networks

The processing of images with FCNNs has a number of drawbacks. FCNNs require a fixed number of inputs and outputs, so that only images of a fixed size can be processed with one FCNN. Object detection or segmentation in images should work regardless of the position of the object in the image. This is not the case with FCNNs, since it cannot be ensured that the learned network weights evaluate different image regions in the same way. Moreover, the pixels in an image have a spatial arrangement, that is being lost in fully-connected layers (FC layers) since all output neurons are connected to all input neurons. Finally, FCNNs have an enormous number of parameters to be learned, making the processing of large 3D images infeasible, even with modern computers.

CNNs tackle these problems by using specific types of network layers designed to process images, which are illustrated in Fig. 2.3. The name-giving convolutional layers contain learnable weights and bias values like the perceptrons in FC layers and are also used to extract features from the data. In contrast to FC layers, convolutional layers have only a small number of parameters that are used as discrete filters applied to small local image regions. Filtering is performed by calculating the scalar product of pixel and filter values. Unlike in FC layers, the multiplication is not performed only once, but is applied to each pixel position of the image in a sliding window fashion (blue and green pixels in Fig. 2.3). Thus, the same weights are used at each position in the image, which is why the weights in convolutional layers are called shared weights. Shared weights inherently lead to translation invariance. The features extracted by the filter kernels are the same regardless of where the object is located in the image. As in FCNNs, the convolutional layer is most often followed by a non-linear activation function.

The second type of layer designed for the processing of images are pooling layers. These subsample the feature maps that result from the convolutional layers. A 2×2-kernel is often used, that is moved over the image with a step size (stride) of two pixels. The four pixels covered by this kernel are reduced to one, usually by keeping

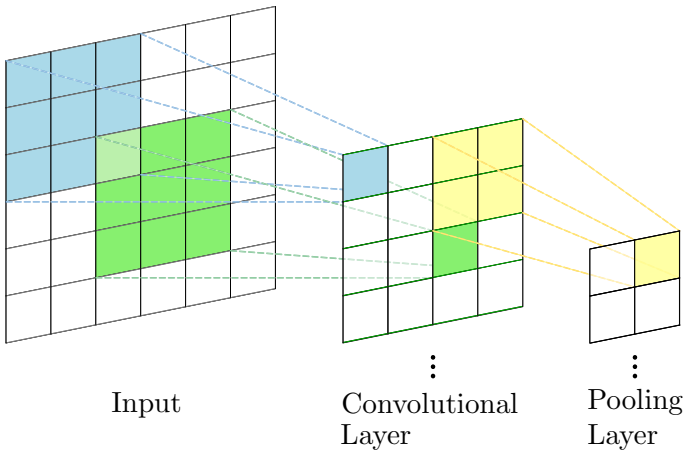


Figure 2.3: Illustration of convolutional and pooling layers typically used in modern CNNs.

the maximum pixel value or by taking the average of all four values (yellow pixels in Fig. 2.3). Typically, pooling layers are applied periodically throughout the CNN after a small number of convolutional layers. Pooling helps to condense the extracted information and to reduce the computational load of CNNs. The reduced size of feature maps enforces meaningfulness of the learned features. Moreover, pooling increases the receptive field, the region in the input image that influences one pixel of the network output.

In deep CNNs, multiple convolutional layers followed by pooling layers are used. With each such layer, the receptive field is successively increased, allowing deep networks to use large image regions to create outputs. The receptive field of deep networks fully covers large objects and allows more context to be considered, which helps many image analysis tasks like object detection, classification or registration. Another advantage of using deep CNNs is that more complex features can be learned in deeper layers. Simple features like edges are extracted early in the network, whereas the combination of these simpler features leads to more complex and defining feature maps later in the network.

2.2.4.1 Classification with Convolutional Neural Networks

Most approaches for image classification with CNNs are based on network architectures with multiple convolutional and pooling layers, followed by FC layers that finally provide the classification label. The typical CNN architecture was proposed as early as 1980 by K. Fukushima under the name Neocognitron [Fukushima, 1980], but the enormous success of CNNs started only in 2012 when the AlexNet proposed by A. Krizhevsky et al. won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [Russakovsky et al., 2015] with a lead of more than 10% over the other competitors [Krizhevsky et al., 2012]. In Fig. 2.4, the AlexNet architecture is shown. Input to the network are RGB images, rescaled and cropped to a resolution of 224×224×3 pixels. These inputs are first

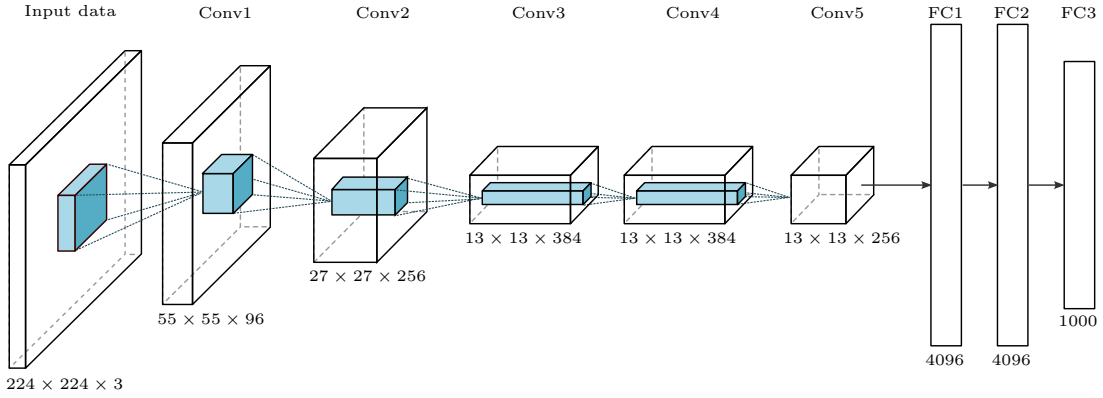


Figure 2.4: The architecture of AlexNet [Krizhevsky et al., 2012], one of the first successful CNNs for image classification. Five convolutional layers are completed by three fully-connected layers that map the input image to one of the 1000 classes of the ImageNet dataset. Maximum pooling is performed after the first two and the last convolutional layers. Numbers below convolutional layers give the sizes of the resulting feature maps, and numbers below fully-connected layers indicate the number of neurons of the respective layers. Illustration adapted from [Han et al., 2017].

processed by a convolutional layer with a filter size of 11×11 and a stride of four pixels. The three color channels of the input images are transformed into 96 feature maps, which are then processed with maximum pooling and fed to a second convolutional layer with a filter size of 5×5. The 256 feature maps generated here are again processed with maximum pooling before being sent through three more convolutional layers. Maximum pooling is not applied between these layers, but only after the last convolutional layer. The last three convolutional layers have 384, 384 and 256 kernels, respectively. The resulting feature maps are flattened into a single vector, that is processed by three FC layers with 4096, 4096 and 1000 neurons. The output of the network is a 1000-component vector that is fed to a softmax activation function to produce a distribution over the 1000 class labels of the ILSVRC. The AlexNet architecture is still a popular choice for image classification tasks. Other successful classification CNNs are, for example, VGG-16 and -19 [Simonyan and Zisserman, 2015], GoogLeNet [Szegedy et al., 2015], ResNet [He et al., 2016] and DenseNet [Huang et al., 2017].

In 2021, Dosovitskiy et al. presented Vision Transformer (ViT) as an alternative to CNNs for image classification tasks [Dosovitskiy et al., 2021]. Transformers originate from language processing tasks, where they surpassed the previously established recurrent neural network (RNN) architectures. RNNs process words one at a time, whereas transformers are non-sequential and use full sentences to perform the task at hand. This allows transformers to access past information while prioritizing the various input components according to their importance. Analogously, in ViTs, input images are treated as a sequence of image patches. Each patch is flattened into a vector which is projected linearly onto a lower-dimensional embedding and combined with a learnable positional

embedding. The embedded sequence is fed to a transformer encoder that is finalized by FC layers that produce the class label. In contrast to CNNs with their local receptive field, ViTs can capture the global context of the images [Shamshad et al., 2023]. ViTs have gained in popularity since they were first presented. In particular, hybrid models, where the input sequence is formed from CNN feature maps, are often used for image classification, e.g. [Dosovitskiy et al., 2021; Chen et al., 2021; Wang et al., 2021b], but also for other image processing tasks such as segmentation and registration, e.g. [Chen et al., 2022; Zhang et al., 2023].

2.2.4.2 Segmentation with Convolutional Neural Networks

Image segmentation requires pixel-wise rather than global class labels. The architecture of segmentation CNNs, therefore, needs to be changed compared to classification networks such that they output images. The fully convolutional U-Net by Ronneberger et al. accordingly consists of one encoding and an inverse decoding path [Ronneberger et al., 2015]. The encoder processes the input images with several convolutional layers and maximum pooling. This leads to ever smaller feature maps, the number of which increases with the convolutional layers. The small feature maps are enlarged again in the decoder and their number is reduced by convolutions. To preserve both high-resolution spatial detail of the early layers, as well as the more meaningful information of deep layers, skip-connections are used to connect features in the encoding path with features in the decoding path. This makes it possible to recognize the context of image objects and perform an exact localization of the objects by using both the abstract feature maps of the deep layers and the simpler features from the beginning of the network at the same time.

The U-Net architecture is displayed in Fig. 2.5. At each resolution level, two convolutional layers with a kernel size of 3×3 followed by maximum pooling are used. The very first convolutional layer produces 64 feature maps, while the second does not increase the number of features but produces again 64 feature maps. Analogously, in the following convolutional blocks, the first layer doubles the number of features and the second retains this number. ReLU is used after each convolution as non-linear activation function. In the decoder of the original U-Net paper, the enlargement of feature maps is done by upsampling, followed by a 2×2 convolution to halve the number of feature maps. Since upsampling may lead to checkerboard artifacts, more sophisticated methods can be used, such as interpolation or transposed convolutions. An often used modification to the original U-Net architecture is the usage of padded convolutions to prevent the size of the feature maps from changing due to the convolutions. In [Ronneberger et al., 2015], segmentation maps were only given for the central part of the input images. Furthermore, it has proven beneficial to use normalization layers after the convolutions (and before applying the activation function), in which the values of the feature maps are re-scaled to have zero mean and unit variance. While in instance normalization, standardization is performed for each feature map of a single input example,

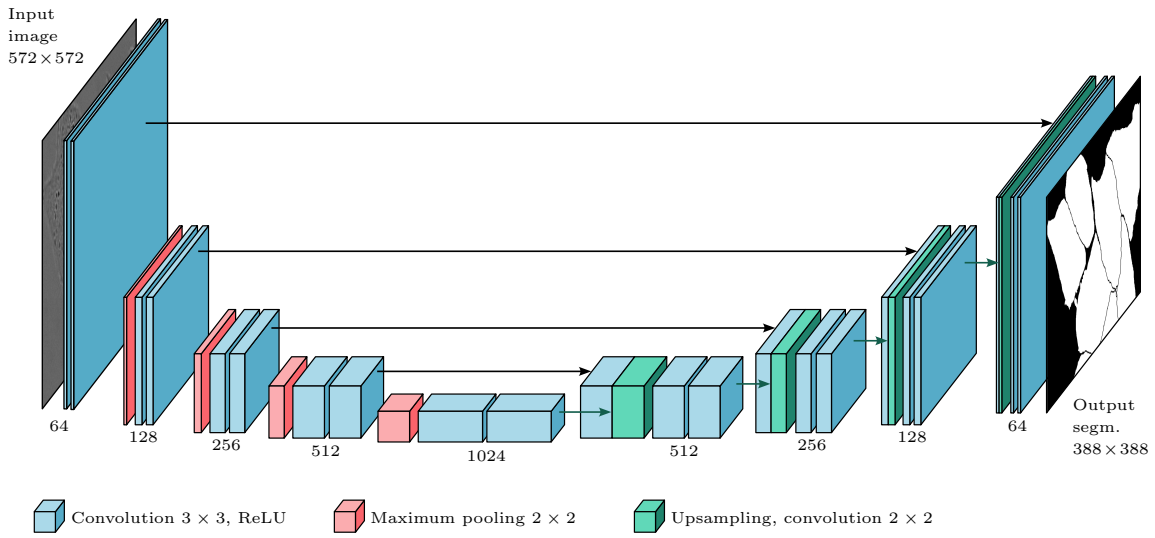


Figure 2.5: The architecture of U-Net [Ronneberger et al., 2015], the most frequently used CNN for image segmentation. The network consists of an encoding path, reducing image resolution and extracting increasingly abstract features, and a mirroring decoding path, that enlarges the images again. Features from both parts are concatenated to preserve both spatial detail and abstract features. Numbers below boxes indicate the number of feature maps, while arrows show the concatenation of features.

in batch normalization (BN), standardization is performed for each feature map of all elements in a mini batch. Normalization can speed up the network's convergence and, in the case of BN, also improve the generalization of the network, since the result for a given instance depends on the other elements in the batch, which introduces regularizing noise into the training process.

Since its development in 2015, U-Net is the most widely adopted CNN architecture not only for segmentation but also for other tasks, such as image-to-image translation or the generation of deformation fields for image registration. While U-Net-based networks proofed very successful in many medical image analysis tasks, their training requires large amounts of (labeled) data. Especially in the medical field, the size of datasets is often limited due to the small number of patients, data protection regulations and, most importantly, the need for expert knowledge to create ground truth labels. This data scarcity can lead to overfitting, where the enormous number of parameters in the network are used to store the input data and the corresponding desired output, rather than recognizing general patterns. This results in such a network performing significantly worse on unknown data than on training data. To prevent overfitting and to robustly train CNNs based on small datasets, data augmentation techniques are widely used in the DL domain. Data augmentation serves to increase the variance in the dataset artificially by applying transformations to the existing images. These transformations can include geometric deformations as well as intensity changes. Most often, geometric deformations comprise rigid transformations such as flipping, translation and rotation. Depending on the application, elastic deformations might also be conceivable. Intensity

changes include, for example, adding random noise, masking out random image areas or adjusting the image brightness and contrast.

2.3 Evaluation Metrics

This section lists and shortly introduces the metrics used throughout this thesis to evaluate the segmentation and registration performance of the proposed algorithms. First, metrics to assess segmentation results are given, followed by metrics for the assessment of image registration accuracy and deformation regularity.

2.3.1 Segmentation Metrics

Given a binary ground truth segmentation $S^{\mathrm{GT}} : \Omega \to \{0, 1\}$ of image $I : \Omega \to \mathbb{R}$, several metrics can be used to evaluate the quality of a predicted segmentation $S^{\mathrm{P}} : \Omega \to \{0, 1\}$ by comparing it to the ground truth. Typically, these metrics relate the number of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) voxels.

The sensitivity or true positive rate

$$\mathrm{sens} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2.9}$$

gives the proportion of voxels that were correctly segmented, i.e. they were assigned value one, out of all voxels that belong to the foreground class in the ground truth. In turn, the specificity or true negative rate

$$\mathrm{spec} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{2.10}$$

describes the proportion of correctly non-segmented voxels, whereas the accuracy

$$\mathrm{acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}. \tag{2.11}$$

measures the percentage of correctly classified voxels. All these metrics range between zero and one, where zero indicates no agreement between ground truth and predicted segmentation and one perfect overlap.

Oftentimes, it is more important to recognize a certain structure in the first place than to detect every single voxel of the structure. In multiple sclerosis, for example, the number of new focal lesions is a decisive biomarker to test the efficacy of drugs, while the small size of these lesions means that even small changes in the segmentation mask cause voxel-based metrics to fluctuate widely. In some applications, therefore, the sensitivity and specificity are not reported on voxel-level, but on pathology- or image-level. The sensitivity on pathology-level $\mathrm{sens}^{\mathrm{patho}}$ gives the proportion of pathologies that were correctly identified out of all pathologies in the ground truth, and the sensitivity

on image-level sens$^{\text{img}}$ gives the proportion of images identified as pathological out of all pathological images. Analogously, the specificity on image-level spec$^{\text{img}}$ describes the proportion of correctly empty predicted segmentations among all images showing no pathology.

Area-based metrics

Let $G = \{\boldsymbol{x} \in \Omega | S^{\text{GT}}(\boldsymbol{x}) = 1\}$ be the set of segmented voxels in the ground truth segmentation and $P = \{\boldsymbol{x} \in \Omega | S^{\text{P}}(\boldsymbol{x}) = 1\}$ be the set of segmented voxels in the predicted segmentation. The Dice similarity coefficient (DSC) then measures the segmentation quality as

$$\text{DSC}(P, G) = \frac{2|P \cap G|}{|P| + |G|} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}. \tag{2.12}$$

Here, the segmentation overlap is set into relation to the total segmented area in both ground truth and predicted segmentations. Similarly to the DSC, the Jaccard index is given by

$$\text{Jac}(P, G) = \frac{|P \cap G|}{|P \cup G|}. \tag{2.13}$$

For some applications in this work, such a voxel-by-voxel comparison is less important than the overall extent of the respective pathologies (e.g. the wound localization in Chapter 3 or changes in the photoreceptor layer of the retina in Chapters 3 and 6). A modification of the DSC is used here, in which the three-dimensional segmentations $S^{\text{GT}}, S^{\text{P}} : \Omega \subset \mathbb{R}^3 \to \{0, 1\}$ are first projected onto 2D using

$$S^{\text{GT}}_{\text{proj}}(x, y) = \begin{cases} 1, \text{ if } \exists z \text{ such that } S^{\text{GT}}(x, y, z) = 1 \\ 0, \text{otherwise} \end{cases} \tag{2.14}$$

$$S^{\text{P}}_{\text{proj}}(x, y) = \begin{cases} 1, \text{ if } \exists z \text{ such that } S^{\text{P}}(x, y, z) = 1 \\ 0, \text{otherwise.} \end{cases} \tag{2.15}$$

That is, a pixel in the 2D images is assigned value one, if any voxel in the respective 3D column is segmented. The projected DSC is then calculated with

$$\text{DSC}_{\text{proj}}(P, G) = \frac{2|P_{\text{proj}} \cap G_{\text{proj}}|}{|P_{\text{proj}}| + |G_{\text{proj}}|}, \tag{2.16}$$

where $G_{\text{proj}} = \{\boldsymbol{x} \in \Omega_{\text{proj}} | S^{\text{GT}}_{\text{proj}}(\boldsymbol{x}) = 1\}$ and $P_{\text{proj}} = \{\boldsymbol{x} \in \Omega_{\text{proj}} | S^{\text{P}}_{\text{proj}}(\boldsymbol{x}) = 1\}$.

Surface Distances

DSC and Jaccard index are area-based metrics, measuring segmentation quality based on voxel overlap. However, they are insensitive to small segmentation errors or form deviations, especially when evaluating large structures. Surface distance metrics are, therefore, often used together with the DSC or the Jaccard index for applications where the exact localization of segmentation boundaries is important. Given two sets of contour points G and P, let

$$d(g, P) = \min_{p \in P} d(g, p), \tag{2.17}$$

where d is the Euclidean distance, then the average symmetric surface distance

$$\mathrm{ASSD}(P,G) = \frac{\sum_{g\in G} d(g,P) + \sum_{p\in P} d(p,G)}{|G| + |P|} \tag{2.18}$$

measures the average of all shortest boundary distances between contour G to any point on contour P and vice versa. The Hausdorff distance

$$\mathrm{HD}(P,G) = \max\Big(\max\big(d(g,P)|g\in G\big), \max\big(d(p,G)|p\in P\big)\Big), \tag{2.19}$$

in turn, uses the maximum distance between the contours as measure, and is, thus, very sensitive to outliers in the segmentation.

2.3.2 Registration Metrics

When registering two or more images, several features of the resulting deformed images and the deformation field itself can be analyzed to evaluate how well corresponding structures are aligned after registration and how plausible the applied deformation field is. The regularity of the deformation field φ is typically assessed using the Jacobian determinant $J(\boldsymbol{x}) = \det(\nabla\varphi(\boldsymbol{x})) = |\nabla\varphi(\boldsymbol{x})|$. When $J(\boldsymbol{x}) = 1$, the deformation is volume-preserving, whereas $J(\boldsymbol{x}) < 0$ indicates foldings in the deformation field. In real deformations inside the human body, inversions do not occur, which is why the (relative) number of voxels with a negative Jacobian determinant is an important indicator for the plausibility of the registration field and used in this thesis to evaluate image registration results.

For evaluation of the registration accuracy, the segmentation metrics listed above can be used, provided that the images to be registered are segmented. In that case, the segmentation metrics are used to measure the overlap of corresponding structures after registration. Additionally, image similarity metrics can be used to evaluate the similarity of the fixed and warped moving images.

Image Similarity Metrics
Given two images $I : \Omega \to \mathbb{R}$ and $J : \Omega \to \mathbb{R}$, the mean absolute error (MAE) measures the absolute intensity difference between the voxels in I and J:

$$\mathrm{MAE}(I,J) = \frac{1}{n}\sum_{\boldsymbol{x}\in\Omega} |I(\boldsymbol{x}) - J(\boldsymbol{x})|, \tag{2.20}$$

where n is the total number of voxels. The mean squared error (MSE) similarly is defined by

$$\mathrm{MSE}(I,J) = \frac{1}{n}\sum_{\boldsymbol{x}\in\Omega} \big(I(\boldsymbol{x}) - J(\boldsymbol{x})\big)^2, \tag{2.21}$$

Using the MSE, the peak signal-to-noise ratio (PSNR) is given by

$$\mathrm{PSNR}(I,J) = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}}{\mathrm{MSE}(I,J)}\right). \tag{2.22}$$

Here, MAX is the maximum possible voxel value of the images. The normalized cross-correlation (NCC) image distance measure is defined by

$$\mathrm{NCC}(I,J) = \frac{\sum_{\boldsymbol{x}\in\Omega}\left(I(\boldsymbol{x})-\bar{I}\,\right)\left(J(\boldsymbol{x})-\bar{J}\,\right)}{\sqrt{\sum_{\boldsymbol{x}\in\Omega}\left(I(\boldsymbol{x})-\bar{I}\,\right)^2\sum_{\boldsymbol{x}\in\Omega}\left(J(\boldsymbol{x})-\bar{J}\,\right)^2}}, \tag{2.23}$$

with $\bar{I}$ being the average voxel intensity of image I and $\bar{J}$ the average intensity of image J, and the local normalized cross-correlation (LNCC) calculates the NCC for small image patches in a sliding window fashion.

These metrics have shown to oftentimes be inconsistent with human visual perception, which is why the structural similarity index measure (SSIM) was proposed in [Wang et al., 2004]. The SSIM is calculated with

$$\mathrm{SSIM}(I,J) = \frac{(2\mu_I\mu_J + c_1)(2\sigma_{IJ} + c_2)}{(\mu_I^2 + \mu_J^2 + c_1)(\sigma_I^2 + \sigma_J^2 + c_2)}, \tag{2.24}$$

where μ_I and μ_J are the mean voxel intensities of I and J, σ_I and σ_J the variances of the intensities and σ_{IJ} is the sample covariance of I and J. Furthermore, $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ are used to stabilize the division numerically, where L is the dynamic range of the voxel-values. The default values of k_1 and k_2 are 0.01 and 0.03. The SSIM is a combination of luminance (the overall brightness of the images), contrast and image structure. The image structure is measured using the correlation of intensity changes. Since human perception is sensitive to structures (e.g. edges) but not so sensitive to errors in single voxels, the SSIM corresponds more closely to the perceived image similarity than, for example, the MSE, which only compares the intensities of the individual voxels.

Similarly, the learned perceptual image patch similarity (LPIPS) was developed to match the human perception of image similarity [Zhang et al., 2018]. Here, a neural network (VGG or AlexNet) is trained on a dataset of human similarity judgments. The images to be assessed are then both passed through the trained network, and the image features that are extracted in the intermediate layers of the network are compared to generate a similarity metric:

$$\mathrm{LPIPS}(I,J) = \sum_l \frac{1}{H_l W_l}\sum_{h=1}^{H_l}\sum_{w=1}^{W_l}\left\| w_l \odot \left(\hat{f}_l^I(h,w) - \hat{f}_l^J(h,w)\right)\right\|_2^2, \tag{2.25}$$

where the index l runs over the CNN layers, w_l are the network weights of layer l and H_l and W_l are the height and width of the feature map f_l^I in layer l. Each feature map is normalized channel-wise using $\hat{f}_l^I = f_l^I / \|f_l^I\|_2$.

2.4 Imaging Modalities and Datasets Analyzed in this Thesis

The algorithms developed within this thesis serve the analysis of two different imaging modalities, namely optical coherence tomography (OCT) and magnetic resonance imaging (MRI). OCT applications include wound monitoring on images of the skin and the analysis of various biomarkers in retinal diseases, primarily focusing on chronic central serous chorioretinopathy (CSCR) and age-related macular degeneration (AMD). The MRI applications consider diseases of the brain, including stroke and multiple sclerosis (MS). In the following, a brief overview of the dataset used throughout this thesis is given.

2.4.1 Optical Coherence Tomography Datasets

Optical coherence tomography (OCT) is a non-invasive imaging technique providing high-resolution cross-sectional images of transparent and semi-transparent materials. The main application of OCT in medicine is in ophthalmology, but also increasingly in dermatology. In OCT, a low-coherence light beam is directed towards the tissue under examination. Different structures within the tissue have different optical properties and reflect the light back towards the source. Similar to ultrasound imaging, the echo time delay and the intensity of the reflected light are then used for depth resolution and visualization of the tissue microstructure. In OCT, one such intensity profile is called an A-scan, and to obtain 2D images (B-scan), the light source is moved over the sample. For 3D volume images, several B-scans are taken one after the other and stacked to form a volume. OCT visualizes individual tissue layers and pathologies of the retina or skin with a resolution in the micrometer range. In Figs. 2.6 and 2.7, examples of a healthy and a diseased retina are shown, highlighting the appearance of healthy retinal layers and pathological biomarkers in OCT images.

The wounds dataset The dataset used in Chapter 3 for wound segmentation has been acquired at the Cutaneous Biology Research Center in the Department of Dermatology at Massachussetts General Hospital, Harvard Medical School. Twelve human skin equivalents (EFT-400, MatTek, Ashland, MA, USA) were wounded and monitored for one week. To do so, the epidermis was detached from each skin equivalent and two circular wounds were created with a 1.5mm biopsy punch. After reattachment of the epidermis, the skin equivalents were cultured in medium (EFT-400-ASY, MatTek). Each wound was imaged with a Thorlabs OCT scanner immediately after the injury and on the 1st, 2nd, 3rd, 4th and 7th day afterward (cf. Fig. 2.2 for an example wound image). The dataset used for wound segmentation, thus, consists of 144 OCT images, each of which spans a field of view of $3.5 \times 3 \times 3$ mm^3 and has an image size of $1024 \times 461 \times 461$ voxels. Out of the 24 wounds, six were manually segmented in the lateral dimension for

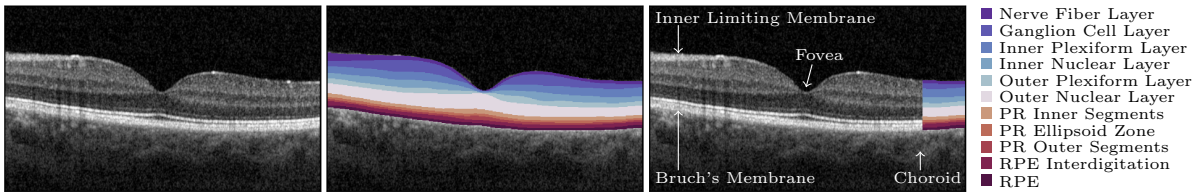


Figure 2.6: Appearance of the healthy retina in OCT images. PR stands for photoreceptor and RPE for retinal pigment epithelium.

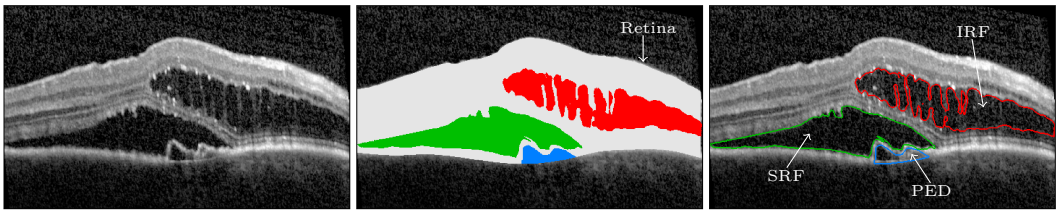


Figure 2.7: A diseased retina on OCT. Shown is an example from the CSCR dataset with pathologies intraretinal fluid (IRF), subretinal fluid (SRF) and pigment epithelial detachment (PED).

all time points, i.e. the labels are given per A-scan, allowing to measure the extent of the wound rather than its depth.

The CSCR dataset The CSCR dataset used in Chapter 3 for pathology segmentation and in Chapters 6 and 7 for image registration is a longitudinal retinal OCT dataset, collected retrospectively from clinical databases at the Eye Clinic of the University Hospital Schleswig-Holstein in Kiel (Universitätsklinikum Schleswig-Holstein (UKSH) Kiel). The overall dataset consists of follow-up data from 126 CSCR patients acquired with a Heidelberg Spectralis scanner. The follow-up times vary from zero months to 14 years, with up to 35 available OCT images per eye. For 20 patients (204 images), the retinal pathologies intraretinal fluid (IRF), subretinal fluid (SRF) and pigment epithelial detachment (PED) were segmented manually. Furthermore, manual annotations of photoreceptor pathologies are available for 66 patients, out of which 26 show atrophy. The dataset is described in more detail in Tabs. A1 - A3 in the appendix, which include for each patient the number of available OCT and segmentation images, as well as the information whether the images were used for the training of the CNNs presented in this thesis.

The IMI dataset The IMI dataset contains images of healthy eyes, acquired at the Institute of Medical Informatics at the University of Lübeck. Both eyes of 50 volunteers were imaged with a Heidelberg Spectralis scanner and a research prototype of a low-cost handheld OCT system [Sudkamp et al., 2016]. For this thesis, only the Spectralis

images are used, with Chapter 6 using images taken with the standard clinical settings and Chapter 7 using an experimental OCT angiography (OCTA) setting. The images taken with the standard setting are expected to match the appearance of the patient data from the other datasets best. These images consist of 49 B-scans with a resolution of 496×512 pixels, covering a field of view of 2×6×6 mm^3. There are no manual annotations for these images. The OCTA images are densely sampled volumes consisting of 512 B-scans, each having a resolution of 496×512 pixels, also spanning a field of view of $2 \times 6 \times 6$ mm^3. For these images, 12 retinal layers were first segmented automatically with the Iowa Reference Algorithms [Li et al., 2005], and the segmentations were then corrected semi-automatically. Chapter 6 uses the standard OCT images for anomaly detection, and Chapter 7 uses the OCTA images for B-scan interpolation and retinal layer segmentation.

AMD datasets Two AMD datasets are used in this thesis, one containing only one time point per patient and the other one being a longitudinal dataset. Both datasets were retrospectively collected at the UKSH Eye Clinic in Kiel and exclusively contain images acquired with a Heidelberg Spectralis scanner. The first AMD dataset consists of 60 OCT images from 60 eyes of 36 AMD patients with manual annotations of IRF, SRF and PED. This dataset is used in Chapter 7 for anomaly detection. The longitudinal AMD dataset consists of 709 OCT volumes from 41 eyes of 41 AMD patients. Follow-up times range from 32.5 to 82.8 months with an average of 58.9 months (nine to 24 images per eye, 17.3 images on average). For each image, the inner limiting membrane (ILM), the retinal pigment epithelium (RPE) and the Bruch's membrane (BM) were segmented. While the RPE and BM delineations allow analyzing PED, no segmentations of further retinal fluids IRF and SRF are given. The longitudinal dataset is used for image registration in Chapter 4.

The RETOUCH dataset The Retinal OCT Fluid Challenge (RETOUCH) [Bogunović et al., 2019] was held in 2017 with the goal to detect and segment pathological fluids in OCT images from different scanners. The publicly available RETOUCH training data consists of 70 OCT volumes, out of which 24, 24 and 22 were acquired with Cirrus, Spectralis and Topcon scanners, respectively. Several diseases are represented in this dataset, including AMD, retinal vein occlusion and diabetic macular edema. Again, the three fluid types IRF, SRF and PED were segmented manually. The RETOUCH dataset is used in Chapter 6 for anomaly detection.

2.4.2 Magnetic Resonance Imaging Datasets

Magnetic resonance imaging (MRI) is a non-invasive imaging technique providing detailed cross-sectional images of the inside of the human body without using ionizing radiation. During MRI, the patient or the part of the body to be examined is placed in the MRI machine, which produces a strong magnetic field. The protons in the body align themselves along this field and generate a magnetization in the direction of the externally applied field. Subsequently, a high-frequency alternating field is applied that deflects the magnetization from the direction of the static field and converts it into a transverse magnetization. The transverse magnetization starts rotating around the field direction of the static magnetic field, generating an electrical voltage that can be measured. Depending on the tissue type, it takes a different amount of time for the protons to realign along the static magnetic field, after switching off the alternating field. The brightness of different tissue types in MR images is determined by these differing relaxation times and the amount of hydrogen atoms (proton density). Which of these parameters dominates the image contrast is influenced by the choice of imaging method (pulse sequence) and the measurement parameters, allowing to generate images with different contrasts within the same imaging test. MRI is particularly suited for the imaging of soft tissues and is, therefore, the most sensitive imaging method to diagnose and monitor brain conditions.

The LPBA40 dataset with phantom stroke lesions The LONI Probabilistic Brain Atlas (LPBA40) dataset is a public dataset of 40 whole-head MRI volumes with given manual segmentations of 56 anatomical brain regions [Shattuck et al., 2008]. The images from 40 healthy volunteers were skull-stripped and co-registered into a common atlas space. The images have a resolution of 181×217×182 voxels with an isotropic voxel spacing of 1 mm. Here, a modified version of the dataset is used where four different stroke lesions extracted from the Ischemic Stroke Lesion Segmentation (ISLES) dataset [Maier et al., 2017] are inserted into the MR images artificially. Each lesion is inserted separately into each LPBA40 image, leading to five versions of each image, i.e. original and corrupted with lesion L1, L2, L3 and L4.

The MSSEG-2 dataset The Multiple Sclerosis New Lesions Segmentation (MSSEG-2) challenge data consists of 100 whole-head FLAIR MR image pairs of MS patients acquired between one and three years apart. 40 image pairs are defined as training data by the challenge organizers, and the remaining 60 image pairs serve as testing data. The dataset comprises images from 15 different MRI scanners (three GE, six Philips and six Siemens scanners), with all images from GE scanners being assigned to the test data in order to check the generalizability of submitted segmentation algorithms. For each patient, baseline and follow-up images have been rigidly pre-aligned and manual segmentation of new MS lesions was done by four neuroradiologists using the pre-aligned scans. Afterward, a consensus was formed by letting a senior expert neuroradiologist decide upon disputed lesions and then fusing the segmentations for the accepted lesions

with majority voting. No new lesions were observed in 39 cases, out of which eleven belong to the training and 28 to the testing datasets. For the images with new lesions, the lesion count ranges from one to 45.

Chapter 3

Invariant Information Clustering for the Unsupervised Segmentation of Pathologies in OCT Image Data

The manual segmentation of three-dimensional medical images is a tedious, difficult and expensive task requiring expert knowledge and a lot of time. Moreover, manual segmentation is error-prone, especially for pathology cases, thus suffering from high inter- and intra-rater variability. Automatic segmentation methods are created to speed up and harmonize medical image segmentation. The development of such methods often requires large amounts of labeled image data. The training of deep learning-based methods such as CNNs is particularly data-intensive. However, the generation of sufficiently large annotated databases is often not feasible, despite the availability of vast amounts of unlabeled image data.

Unsupervised image segmentation aims to divide images into distinct regions without relying on expert annotations. Segmentation masks are generated based on the structures or patterns observed in the input images. Both non-learning-based clustering and patch-based contrastive learning methods achieve a separation of semantically meaningful image areas by grouping similar and separating dissimilar image parts. Contrastive learning offers particularly interesting possibilities here, since the discriminative and common image features are learned over a large dataset, which reduces the strong dependence on the image information of individual images in classical clustering.

This chapter uses invariant information clustering (IIC) for the unsupervised segmentation of pathologies in OCT images. IIC is a DL-based framework for image classification and segmentation that originally has been proposed in [Ji et al., 2019]. Ji et al. segment images by maximizing the mutual information (MI) of neighboring image patches. In this thesis, IIC is applied to image patches in a contrastive learning setup to improve segmentation performance and generalizability. The chapter is organized as follows. Sec. 3.1 summarizes the recent literature on unsupervised image segmentation. In Sec. 3.2, the theoretical foundations necessary for the upcoming experiments are laid. In particular, IIC is explained. The IIC extensions that were introduced as part of this thesis for the segmentation of OCT images are presented in Sec. 3.2.1. In Sec. 3.3, IIC is used for the unsupervised segmentation of wounds in OCT images of the skin. This work has partially been published in [Andresen et al., 2022b]. Additionally, in Sec. 3.4, IIC is used to segment atrophy in retinal OCT images. Here, partially supervised training is performed, allowing the network to learn from few annotated samples while profiting from large amounts of unlabeled data that would not be usable in fully supervised approaches. Both Secs. 3.3 and 3.4 first give a short motivation and literature review of the

respective use case before describing the application-specific methodologies, experiments and results.

3.1 Related Literature

Unsupervised image segmentation methods can broadly be classified into classical, i.e. non-learning-based, and learning-based approaches. Classical techniques usually work by grouping similar voxels, that are assumed to be part of the same object. The earliest classical methods include thresholding, edge- or contour-based methods, clustering-based and graph-based approaches, among others [Rodrigues et al., 2024]. In Watershed-based approaches, grayscale images are viewed as topographic surfaces where the voxel intensities define the height of the surface. Image segmentation is then modeled as a flooding of this landscape. Starting from local minima, the landscape is flooded by enlarging the water basins with the neighboring voxels based on their intensity, and wherever water from two different minima meets, a segmentation border is placed [Beucher, 1979; Vincent and Soille, 1991]. Many variants and extensions of this algorithm have been proposed to overcome the susceptibility to noise and to improve adherence to the semantic content of the image, e.g. [Beucher and Meyer, 1992; Meyer, 1994; Najman and Schmitt, 1996].

Clustering-based methods can be subdivided into partitioning and hierarchical approaches. In partitioning clustering, images are divided into a pre-defined number of segments, referred to as superpixels. For example, k-means clustering works by using k random initial cluster centers and assigning pixels to the cluster whose center has the smallest Euclidean distance to the pixel in the feature space. The cluster centers are then recalculated as the average location of the pixels assigned to the respective center. This procedure is iteratively repeated until convergence is reached [MacQueen, 1967; Lloyd, 1982]. Hierarchical clustering, in turn, builds clusters by either starting with each voxel as a separate cluster and iteratively merging the most similar clusters, e.g. [Arifin and Asano, 2006; Bruse et al., 2017], or by starting with a single cluster that is subdivided iteratively, e.g. [Antonio et al., 2010].

Graph-based methods model the images as graphs, where each voxel is a node and the edges between neighboring nodes are defined based on a similarity measure between the voxels. This similarity measure may include color, texture or location information. The graph is then divided into two or more subsets, defining the segmenting clusters, based on the edge costs. Some of the most prominent works in this field are [Shi and Malik, 2000; Felzenszwalb and Huttenlocher, 2004; Grady, 2006; Couprie et al., 2011]. Similarly, in density-based approaches, e.g. [Ester et al., 1996; Hinneburg and Keim, 1998; Ankerst et al., 1999; Comaniciu and Meer, 2002; Vedaldi and Soatto, 2008], voxels are assigned to clusters based on image features. Here, the voxels are grouped by shifting them to the mode of a computed image distribution in the feature space [Rodrigues et al., 2024].

More recent approaches for unsupervised image segmentation most often are based on DL. W-Net, for example, combines two U-Nets into an autoencoder [Xia and Kulis,

2017]. The first U-Net segments the input images, and the second one receives these segmentations to reconstruct the original images. Both networks are trained jointly using the reconstruction error and a differentiable approximation of the normalized cut applied to the output of the segmentation network. In [Kanezaki, 2018], a CNN is proposed that outputs a probability distribution over k classes for each pixel. The segmentation is generated by assigning the class with the highest probability to the pixel. The network is trained with a specialized loss function that (1) favors pixels with similar features being assigned to the same class, (2) favors neighboring pixels to be assigned to the same class and (3) favors the number of labels used to be large.

In the medical field, anomaly detection methods are frequently used for the unsupervised segmentation of pathologies. Here, the networks are trained on healthy images and applied to pathological images to find everything that deviates from the "normal" appearance. For example, autoencoder-based methods are assumed to only be able to reconstruct the healthy anatomy they were trained on, and pathologies are segmented as areas with high reconstruction error [Schlegl et al., 2019; Uzunova et al., 2019; Li et al., 2021]. The GAN-based methods [Bielski and Favaro, 2019; Chen et al., 2019] are able to detect, extract and redraw the main object of an image, giving segmentations of foreground and background. However, these methods are limited to binary segmentation problems. [Van Gansbeke et al., 2021] uses an unsupervised saliency detector to generate mask proposals, which are used to generate negative and positive pairs for a contrastive pre-training of a CNN. The pre-trained CNN already can separate images into semantically meaningful segments, which are refined further using either classical clustering methods or fine-tuning of the network on a small dataset. [Melas-Kyriazi et al., 2022] also train a network in a self-supervised manner and use the feature maps that the network extracts to generate semantic segmentations. Finally, [Ji et al., 2019] and [Ouali et al., 2020] both train CNNs to maximize the mutual information between network outputs for paired input samples to achieve unsupervised image segmentation. In [Ji et al., 2019], neighboring image patches are passed through an encoder network, whereas [Ouali et al., 2020] uses a feature extractor, an autoregressive encoder and a decoder that output segmentations for different views of the same input image. The approach presented by Ji et al. is extended in this chapter. This method, first presented in [Andresen et al., 2022b], appears to be the only one to date that performs a completely unsupervised segmentation of cutaneous wounds in human skin equivalents.

3.2 Methods

To enable the unsupervised clustering of images and image patches, invariant information clustering (IIC), proposed in [Ji et al., 2019], uses paired data samples and trains a classification CNN to produce the same outputs for both instances in a pair. As a novel clustering objective, it is proposed to maximize the mutual information (MI) between paired network outputs. For image clustering, i.e. image classification, getting paired data samples can be done by simply selecting two images from the same class. Since

this requires given ground truth labels, the authors of [Ji et al., 2019] instead follow a self-supervised approach using random augmentations of the same image as paired data samples. To apply the approach for image patch classification, and like this for semantic segmentation, paired data samples are generated by using neighboring image patches. IIC is shown to achieve state-of-the-art performance for both image clustering and unsupervised semantic segmentation on several datasets.

More formally, given a paired data sample $\boldsymbol{x}$ and $\boldsymbol{x}'$, be it images or patches, IIC requires a classification CNN Φ with parameters θ to output the same class assignments $\Phi(\theta; \boldsymbol{x})$ and $\Phi(\theta; \boldsymbol{x}')$ for inputs $\boldsymbol{x}$ and $\boldsymbol{x}'$. The goal of IIC is, thus, to preserve what is in common between $\boldsymbol{x}$ and $\boldsymbol{x}'$ and to discard features that are instance-specific. The training objective of IIC is

$$\max_{\theta} \ \mathrm{MI}\big(\Phi(\theta; \boldsymbol{x}),\ \Phi(\theta; \boldsymbol{x}')\big). \tag{3.1}$$

Since Eq. (3.1) is equivalent to maximizing the predictability of $\Phi(\boldsymbol{x})$ from $\Phi(\boldsymbol{x}')$ and vice versa, the representations of paired data samples are favored to be the same. Therefore, network training with IIC results in the desired preservation of common information. If Φ has a small output capacity, which for clustering typically is the case, since the representation space is a finite set of class assignments, Eq. (3.1) also fulfills the second desired property of discarding instance-specific information irrelevant to the classification task. The networks used for IIC perform soft clustering, i.e. the last network layer is a softmax activation, resulting in network outputs being interpretable as class probabilities.

Clustering can lead to degenerate solutions, e.g. to the prediction of only a single class for all samples or, conversely, to an equally likely prediction for each class. Maximizing the MI avoids both of these pitfalls as the MI expands to

$$\mathrm{MI}(z, z') = \mathrm{H}(z) - \mathrm{H}(z|z'). \tag{3.2}$$

As the authors of [Ji et al., 2019] explain, maximizing MI compromises minimizing the conditional cluster assignment entropy $\mathrm{H}(z|z')$ and maximizing the individual cluster assignment entropy $\mathrm{H}(z)$. While $\mathrm{H}(z)$ reaches its maximum $\ln(C)$ (for a clustering into C classes) when all classes are predicted with equal probability, the predictability of $\Phi(\boldsymbol{x})$ from $\Phi(\boldsymbol{x}')$ is highest when class assignments are always the same ($\mathrm{H}(z|z') = 0$) and one-hot, i.e. the predicted cluster is predicted with probability 1. The maximization of the individual cluster assignments, thus, favors mass equalization between classes, whereas the minimization of the conditional cluster assignment entropy favors one-hot predictions.

3.2.1 IIC Extensions for OCT Image Segmentation

IIC has been successfully used for unsupervised image segmentation by using neighboring image patches as paired data samples. Still, maximizing the MI for neighboring

image patches that are not in the same object, that is they do not belong to the same semantic label, is not actually a suitable objective function but only works in expectation over images and patches. Therefore, a different approach is chosen here for OCT segmentation and paired patches are generated by random augmentations of the same patch. Furthermore, to avoid inhomogeneous clusters, a second term is added to the loss function that pushes spatially close patches to similar encodings, while it allows distant patches to be clustered independently. Based on [Huang et al., 2014; Aljalbout et al., 2018], the locality preservation loss (LP loss) is defined as

$$\mathcal{L}_{\mathrm{LP}}(\theta; \boldsymbol{x}_i, \boldsymbol{x}_j) = \omega\left(\boldsymbol{l}_i, \boldsymbol{l}_j\right) \cdot \mathrm{D}(\boldsymbol{x}_i, \boldsymbol{x}_j) \cdot ||\boldsymbol{z}_i - \boldsymbol{z}_j||_2^2 \tag{3.3}$$

for two patches $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ from the same image volume at locations $\boldsymbol{l}_i$ and $\boldsymbol{l}_j$. The encoded representations $\boldsymbol{z}_i$ and $\boldsymbol{z}_i$ are feature maps generated within the IIC network when passing $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ through the network. In the following section, an ablation study will be performed to analyze the influence of the selected network layer from which $\boldsymbol{z}_i$ and $\boldsymbol{z}_i$ are selected. The weighting function ω assigns small distances between the patch locations a large weight and sets large distances to zero. Spatially close patches are, thus, pushed to similar representations. While this is similar to the application of IIC on neighboring image patches, here, an image similarity measure D is used that reinforces locality preservation for visually similar image patches and suppresses it for dissimilar image patches. The full loss function for OCT segmentation is given by

$$\mathcal{L}(\theta; \boldsymbol{x}_i, \boldsymbol{x}_j) = \mathcal{L}_{\mathrm{IIC}}(\boldsymbol{x}_i, \boldsymbol{x}_i') + \mathcal{L}_{\mathrm{IIC}}(\boldsymbol{x}_j, \boldsymbol{x}_j') + \mathcal{L}_{\mathrm{LP}}(\boldsymbol{x}_i, \boldsymbol{x}_j) + \mathcal{L}_{\mathrm{LP}}(\boldsymbol{x}_i', \boldsymbol{x}_j') \tag{3.4}$$

with

$$\mathcal{L}_{\mathrm{IIC}}(\boldsymbol{x}, \boldsymbol{x}') = -\mathrm{MI}\big(\Phi(\theta; \boldsymbol{x}),\ \Phi(\theta; \boldsymbol{x}')\big) \tag{3.5}$$

and $\boldsymbol{x}'$ being a random augmentation of $\boldsymbol{x}$.

In the following two sections, depending on the application, data preprocessing and the choice of network architecture varies due to the differing dataset characteristics. For wound segmentation, dense 3D volumes are analyzed whereas for retinal OCT images, 2D analyses are preferred to handle large inter-slice distances. The implementation details per application are, therefore, described in the respective upcoming sections. In all applications, $\mathrm{D}(\boldsymbol{x}_i, \boldsymbol{x}_j) = 1 - \mathrm{MSE}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is used as the image similarity measure and $\omega(\boldsymbol{l}_i, \boldsymbol{l}_j) = \exp(-||\boldsymbol{l}_i - \boldsymbol{l}_j||_2^2/d)$ for the weighting function, where d is set to 0.01. This means that positions that lie within a circle with a radius of approx. 3% of the image width or height are assigned a weight of more than 0.9, while at a distance of more than a quarter of the image width/height the LP loss has practically no influence (weight ≤ 0.001).

3.3 Wound Segmentation in OCT Images

Cutaneous wounds are defined as a disruption of the normal anatomic structure and function of the skin caused by an injury. Natural wound healing is a complex process

consisting of overlapping phases hemostasis, inflammation, tissue formation and tissue remodeling. Evaluating wound healing processes is a very difficult task due to the high number of factors that influence the course of individual wounds, such as wound size and depth, the cause of the injury, the environmental conditions and also the age and state of health of those affected. Typically, wounds are assessed by visual observation and surface measurements [Wysocki, 1996; Deegan et al., 2018]. However, these assessments are heavily dependent on the experience of the treating physicians and do not provide any information about the structural changes in the deeper tissue layers [Sullivan et al., 2004]. The Vancouver scar scale and the Manchester scar proforma are attempts to quantify such measurements, but to date the best objective measurement of wound recovery are biopsies [Deegan et al., 2018]. These, however, are invasive, leading to additional tissue destruction and increasing the risk of infection [Yuan et al., 2010].

Given that, there is currently no standardized procedure for the development of ointments or medications to support wound healing. OCT offers a noninvasive and noncontact tool to visualize the tissue layers of the skin and is used in dermatology to evaluate skin tumors, inflammatory skin diseases and various other skin changes [Welzel, 2001; Wan et al., 2021]. Therefore, the use of OCT is being investigated as a promising way to objectively assess wounds without having to perform a biopsy. For example, OCT has successfully been used to monitor wound healing in mice and human skin [Kuck et al., 2013; Greaves et al., 2014; Yousefi et al., 2014], to stage burns [Park et al., 2001] and to observe scar formation [Greaves et al., 2015; Ghosh et al., 2020].

To date, only a few works exist that tackle the segmentation of skin structures such as the epidermis, dermis, vessels, hair follicles or subcutaneous fat on OCT images and even less address the segmentation of cutaneous wounds. Skin layer segmentation of normal skin has been done with intensity-based image analysis [Hori et al., 2006], graph-based methods [Srivastava et al., 2018] and convolutional neural networks [Calderon-Delgado et al., 2018; Kepp et al., 2019; Del Amor et al., 2020]. A semi-automatic segmentation is used to assess tissue injury depth in [Deegan et al., 2019]. [Ji et al., 2022] use five different deep neural networks to segment the epidermis and scab, and wound quantification based on skin layer segmentation was done in [Kumar et al., 2023]. Here, a U-Net ensemble was trained supervised to segment the epidermis, dermis and subcutaneous tissue. The extent of the wound is determined by the edges of the intact epidermis, allowing to establish a wound healing progression graph.

Despite the possibility to quantify wound extension and skin layers on OCT, the evaluation of treatment effects is complicated by a lot of external factors. Skin equivalents offer the possibility to induce wounds standardized in shape and depth and to monitor these wounds in a controlled setting, as for example done in [Yeh et al., 2004; Xie et al., 2010]. In this controlled environment, it is possible to isolate the effects of different treatments. This could advance and improve future research into wound healing. The goal of this section, thus, is to develop an automatic method to segment wounds in human skin equivalents. Since hardly any manual annotations are available for the given dataset, unsupervised clustering with the previously described extension of IIC is used.

3.3.1 Data Preprocessing

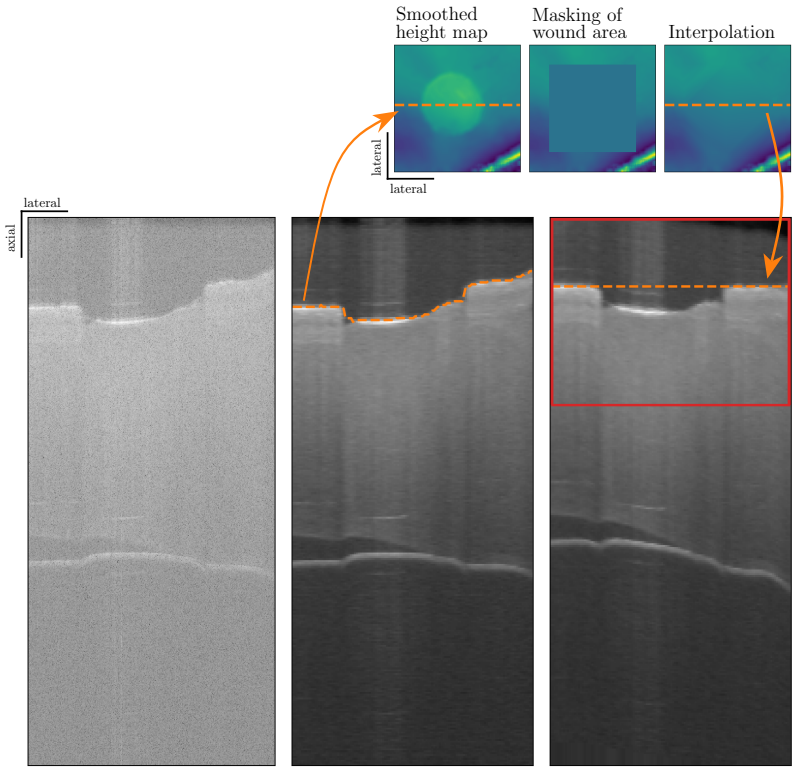
The wound dataset (c.f. Sec. 2.4.1) contains images from 24 artificially introduced wounds in human skin equivalents monitored over one week. It comprises 144 OCT images with a field of view of $3.5 \times 3 \times 3$ mm^3 and an image size of $1024 \times 461 \times 461$ voxels. For six wounds, manual wound delineations are given in the lateral dimension. The OCT images exhibit a poor contrast and are, therefore, preprocessed using average-smoothing on each horizontal slice with a window size of 11×11 pixels. To reduce storage requirements, the images are resampled to a resolution of 291×128×128 voxels. The next preprocessing step is a flattening of the upper skin border to a fixed height. Flattening is performed by

1. calculating the image gradient magnitude
2. extracting the maximum value of the gradient per A-scan and entering the values in a height map
3. smoothing the height with a mean filter of 5×5 pixels size
4. masking out the wound area in the height map using a square-shaped mask that has been defined manually such that it covers all wounds in the dataset entirely
5. filling the masked area via bilinear interpolation
6. shifting each A-scan according to the filled-in height map such that the maximum gradient position (corresponding to upper skin border) is shifted to a pre-defined height.

This procedure assures that the flattening is guided only by the intact skin and not the wounds. Finally, images are cropped to the upper 96×128×128 voxels, discarding image parts not showing the skin. An overview of the preprocessing steps is given in Fig. 3.1.

3.3.2 IIC Network Architecture and Implementation Details for Wound Segmentation

To use IIC for unsupervised segmentation of wound extension in lateral direction, patches of the cropped OCT images are assigned to one of two clusters corresponding to wound and healthy skin. Each patch extends over the entire height of the image while covering a small square area in the lateral direction, which requires a network architecture tailored towards the column-like shape of the inputs. The exact architecture is shown in Fig. 3.2 exemplarily for input patches of size $96 \times 11 \times 11$. In the first network layers, $3 \times 3 \times 3$ convolutions are performed, whereas in deeper layers 3×1×1 kernels are used, performing convolution only in axial direction. The number of full 3D or axial convolutional layers depends on the size of the input patches. As long as the resulting feature maps are larger than 3×3 voxels in the lateral direction, 3D convolutions are performed and then axial



(a) Original image **(b)** Average smoothing and subsampling **(c)** Flattening and cropping

Figure 3.1: Overview of the preprocessing steps for wound segmentation via IIC. Average smoothing is performed on each horizontal slice before subsampling. The healthy skin is then flattened to a fixed height by analyzing the gradient magnitude of the preprocessed images to extract the upper skin border (dashed orange line in Subfig. b)). The skin borders are extracted per B-Scan and combined to a height map in the lateral plane. The wound area is generously masked out and then interpolated so that the processed height map is only based on the healthy skin. The difference between height map and desired position of the healthy skin is used to shift each A-Scan vertically. Finally, images are cropped, as indicated by the red box in Subfig. c).

convolutions. The network uses four convolutional blocks, which means that patches with a size of 96×3×3 to 96×19×19 voxels can be used as input, with the output of the last convolution layer having a size of 4×3×3 voxels for all sizes. After the convolutional backbone, the resulting feature maps are flattened into a single vector and concatenated with the position vector of the respective input patch. Since input patches span the entire height of the images, position vectors are 2D, indicating the patch position in the lateral plane. Finally, three FC layers process the enlarged feature vector and perform the classification into wound and healthy. A second output head is used to allow for auxiliary overclustering [Caron et al., 2018; Ji et al., 2019] generating ten instead of two clusters.

Network training is performed with a six-fold cross-validation, whereby each fold uses 120 images of 10 skin equivalents for training and the remaining 24 images for testing.

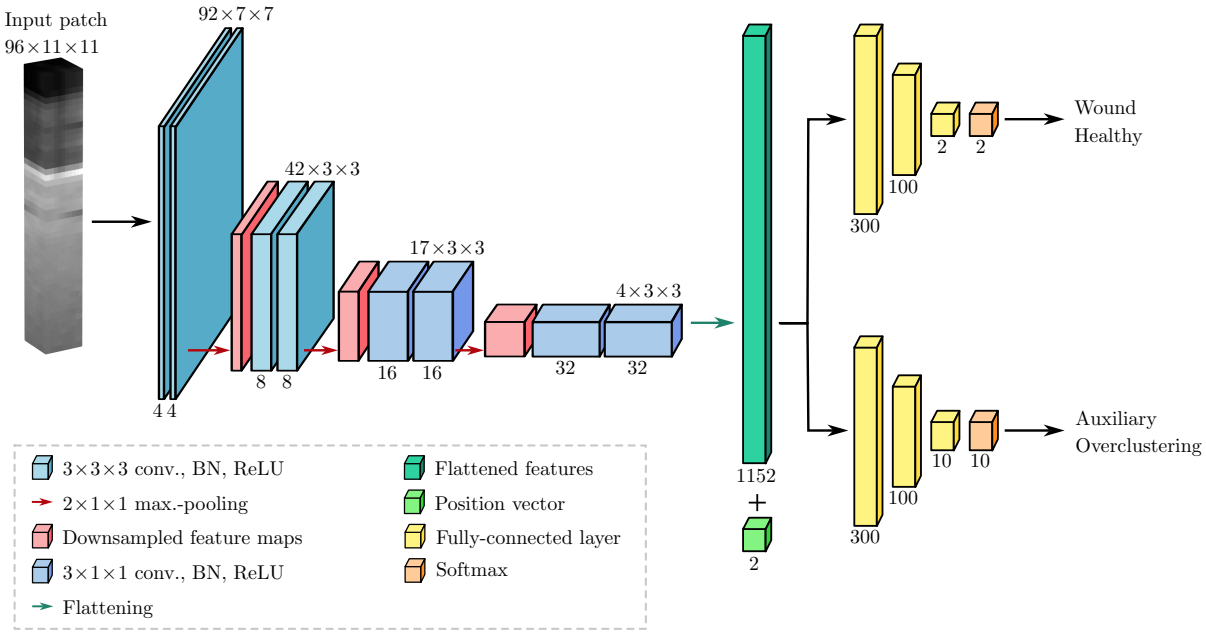


Figure 3.2: Network architecture for OCT wound clustering. In the first two convolutional blocks, $3 \times 3 \times 3$ kernels are used, in the following two blocks convolution is performed in axial direction only using $3 \times 1 \times 1$ kernels. The resulting feature maps are flattened and passed to two output heads, performing binary clustering and auxiliary overclustering. Numbers above convolutional layers indicate the size of the feature maps, whereas numbers below give the number of feature maps. Numbers below FC layers describe the number of output neurons of the respective layer.

To allow quantitative evaluation, each test set contains exactly one of the six segmented wounds. Training is performed for 200 epochs, using Adam optimization, a learning rate decay of 0.9 every 20 epochs and an initial learning rate of 1e-6. During training, the image volumes are augmented using random flipping and scaling. Scaling is used to prevent the network from concentrating solely on the upper border of the imaged tissue. Patches are augmented by randomly applying Gaussian noise, intensity shifts and intensity scaling. For each image, 150 patch pairs are randomly drawn in each epoch, using a uniform distribution across the lateral dimension. The two patches are augmented independently of each other, and the four resulting patches - two original and two randomly modified - are passed through the network. At inference, only original image patches serve as network input and for the generation of dense image segmentations, the clustering results are kept only for the central A-scan in each patch, which is assigned the cluster with the highest output probability.

3.3.3 Experiments and Results

Several ablation studies are performed to analyze the influence of input patch size, the network level on which locality preservation is applied and the proposed contrastive learning setup. First, the clustering network is trained using only the clustering loss for training, i.e., no locality preservation is applied, once with the proposed contrastive learning-like approach (IIC_{cont}) and once with neighboring image patches as proposed in [Ji et al., 2019] (IIC_{Ji}). This means that the loss function used for IIC_{cont} is $\mathcal{L}(\theta; \boldsymbol{x}_i, \boldsymbol{x}_j) = \mathcal{L}_{\text{IIC}}(\boldsymbol{x}_i, \boldsymbol{x}_i') + \mathcal{L}_{\text{IIC}}(\boldsymbol{x}_j, \boldsymbol{x}_j')$ for two randomly selected patches $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ from the same image volume and for IIC_{Ji} $\mathcal{L}(\theta; \boldsymbol{x}_i, \boldsymbol{x}_j) = \mathcal{L}_{\text{IIC}}(\boldsymbol{x}_i, \boldsymbol{x}_j) + \mathcal{L}_{\text{IIC}}(\boldsymbol{x}_i', \boldsymbol{x}_j')$ for neighboring image patches $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ with random overlap. For IIC_{Ji}, the augmentations applied are the same for patches from the same image, while for IIC_{cont} these are selected independently for each patch. Apart from this, all other hyperparameters such as the learning rate, the number of training steps, the number of patches used per epoch, etc. are retained.

In the second ablation study, the locality preservation loss is applied on different levels of the network. Precisely, the LP loss is applied to the outputs of the convolutional backbone (LP-Conv) and of the first, second and last fully-connected layer (LP-FC1, LP-FC2 and LP-FC3). Third, different input patch sizes are used for the training of the proposed framework using the best-performing version from the first ablation study (LP-Conv). Due to the different patch sizes, different sized border areas of the images cannot be classified. For a fair comparison, the results of all experiments are, therefore, evaluated only in the central image area that is covered by all variants. Finally, the auxiliary overclustering head is shown to provide useful subclusters that can be used to improve over the main head's clustering results. For this, cluster labels corresponding to the wound are found automatically using the manually annotated test images. For each test image and the corresponding output of the overclustering head, all 10 labels

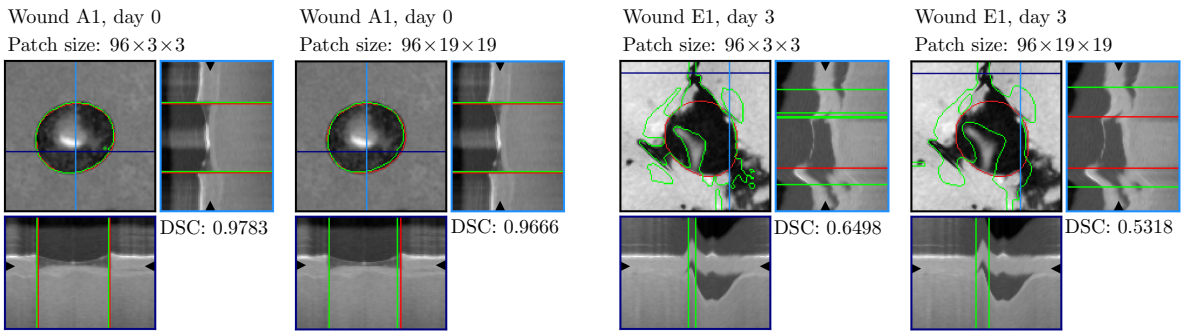


Figure 3.3: Exemplar clustering results using patch sizes $96 \times 3 \times 3$ and $96 \times 19 \times 19$ to segment wounds A1 and E1. Wound A1 is generally well delineated, whereas wound E1 is one of the worst performing examples. The contour of the manually segmented wound is shown in red and the predicted contour in green. The positions of the depicted image slices are indicated with blue lines and black triangles. Results are similar irrespective of the used patch size, but smaller patches tend to locate the wound border more precisely, whereas larger patches deliver more consistent results.

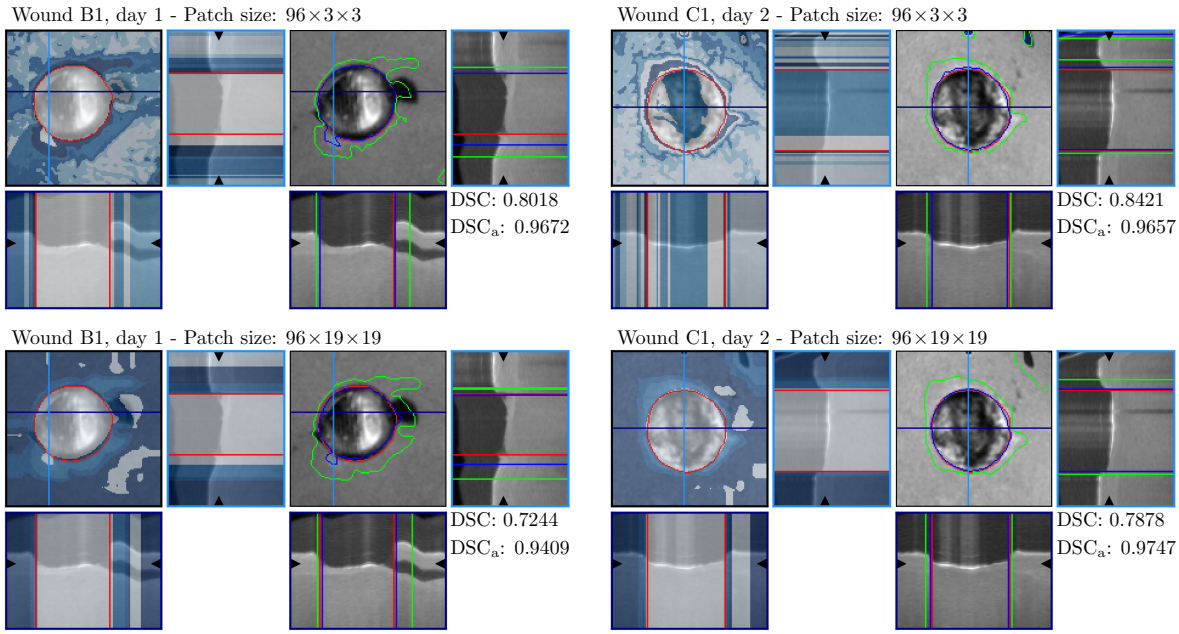


Figure 3.4: Exemplar clustering results using patch sizes $96\times3\times3$ and $96\times19\times19$ to segment wounds B1 and C1. For each wound and patch size, the clustering result of the auxiliary overclustering head is overlaid on the wound image on the left, with different shades of blue indicating different cluster assignments. On the right side, the result of the main output head is shown with green segmentation contours and the result of the overclustering with blue contours. The ground truth segmentation contour is shown in red. DSCs for main and overclustering head are indicated with DSC and DSC_a. Auxiliary clusters deliver more precise wound delineations than the main head.

are considered, and the number of pixels that overlap with the segmented wound is determined for each label. A label is included in the list of possible wound labels if more than 75% of the label's pixels lie within the wound. If this is not the case for any label, the label with the highest number of wound pixels is included in the list. After analyzing all six manually segmented test images of a fold, the labels that appear at least four times in the list are retained and used to generate binary segmentations for the six images. This procedure is repeated for each fold.

Quantitative results for unsupervised wound segmentation are given in Tab. 3.1, listing the DSC and the accuracy when comparing the predicted cluster maps with the ground truth wound segmentations (cf. Sec. 2.3). In addition, the number of images for which the clustering failed is reported. Failed clustering is defined here as cases in which all pixels are assigned to a single label. The evaluation metrics are averaged over successful clusterings only. Results show that the proposed unsupervised methods manage to separate wound and healthy skin well, with average DSCs above 72% for each configuration. The contrastive learning approach IIC_cont improves over the training with neighboring patches (IIC_cont), achieving a higher DSC of 0.7512 compared to 0.7265 (averaged over non-failure cases only), and stabilizes training. IIC_cont fails for two of the 36 test images and shows much higher variance in the metrics, whereas IIC_cont

Table 3.1: Quantitative results for unsupervised wound clustering. In the upper part of the table, the proposed IIC_{cont} is compared against IIC_{Ji}. Two ablation studies are performed, comparing performances when applying the LP loss on different levels of the network (second part of the table) and when using differently sized input patches (third part of the table plus last row of the second part). Finally, for the networks trained with IIC_{cont} and LP applied to the output of the last convolutional layer, results are additionally reported for the auxiliary overclustering head (marked with subscript "aux"). For each configuration, DSC and accuracy are reported averaged over all test images for which clustering was successful. Best results per ablation are underlined, and the best overall results are given in bold font.

Method	Patch size	DSC ↑	acc ↑	Failed ↓
IIC_{Ji}	96×11×11	0.7265 (0.1670)	0.7975 (0.1674)	2
IIC_{cont}	96×11×11	<u>0.7512</u> (0.1035)	<u>0.8045</u> (0.0931)	0
IIC_{cont} + LP-FC3	96×11×11	0.7512 (0.1038)	0.8043 (0.0934)	0
IIC_{cont} + LP-FC2	96×11×11	0.7375 (0.1024)	0.7907 (0.0949)	0
IIC_{cont} + LP-FC1	96×11×11	0.7429 (0.0989)	0.7971 (0.0902)	0
IIC_{cont} + LP-Conv	96×11×11	<u>0.7589</u> (0.1053)	<u>0.8120</u> (0.0940)	0
IIC_{cont} + LP-Conv	96×3×3	<u>0.8322</u> (0.0811)	<u>0.8833</u> (0.0615)	0
IIC_{cont} + LP-Conv	96×5×5	n/a	n/a	36
IIC_{cont} + LP-Conv	96×7×7	0.7743 (0.1057)	0.8275 (0.0924)	0
IIC_{cont} + LP-Conv	96×9×9	0.7680 (0.1662)	0.8096 (0.1722)	0
IIC_{cont} + LP-Conv	96×13×13	n/a	n/a	36
IIC_{cont} + LP-Conv	96×15×15	0.7903 (0.0956)	0.8452 (0.0779)	0
IIC_{cont} + LP-Conv	96×17×17	0.7749 (0.0983)	0.8300 (0.0830)	0
IIC_{cont} + LP-Conv	96×19×19	0.7904 (0.0935)	0.8465 (0.0752)	0
IIC_{cont} + LP-$Conv_{aux}$	96×3×3	0.8728 (0.1350)	0.9389 (0.0542)	0
IIC_{cont} + LP-$Conv_{aux}$	96×5×5	0.8642 (0.0927)	0.9324 (0.0395)	0
IIC_{cont} + LP-$Conv_{aux}$	96×7×7	0.8645 (0.1193)	0.9323 (0.0499)	0
IIC_{cont} + LP-$Conv_{aux}$	96×9×9	0.7729 (0.2005)	0.8934 (0.0776)	0
IIC_{cont} + LP-$Conv_{aux}$	96×11×11	0.8979 (0.0776)	0.9466 (0.0383)	0
IIC_{cont} + LP-$Conv_{aux}$	96×13×13	0.8767 (0.1187)	0.9400 (0.0470)	0
IIC_{cont} + LP-$Conv_{aux}$	96×15×15	0.8811 (0.1179)	0.9415 (0.0485)	0
IIC_{cont} + LP-$Conv_{aux}$	96×17×17	0.8856 (0.1167)	0.9533 (0.0488)	0
IIC_{cont} + LP-$Conv_{aux}$	96×19×19	**0.9063** (0.0711)	**0.9504** (0.0350)	0

delivers meaningful results for all test images. Moreover, the LP loss helps to generate semantically meaningful clusters, improving slightly over networks that were trained using IIC only when applied to the output of the last convolutional layer (average DSC of 0.7589). Compared to the results published in [Andresen et al., 2022b], this effect is less pronounced here, which might be explained by an improved pre-processing of the images carried out for this thesis. Still, as observed previously, results tend to be better when locality preservation is applied on earlier network layers.

The size of the input patches has a greater influence, though no clear trend can be observed for the main head. While best results are achieved for the smallest patch

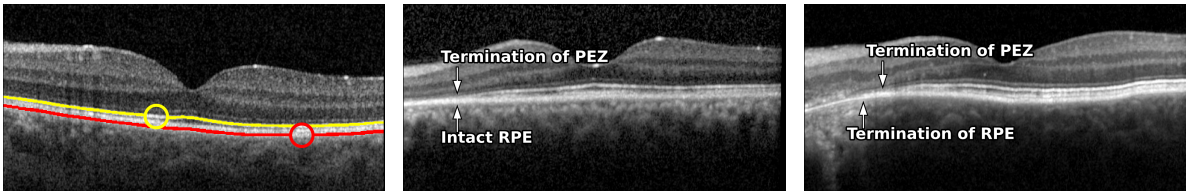
size (DSC of 0.8322), second-best results are achieved when using the largest patch size (DSC of 0.7904). Small patches seem to help to delineate the wound edges more precisely, whereas the context provided by the larger patches leads to smoother and more consistent segmentations, as shown exemplarily in Fig. 3.3. Generally, results are more stable if the same kernel sizes are used within the convolutional blocks, i.e. if the same sizes are used in the two subsequent convolutions between the maximum pooling layers. Mixed kernel sizes are used for patch sizes 5, 9, 13 and 17 in lateral direction and for sizes 5 and 13, all the cross-validation networks fail to separate wound from healthy skin, predicting only a single label for all input patches.

Despite these failure cases for the main head, the overclustering head still generates useful results for all considered patch sizes. In fact, the auxiliary overclustering delivers much better results than the clustering with only two classes. The best performing method here uses patch size $96 \times 19 \times 19$ and delivers an average DSC of 0.9063 and accuracy of 0.9504. Exemplary results comparing the outputs of main and overclustering head are shown in Fig. 3.4. It can be seen that the higher number of classes allows the network to separate different areas within the healthy skin and the wound, while aligning well with the output of the main head. Misclassifications typically occur in areas where the flattening failed. Most of these areas are caused by a partial detachment of the epidermis, a problem caused by the use of artificial skin equivalents.

3.4 Atrophy Segmentation in Retinal OCT Images

In the context of eye diseases, atrophy is an irreversible loss of retinal tissue that often occurs in the late stages of chronic diseases such as AMD or CSCR. Depending on the affected retinal layers, different atrophy phenotypes and stages are distinguished. ORA (abbreviation for outer retinal atrophy) describes atrophy of the outer retinal layers without the RPE, while RORA (abbreviation for RPE and outer retinal atrophy) describes the death of all outer retinal layers including the RPE. Further distinctions into incomplete or complete (R)ORA can be made based on the extent of the observed pathologies. Typically, the atrophy extension is evaluated on fundus autofluorescence, providing high contrast between atrophic and normal regions [Sadda et al., 2018]. The changes introduced by atrophy are less pronounced in OCT images, but the cross-sectional scans provided by OCT allow a more detailed analysis of which specific retinal layers are affected by the tissue loss. Fig. 3.5 shows an example for both ORA and RORA. In ORA, the RPE remains intact but an absence of external limiting membrane (ELM), ellipsoid zone (EZ) and interdigitation zone (IZ) can be observed. The disappearance of the RPE layer in RORA makes the underlying choroid appear brighter on OCT images. In both ORA and RORA, a thinning of the outer retina can be observed.

As soon as atrophic processes affect the foveal center, severe vision loss occurs. Thus, atrophy is a crucial biomarker for the treatment of chronic eye diseases. Its pathophysiology, however, is not yet fully understood. The close observation of affected eyes is a key component for the development of therapeutic approaches, as it provides an un-



(a) Healthy **(b)** ORA **(c)** RORA

Figure 3.5: A healthy photoreceptor ellipsoid zone (PEZ) and two examples from the CSCR dataset of different atrophy types. (a) The healthy PEZ and RPE (segmentations overlaid in yellow and red, respectively) appear as bright bands in OCT images. (b) In ORA, the ELM, PEZ and IZ are damaged while the RPE remains intact. (c) In RORA, also the RPE dies off, which leads to bright shadows in the choroid area on OCT images. For both atrophy examples, the atrophic region is located on the outer left side of the B-scans.

derstanding of how atrophy develops and progresses. Many existing works for atrophy segmentation on OCT images are developed on small in-house datasets, e.g. [Wu et al., 2019; Derradji et al., 2021; Chu et al., 2022; Mai et al., 2023] report that they are using images from 56, 57, 115 and 100 patients, respectively. Nonetheless, all these works use CNN networks that are trained in a supervised manner. Lachinov et al. propose projective skip-connections to directly segment atrophy extension in the en-face plane while using 3D OCT volumes as input [Lachinov et al., 2021]. Again, the network is trained in a supervised manner.

To date, only very few works exist that propose weakly supervised or unsupervised methods for atrophy segmentation. In [Ma et al., 2020], a classification network is trained to distinguish healthy and geographic atrophy images. Subsequently, multi-scale class activation maps are generated and post-processed to form atrophy segmentations. Seeböck et al. propose to train a Bayesian U-Net to segment retinal layers on healthy OCT images and to exploit the epistemic uncertainty of the network for unsupervised anomaly detection. They successfully apply this approach for atrophy segmentation [Seeböck et al., 2020].

In the following, the presented contrastive IIC approach is applied for the partially supervised segmentation of atrophic eye regions. Similar to the wound segmentation, atrophy is segmented A-scan wise, as its extension typically is measured in the en-face plane. Since fully unsupervised segmentation of retinal pathologies using IIC has proven to be infeasible in the experiments carried out for this thesis, partial supervision is introduced using manual segmentations where available. Still, making use of the clustering abilities of IIC, the method allows exploiting also the unlabeled data for network training, which would otherwise not be usable.

3.4.1 Methods

Retinal OCT images are taken with high inter-slice distances in clinical practice, making 3D evaluations infeasible. The IIC method presented in the previous section is, therefore, adapted to perform on 2D images. As with the wounds, the main focus here is on the extent of the pathology, which is why clustering is again performed on A-scan level using patches that consist of eleven A-scans as network input. The network architecture remains the same, with 3D convolutions and max pooling layers being replaced by their corresponding two-dimensional variants. For example, filters of size $3\times1\times1$ are replaced by filters of size 3×1. The additional input, which describes the position of the patch, now consists of just one value instead of two. The number of feature maps is doubled compared to the wound network. During training, random intensity changes are applied to the image patches as before online, but extended by shadow removal and contrast limited adaptive histogram equalization. Another difference to the previous section is that the predicted segmentations are post-processed by morphological operations to fill small gaps and remove narrow segmented areas. For this purpose, first a binary closing and then an opening is performed. A structural element with a width of eleven pixels is used for the closing and seven pixels for the opening.

The dataset used for atrophy clustering is a small subset of the larger CSCR dataset presented in Sec. 2.4.1. Manual atrophy annotations are available for 26 patients. For most of these patients, annotation was done for the last available visit only, except for one patient, for whom six time points were segmented. For four patients, atrophy was observed in both eyes. The number of images segmented for atrophy, thus, sums up to 35, out of which six had to be excluded as they cover a different field of view than the majority of the images. Manual annotation of atrophy was performed in two rounds by focusing on changes in the photoreceptor ellipsoid zone (PEZ). No distinction is made between ORA and RORA. For images segmented during the first round (eleven images from six patients), only the five central B-scans are annotated whereas for images that were segmented during the second round, all B-scans are annotated (24 images from 20 patients). For details on the CSCR dataset, please refer to the Tabs. A1 to A3 in the appendix. All images used for atrophy clustering have a B-scan resolution of 496×512 pixels and consist of either 25 or 49 B-scans.

The number of pixels affected by atrophic processes is very small compared to the size of the OCT image. Therefore, successful atrophy clustering requires several supporting mechanisms that make it easier for the network to recognize the relevant image areas. First, a careful data selection is necessary to ensure that only volumes in which atrophy is actually observed are used. Second, a flattening at the Bruch's membrane (BM) is performed, i.e. all A-Scans are shifted up or down such that the BM is brought to a predefined height. Since no manually revised segmentations of the retina or the BM are available for the images segmented for atrophy, a U-Net is trained for retina segmentation. For this purpose, the entire CSCR dataset is used, selecting those images for training for which manual retina segmentations are available. The architecture of the U-Net is described in Fig. A3 in the appendix, and training is performed for 200

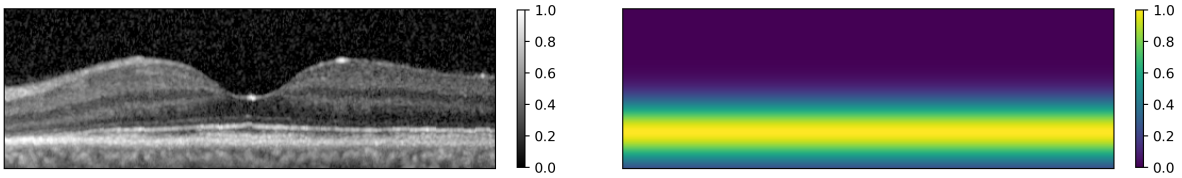


Figure 3.6: Inputs given to the atrophy clustering network: The flattened and cropped OCT image, as well as a weight map highlighting the PEZ region in the retina.

epochs using the Dice loss function, Adam optimization, an initial learning rate of 1e-4 and a learning rate decay with factor 0.99 in each epoch. A five-fold cross-validation is performed using approximately 80% of images for training and 20% for testing. Splitting is performed on patient-level, and an average DSC of 0.9887 ($\pm$0.0627) is achieved on the test images.

The retina segmentation of the atrophy cases is performed by using all five networks and segmenting only the pixels that were marked as retina by all networks. The BM is extracted as the lower border of the retina segmentations and smoothed using an average filter with a width of 21 pixels. The smoothed BM is then used for flattening. To ensure the adequacy of the data for atrophy clustering, the smoothed images were manually reviewed by the author of this thesis, and B-scans where flattening failed were excluded. After flattening, the images are cropped to 150×462 pixels to simplify the training process by removing most of the background and by cutting off flattening artifacts at the borders of the images. Since the aim is not to distinguish between different types of atrophy, but to recognize any type of photoreceptor atrophy, most of the choroid is also cut away to prevent the bright shadows, that only occur with RORA, from distorting the results. A representative pre-processed image is shown in Fig. 3.6. Finally, to further support finding the relevant image areas, a weight map is defined that highlights the outer retina in the flattened image space (also shown in Fig. 3.6). This weight map is concatenated with the OCT B-scans, and the resulting two-channel images are used as input for the network.

3.4.2 Experiments and Results

Several versions of the clustering network are trained to assess the influence of dataset size and selected loss function. First, the network is trained in an unsupervised manner with the same loss function as for the wound clustering, using only the 29 manually annotated images. Second, the same 29 images are used for training, but a partially supervised training is performed in which the IIC and LP losses are supplemented by the binary cross entropy (BCE) loss for patches for which the ground truth label is known. For patches for which the label is not known, only the clustering loss terms are used. For comparison, a fully supervised training is also performed using only the annotated B-scans to train a classification network with the BCE loss. Apart from the removal of the overclustering head, the architecture of the network remains the same as

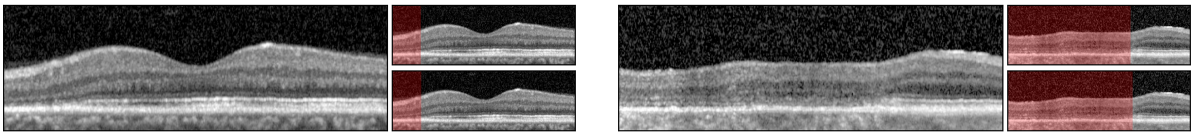
for clustering, as do the augmentations and hyperparameters used. Third, the database used for training is extended by unannotated images from the full CSCR dataset. To avoid worsening the class imbalance, the additional images are selected manually so that only images in which atrophy is observed continue to be used. 145 images are selected, out of which 68 stem from the previously used 26 patients and 77 from eleven further patients. The partially supervised training is repeated for the extended dataset. The networks are trained in a five-fold cross-validation and the results, reported in Tab. 3.2, are averaged over all test images.

Results show that despite the carefully selected training dataset, which contains only atrophy cases and only B-scans where flattening was successful, unsupervised training (IIC) only provides meaningful results in rare cases. For most images, no clustering takes place at all and the network outputs a single class, resulting in very low sensitivity and an average DSC_{proj} of 0.0662. Results for the supervised (BCE) and partially supervised (IIC + BCE) training on the small dataset are very similar. On average, the networks achieve a sensitivity of 0.76 and a DSC_{proj} of 0.71, which is remarkable given the small number of training images. The fully supervised training achieves marginally better results, but does not allow the unannotated part of the data to be utilized, which is possible with the proposed partially supervised method. When this method is used for training on the much larger dataset, the results can be improved to 0.8198 sensitivity and 0.7450 DSC_{proj}. The combination of patch clustering and patch classification benefits from the information implicitly contained in the unannotated images and leads to a good segmentation performance, while requiring only very small amounts of manually segmented images.

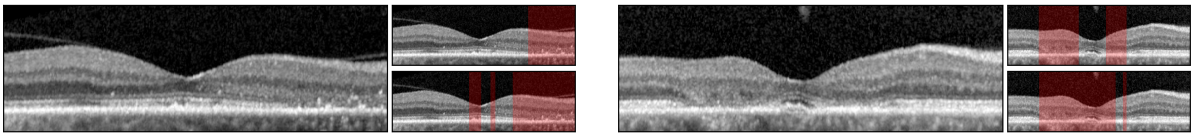
Exemplary atrophy segmentation results are shown in Fig. 3.7 for the training with the extended dataset. The examples are chosen to show the maximum achievable performance, failure cases and cases close to the average DSC_{proj} to give an impression of the overall performance. It can be seen that for images with no or small retinal fluids, the proposed algorithm manages to reliably delineate atrophic eye regions. Sometimes, changes in the photoreceptor layer that are not atrophy are incorrectly segmented (see the first example in Subfig. 3.7 (b)). Cases with additional pathologies, such as large IRF or SRF, seem to pose problems to the framework. There are two explanations for this behavior. The first is the fact that there are almost no segmented atrophy cases with severe fluid accumulation in the training data. The second is that the retinal fluids strongly deform and displace the retinal layers, which can change the appearance of the outer retina, making the problem of atrophy delineation even harder for these cases.

Table 3.2: Atrophy segmentation results using different loss combinations and datasets. All experiments using IIC loss are performed with locality preservation.

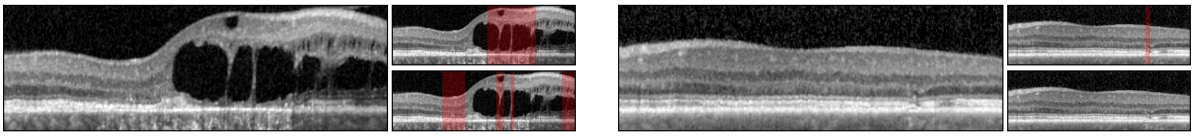
Dataset	Loss	DSC_{proj} ↑	sens ↑	spec ↑
small	IIC	0.0662 (0.1931)	0.0518 (0.1545)	**0.9744** (0.0900)
small (subset)	BCE	0.7121 (0.2331)	0.7634 (0.2793)	0.9025 (0.1207)
small	IIC + BCE	0.7076 (0.2407)	0.7601 (0.2795)	0.9034 (0.1158)
extended	IIC + BCE	**0.7450** (0.2197)	**0.8198** (0.2521)	0.8970 (0.1195)



(a) Best performing images (Dice scores of 0.9932 (left) and 0.9920 (right)).



(b) Images with performance close to the average DSC_{proj} (0.7438 and 0.7528).



(c) Failure cases (Dice scores of 0.2449 and 0.0000).

Figure 3.7: Atrophy segmentation results. For each example, the original B-scan is shown in large size and to the right of it the ground truth is shown at the top and the predicted segmentation at the bottom, both overlaid in red.

3.5 Discussion

This chapter used invariant information clustering for the unsupervised segmentation of wounds and the partially supervised segmentation of photoreceptor atrophy in OCT images. A sophisticated network architecture was used to exploit the information contained in entire A-scans and to measure the pathology extension in the en-face plane. Additionally, a new combination of loss functions, adding locality preservation to IIC, was proposed. It was shown that this extended loss function helps to give semantically meaningful clusters by favoring similar encodings for spatially close and similar looking image patches, while allowing distant patches to be clustered independently.

The wound images analyzed in the first part of the chapter show artificially introduced pathologies, that are largely standardized in terms of their size and depth. These can successfully be delineated in a fully unsupervised manner using IIC, achieving an average DSC of 0.91 for the best performing network configuration. Simpler approaches that do not require deep learning, such as fitting a circle or an ellipse to the height maps used to flatten the images, would certainly also be suitable for segmenting the wounds. However, these approaches would not be directly transferable to real wounds, which have more irregular shapes, while this should more easily be possible with the clustering approach presented. Furthermore, IIC is not limited to binary segmentations, but can potentially also be used to distinguish different wound healing stages, as shown by the results of the auxiliary overclustering, which were qualitatively shown to be consistent with different wound areas.

The photoreceptor atrophies observed in the second part of the chapter are much harder to segment than the wounds, resulting in the unsupervised approach to fail for these images. However, the proposed combination of partially supervised classification with IIC manages to detect atrophic regions in retinal OCT images with a promising accuracy based on only very few annotated cases. The proposed method can, thus, reduce the problem of data scarcity by allowing unannotated data to be used profitably for training. A limitation of the method is that no distinction is made between ORA and RORA. Since CSCR mainly leads to ORA rather than RORA, the manual annotation focused on PEZ atrophy and did not include RPE evaluations. The achieved segmentations might potentially be subdivided further into atrophy with and without RPE atrophy, given RPE annotations and more data to learn from. Another solution might be achieved without DL by analyzing the choroid in A-scans marked as atrophic. The choroid often appears hyperreflective when RPE atrophy occurs, whereas no change in reflectivity is observed for ORA. Since no annotations are available to date that could be used to evaluate such an analysis, this idea is not pursued further in this work.

For both applications, large annotated datasets are missing to investigate what performance can be achieved when supervised methods are used. Also, longitudinal investigation of changes or segmentation consistency cannot be done with the given data; For the wounds, no incipient healing could be observed with the skin equivalents used and for most of the atrophy cases only one time point was segmented manually. Still, given that wound segmentation was done completely unsupervised and that the segmentation of PEZ atrophy is a very difficult task, even manually, the achieved results seem promising. IIC combined with LP and partial supervision proofed especially useful for the segmentation of small datasets and the integration of knowledge hidden in unannotated images.

Chapter 4

Unsupervised Segmentation of Non-Correspondences during Image Registration for the Analysis of Pathological Changes

Segmentation and cluster analysis methods enable precise volumetric analyses of individual images. In medicine, however, it is often necessary to analyze several images together to enable diagnosis, treatment planning or follow-up examinations. This may involve images from different points in time, e.g. before and after an operation, but also from different modalities, such as CT and MRI, or patients, e.g. for the creation of anatomical atlases. These images are usually not aligned with each other, which makes direct comparison or fusion impossible. Image registration is, therefore, one of the most important topics in medical image processing.

A major challenge in the application of registration procedures is the presence of non-corresponding regions between the images. These can include inter-individual or age-related anatomical differences, but also pathologies that often cause even greater deviations between images. For approaches that find the registration deformation field based on global intensity metrics, the resulting differences in intensity can lead to registration errors. Unrealistic deformations will be introduced that attempt to compensate for the differences by dragging voxels that are similar in intensity but do not correspond structurally into the non-correspondent area. As a result, the predicted deformation fields are unsuitable for subsequent volume change analyses and the warped images cannot be used for atlas generation or image fusion.

In this chapter, NCR-Net is presented, one of the first deep learning-based approaches for the simultaneous registration and unsupervised non-correspondence segmentation of medical images with pathologies that has been published in [Andresen et al., 2022a]. The network is trained with a loss function that uses a masked image distance measure and at the same time favors the masked area to be small and compact. Like this, non-correspondences between moving and fixed images are found as outliers in the image distance measure, as proposed in [Chen et al., 2015] for a classical image registration method. The built-in differentiation engine of modern DL libraries makes it comparatively easy to split up the optimization of the CNN parameters into several phases, or to extend the loss function by further loss terms. Here, a two-stage training procedure is proposed to facilitate the disentanglement of spatial alignment and outlier segmentation. First, only the parts of the network that are used to create the deformation field are trained, before also training the non-correspondence segmentation part of the network.

Furthermore, weak supervision is used to guide the registration task by measuring the overlap of the main anatomical structures after deformation. No expensive manual segmentations of pathologies are required, and the method is additionally shown to perform similarly when trained in a completely unsupervised manner.

The remainder of the chapter is organized as follows. Sec. 4.1 gives an overview of the literature tackling the problem of registering medical images with pathologies. Sec. 4.2 presents the objective function, training, architecture and other implementation details of the proposed method. In Sec. 4.3, NCR-Net is used for the 2D registration of retinal OCT images from AMD patients (4.3.1) and for the registration of brain MRIs with artificially inserted lesions, both in 2D (Sec. 4.3.2.1) and 3D (Sec. 4.3.2.2). Performances are compared to several state-of-the-art registration frameworks. Additionally, the non-correspondence segmentations generated by NCR-Net are shown to be usable for the unsupervised segmentation of pathologies in Secs. 4.3.1.2 and 4.3.2.3.

4.1 Related Literature

The most intuitive solution to handle non-correspondences in image registration is to mask out or weigh down the respective areas in the image similarity measure during the registration process, e.g. [Brett et al., 2001; Nachev et al., 2008; Xie et al., 2021; Schaffland and Schöning, 2024]. This requires that non-correspondent image regions be segmented before registration. While the segmentation of pathologies already is cumbersome, the segmentation of non-correspondences is often impossible, since the unaligned images cannot be compared directly. For example, the distinction of existing pathology and developing new pathological tissue is very difficult without exact spatial alignment.

Several iterative approaches exist that detect non-correspondent image regions during the registration process [Periaswamy and Farid, 2006; Chitphakdithai and Duncan, 2010; Ou et al., 2011; Chen et al., 2015; Werner et al., 2016; Krüger et al., 2019, 2020]. Chen et al. use a sophisticated optimization functional to both segment non-correspondences and find the registration deformation field [Chen et al., 2015]. Non-correspondences are found as areas with large image distances, which are masked out and favored to be small and have a smooth boundary. In [Ou et al., 2011], DRAMMS is introduced, a registration method in which continuously valued mutual-saliency maps are used to weigh the influence of individual voxels onto the deformation. This is achieved by extracting attribute vectors reflecting the geometrical and anatomical context around each voxel and assigning higher weights to voxels that establish unique correspondences across images. Like this, outliers are given less weight. Similarly, Krüger et al. weigh the image distance during registration based on correspondence probabilities rather than assuming one-to-one correspondences [Krüger et al., 2019]. These probabilities are estimated based on sparse image representations that describe shape and appearance features of the images.

Recently, with the rise of DL-based image registration frameworks, the problem of aligning non-correspondent images has gained more attention. Sedghi et al. use a classifier network to patch-wise predict class probabilities for either registered, unrelated or 18 different transformations [Sedghi et al., 2019]. The deformation field is regularized by setting it to zero in patches for which the unrelated class is predicted. While this method allows to estimate non-correspondences directly, it only works on patch-level and relies on an iterative scheme. The method described in the upcoming section and published in [Andresen et al., 2022a] was at the time one of the first works to tackle the joint estimation of non-correspondences and dense deformation fields with DL methods. It builds on the idea to find non-correspondent image areas as outliers in the image distance measure described in [Chen et al., 2015]. The objective function proposed there is transferred into a differentiable loss function usable for CNN training, and the functionality of modern CNNs is exploited for improved shape and appearance disentanglement.

In 2022, the BraTS-Reg Challenge was hosted with the goal to register pre- and post-operative brain MRIs of glioma patients [Baheti et al., 2024]. Most challenge submissions rely on DL to solve this task [Großbröhmer et al., 2023; Mok and Chung, 2023; Wodzinski et al., 2023; Meng et al., 2023; Almahfouz Nasser et al., 2023; Zeineldin et al., 2023; Abderezaei et al., 2023]. Oftentimes, prior semantic information is given to the registration networks by either predicting tumor segmentations [Großbröhmer et al., 2023; Almahfouz Nasser et al., 2023] or using the manually annotated landmarks provided with the data [Meng et al., 2023]. [Mok and Chung, 2023] and [Wodzinski et al., 2023] use LapIRN [Mok and Chung, 2020] as baseline model to explicitly estimate the areas of non-correspondence. For this purpose, baseline and follow-up are passed through the network twice, with each image being used once as a fixed and once as a moving image. The displacements resulting from both runs are used to calculate the forward-backward error, which is large for non-correspondent regions. These regions are masked out in the image similarity measure of the loss function during training, similar to the method proposed in the following section.

Uncertainty estimation of CNN predictions opens a new way to find image areas where registration is more difficult and where non-correspondences occur. The first papers to perform such studies use Monte Carlo dropout and visualize the uncertainty after the network has been trained [Yang et al., 2017; Sentker et al., 2018]. More recent works let the networks predict uncertainty directly to improve or guide the network training [Gong et al., 2022; Xu et al., 2022; Chen and Yang, 2025]. Xu et al. use a mean-teacher model that allows to dynamically adjust the weighting parameter of the spatial regularization for individual registration problems based on the estimated uncertainty [Xu et al., 2022]. Chen et al. propose to couple a segmentation and a registration network, which are both trained to perform the respective task and predict the uncertainty. Both networks are first trained separately and then alternately, connecting the networks by comparing the deformed segmentation masks with those of the other images in the loss. In this way, both networks can benefit from each other, but the training requires manual segmentations [Chen and Yang, 2025].

4.2 Methods

The basic idea to achieve simultaneous image registration and non-correspondence segmentation is to extend the registration problem (2.1) from

$$\mathrm{D}\left[F, M \circ \varphi; \Omega\right] + \alpha \mathrm{R}\left[\varphi\right] \xrightarrow{\varphi} \min \tag{4.1}$$

to

$$\mathrm{D}\left[F, M \circ \varphi; \Omega \setminus \mathcal{O}\right] + \alpha \mathrm{R}\left[\varphi\right] + \beta \mathrm{Per}[\mathcal{O}] + \gamma \mathrm{Vol}[\mathcal{O}] \xrightarrow{\varphi,\, \mathcal{O}} \min . \tag{4.2}$$

While in Eq. (4.1) the distance measure is evaluated on the entire image domain Ω, the extended registration objective (4.2) requires image similarity only in $\Omega \setminus \mathcal{O}$, where $\mathcal{O}$ describes the region in F that has no correspondence in M. At the same time, to avoid the trivial solution of masking out the image distance on the entire image domain, the non-corresponding region is regularized to be smooth-edged and small by minimizing its perimeter ($\mathrm{Per}[\mathcal{O}]$) and volume ($\mathrm{Vol}[\mathcal{O}]$). Solving Eq. (4.2), thus, leads to non-correspondence regions that correspond to outliers in the image distance measure [Chen et al., 2015].

In [Andresen et al., 2022a], this approach is transferred to the field of DL. For the registration of the moving image $M : \Omega \to \mathbb{R}$ to the fixed image $F : \Omega \to \mathbb{R}$, an extended U-Net is used that consists of one encoder and two separate decoders that output a diffeomorphic deformation field $\varphi = \exp(v) : \mathbb{R}^d \to \mathbb{R}^d$ and a segmentation of non-correspondences $N : \Omega \to \{0, 1\}$. The network is trained with the loss function

$$\mathcal{L}(\theta; M, F) = \mathcal{L}_{\mathrm{MSE}}^{\mathrm{masked}}(F, M \circ \varphi; N) + \alpha \mathcal{L}_{\mathrm{Reg}}(\varphi) + \beta \mathcal{L}_{\mathrm{Shape}}(N), \tag{4.3}$$

consisting of a masked MSE loss

$$\mathcal{L}_{\mathrm{MSE}}^{\mathrm{masked}}(F, M \circ \varphi; N) = \frac{1}{n} \sum_{\boldsymbol{x} \in \Omega} \big(1 - N(\boldsymbol{x})\big) \cdot \big(F(\boldsymbol{x}) - (M \circ \varphi)(\boldsymbol{x})\big)^2,$$

with n being the number of voxels in Ω, and regularizers

$$\mathcal{L}_{\mathrm{Reg}}(\varphi) = \sum_{\boldsymbol{x} \in \Omega} \|\nabla v(\boldsymbol{x})\|_2^2$$

$$\mathcal{L}_{\mathrm{Shape}}(N) = \sum_{\boldsymbol{x} \in \Omega} N(\boldsymbol{x}) + \gamma \sum_{\boldsymbol{x} \in \Omega} \tanh\left(\Big(\big(\mathrm{G}_x * N\big)(\boldsymbol{x}) + \big(\mathrm{G}_y * N\big)(\boldsymbol{x})\Big)^2\right).$$

The image distance is masked with the segmentation output N of the CNN and the loss components of $\mathcal{L}_{\mathrm{Shape}}(N)$ differentiably approximate the volume and the perimeter of the non-correspondence segmentation. The perimeter is calculated by converting the non-correspondence segmentation into an edge image through convolution with Sobel kernels G_x and G_y, and summing its values. Inspired by [Hering et al., 2019], a multi-resolution approach is pursued, using additional output heads on different levels of the

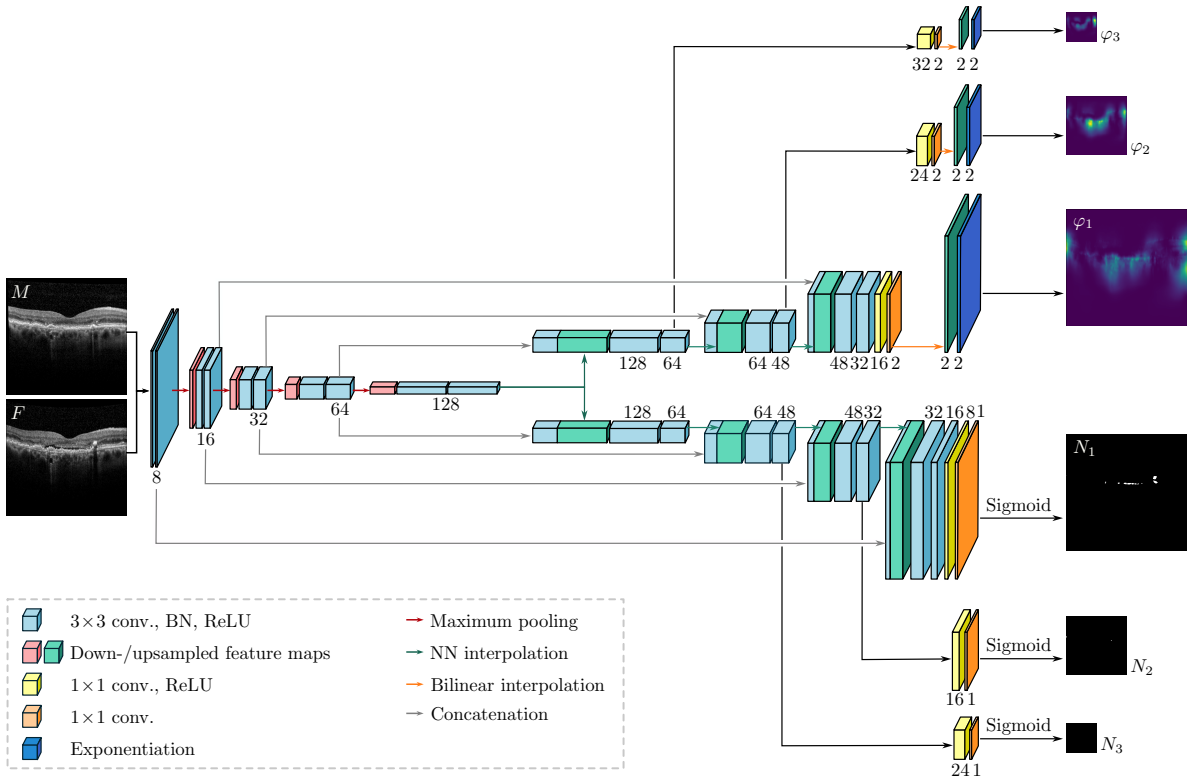


Figure 4.1: The architecture of NCR-Net. Moving and fixed images are passed as a two-channel input to the network and encoded. Two separate decoders generate deformation fields and non-correspondence maps on three levels of resolution.

network. The detailed network architecture of the non-correspondence segmentation and registration network (NCR-Net) is displayed in Fig. 4.1. Similar to the U-Net, the two decoders are connected to the encoder via skip-connections. The segmentation decoder outputs non-correspondence maps N_1, N_2 and N_3 for the original, half and quarter resolution of the input images. For the generation of diffeomorphic deformation fields $\varphi_i = \exp(v_i)$, the output heads of the deformation decoder first produce stationary vector fields v_i, which are then passed through an exponentiation layer [Krebs et al., 2018]. For the exponentiation, the implementation provided by [Sandkühler et al., 2018] is used. The deformation fields are created with a resolution of one eighth, one quarter and one half of the original resolution and then interpolated to the same resolution levels as the non-correspondence maps. This leads to a further regularization of the deformations.

Using the outputs on three resolution levels, the loss function (4.3) is evaluated for each level and combined into one loss that is used for backpropagation. Let $\mathcal{L}^i(\theta; M, F) = \mathcal{L}(\theta; M_i, F_i)$ be the loss function evaluated for downsampled moving and fixed images, denoted as M_i and F_i for resolution level i, then the final loss is given by

$$\mathcal{L} = \omega_1 \mathcal{L}^1 + \omega_2 \mathcal{L}^2 + \omega_3 \mathcal{L}^3. \tag{4.4}$$

Finer resolution levels are assigned a higher weight ω_i. For the OCT registration, $\omega_1 = 0.5, \omega_2 = 0.3$ and $\omega_3 = 0.2$ is used, whereas for the MRI registration, $\omega_1 = 0.7, \omega_2 = 0.2$ and $\omega_3 = 0.1$ is chosen empirically.

Compared to iterative approaches for image registration, CNN-based methods make it comparatively easy to add additional loss terms. This is exploited below to introduce weak supervision of the learning process by adding the Dice loss $\mathcal{L}_{\text{Dice}}(S^F, S^M \circ \varphi)$ of anatomical structures to the loss function. Pathology labels are not used for training, but only segmentations of large structures or tissues, which are much easier to obtain. For OCT applications, S^F and S^M describe the ground truth segmentations of the retina and for MR applications those of the brain.

Finally, to prevent spatial misalignments from being masked out or non-correspondences leading to registration errors, a two-step training scheme is introduced. In the first half of training, the encoder and the deformation decoder are pre-trained with

$$\mathcal{L}^{\text{pre}} = \sum_{i=1}^{3} \omega_i \cdot \Big(\mathcal{L}_{\text{MSE}}\big(F_i, (M_i \circ \varphi_i)\big) + \alpha \mathcal{L}_{\text{Reg}}(\varphi_i) + \lambda \mathcal{L}_{\text{Dice}}(S_i^F, S_i^M \circ \varphi_i) \Big) \tag{4.5}$$

as loss function. This corresponds to the "classical" registration problem (4.1), except for the label supervision. In the second half of training, the non-correspondence segmentation is enabled, and the entire network is trained with loss function (4.3), evaluated on three resolution levels and extended with the weak supervision loss term:

$$\begin{aligned} \mathcal{L}^{\text{main}} = \sum_{i=1}^{3} \omega_i \cdot \Big(& \mathcal{L}_{\text{MSE}}^{\text{masked}}\big(F_i, M_i \circ \varphi_i; N_i\big) + \alpha \mathcal{L}_{\text{Reg}}(\varphi_i) + \beta \mathcal{L}_{\text{Shape}}(N_i) \\ & + \lambda \mathcal{L}_{\text{Dice}}(S_i^F, S_i^M \circ \varphi_i) \Big). \end{aligned} \tag{4.6}$$

Introducing the non-correspondence detection only when the network is already able to perform meaningful registration helps the complicated disentanglement of non-correspondence segmentation and deformation, as shown in the upcoming experiments.

NCR-Net is implemented in the PyTorch framework and training is done for 500 epochs with Adam optimization and a learning rate of 1e-4. The weighting parameters are found empirically and set to $\alpha = 1$, $\beta = 0.005$, $\gamma = 1.2$ and $\lambda = 1$ for OCT registration. For MRI registration, parameters are set to $\alpha = 0.4$, $\beta = 0.05$, $\gamma = 0.03$ and $\lambda = 1$, and the masked MSE loss is replaced with a masked normalized cross correlation (NCC) image distance loss.

4.3 Experiments and Results

In the following, NCR-Net is used to perform longitudinal intra-patient registration of OCT B-scans of AMD patients. Several ablation studies are performed to show the influence of the proposed two-step training procedure and the weak label supervision. Applicability to a wide variety of images is shown by subsequently applying NCR-Net

for the inter-patient registration of horizontal MR image slices of the brain with artificially inserted pathologies (LPBA40 dataset with phantom stroke lesions, cf. Sec. 2.4.2). Registration performance is evaluated by comparing to several state-of-the-art registration frameworks. Finally, NCR-Net is adapted to 3D and used to register the full MR volumes. For both OCT and MRI, the non-correspondence segmentations are shown to be usable for the unsupervised segmentation of pathologies.

4.3.1 OCT Image Registration

For OCT image registration, the AMD follow-up dataset from Sec. 2.4.1 is used. The full dataset consists of 709 OCT volumes from 41 eyes of 41 AMD patients, but for network training, only images with a field of view of $2 \times 6 \times 6$ mm^3 and a resolution of 496×512×25 voxels are used. In addition, only image pairs whose recording dates are no more than five months apart are used for registration. This results in 193 image pairs from 40 patients, with the follow-up scan serving as reference image and the baseline scan as moving image to allow pathology tracking. The OCT images were recorded using eye tracking so that images from different points in time show the same section of the eye. It can, therefore, be assumed that corresponding B-scans also correspond spatially, and that differences between images are mainly due to changes in the disease. The large slice spacing makes 3D registration of OCT images obsolete, as the individual B-scans are too far apart for any deformation between the B-scans to be expected. Thus, registration is done on 2D B-scans, resulting in $913 \cdot 25 = 4825$ image registration pairs being considered in total. Five-fold cross-validation is performed, with each network trained on approximately 80% of the image data. The data is split at the patient-level so that each training effectively uses approximately 3860 2D registration problems from 32 patients.

In most cases, the B-scans are already well aligned globally by the OCT device. In rare cases, however, individual B-scans may be misaligned due to sudden eye movements. To prevent the influence of these outliers, a spatial transformer network (STN, introduced by [Jaderberg et al., 2015]) is trained that rigidly pre-registers the B-scans. The STN is trained on Gaussian-smoothed B-scans using the MSE image distance as loss function. The pre-aligned B-scans then serve as input for the proposed NCR-Net. As data augmentation during training of both the STN and NCR-Net, flipping, rotations between -7° and +7° and vertical shifting up to a quarter of the image height is performed at random. The images are cropped to the central 384 A-scans to assure that the imaged tissue spans the entire width of the images. At inference time, the images can be input at full size.

4.3.1.1 Ablation Studies

In this section, an ablation study is performed to analyze the influence of the proposed two-phase training scheme, the non-correspondence detection and of the weak supervision using the Dice loss of anatomical segmentations as additional loss term. To do so, four CNNs are trained:

1. R-Net: Registration-only network with encoder and deformation decoder architecture identical to NCR-Net but without the non-correspondence segmentation decoder, trained with loss function (4.5)

2. uNCR-Net: Unsupervised training of the non-correspondence detection and registration network using loss function (4.3), evaluated on three resolution levels

3. sNCR-Net: No pre-training is used and the network is trained with loss function (4.6), that uses weak supervision, throughout the entire training

4. NCR-Net: The proposed non-correspondence detection and registration network, trained with weak supervision and two-phase training.

The ablation study results are reported in Tab. 4.1 giving the average symmetric surface distance (ASSD) and the Hausdorff distance (HD) of the given three anatomical surfaces ILM, RPE and BM before and after registration with the different network versions. Results show that all networks manage to achieve a good alignment of the retina despite large appearance differences between time points. The registration-only network, R-Net, thereby achieves best results but at the cost of unrealistic deformations as shown in Fig. 4.2. While R-Net tends to stretch or compress retinal layers to compensate for non-correspondences, NCR-Net manages to generate smoother deformation fields in these areas, as can also be seen from the respective Jacobian determinant. However,

Table 4.1: Joint non-correspondence detection and OCT image registration results for unsupervised (uNCR-Net) and weakly supervised (sNCR-Net) training, as well as for two-phase training with weak supervision (NCR-Net). To analyze the influence of the non-correspondence detection, a registration-only network is trained additionally (R-Net). ASSD and HD are specified in pixels. Best results are written in bold font. Significantly best performing NCR-Net versions are marked with *.

	ILM		RPE		BM	
Method	ASSD ↓	HD ↓	ASSD ↓	HD ↓	ASSD ↓	HD ↓
w/o reg.	6.83 (7.51)	15.15 (14.01)	6.27 (7.28)	13.94 (12.18)	6.64 (8.42)	12.22 (12.42)
R-Net	**0.85** (1.79)	**3.48** (5.98)	**1.86** (1.82)	**7.31** (6.02)	**2.80** (4.17)	**8.48** (9.37)
uNCR-Net	1.09 (3.00)	4.42 (8.65)	2.19 (3.27)	8.16 (7.91)	3.10 (5.13)	9.10 (10.41)
sNCR-Net	1.07 (2.95)	4.37 (8.52)	2.20 (3.25)	8.21 (7.99)	3.10 (5.12)	9.14 (10.46)
NCR-Net	0.93* (2.17)	3.77* (6.90)	2.10* (2.74)	7.89* (7.10)	3.02* (4.76)	8.95 (10.00)

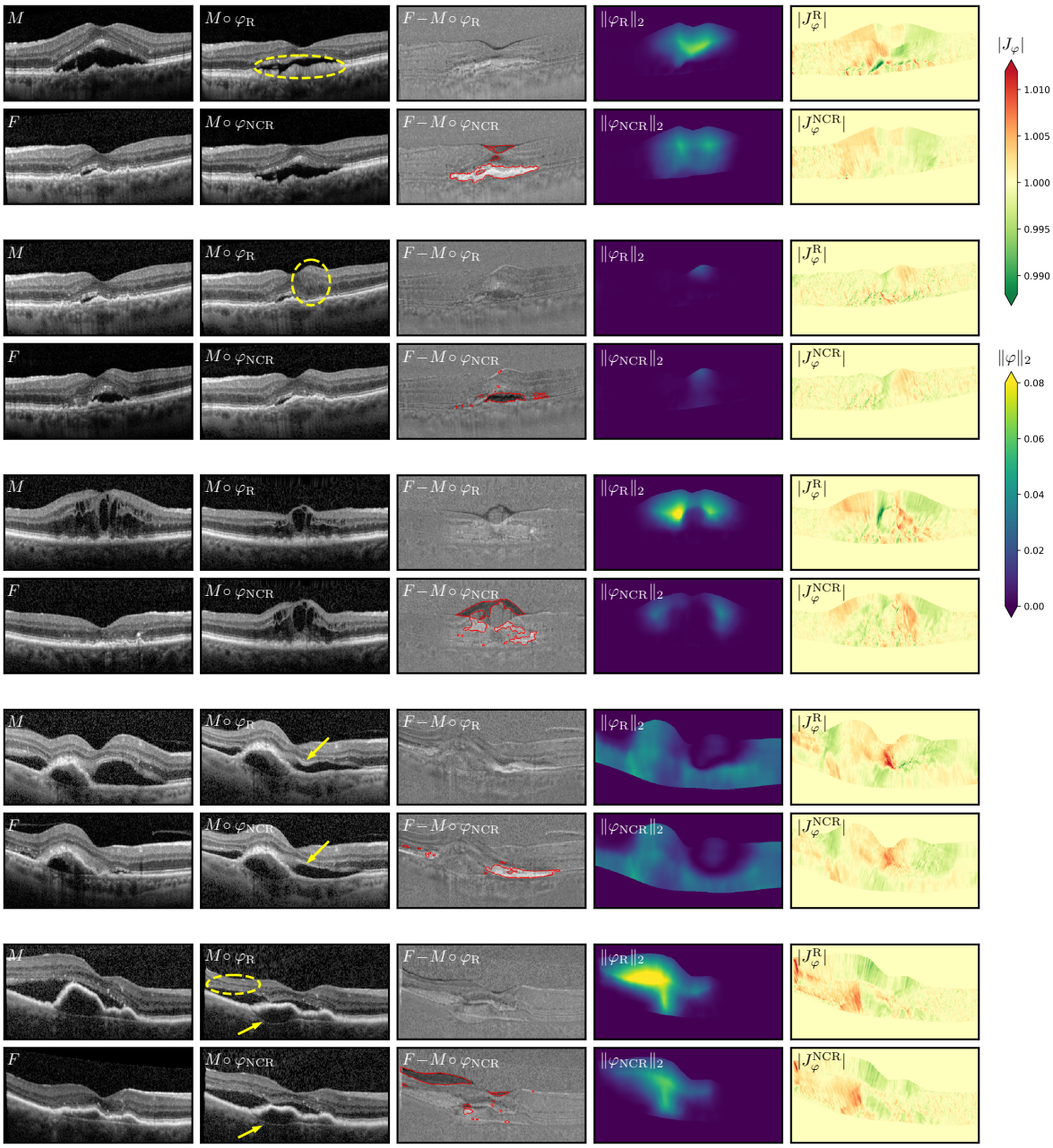


Figure 4.2: Exemplary OCT registration results of R-Net and NCR-Net. For each example, the fixed and moving images (F and M) are shown, as well as the deformed image $M \circ \varphi$ after registration with the two networks. Additionally, the difference images after registration, the amplitude and the Jacobian determinant of the deformation fields are displayed masked to the retina. For NCR-Net, the contours of the non-correspondence segmentation are overlaid onto the difference image in red. Note that the first two examples show images of three successive visits of the same patient. Areas highlighted in yellow indicate more uniform deformations predicted by NCR-Net and unrealistic stretching or compression of retinal layers introduced by R-Net.

for very large retinal deformations, masking the non-matching areas can increase registration errors. In these rare cases, NCR-Net tends to oversegment non-correspondent areas, while introducing too small deformations. This can be seen in the third example in Fig. 4.2, where both networks cannot align the ILM. Here, the masking of the non-corresponding areas leads to larger misalignments for NCR-Net.

Several observations stand out when comparing the NCR-Net versions. First, the unsupervised network version performs very similar to the supervised version (ASSDs of 1.09, 2.19 and 3.10 for uNCR-Net vs. 1.07, 2.20, 3.10 for sNCR-Net for ILM, RPE and BM, respectively), indicating that NCR-Net does not require any supervision to achieve good registration performances. Despite that, a slight improvement in the alignment of the ILM can, however, be observed, indicating that supervision can help to find far away correspondences. Second, the two-phase training significantly (one-sided Wilcoxon signed rank test with significance level 0.05) improves registration performances to 0.93, 2.10 and 3.02 ASSD for the ILM, RPE and BM. Pre-training on the registration task, thus, helps the disentanglement of spatial deformation and cost function masking.

4.3.1.2 Unsupervised New Fluid Segmentation

For longitudinal registration of retinal images, the main source for non-correspondences is disease progression, assuming that age-related changes of the retina can be ignored due to the short time interval between examinations. This can be exploited by using the non-correspondence segmentations generated by NCR-Net as pathology segmentation, as shown in Fig. 4.3 for six AMD patients. From the examples, it can be seen that NCR-Net produces very detailed non-correspondence segmentations that align well with areas of disease progression. Especially, the detection of newly emerging or disappearing fluids is enabled through outlier detection in the image distance measure.

The AMD dataset that was used for NCR-Net training includes manual segmentations of the ILM, the BM and the RPE but no delineations of IRF or SRF. A quantitative evaluation of the segmentation capacity of NCR-Net is, therefore, performed with the manually segmented CSCR dataset described in Sec. 2.4.1 by comparing the non-correspondence segmentation of NCR-Net with manual segmentations of newly emerging fluids. For the CSCR data, a good pre-alignment of the B-scans is ensured by a flattening of the input images at the Bruch's membrane. Each B-scan is passed through all five networks from the cross-validation, and the resulting non-correspondence segmentations are combined via logical OR, which gave better results than majority voting. During this evaluation, the non-correspondence maps are post-processed by masking them to the retina. The ground truth segmentations of new fluids are generated per B-scan from the manual fluid annotations as follows:

- Case 1: There is no fluid in the baseline image, but in the follow-up image.
  → All fluids in the follow-up are new fluids.
- Case 2: There are pathologies in both time points.
  → Dilation is performed on the fluid segmentations for both images. Each dilated

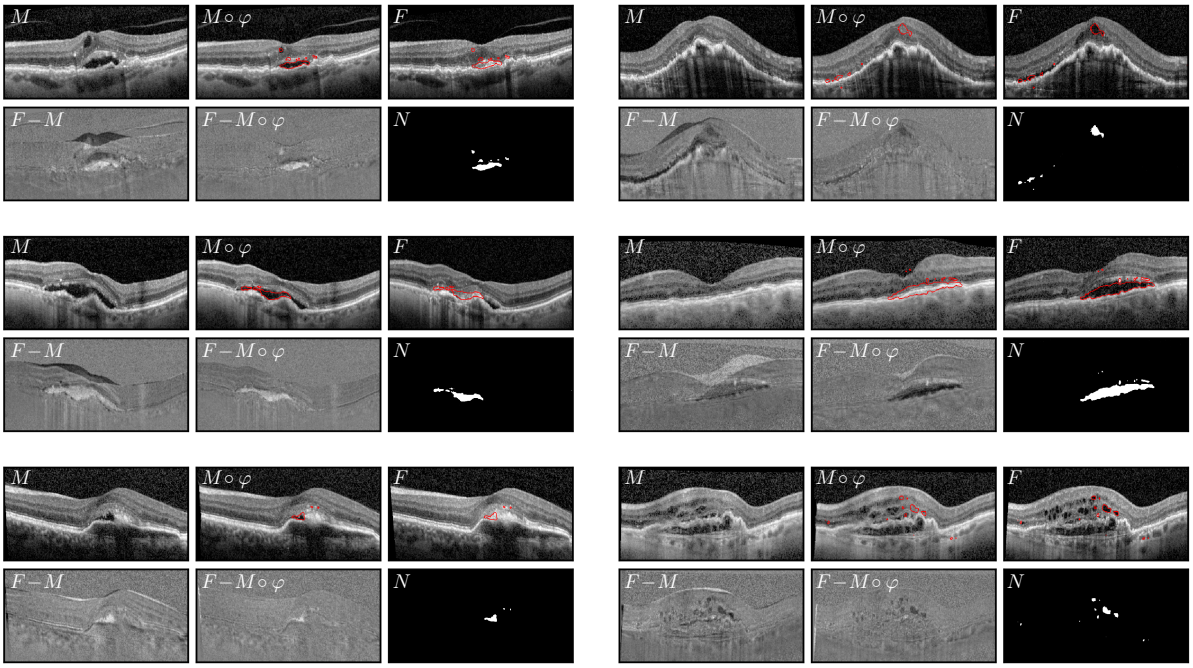


Figure 4.3: Non-correspondence segmentation results of NCR-Net align with areas of disease change. For each example, moving image M and fixed image F are shown. The registered moving image (warped baseline scan, $M \circ \varphi$) is also displayed, as well as the difference images before and after registration. The non-correspondence map N is shown on the bottom right for each example and as a red contour overlay onto fixed and warped moving images.

pathology in the follow-up examination that does not overlap with the expanded fluids in the baseline is defined as a new lesion.

- Case 3: A fluid type is present in the follow-up image that was not observed in the baseline.
  → Each pathology of this fluid type is assigned to the new fluid segmentation.

- The remaining fluids in the follow-up image are defined as "not new" and not added to the ground truth segmentation of new fluids.

In Tab. 4.2, new lesion segmentation and detection results are reported both on B-scan- and on lesion-level. The non-correspondence segmentation of NCR-Net does not distinguish between different fluid types (or other sources of non-correspondences, such as further pathologies or imaging artifacts). Therefore, results are reported for binarized ground truth fluid segmentations first. The detection rate of new fluids is reported (sens) as well as the Dice score, averaged over detected fluids (DSC_{det}). A B-scan containing new lesions or an individual new lesion is counted as detected if there is an overlap with the predicted segmentation. Second, results are reported per lesion type. To do so, each fluid class in the ground truth is considered individually, and all segmented

non-correspondent areas that overlap with the corresponding class are used to calculate the DSC.

The quantitative results show high detection rates for IRF and SRF, with sens^{img}= 29/37 = 0.78 for IRF and sens^{img}=174/212=0.82 for SRF. IRF and SRF both typically have low intensities in OCT images and high contrast to the surrounding retina. For PED, detection rates are lower, with only 19 out of 60 B-scans identified as containing new PED, which can be explained by the lower contrast to the surrounding tissue. Segmentation performance also varies with fluid type. SRF is best segmented with an average DSC_{det} of 0.57 on image- and 0.46 on lesion-level, which may be explained by the high contrast and the fact that subretinal fluids typically consist of only one or very few individual fluid inclusions below the retina. IRF, on the other hand, often consists of many small fluid inclusions and is widely scattered within the retina. The segmentation of IRF is, therefore, much more difficult than that of SRF, resulting in an average lesion-level DSC_{det} of 0.34. Finally, the comparably low-contrasted PED typically consists of only a few individual fluids that occur in the lower part of the retina and can, thus, slightly better be segmented with an average lesion-level DSC_{det} of 0.37. Overall, given that no fluid segmentations are used for network training, segmentation accuracy is high even at lesion-level, with an average DSC_{det} of 0.41 (see also anomaly detection results in Sec. 6.3.2).

Table 4.2: Using the non-correspondence maps of NCR-Net to segment new fluids in the CSCR dataset. The detection rate (sens) of new fluids is given on B-scan- and lesion-level. The DSC averaged over all detected fluids/B-scans is also reported. This evaluation is carried out for the individual fluid types and in binary form by combining all fluids in the ground truth into a single pathology class.

	All fluids		IRF		SRF		PED	
Eval. level	sens ↑	DSC_{det} ↑	sens ↑	DSC_{det} ↑	sens ↑	DSC_{det} ↑	sens ↑	DSC_{det} ↑
B-Scan	213/293	0.4579	29/37	0.4777	174/212	0.5744	19/60	0.4060
Lesion	494/955	0.4068	203/415	0.3362	269/448	0.4633	22/92	0.3676

4.3.2 MRI Registration

MRI registration is performed in 2D and 3D using the LPBA40 dataset [Shattuck et al., 2008] with artificially introduced stroke lesions from the ISLES dataset [Maier et al., 2017]. Four different lesions are inserted separately into each LPBA40 image, leading to five versions of each image (original, corrupted with lesion L1, L2, L3 and L4, that can exemplarily be seen in Fig. 4.4) with exactly known lesion location and appearance. This modified dataset is described in [Krüger et al., 2019, 2020] and used there for evaluation purposes. Here, the images are used for network training. Image registration is performed between patients, with the original LPBA40 images serving as moving images which are registered to the other patients' images, both lesion-free and corrupted

versions. For MRI registration, the NCC is used as image distance in the loss calculation. Online data augmentation includes random rotation of one of the images between -3° and +3° or of both images between -8° and +8°. In 3D, rotations are applied in the horizontal plane only. Additionally, Gaussian noise is added and random shifting in each direction by up to four pixels is performed. The 2D registration is performed on the central horizontal MR slices, where two of the four inserted lesions are small, whereas the other two corrupt large areas of the brain.

4.3.2.1 MRI Registration in 2D

In this section, NCR-Net is trained on the central horizontal slices of the modified LPBA40 dataset. First, an ablation study is performed to compare registration results of NCR-Net and R-Net as done for the OCT data. Subsequently, NCR-Net is compared to three state-of-the-art image registration frameworks.

For the ablation study, NCR-Net is trained with the proposed two-phase training scheme that performed best for the OCT data. Registration performance is measured as in [Krüger et al., 2019] by calculating the average Jaccard index for 19 anatomical labels that are present in all central horizontal scans of the training data. The results are reported in Tab. 4.3 and show that R-Net and NCR-Net perform comparable for uncorrupted images and images with small lesions L2 and L3. Since brain images generally require smaller deformations for registration than OCT images, there are no failure cases for either network. However, NCR-Net delivers much better results for large lesions L1 and L4 with average Jaccard values of 0.600 and 0.595 compared to 0.509 and 0.528 for R-Net. Here, NCR-Net benefits a lot from the non-correspondence detection, leading to plausible deformation fields also inside the pathology area. R-Net, on the other hand, tends to pull background pixels into the brain to compensate for the dark intensities inside lesions, as shown exemplarily in Fig. 4.4. There it can also be seen that R-Net completely removes the cerebellum visible in the moving image only, while NCR-Net better preserves such normal inter-individual differences.

To evaluate the registration accuracy of NCR-Net compared to state-of-the-art image registration methods, three competitive algorithms are evaluated. First, deformable registration via attribute matching and mutual-saliency weighting (DRAMMS) by Ou

Table 4.3: 2D MRI registration results for R-Net and NCR-Net. Reported are the Jaccard indices averaged over 19 anatomical labels that are given in the central horizontal slices of all 40 LPBA40 images. Registration performance is measured separately for the original, uncorrupted images and for the images with four different artificially inserted stroke lesions (L1-L4).

Method	Original	L1	L2	L3	L4
w/o reg.	0.480 (0.063)	0.480 (0.063)	0.480 (0.063)	0.480 (0.063)	0.0480 (0.063)
R-Net	**0.620** (0.034)	0.509 (0.032)	0.610 (0.033)	**0.620** (0.033)	0.528 (0.034)
NCR-Net	0.619 (0.034)	**0.600** (0.036)	**0.619** (0.034)	0.619 (0.034)	**0.595** (0.035)

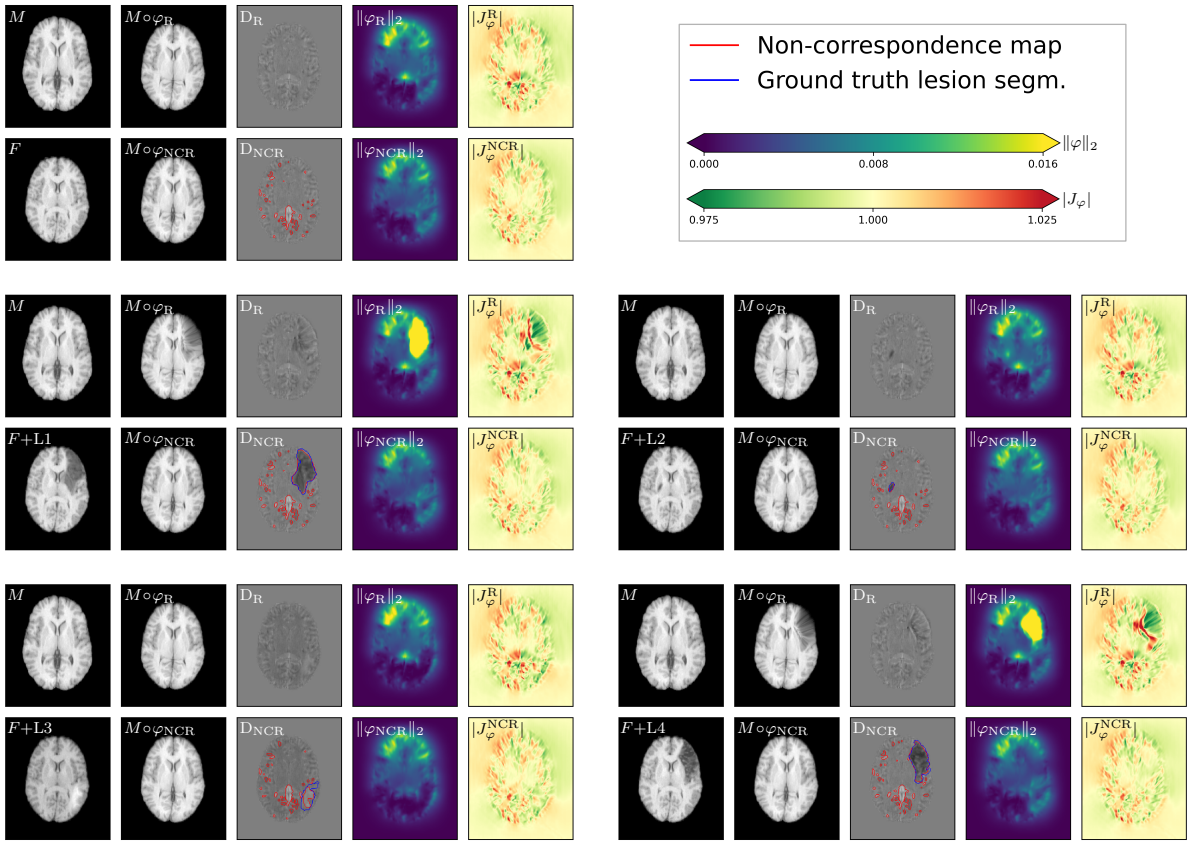


Figure 4.4: Exemplary MRI registration results of R-Net and NCR-Net. Shown are the moving image M and the five versions of the fixed image F. For each example, the warped moving image, the difference image after registration (noted as $\mathrm{D}_{\mathrm{NCR}} = M - F \circ \varphi_{\mathrm{NCR}}$), the amplitude and the Jacobian determinant of the deformation are shown in the top row for R-Net and in the bottom row for NCR-Net. For NCR-Net, the predicted non-correspondences are overlaid onto the difference image in red as well as the ground truth lesion segmentation in blue. NCR-Net generally leads to smoother and more plausible deformation fields.

et al. is used. DRAMMS is similar to NCR-Net in that the impairment done by non-corresponding image regions is reduced via a weighting of the image distance measure [Ou et al., 2011]. For DRAMMS, the default parameters of the algorithm are used. Second, the variational registration toolbox by Ehrhardt et al. (VarReg) is run with NCC distance measure and curvature regularization [Ehrhardt et al., 2015]. These two competitive algorithms are chosen as they are among the best performing registration methods on the LPBA40 dataset and open-source. Both are iterative approaches, whereas for the third competitive algorithm VoxelMorph, a deep learning-based method, is used. The diffeomorphic extension of VoxelMorph [Balakrishnan et al., 2019; Dalca et al., 2019] is trained with loss function (4.5) with the same optimizer and learning rate and for the same number of epochs as NCR-Net. The network architecture of VoxelMorph is not changed from the original implementation.

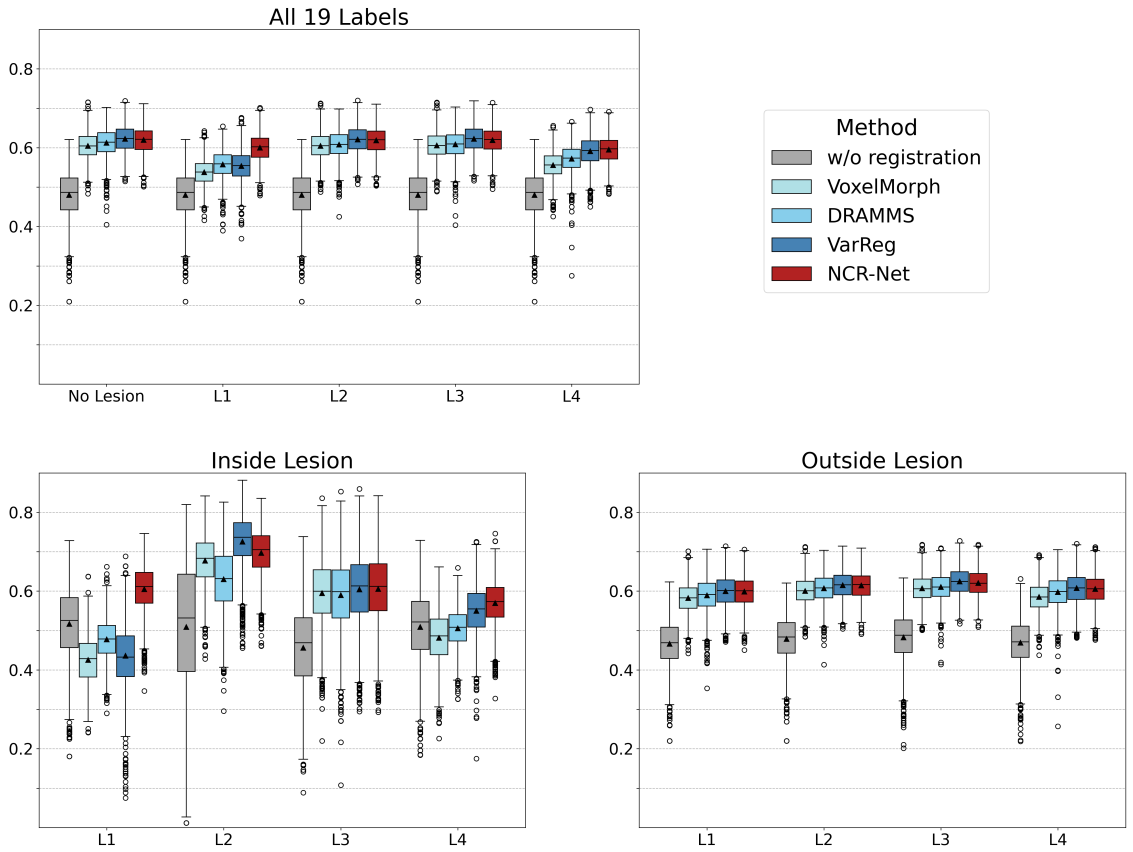


Figure 4.5: Results for 2D MRI registration. Average Jaccard indices are reported for the 19 anatomical labels that are present in all 40 images. Additionally, results are shown for the labels lying inside and outside lesions L1-L4.

Registration performance is measured by calculating the Jaccard index for the previously considered 19 anatomical labels. To analyze the influence of non-correspondences on the registration performance, results are additionally reported for anatomical structures that lie inside and outside the inserted stroke lesions. The results are averaged over the respective labels and given in Fig. 4.5. All registration methods considered perform similar on healthy images, but NCR-Net is the only method that delivers consistently good results for disturbed images (average Jaccard indices over all 19 labels of 0.6194, 0.5999, 0.6187, 0.6192 and 0.5950 for lesion-free images and images with lesions L1 to L4), no matter the size and intensity of the inserted lesions. Even for very large lesions L1 and L4, NCR-Net achieves good alignment of anatomical structures, whereas the performance of the competitive methods drops substantially, indicating that the introduced non-correspondences lead to implausible deformations. For images containing no or only subtle lesions (L2 and L3), NCR-Net performs better than DRAMMS and VoxelMorph, but is slightly yet significantly outperformed by VarReg, as evaluated with a one-sided Wilcoxon signed rank test with significance level 0.05. However, looking at performances inside the lesions, NCR-Net gives better results than all other methods, except for lesion

L2. NCR-Net, thus, proofs especially useful and robust for the registration of images with large pathologies. In addition, NCR-Net provides segmentations of areas that do not match between images, which is qualitatively shown to be a powerful indicator of individual anatomical variations in Sec. 4.3.2.3, where the non-correspondence maps are used for unsupervised stroke lesion segmentation.

4.3.2.2 MRI Registration in 3D

In this section, registration performance is evaluated in 3D using the full LPBA40 volumes for registration. The architecture of NCR-Net is extended for 3D MRI registration by replacing 2D convolutions with 3D convolutions using the same kernel sizes as before but in all three directions. Maximum pooling is also performed in 3D instead of 2D and the bilinear upsampling in the decoding path is replaced by trilinear interpolation. For network training, the healthy LPBA40 image volumes are used as moving images and the images with inserted pathologies serve as reference images. The weighting parameters in the loss function need to be adapted to the relatively smaller area occupied by the inserted lesions and to the fewer interindividual deviations, since the full image volumes are now used. The parameters are set to $\alpha = 0.04$, $\beta = 21000$, $\gamma = 0.14$, $\lambda = 1$. For 3D registration, the images are downsampled to $96 \times 112 \times 96$ voxels so that they fit into the memory.

Registration performance is again measured with the average Jaccard indices of anatomical labels, this time using all 56 labels given in the LPBA40 dataset. Results are reported in Tab. 4.4, once averaged over all labels and once averaged over labels inside and outside lesions. For comparison, 3D VoxelMorph is trained with the same loss function and hyperparameters as NCR-Net. Similar results can be observed as in the 2D setting. NCR-Net outperforms VoxelMorph and, more importantly, registration performance is stable for most inserted lesions (average Jaccard indices of 0.489, 0.486, 0.489, 0.476 and 0.486 for images without lesion and with lesions L1 to L4, respectively).

Interestingly, the registration performance of NCR-Net is impaired most by the hyperintense lesion L3. This lesion had hardly any effect on the 2D registration, as it is rather small and has low contrast to the surrounding tissue in the selected image slice. In 3D, however, the actual large extent of the lesion comes into play. Additionally, it is the only lesion with a hyperintense appearance and with blurry borders. The comparably low registration performance of NCR-Net for this lesion might, therefore, be explained by the unclear delineation of this lesion, which prevents the non-correspondence detection from completely masking it out. Still, NCR-Net outperforms VoxelMorph for all labels considered and achieves a promising alignment of healthy and pathological inter-patient brain images, as shown exemplarily in Fig. 4.7.

Table 4.4: 3D MRI registration results for VoxelMorph and NCR-Net. Reported are the average Jaccard indices of 56 anatomical labels before and after registration, given separately for the registration of uncorrupted images and images with artificially inserted lesions L1 - L4. For artificially corrupted images, Jaccard indices are additionally reported for labels lying inside and outside the inserted stroke lesions. Significantly best results are presented in bold font (one-sided Wilcoxon signed rank test with significance level 0.05).

	Method	All 56 labels	Inside lesion	Outside lesion
Original	w/o reg	0.373 (0.048)	-	-
	VoxelMorph	0.480 (0.056)	-	-
	NCR-Net	**0.489** (0.033)	-	-
L1	w/o reg	0.373 (0.048)	0.418 (0.077)	0.365 (0.047)
	VoxelMorph	0.466 (0.052)	0.432 (0.050)	0.472 (0.056)
	NCR-Net	**0.486** (0.032)	**0.521** (0.042)	**0.480** (0.033)
L2	w/o reg	0.373 (0.048)	0.416 (0.132)	0.372 (0.048)
	VoxelMorph	0.480 (0.056)	0.536 (0.104)	0.479 (0.055)
	NCR-Net	**0.489** (0.032)	**0.556** (0.069)	**0.487** (0.033)
L3	w/o reg	0.373 (0.048)	0.353 (0.078)	0.374 (0.048)
	VoxelMorph	0.474 (0.053)	0.478 (0.055)	0.474 (0.054)
	NCR-Net	0.476 (0.031)	**0.480** (0.049)	0.476 (0.031)
L4	w/o reg	0.373 (0.048)	0.413 (0.075)	0.366 (0.047)
	VoxelMorph	0.464 (0.052)	0.426 (0.052)	0.471 (0.057)
	NCR-Net	**0.486** (0.032)	**0.516** (0.041)	**0.481** (0.033)

4.3.2.3 Unsupervised Stroke Lesion Segmentation

Different from the longitudinal registration of OCT images, the non-correspondence maps resulting from inter-patient MRI registration do not primarily contain pathologic regions, but also a lot of normal inter-individual differences, as seen in Figs. 4.4 and 4.6. Therefore, two experiments are conducted to use the non-correspondence maps for the unsupervised segmentation of the phantom stroke lesions L1 to L4. First, the segmentation outputs are binarized with a threshold of 0.5 and compared directly to the ground truth masks of the inserted lesions. Second, it is assumed that a seed point inside the lesions is given, that could be easily and quickly generated during application of NCR-Net. Using this seed, region growing can be performed on the binarized non-correspondence segmentation to keep only the pathology segmentation and discard other segmented areas.

For evaluation, the seed points are not set manually but determined automatically by calculating the overlap between the non-correspondence map and the ground truth lesion segmentations and by randomly selecting one of the overlapping voxels as seed. Qualitative segmentation results are shown in Fig. 4.6 for a 2D image registration example performed with the four different pathological versions of the fixed image. It can be seen that the segmented non-correspondence regions that do not belong to the pathology are very similar for the different image variants and align with anatomical differences be-

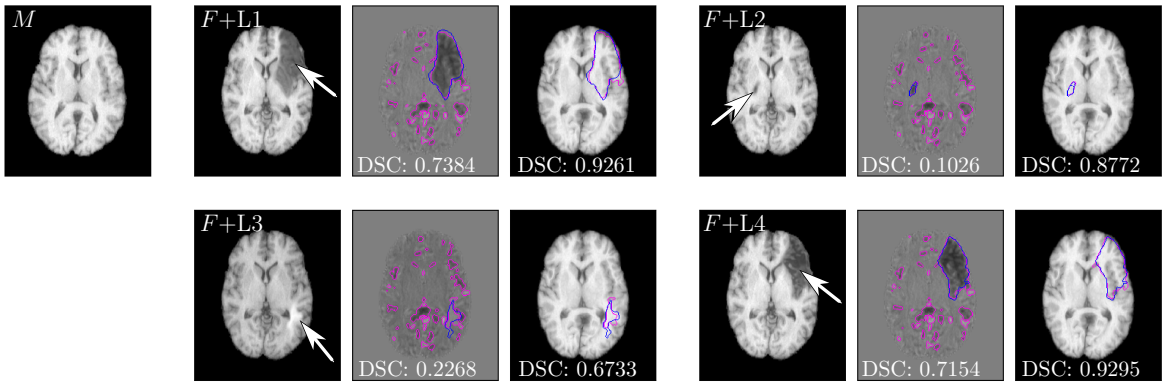


Figure 4.6: Stroke lesion segmentation results using thresholding and region growing on the segmentation output of NCR-Net. The moving image M is registered to four versions of the fixed image F with artificially introduced lesions L1 to L4 (shown in the second and fifth column). For each image pair, the difference image after registration is shown, with the segmentation of non-correspondent areas overlaid in pink and the ground truth lesion segmentation in blue. In addition, the deformed moving image is shown with the segmentation contour created by region growing on the binary non-correspondence map.

Table 4.5: Unsupervised (thresholding on (non-) correspondence maps) and weakly supervised (region growing with seed point inside lesion) stroke lesion segmentation performances for ProbReg [Krüger et al., 2019] and NCR-Net. Reported are the DSCs between ground truth and predicted segmentations averaged over images containing lesions L1 - L4.

Method	Post-proc.	L1	L2	L3	L4
ProbReg	Thresholding	0.693 (0.174)	**0.278** (0.174)	0.126 (0.092)	-
NCR-Net	Thresholding	**0.742** (0.067)	0.225 (0.086)	**0.362** (0.107)	0.737 (0.067)
ProbReg	Region growing	0.865 (0.111)	0.629 (0.172)	**0.754** (0.156)	-
NCR-Net	Region growing	**0.871** (0.049)	**0.870** (0.040)	0.630 (0.157)	0.880 (0.041)

tween patients. NCR-Net can, therefore, reliably detect inter-individual differences, even when the appearance of pathologies varies greatly. The pathology regions are sharply delineated, and region growing can deliver very accurate and lesion-specific segmentations.

The segmentation performance is measured quantitatively for the 2D NCR-Net by calculating the DSC between the predicted and the ground truth segmentations. Results are averaged over all images and reported in Tab. 4.5. A similar analysis has been performed for lesions L1 to L3 by [Krüger et al., 2019], whose results are cited in the table as well. Krüger et al. estimate correspondence probability maps that have continuous values, that are processed with thresholding to achieve binary lesion segmentations. To do so, two different thresholds are used and the better result is kept for each image.

For the unsupervised lesion segmentation using thresholds, NCR-Net achieves comparable or better results than ProbReg, indicating that the non-correspondences predicted

by NCR-Net are more focused on pathology-related image differences. For the weakly supervised lesion segmentation with region growing, NCR-Net achieves similarly high segmentation DSCs of 0.871, 0.870 and 0.880 for lesions L1, L2 and L4. These lesions are characterized by hypointense appearance and relatively sharp borders to the healthy tissue. For lesions L1 and L2, which were also used in [Krüger et al., 2019], NCR-Net can outperform the results reported there. Thereby, the standard deviations for NCR-Net are smaller than for ProbReg, which shows a more stable performance of the proposed method. However, the rather small and subtle lesion L3 is segmented better by ProbReg, with NCR-Net achieving an average DSC of 0.630 only. As observed in the 3D registration results, lesion L3 is the only hyperintense lesion that is visually difficult to distinguish from the surrounding tissue. The underperformance of NCR-Net might, thus, again be explained by the blurry borders of the lesion. Similar results can be observed when evaluating segmentation performances for the 3D version of NCR-Net. Here, average DSCs of 0.584, 0.039, 0.028 and 0.679 are achieved for lesions L1 to L4 when using simple thresholding (cf. Fig. 4.7 for exemplary results). Region growing improves results to 0.713, 0.675, 0.071 and 0.760. All results taken together, NCR-Net shows a promising ability to segment pathologies with little or no manual supervision, while providing consistent segmentations of other non-correspondences caused, e.g., by normal inter-individual deviations.

4.4 Discussion

In this chapter, NCR-Net was introduced, one of the first deep learning-based frameworks capable of simultaneous image registration and non-correspondence segmentation. NCR-Net provides state-of-the-art registration accuracy with improved plausibility of the deformation fields through outlier detection in the image distance and cost function masking. Without the need for expensive manual pathology annotations, the proposed training scheme enables the network to generate detailed non-correspondence maps that align well with areas of disease change, as shown with the longitudinal OCT data. For inter-patient registration, the non-correspondence maps additionally highlight individual anatomical differences between images. Most experiments in this work were conducted with weak supervision of the network training via an additional loss term that measures the overlap between the tissues shown on the images. However, it was also shown that the fully unsupervised NCR-Net achieves almost the same performance, indicating the usability of the framework for unannotated datasets.

Especially when rather small deformations are required to align the images, non-correspondence detection leads to increased robustness of the registration and smoother deformation fields, as shown for MRI registration. For OCT registration, deformations of almost unlimited size can occur. In such extreme cases, the non-correspondence detection can lead to registration errors being amplified, as a non-correspondence is marked at voxels where a deformation should actually occur. The training of NCR-Net, therefore, requires tuning of the weighting parameters in the loss in order to achieve a

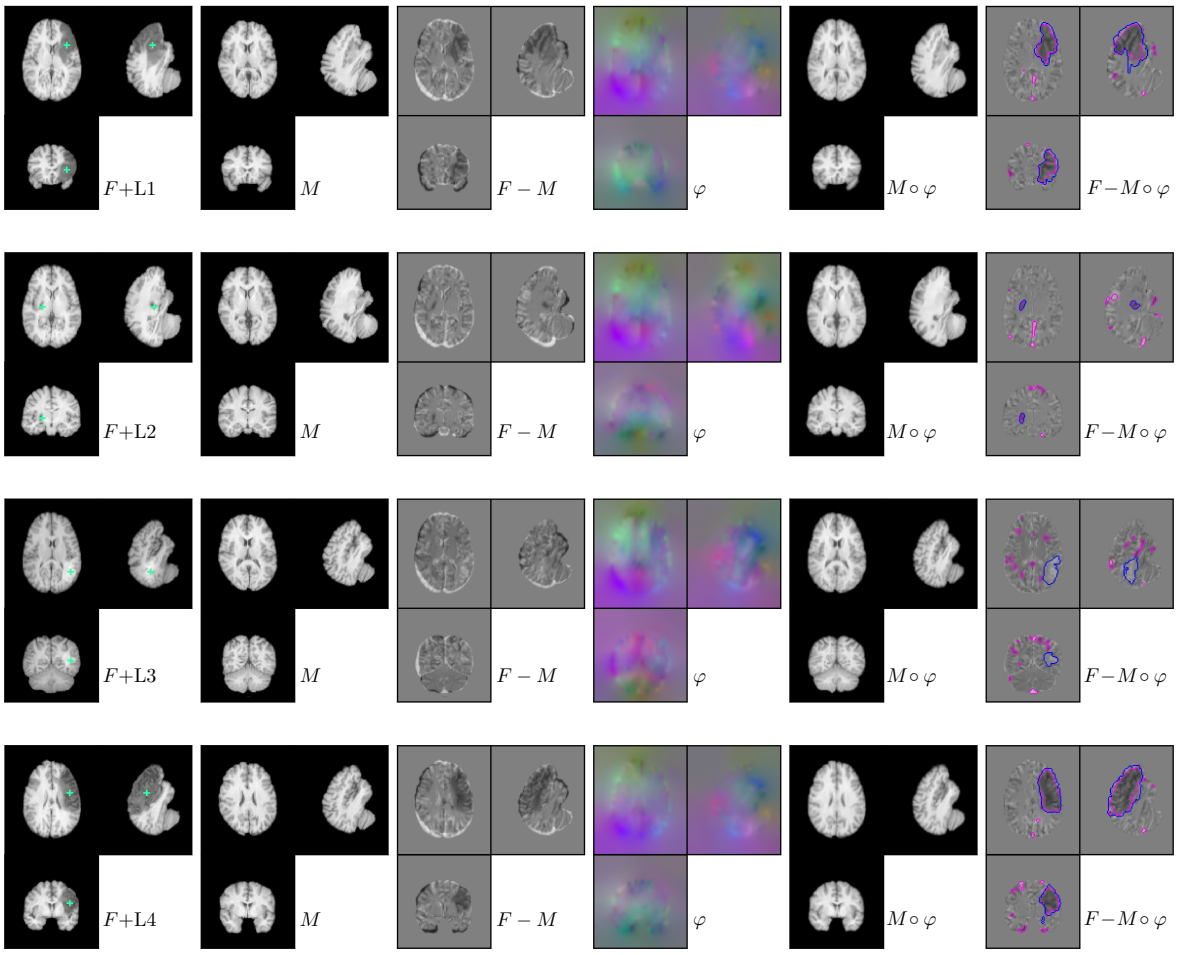


Figure 4.7: MRI registration results for 3D volumes. Shown are the results for registering the same moving image to four versions (L1-L4) of the same fixed image. The depicted slices are selected to lie at the center of the inserted lesions, the locations are indicated with turquoise crosses. The non-correspondence segmentation results are shown with pink contours on the difference image after registration and the ground truth lesion segmentations in blue.

good compromise between deformation accuracy and non-correspondence segmentation. The proposed two-stage training reduces this problem and can also be used to find good hyperparameters: During the first phase of training, in which only the registration part of the network is trained, suitable weightings can be determined for image distance and deformation regularization. After this first phase, when the CNN already provides meaningful deformations, only the parameters for the non-correspondence map need to be found.

Further research might focus on automating this hyperparameter search to ease adaptation to different datasets, which is often difficult with problem-specific image registration solutions. Recently, the first foundation model for image registration, uniGradICON [Tian et al., 2024b] was published. Foundation models are a promising field of research

for highly generalized, easily adaptable models that are pre-trained with huge amounts of data and can be quickly fine-tuned to new problems. Using such a foundation model as a starting point for image registration and non-correspondence segmentation CNNs, like NCR-Net, could potentially boost performance and improve robustness compared to networks being trained from scratch as done in this work.

A limitation of 3D NCR-Net is that it is currently only usable for comparably small image volumes. Patch-based approaches such as [Cao et al., 2017; Lu et al., 2024; Wu et al., 2024] could circumvent this problem. Finally, additional improvements might be achieved regarding the disentanglement of cost function masking and deformation. Although the plausibility of the deformation fields has been improved, their realism is still questionable, especially in the case of OCT images. As fluids can appear very suddenly in AMD or CSCR, deformations alone are not sufficient to model the development of the pathologies. Chapter 6, therefore, presents an approach that makes it possible to integrate new fluids into the registration. A pathology seed point is set, which is extended with the deformation to the size of the new fluid. Both the registration and the creation of the seed points are automatically learned by a CNN, similar to NCR-Net, leading to more realistic deformation fields.

Chapter 5

Detection and Segmentation of New Multiple Sclerosis Lesions Using Image Registration CNNs

Multiple sclerosis (MS) is a chronic immune-mediated disease that is estimated to affect more than 200,000 people in Germany and 2.8 million worldwide [Petersen et al., 2014; Walton et al., 2020]. Affecting the entire central nervous system (CNS), MS is the main cause of neurological disability in young adults [WHO, 2008]. Various inflammatory processes, which are not yet fully understood, lead to a loss of CNS tissue during the course of the disease. In focal lesions, the body's own immune cells attack the myelin sheaths of the nerves in the white matter of the brain and spinal cord, which initially leads to a slowing of the nerve conduction velocity and subsequently to destruction of the axons. Focal lesions are considered to be the trigger for the typical relapsing neurological disorders that occur in MS. However, autopsies and magnetic transfer imaging have shown that neural destruction is not limited to the inflammatory foci, but that microlesions occur diffusely throughout the entire CNS [Bernhardt, 2010]. MRI is the standard imaging modality for MS diagnosis and therapy monitoring. On T2/FLAIR, focal lesions appear as localized, hyperintense areas, while the diffuse micro-lesions cannot be visualized.

The size, number and location of inflammatory lesions are established MS biomarkers. However, in the growing literature, the development of new lesions is considered an even more important biomarker to evaluate. In fact, the only indicator of drug efficacy is the absence of new CNS lesions on T2 images [Commowick et al., 2021]. Since MS lesions are small and difficult to detect, the manual assessment of MS lesions is a very time-consuming and complex task. Geurts et al. reported that only 40% of lesions found on histopathology were also found on FLAIR MR scans in a postmortem study [Geurts et al., 2005]. Even more challenging is the assessment of new MS lesions, as it requires the joint analysis of images from two time points. Spatial misalignment and non-correspondences between time points, such as imaging artifacts or physiological changes in the brain, complicate the differentiation of new from existing, possibly progressive MS lesions. Automating the detection and delineation of such lesions would, therefore, be a major advance for MS monitoring.

In this chapter, different methods are presented to solve the task of new MS lesion segmentation and detection. These methods are all based on deep learning image registration networks with simultaneous non-correspondence detection. First, an extension of NCR-Net is proposed, that models new lesions as appearance offsets in non-correspondent image areas. This allows analyzing different regions of individual focal lesions. The CNN is called ANCR-Net for appearance-adaptation in non-correspondent

regions and image registration network and published in [Andresen et al., 2022c]. Second, the NCR-Net approach from the previous chapter is used and complemented with an additional U-Net that refines the non-correspondence segmentations generated by NCR-Net. This approach was submitted to the MICCAI 2021 - Longitudinal Multiple Sclerosis Lesion Segmentation (MSSEG-2) challenge [Commowick et al., 2021], which focuses on the detection and segmentation of new MS lesions, and published in [Andresen et al., 2021].

The remainder of this chapter first gives an overview of the literature related to new MS lesion segmentation (Sec. 5.1). In Sec. 5.2, the architectures and training procedures of the proposed methods are explained. New lesion detection and segmentation performances are evaluated in Secs. 5.3.1 and 5.3.2. The appearance offset maps generated by ANCR-Net are used to model new lesions in Sec. 5.3.3. A summary and discussion of the presented approaches and results are given in Sec. 5.4.

5.1 Related Literature

The existing literature, tackling the automatic examination of longitudinal MS data, can broadly be classified into lesion detection and change detection methods [Lladó et al., 2012]. Lesion detection approaches analyze images from different time points separately and only subsequently detect changes over time by comparing the resulting segmentations. In [Köhler et al., 2019], for example, all scans of a patient are affinely registered to a reference scan and lesions are categorized into static, dynamic and new, based on the overlap of lesion masks from the individual time points. In change detection approaches, images from subsequent time points are analyzed jointly to directly detect changes between baseline and follow-up. These approaches can further be classified into deformation- and intensity-based approaches [Salem et al., 2020]. Typically, registration is a pre-processing step for intensity-based approaches and image intensities are compared voxel-wise to generate segmentations of new lesions [Moraal et al., 2010; Battaglini et al., 2014; Ganiler et al., 2014; Battaglini et al., 2014; Jain et al., 2016; Fartaria et al., 2019]. Non-rigid registration is used in deformation-based approaches, where deformation fields are used to improve new lesion segmentation [Rey et al., 2002; Cabezas et al., 2016; Salem et al., 2018, 2020].

Since 2020, the majority of works on new MS lesion segmentation has been based on DL, e.g. [Krüger et al., 2020; McKinley et al., 2020; Salem et al., 2020; Combès et al., 2021]. In [Salem et al., 2020], image pairs from four different MRI modalities are used and registered in a CNN, before the generated deformation fields are passed on to the second part of the network, which segments new lesions. Krüger et al. use FLAIR image pairs to train a U-Net-like CNN for the segmentation of new or enlarged MS lesions [Krüger et al., 2020]. [Combès et al., 2021] propose a complete workflow for the integration of computer-aided new MS segmentation into the clinic. The automatic segmentation module uses nnU-Net to segment new lesions in FLAIR, T1-w and T2-w MRIs and the network is purposely tuned to be oversensitive in order to not miss any

new lesions and removal of false positives is left to users. In contrast, false positives are reduced automatically using segmentation uncertainty in [McKinley et al., 2020].

In 2021, the MSSEG-2 challenge took place [Commowick et al., 2021], as part of which the paper [Andresen et al., 2021] was published. The challenge provided a longitudinal dataset of 100 FLAIR MR image pairs of MS patients taken one to three years apart and manual segmentations of new MS lesions in the follow-up image. Four medical experts segmented the images, and a consensus ground truth was built that served as a reference for evaluating submissions that addressed the task of new MS lesion segmentation. 24 teams submitted 30 pipelines, only one of which was not based on DL.

The U-Net architecture is predominant, whereby different 3D [Andresen et al., 2021; Ashtari et al., 2021; Basaran et al., 2021; Fenneteau et al., 2021; Kamraoui et al., 2021b,c,a; Kang et al., 2021; Löhr et al., 2021; Macar et al., 2021; Masson et al., 2021; Schmidt-Mengin et al., 2021] and 2D [Dalbis et al., 2021; de Sousa and de Oliveira, 2021; Sarica and Seker, 2021; Zhang et al., 2021] variants and ensembles [Efird et al., 2021; Gibicar et al., 2021; Nichyporuk et al., 2021; Siddiquee and Myronenko, 2021] are used. Promising results are achieved by Dalbis et al. and Zhang et al. that both rely on a 2.5D approach, using 2D slices from all three directions as network input. Several strategies are implemented to overcome the class imbalance problem, inherent to the MSSEG-2 training data, due to the small size and number of new lesions. Most methods rely on patch sampling strategies, either selecting only patches within the brain [Dalbis et al., 2021; Efird et al., 2021; Fenneteau et al., 2021] or oversampling positive patches using manual ground truth, e.g. [Andresen et al., 2021; Ashtari et al., 2021; Basaran et al., 2021; Gibicar et al., 2021; Kang et al., 2021; Macar et al., 2021; Preloznik and Žiga Špiclin, 2021; Sarica and Seker, 2021]. Some works automatically select ROIs based on the subtraction image of both time points [Rosa et al., 2021; Prados and Kanber, 2021]. [Hamzaoui et al., 2021] is the only work that analyzes images from both time points separately. Also, additional data, either generated [Kamraoui et al., 2021a] or real [Basaran et al., 2021; Cabezas et al., 2021; Dalbis et al., 2021; Löhr et al., 2021; Masson et al., 2021], is used to pre-train networks or to expand the training set. Nichyporuk et al. use a lesion reweighting strategy to adapt the BCE loss to give more weight to lesion voxels [Nichyporuk et al., 2021], and Schmidt-Mengin et al. perform weighted patch sampling and hard example mining to improve training [Schmidt-Mengin et al., 2021].

A common approach is to use a two-stage process, where the first stage is used to identify potential lesion regions and the second to refine the result of the first stage. This strategy is also used in [Andresen et al., 2021], which uses NCR-Net to register the images from both time points and simultaneously segment non-correspondent regions. The non-correspondence map serves as a segmentation of new lesion candidates and is refined with a segmentation U-Net that receives the registered images, their subtraction image and the non-correspondence map as input. [Prados and Kanber, 2021] detect potential lesions in the subtraction image of the two time points, and use a gradient boosting classifier to separate positive and negative lesion candidates. [Kang et al., 2021] tune a U-Net to be oversensitive and reduce false positives with a ResNet first, and by comparing baseline and follow-up in the potential lesion area second. Methods

proposed in [Basaran et al., 2021; Fenneteau et al., 2021; McKinley et al., 2021] all use a CNN to segment all lesions for both time points separately and pass the result as input to a refinement network to generate segmentations of new lesions. [Löhr et al., 2021] also train a network to segment all lesions in both time points, but generate segmentations of new lesions by simply removing all lesions present in the segmentation of the first time point from the segmentation of the second time point using connected components analysis. Similarly, [Nichyporuk et al., 2021] cascade five U-Nets, out of which four are trained on the segmentation masks of individual raters and the fifth receives the outputs of the other four networks plus the MRIs as input and is trained to generate the consensus segmentation.

Following the MSSEG-2 challenge, numerous follow-up papers, including ANCR-Net, the method presented in this chapter, were published in the Research Topic [Commowick et al., 2023], addressing the problem of detecting and segmenting new MS lesions. Again, U-Net variants are predominant [Andresen et al., 2022c; Basaran et al., 2022; Hitziger et al., 2022; Kamraoui et al., 2022a; Sarica and Seker, 2022; Schmidt-Mengin et al., 2022], many of which use residual blocks [Ashtari et al., 2022; Hitziger et al., 2022; Sarica and Seker, 2022] and are further improved with attention gates [Sarica and Seker, 2022], deep supervision [Ashtari et al., 2022] or 2.5D inputs [Hitziger et al., 2022]. Best performance is achieved by Basaran et al. who use deformable registration for pre-processing and train an nnU-Net with a sophisticated data augmentation that mimics imaging artifacts and uses CarveMix to increase the number of new lesions in the training data [Basaran et al., 2022].

Several works use joint image registration and segmentation [Andresen et al., 2022c; Dufresne et al., 2022; Salem et al., 2022] to improve over pure intensity- or deformation-based approaches. The here presented method follows this direction by using an extension of NCR-Net, that introduces appearance matching [Andresen et al., 2022c]. This work manages to achieve state-of-the-art performance compared to the other methods published in [Commowick et al., 2023]. [Salem et al., 2022] modify their CNN for registration and subsequent segmentation presented in [Salem et al., 2020] to work with only the FLAIR modality, while [Dufresne et al., 2022] use an iterative, alternating optimization scheme to jointly register images and segment changes between time points.

Since then, research into the automatic segmentation of MS lesions has produced many more papers, most of which focus on the segmentation of all, rather than new, lesions in multimodal MR data, e.g. [Hashemi et al., 2022; Tran et al., 2022; Kamraoui et al., 2022b; Rondinella et al., 2023; Sarica et al., 2023]. In line with this, another challenge took place at ICPR 2024 that tackled the segmentation of MS lesions on T1-w, T2-w and FLAIR images [Rondinella et al., 2025]. Here, up to four time points per patient are given to allow for analysis of segmentation robustness over time points. Recent works covering the segmentation of new MS lesions are given in [Martínez-Heras et al., 2023; Wu et al., 2023b; Tahghighi et al., 2024; Basaran et al., 2025, 2024; Nasheeda and Rajangam, 2024]. [Basaran et al., 2024] propose SegHeD, an nnU-Net that can be used for the segmentation of all, new and vanishing MS lesions. The network is trained by incorporating domain knowledge into the loss function using longitudinal, anatomical

and volumetric constraints. The CoactSeg network, described in [Wu et al., 2023b], improves over previous baselines using a network with two output heads that generate segmentations of all and only new lesions. Input to the network are both the original MRIs and a brain difference map. Current reviews on the usage of DL and machine learning algorithms for the diagnosis, analysis and prognosis of MS can be found in [Vázquez-Marrufo et al., 2023; Amin et al., 2024; Pilehvari et al., 2024].

5.2 Methods

Borrowing ideas from metamorphic approaches such as [Uzunova et al., 2021], ANCR-Net models new MS lesions as appearance offsets between baseline and follow-up image. As in the previous chapter, the goal is to perform image registration and simultaneous non-correspondence segmentation within the same network. Here, non-correspondences should only include the new lesions. Therefore, rather than masking out non-corresponding regions in the image distance calculation, the appearance of the baseline is adapted to directly model the newly appearing lesions. The loss function used to train ANCR-Net for the registration of the moving baseline image $M : \Omega \to \mathbb{R}$ and the fixed follow-up image $F : \Omega \to \mathbb{R}$ is

$$\mathcal{L}(\theta; M, F) = \mathcal{L}_{\text{Dist}}\big(F, (M + N \cdot A) \circ \varphi\big) + \alpha \mathcal{L}_{\text{Reg}}(\varphi) + \beta \mathcal{L}_{\text{Dice}}(N \circ \varphi, S^{\text{GT}}), \qquad (5.1)$$

with non-correspondence segmentation N and appearance offset map A. While the spatial displacement φ accounts for general misalignments between visits as well as for existing lesions changing size and shape, the intensity transformations are masked with the non-correspondence segmentation and inserted into the baseline image to compensate for new lesions. Normalized cross-correlation is used as image distance measure and diffeomorphic deformation fields $\varphi = \exp(v)$ with stationary velocity fields v are generated. The regularizer $\mathcal{L}_{\text{Reg}} = \frac{1}{n} \sum_{\boldsymbol{x} \in \Omega_{\text{patch}}} \|\nabla v\|_2^2$ enforces smoothness of the velocity field, with n being the number of voxels of the input patches. To guide the segmentation task, the Dice loss is used that compares the deformed non-correspondence segmentation N with the ground truth segmentation of new MS lesions S^{GT}. The weighting parameters α and β are found empirically and set to $\alpha = 60$ and $\beta = 0.1$.

The architecture of ANCR-Net is U-Net-based and described in detail in Fig. 5.1. The network consists of one encoder and two separate decoders. Input to the network are the concatenated fixed and moving images and their difference image. The two separate inputs are passed through two convolutional layers each, before being concatenated and processed together. The first decoder is the deformation decoder that generates the deformation field for three resolution levels. The second decoder predicts both the appearance offsets and the non-correspondence segmentation output for the three resolution levels. At each level, two separate output heads consisting of two 1×1 convolutional layers are used to generate non-correspondence segmentation N and appearance offset A. Skip-connections are used between encoder and decoders.

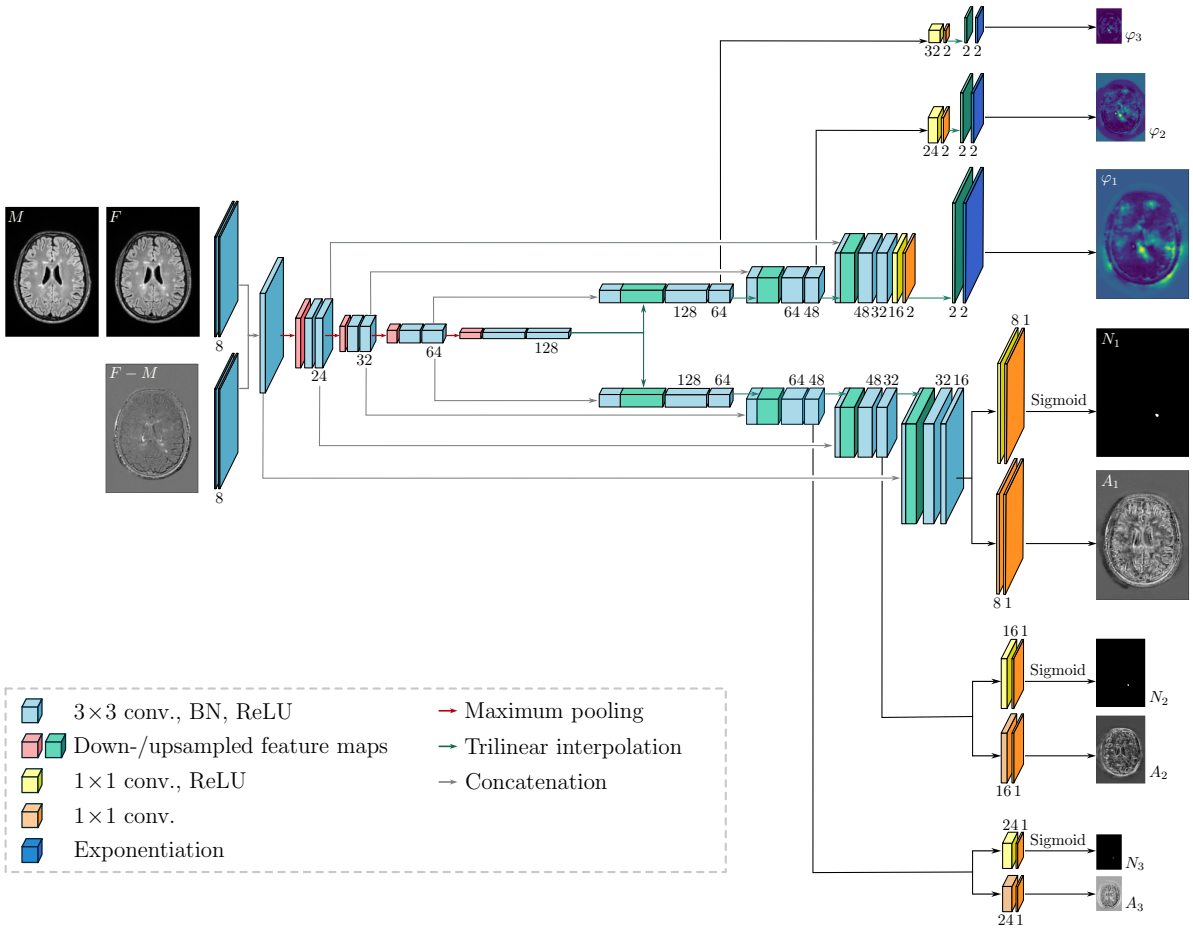


Figure 5.1: The architecture of ANCR-Net. Two separate input blocks process the MRIs to be registered and their difference image. Two separate decoders generate the deformation field φ, the segmentation of new MS lesions N and the appearance offset map A. The numbers below/above the boxes indicate the number of feature maps resulting from the respective convolutions.

5.2.1 Network Training

Network training is performed with the 40 training images of the MSSEG-2 challenge dataset (c.f. Sec. 2.4.2), which is split into training and validation subsets in a five-fold cross-validation. Eleven out of the 40 training images show no new lesions, whereas the new lesion count for the remaining 29 cases ranges from one to 33. Furthermore, the training data comprises images from 12 different MRI scanners with image sizes ranging from $144\times192\times192$ to $300\times576\times576$ voxels. For training, the horizontal slices of the images are resampled to an isotropic pixel spacing and, if necessary, padded to have an image size of 368×512 pixels. An error in the segmentation of patient 12 only became apparent after the challenge. In [Andresen et al., 2022c], evaluation results were, therefore, reported for only 59 of the 60 test patients. Since challenge submissions and

further related works evaluated on all 60 cases, the evaluations in this thesis are carried out on the full test dataset to ensure fair comparability to other methods.

As done with NCR-Net, deep supervision is used, i.e. loss calculation is performed on each resolution level and the weighted sum of the three losses is used for backpropagation during training. Finer resolution levels are again given higher weight and the weighting factors are chosen to be 0.7, 0.2 and 0.1. New MS lesions are rare, often subtle and small, resulting in a severe underrepresentation of the lesion class compared to the background class. Thus, a high image resolution, especially in the horizontal plane, is required for successful network training. However, memory limitations prevented the entire image volumes from being passed on to the network. Consequently, network training is performed on patches consisting of five successive horizontal slices. The inherently high class imbalance caused by the small number and size of new MS lesions is tackled two-fold. First, careful training data selection is done by sampling training patches so that they all contain new lesions. This is necessary to avoid the network from simply ignoring the segmentation task and outputting an all-zero non-correspondence prediction. During inference, the entire image volumes are segmented by slice-wise iteration through the volume and keeping the segmentation result for the central slice of the stacked input patches only.

Second, new lesions are simulated to pre-train the network with the images from the training data without real new lesions. New lesions are simulated as Gaussian ellipsoids that are placed at random locations inside the brain by adding their values to the image intensities. To do so, a brain mask is generated by performing brain extraction on both time points separately and calculating the union of both masks. Brain extraction is done using the default pre-processing pipeline provided with the challenge data[1]. To exclude the ventricles from the mask, baseline and follow-up images are normalized to intensity range $[0, 1]$, thresholded above 0.1 and multiplied with the brain mask. Finally, the mask is shrunk with morphological erosion to make sure inserted lesions do not protrude the brain. Lesion insertion is done on the fly during pre-training, and for each image, a random number of inserted lesions is chosen between one and five.

The manipulated follow-up is then subjected to a random elastic deformation. This modified image serves as reference scan and the baseline image as moving scan. Pre-training is done for 200 epochs in a fully supervised manner using the Dice loss of the inserted and predicted lesions as well as the MSE between ground truth and predicted transformations. Adam optimization is used with an initial learning rate of $1e-4$ and learning rate decay with a factor of 0.8 every 20th epoch.

For both pre-training and the subsequent main training, several augmentation techniques are applied. These include spatial augmentations (random rotations between $\pm 5°$ and random shifts up to three pixels in the axial plane performed on both images) and intensity augmentations. Most intensity augmentations are applied to the brain region only (Gaussian noise, overall brightness change, artificial brightness gradient) except for adaptive histogram equalization that is applied to the entire image patches. During

[1] https://anima.irisa.fr/

main training, the dataset size is doubled by using patches containing new lesions twice, once in the original orientation and once flipped horizontally. To assure full convergence of the network, no early stopping is performed, running the training for 400 epochs with the same optimizer and learning rate schedule as for the pre-training.

5.2.2 Application of NCR-Net for New MS Lesion Segmentation

With the preliminary work described in Chapter 4, the question arises as to how well the non-correspondence maps generated with NCR-Net can delineate new MS lesions and how it compares to ANCR-Net. Therefore, NCR-Net is trained as described in the previous chapter to generate diffeomorphic registration deformation fields and segmentations of non-correspondent image areas. The only modification done to NCR-Net is the addition of a second input block, consisting of two convolutional layers with batch normalization and ReLU activation, that processes the difference image of baseline and follow-up as done with ANCR-Net. The unsupervised non-correspondence maps are not exclusive for new lesions but also contain other sources of mismatch between the images. The non-correspondence segmentations are, therefore, refined with a second CNN, that receives the registered images, their difference image and, in a second experiment, also the non-correspondence map as input. The architecture of this network is a U-Net with separate input blocks for each input modality. This U-Net is trained supervised using the Dice loss to compare predictions to ground truth new lesion segmentations. The intuition behind this method is two-fold. First, the better aligned input images should ease the identification of altered regions. Second, the spatial context provided by the non-correspondence maps should guide the network to the relevant image areas and simplify the task of new lesion detection and segmentation to a refinement of the provided non-correspondence maps. The proposed pipeline is depicted in Fig. A2 in the appendix.

This procedure corresponds to the methodology described in [Andresen et al., 2021] and has been submitted to the MSSEG-2 challenge. However, several improvements have been introduced for the experiments performed in this thesis. Most importantly, the pre-processing of the images was adapted with an improved resampling that requires only one interpolation rather than two. Instead of using three successive horizontal image slices as input patches now five slices form one patch. Also, in [Andresen et al., 2021], the U-Net received the original baseline image as input and used the registered image only for the difference image input, whereas now the deformed baseline is used as input. Finally, in the previous implementation, training patches included all patches containing new lesions and one percent of patches containing no new lesions. Here, only patches with new lesions are used for training. These changes are consistent with the implementation used for ANCR-Net to make results directly comparable.

5.3 Experiments and Results

Five-fold cross-validation is performed for all experiments run in this section, using 32 of the 40 MSSEG-2 training images for network training and the remaining eight for validation in each fold. For segmentation evaluation on the test data, the predictions of the five networks are combined using majority voting. Post-processing of the predicted segmentations includes masking to the brain and discarding lesion with a volume smaller than 3 mm^3.

5.3.1 New Lesions Detection

In this section, new lesion detection performance of ANCR-Net is evaluated separately for stable and progressing patients. For stable patients with no new lesions in the ground truth, the average number and volume of false positive predicted lesions are reported. For progressing patients, the lesion sensitivity, lesion positive predictive value and the F_1 score are reported. The lesion sensitivity sens^{L} describes the proportion of correctly identified lesions out of all lesions in the consensus segmentation. The lesion positive predictive value PPV^{L}, in turn, gives the proportion of true positive lesions out of all predicted lesions. In the F_1 score

$$F_1 = \frac{2 \cdot \text{sens}^{\text{L}} \cdot \text{PPV}^{\text{L}}}{\text{sens}^{\text{L}} + \text{PPV}^{\text{L}}}, \tag{5.2}$$

both metrics are combined to describe the trade-off between detection rate and the generation of few false positive predictions in one value. Additionally, for both stable and progressing patients, the proportion of correctly classified patients is reported (sens^{P}). Progressing patients are counted as correctly classified if at least one ground truth lesion is correctly detected. Stable patients are counted as correctly classified if the predicted segmentation is empty.

In the following, ANCR-Net results are compared to human medical expert performance, the best performing challenge submissions, results reported in [Basaran et al., 2022; Dufresne et al., 2022; Salem et al., 2022; Sarica and Seker, 2022; Wu et al., 2023b], NCR-Net, and three different U-Nets. The metrics reported for the medical experts have been calculated for each of the four manual segmentations compared to the consensus and then averaged. Three competitive methods were chosen from the challenge submissions: MedICL [Zhang et al., 2021] achieved the highest DSC, Mediaire-B [Dalbis et al., 2021] the highest F_1 score and LYLE [Ashtari et al., 2021] performed best for stable patients, with lowest number and volume of false positive lesions.

Additionally, five papers were selected from the relevant literature for comparison. Two papers were selected because they follow methodologically similar approaches to this work. [Dufresne et al., 2022] use an iterative non-learning-based approach for joint change detection and image registration, and [Salem et al., 2022] use two cascaded segmentation U-Nets, the first of which is trained to be hypersensitive to find lesion

Table 5.1: New lesion detection results for images with and without new lesions in ground truth. Reported are average F_1 score, lesion sensitivity $sens^L$ and positive predictive lesion value PPV^L for images containing new lesions. For stable patients, the average number of erroneously detected lesions and their volume are reported (n_{FP} and V_{FP}). The results are given for the medical experts who generated the manual ground truth data, as well as for ANCR-Net with (ANCR-Net_{PT}) and without (ANCR-Net) pre-training. Performance is compared to NCR-Net, NCR-Net plus additional U-Nets using unregistered images (U-Net), registered images (U-Net_{reg}) and registered images plus non-correspondence segmentation (U-$Net_{reg,N}$) as input, the three pipelines that performed best in the MSSEG-2 challenge (MedICL, Mediaire-B and LYLE) and five selected methods from the literature. Metrics not given in the literature are marked with n/a. Best results are given in bold font and second best are underlined.

		Progressing patients				Stable patients		
Method	Data	$F_1 \uparrow$	$sens^L \uparrow$	$PPV^L \uparrow$	$sens^P \uparrow$	$n_{FP} \downarrow$	$V_{FP} \downarrow$	$sens^P \uparrow$
Expert 1	Val.	0.808	0.808	0.853	0.931	0.000	0.000	1.000
Expert 2	Val.	0.691	0.665	0.778	0.828	0.000	0.000	1.000
Expert 3	Val.	0.718	0.667	0.825	0.862	0.000	0.000	1.000
Expert 4	Val.	0.721	0.662	0.839	0.862	0.000	0.000	1.000
NCR-Net	Val.	0.170	0.402	0.160	0.690	9.636	970.158	0.000
U-Net	Val.	0.578	0.605	0.635	0.828	0.182	0.944	0.818
U-Net_{reg}	Val.	0.642	0.704	0.641	0.897	0.273	6.746	0.727
U-$Net_{reg,N}$	Val.	0.543	0.572	0.591	0.828	0.091	1.998	0.909
ANCR-Net	Val.	0.591	0.634	0.624	0.828	0.545	6.959	0.636
ANCR-Net_{PT}	Val.	0.622	0.666	0.623	0.862	0.455	6.948	0.636
Expert 1	Test	0.713	0.717	0.719	0.812	0.036	1.453	0.964
Expert 2	Test	0.637	0.614	0.738	0.844	0.107	3.981	0.893
Expert 3	Test	0.638	0.615	0.730	0.844	0.000	0.000	1.000
Expert 4	Test	0.526	0.470	0.660	0.688	0.036	0.623	0.964
MedICL	Test	0.500	<u>0.736</u>	0.450	**0.844**	0.536	12.713	0.643
Mediaire-B	Test	0.541	0.685	0.491	0.781	0.536	29.235	0.643
LYLE	Test	0.441	0.418	0.506	0.656	**0.036**	<u>0.470</u>	**0.964**
Dufresne et al.	Test	-	**0.872**	0.111	n/a	n/a	n/a	n/a
Salem et al.	Test	0.500	0.527	0.519	n/a	n/a	n/a	n/a
Basaran et al.	Test	0.552	n/a	n/a	n/a	**0.036**	**0.192**	n/a
Sarica et al.	Test	0.480	n/a	n/a	n/a	0.148	1.488	n/a
Wu et al.	Test	**0.620**	n/a	n/a	n/a	n/a	n/a	n/a
NCR-Net	Test	0.116	0.246	0.133	0.562	13.750	8996.273	0.179
U-Net	Test	0.523	0.530	<u>0.609</u>	0.781	<u>0.107</u>	2.069	<u>0.893</u>
U-Net_{reg}	Test	0.545	0.606	0.545	0.781	0.143	2.882	0.857
U-$Net_{reg,N}$	Test	0.529	0.502	**0.654**	0.781	**0.036**	1.304	**0.964**
ANCR-Net	Test	0.549	0.605	0.592	<u>0.812</u>	0.250	4.443	0.750
ANCR-Net_{PT}	Test	<u>0.565</u>	0.615	0.564	0.781	<u>0.107</u>	2.039	<u>0.893</u>

candidates and the second refines the segmentation of the first network. The other three papers were selected for their high performances [Basaran et al., 2022; Wu et al., 2023b]

and because they are the only papers that report metrics for both stable and progressing patients [Basaran et al., 2022; Sarica and Seker, 2022]. All use U-Net variants and circumvent the class imbalance problem with sophisticated patch selection [Sarica and Seker, 2022], specialized augmentation [Basaran et al., 2022] or by combining several MS datasets [Wu et al., 2023b].

NCR-Net is trained unsupervised, i.e. the ground truth segmentations are only used for the selection of training patches, but not in the loss. Results are reported for the non-correspondence maps directly, with the only post-processing being a masking to the brain. These maps, thus, contain all sorts of non-correspondences rather than new lesions exclusively. Subsequently, three versions of U-Net are trained supervised for new lesion segmentation. The baseline U-Net is trained with two inputs, namely the stacked baseline and follow-up images and their subtraction image. The same U-Net is trained again, but using the registration results of NCR-Net as input ($\text{U-Net}_{\text{reg}}$). Finally, the U-Net is extended with a third input block that additionally receives the non-correspondence maps as input to provide guidance to the network ($\text{U-Net}_{\text{reg,N}}$, the version shown in Fig. A2).

The results are given in Tab. 5.1 and show that most automatic methods achieve detection rates close to or even surpassing the performance of the medical experts, both on lesion- and patient-level. However, most methods achieving high detection rates also tend to produce quite a lot of false positives, e.g. MedICL and Mediaire-B. This behavior can partially be explained by spurious segmentations due to spatial misalignments (remaining after registration) and old growing lesions. The tendency to be oversensitive shows in low PPV^{L} values and high numbers of predicted lesions for stable patients. Methods that provide few false positive results, in turn, tend to be undersensitive, missing many lesions. This is the case for LYLE, for example. The methods proposed in this thesis, as well as the one presented by [Basaran et al., 2022] provide best compromises here, achieving F_1 scores of 0.565 (ANCR-Net) and 0.552 (Basaran et al.), and generating only few false positives (average number of false positive lesions being 0.107 and 0.036, respectively).

ANCR-Net manages to reliably separate progressing from stable patients, correctly detecting at least one ground truth lesion for 25 out of 32 progressing patients and producing no false positives for 25 out of 28 stable patients. When considering all predicted lesions, not only those overlapping with ground truth lesions, ANCR-Net correctly identifies 29 of the progressing patients. The average deviation in the estimate of the number of new lesions for ANCR-Net is only 1.322 lesions.

Following the analysis by Basaran et al., in Fig. 5.2, the F_1 score for new lesion cases is contrasted to metrics for no new lesion cases. In this graphical representation, methods that recognize all lesions without marking any false areas would be entered at the top left, whereas oversensitive methods appear further to the right and undersensitive methods further down. Methods reported in Tab. 5.1 are displayed with a star, challenge submissions are shown with blue markers, the proposed methods with red or orange markers and the methods from [Basaran et al., 2022] and [Sarica and Seker, 2022] with green markers. Only those methods are shown that surpass an F_1 score of 0.35. The

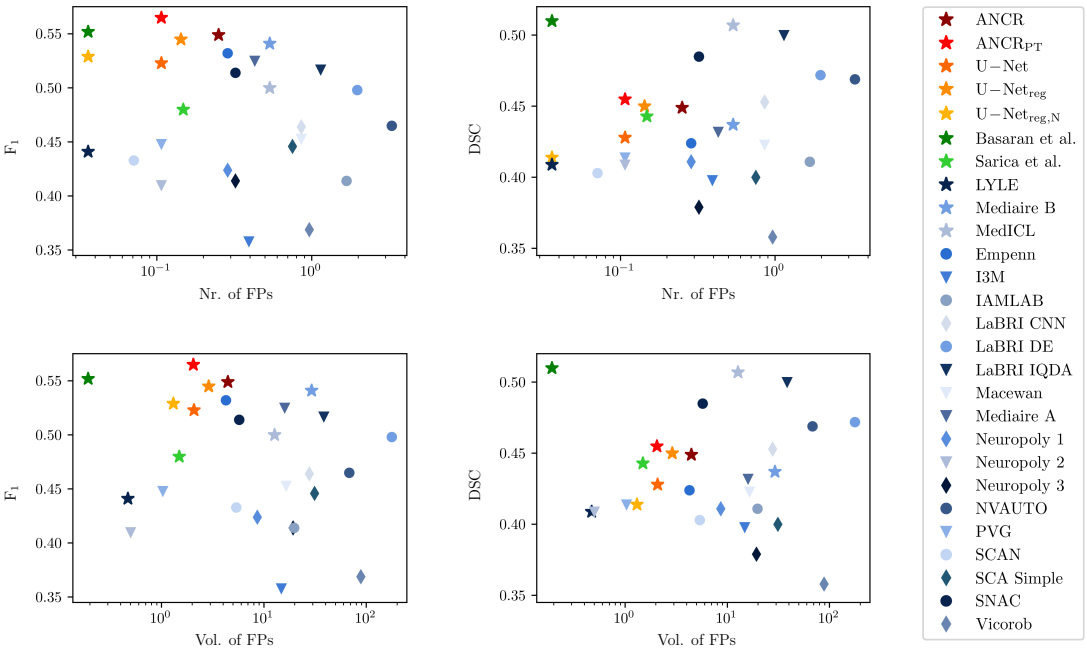


Figure 5.2: The trade-off between performance for progressive and stable patients. Methods participating in the MSSEG-2 challenge are displayed with blue markers, methods from the follow-up literature are shown with green markers, and the methods proposed in this thesis are given with red/orange markers. Most methods are biased to be over- or undersensitive, resulting in good performances for stable or progressing patients only. ANCR_{PT} and [Basaran et al., 2022] manage to reduce this tendency.

figure confirms the observation that most automatic methods are either under- or oversensitive, i.e. they either perform well for cases with new lesions or for stable cases without new lesions. Both ANCR-Net and the nnU-Net by Basaran et al. manage to ease this problem, performing well on both new lesion and no new lesion cases. The method by Basaran et al. thereby achieves less false positives but a slightly lower F_1 score.

The additionally proposed methods based on NCR- and U-Net provide similar results to ANCR-Net, but show slightly reduced lesion sensitivity (0.615 for ANCR-Net vs. 0.530, 0.606 and 0.502 for baseline U-Net, $\text{U-Net}_{\text{reg}}$ and $\text{U-Net}_{\text{reg,N}}$). NCR-Net alone produces many false positives, but still misses a lot of lesions. This can be explained by the oftentimes subtle appearance of MS lesions and comparably strong imaging artifacts. It was also tested to use images pre-processed with skull stripping and N4 bias field correction. While this reduced false positive predictions, a degradation in registration performance was observed. Using the images registered with NCR-Net as input for the segmentation U-Net improves lesion sensitivity, whereas the additional input of the non-correspondence map increases the slight undersensitivity observed in the baseline U-Net.

5.3.2 New Lesions Segmentation

To evaluate segmentation performance, three metrics are considered. The DSC is used for pixel-wise comparison, whereas the average symmetric surface and Hausdorff distances are used to focus evaluation on the boundary of lesions (cf. Sec. 2.3). The surface metrics are calculated with the `animaSegPerfAnalyzer` tool provided by the challenge organizers [Commowick et al., 2018]. Results are given in Tab. 5.2, comparing ANCR-Net to the same competitive methods as for the new lesion detection evaluation. Since the surface metrics can only be calculated for non-empty predictions, averaging of ASSD and HD results is performed over non-empty predictions only, and patients for whom the predicted segmentations are empty are listed in the table.

Segmentation results show the difficulty of the MS lesion segmentation, aggravated by the blurry lesion borders. Even the medical experts achieve mean Dice values of only 0.631, 0.536, 0.598 and 0.461. The best performing automatic methods MedICL and the nnU-Net by Basaran et al. achieve average DSCs of 0.507 and 0.510, respectively, surpassing the performance of expert 4 but lacking behind the other experts. ANCR-Net scores third with an average DSC of 0.455 close to expert 4. Both experts and automatic segmentation methods struggle with the same images, as can be seen from the missed cases.

The global DSC does not reflect how well individual lesions are delineated. Therefore, a local DSC is calculated for ANCR-Net$_{\mathrm{PT}}$ and the medical experts by calculating the DSC for individual lesions (found with connected component analysis) and averaging over detected lesions only. ANCR-Net$_{\mathrm{PT}}$ achieves a local DSC of 0.631 showing that the delineation of detected lesions is more exact than can be expected from the global scores in Tab. 5.2, but the gap to the experts is still large, who achieve an average local DSC of 0.817.

Furthermore, segmentation metrics are negatively affected by false positive lesions. In total, ANCR-Net generates 71 false positive lesions for the 60 test cases. Oftentimes, however, false positives predicted by ANCR-Net$_{\mathrm{PT}}$ correspond to image regions that were controversial among experts, too. 28 of the 71 falsely segmented lesion were also identified as new lesion by at least one expert, but not included in the ground truth. Examples of such disputed lesions are shown in Fig. 5.3. The false positive predictions might, thus, still be useful to identify conspicuous regions that need further investigation by experienced human experts.

Finally, the robustness of ANCR-Net to changes in overall image appearance is evaluated to show that new lesion detection and segmentation focuses on local differences between baseline and follow-up images. To do so, input images are artificially disturbed to an increasing extent and the performance of ANCR-Net is evaluated for different perturbation levels. Disturbances include Gaussian noise, reduced image resolution and typical MR imaging artifacts ghosting and intensity bias fields. Reduced image resolution is achieved via anisotropic downsampling to simulate different imaging protocols. For each of the four introduced disturbances, a manually chosen level is defined that is considered to be the level that no longer allows proper examination of the images (ex-

Table 5.2: New lesion segmentation results for images with new lesions in ground truth. Reported are average Dice score (DSC), average symmetric surface distance (ASSD) and Hausdorff distance (HD). The results are given for medical experts, ANCR-Net with and without pre-training, the three pipelines that performed best in the MSSEG-2 challenge (MedICL, Mediaire-B and LYLE), methods proposed in [Dufresne et al., 2022; Salem et al., 2022; Basaran et al., 2022; Sarica and Seker, 2022; Wu et al., 2023b] and several U-Nets. Values not given in the literature are marked with n/a. Best results are presented in bold font, and second best are underlined. Surface distances can only be calculated for non-empty segmentations; results are averaged over all predictions that are non-empty and patients for whom no lesion was predicted are given in the last column.

Method	Data	DSC ↑	ASSD ↓	HD ↓	Missed patients
Expert 1	Val.	0.730	5.075	24.215	32, 47
Expert 2	Val.	0.626	9.110	30.577	13, 32, 61, 88
Expert 3	Val.	0.654	9.824	32.671	20, 47, 88
Expert 4	Val.	0.644	7.370	31.074	13, 29, 77, 88
NCR-Net	Val.	0.178	46.186	84.998	-
U-Net	Val.	0.456	15.539	43.012	32, 88, 99
U-Net$_{\text{reg}}$	Val.	0.515	13.129	42.863	88
U-Net$_{\text{reg,N}}$	Val.	0.464	16.743	47.429	18, 88, 99
ANCR-Net	Val.	0.512	14.103	41.502	29, 32
ANCR-Net$_{\text{PT}}$	Val.	0.502	12.171	37.688	32, 47
Expert 1	Test	0.631	10.288	25.922	25, 31, 67
Expert 2	Test	0.536	13.247	34.330	22, 31, 67, 78, 85
Expert 3	Test	0.598	13.972	40.060	12, 25, 78, 85
Expert 4	Test	0.461	15.754	37.839	5, 12, 22, 25, 31, 44, 78, 85
MedICL	Test	0.507	19.212	61.821	-
Mediaire-B	Test	0.437	17.236	44.638	31, 78
LYLE	Test	0.409	12.975	37.919	22, 25, 31, 78, 85
Dufresne et al.	Test	0.356	n/a	n/a	n/a
Salem et al.	Test	0.420	n/a	n/a	n/a
Basaran et al.	Test	<u>0.510</u>	n/a	n/a	n/a
Sarica et al.	Test	0.443	n/a	n/a	n/a
Wu et al.	Test	**0.638**	n/a	n/a	n/a
NCR-Net	Test	0.097	49.765	84.984	-
U-Net	Test	0.428	20.286	49.851	22, 31
U-Net$_{\text{reg}}$	Test	0.450	16.972	44.380	31, 67
U-Net$_{\text{reg,N}}$	Test	0.414	21.825	48.756	22, 31
ANCR-Net	Test	0.449	21.641	48.241	-
ANCR-Net$_{\text{PT}}$	Test	0.455	14.837	41.774	31, 78

amples can be seen in Fig. 5.4). Three experiments are performed for all disturbances, manipulating either the baseline, the follow-up or both images before passing the images to ANCR-Net. Directional disturbances (ghosting, bias fields) are applied in different directions when disturbing both images. New lesion detection and segmentation performances of ANCR-Net for the different manipulated images are given in Fig. 5.4. For

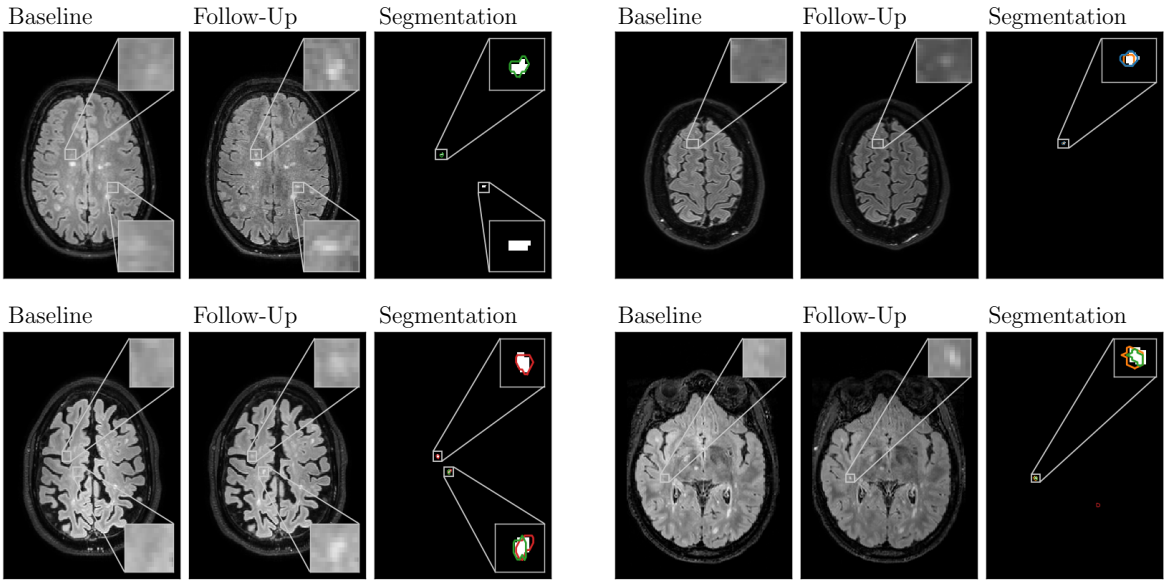


Figure 5.3: False positive lesions predicted by ANCR-Net that also were segmented by at least one radiologist but not included in the ground truth. Shown are patients 2 (upper left, slice 173), 4 (upper right, slice 316), 36 (lower left, slice 189) and 81 (lower right, slice 263). For each patient, the baseline and follow-up scans are shown as well as the segmentation generated by the proposed ANCR-Net. Experts' segmentations are overlaid in red, orange, green and blue for expert 1, 2, 3 and 4, respectively.

progressing patients, the average DSC is shown and for stable patients, the number of false positive lesions is shown. For both stable and progressing patients, the amount of correctly classified patients is given.

The results indicate that ANCR-Net reacts well to disturbed input images and delivers stable detection rates for most of the considered artifacts. ANCR-Net is able to separate stable and progressing patients, and the segmentation accuracy decreases only slowly, even with significant appearance differences between time points. This shows that the proposed network can directly be used in more realistic settings than in the challenge dataset, where different scanners might be used for different visits or imaging artifacts might be present. Ghosting and strong Gaussian noise lead to the greatest visual change in the images, and consequently also have the greatest impact on the detection performance. Thereby, disturbances in the baseline or follow-up image manifest themselves differently. While changes in the baseline generally lead to higher false positive rates in stable patients, changes in the follow-up worsen the detection rate in progressive patients. Likewise, the segmentation accuracy is impaired more by artifacts in the follow-up image. Since new lesions are visible in the follow-up image only, this corresponds to the expected behavior.

Artifacts covered during network training tend to have less of a negative impact on the network's performance. The training dataset comprises images with very different resolutions and image sizes, that have been rescaled to the same horizontal resolution.

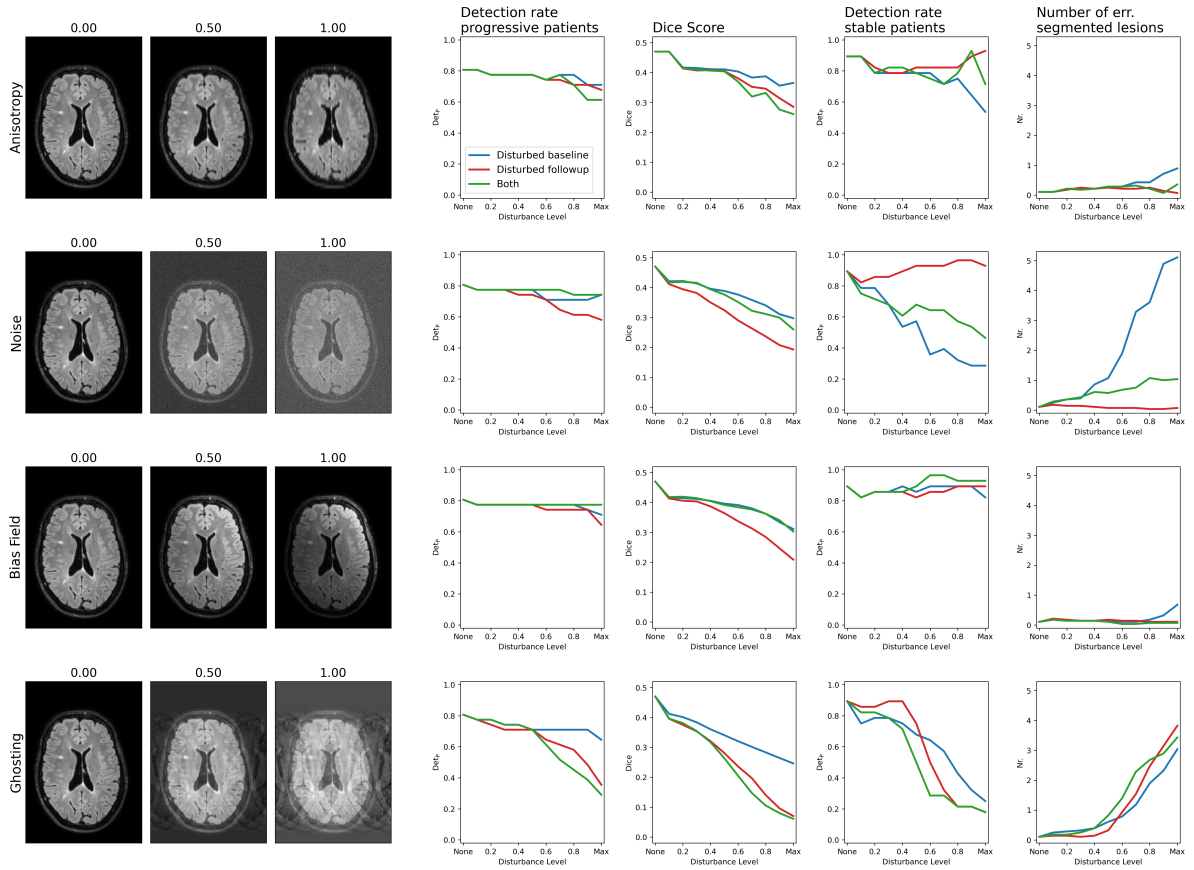


Figure 5.4: Influence of imaging artifacts on the performance of ANCR-Net. Images are artificially disturbed at increasing level before being passed to the network. For each artifact considered, an example image is shown for no, moderate and severe disturbance (disturbance levels 0.0, 0.5 and 1.0). For progressive and stable patients, the patient-level detection rate is shown. For progressive patients, the DSC is given additionally, and the number of false positive lesions for stable patients. Blue lines show results for disturbed baseline images, the red lines for disturbed follow-up images, and the green lines for both images being disturbed.

The artificially introduced anisotropy, that reflects reduced sampling rates during image acquisition, thus, leads to almost no impairment in the detection performance of ANCR-Net. Likewise, the artificial bias fields were covered with random intensity gradients as augmentation technique during training. Here, again, a very stable performance can be observed. Thus, the method's robustness might potentially further be improved by integrating typical MR imaging artifacts like ghosting in the data augmentation.

Overall, ANCR-Net proofs robust against altered input images, despite relying on appearance differences between time points. Even in the presence of heavy disturbances, the proposed method manages to separate stable and progressive patients and delivers stable segmentation accuracy for most artifacts. This highlights the applicability of

ANCR-Net for real clinical data that exhibits an even greater variety than the challenge data.

5.3.3 Modeling of New Lesions

New MS lesions can take different courses after a first phase of inflammatory demyelination. Histopathologically, four main MS lesion subtypes have been identified: early active, chronic active, inactive and remyelineated. Chronic active lesions are characterized by an inactive demyelinated center and persistent inflammation at the lesion edges, and are associated with higher overall lesion load and worse clinical outcome [Calvi et al., 2022; Kolb et al., 2022]. The identification of imaging biomarkers for the examination of chronic active lesions is ongoing research. For example, chronic active lesions can be identified with quantitative susceptibility mapping, in which they show a hypointense rim surrounding the lesion core [Voon et al., 2024]. However, MRIs taken routinely in clinical practice are "highly sensitive in detecting white matter lesions but not specific for histological subtypes/stages of lesions" [Calvi et al., 2022].

ANCR-Net allows to simultaneously register baseline and follow-up images, to segment new lesions and to model the appearance of the newly formed lesions. While this is exploited during training to compensate for outliers in the image distance measure to assist the joint image registration and segmentation task, the appearance outputs can also be used during inference to investigate intensity patterns of new MS lesions. Fig. 5.5 shows qualitative results for image registration and appearance adaptation using ANCR-Net. It is shown exemplarily that the appearance offset maps successfully manipulate the baseline image to resemble the follow-up image. The intensity differences observed inside and outside the introduced lesions are comparable, showing that the network does not overcompensate for overall intensity differences between baseline and follow-up but fits the intensity distribution of the baseline image.

As can be seen in Fig. 5.5, MS lesions primarily appear as bright spots in FLAIR MR images, making it difficult to distinguish different pathology areas. However, the appearance maps model the newly created lesions as changes that need to be made to the baseline image in order to look like the follow-up image, making it possible to recognize different intensity patterns within the lesions. This is visualized with the appearance offset maps masked to the lesion area in the third image in the upper row of each example. Here, some of the inserted lesions show a uniform intensity, while others show irregular intensity patterns with higher values in the border areas (patient 1 and large lesions of patients 60 and 66). Therefore, the modeled lesions can potentially be used as a new imaging biomarker to distinguish between different types of lesions on FLAIR MR images and to separate active from inactive areas within individual lesions. As there is no ground truth data to differentiate the lesions included in the MSSEG-2 dataset, this claim cannot be verified and is left for further research.

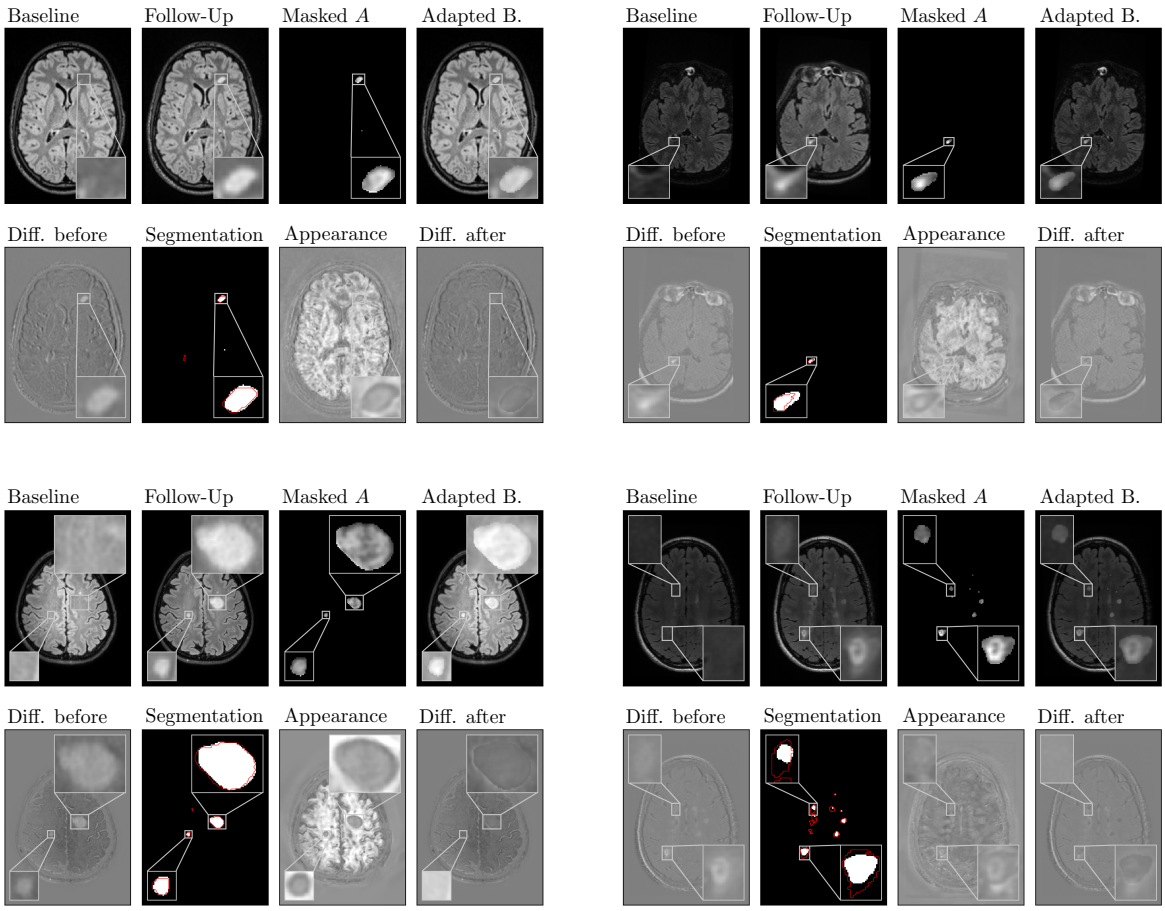


Figure 5.5: Examples of modeled new MS lesions. For each patient, the baseline and follow-up MR images are shown, as well as their subtraction image. The segmentation and appearance offset map output by ANCR-Net are shown in the bottom row of each example. The ground truth segmentation is overlaid in red onto the predicted segmentation. The appearance map is masked and added to the baseline image before registering the images with the deformation field that is also generated by ANCR-Net. The lesions modeled in this way make the adapted baseline image look like the follow-up image (top right in each case). Some of the modeled lesions show homogeneous intensity, whereas others allow the visualization of heterogeneous intensity patterns, visible in the masked appearance maps. Results are shown for patients 1, 6, 60 and 66.

5.4 Discussion

In this chapter, the challenging task of detecting and segmenting new MS lesions was tackled with an image registration CNN, that simultaneously with the registration segments new lesions and models their appearance as offsets from the baseline image. This model is called ANCR-Net and is a direct extension of NCR-Net from Chapter 4. ANCR-Net is compared to the best-performing submissions from the MSSEG-2 challenge, state-

of-the-art methods from the literature and several U-Nets that use the unsupervised non-correspondence maps and registration results of NCR-Net for guidance.

From a clinical point of view, detection performance is the most important outcome measure in the automatic examination of new MS lesions. Detection methods should reliably identify new lesions while simultaneously generating few false positives to correctly separate progressing from stable patients. For progressive patients, many automatic methods achieved performances close to the medical experts but generated high number of false positives for stable patients. ANCR-Net showed the highest F_1 score for new lesion cases and second-best score for the number of false positives in stable cases. Except for the state-of-the-art method by Basaran et al., ANCR-Net, thus, shows the best compromise between correct lesion detection and a low number of false positives and allows a reliable separation of stable and progressive patients. Only two progressive patients are overlooked, which is better than all medical experts, and three stable patients are classified as progressing. Interestingly, the automatic methods struggle with the same images as the medical experts (missed patients in Tab. 5.2), highlighting the possibility to improve results by focussing on especially difficult cases using e.g. online hard example mining as done in [Schmidt-Mengin et al., 2022].

Regarding segmentation performance, ANCR-Net scored third with an average DSC of 0.455. Even the best performing medical expert achieved an average DSC of only 0.631 showing that MS lesion delineation is a highly subjective task exhibiting high inter-rater variability. Since lesion numbers vary a lot between patients and lesions are small, the global DSC is not a suitable metric to assess how well individual lesions are delineated. For a better evaluation of segmentation accuracy, the DSC was also calculated for individual lesions and averaged over detected lesions only. Here, ANCR-Net achieves an average score of 0.631 which shows that detected lesions are segmented well. Still, there was a large gap to medical experts, who achieve an average local DSC of 0.817.

The MSSEG-2 dataset comprises a large variety of image sizes, resolutions, MR scanners and suffers from a severe class imbalance problem. Thus, in order for DL models to learn to segment new MS lesions, a sophisticated data preprocessing and training procedure is required. In this work, all images were resized to the same horizontal resolution and heavy data augmentation was used during training to help generalization. To prevent empty segmentations, network training was performed on image patches containing new lesions only. This data selection process leads to about a third of the given images to not being used during training (the stable patients without new lesions). To still make use of these cases, a pre-training step is introduced that uses the stable patients' images with artificially inserted lesions. In addition, ANCR-Net was shown to be robust against several typical MR imaging artifacts and can directly be used for real clinical data, that might exhibit even more variations and lower image quality than the challenge dataset.

Another problem associated with the nature of MS lesions is that their small size makes it difficult to select good models. Evaluation metrics can change drastically for only slightly modified segmentations, making the interpretation of results difficult and sometimes contradictory. Many published papers report only a few of the metrics

considered in the MSSEG-2 challenge, which further complicates the interpretability of the results. Therefore, a comprehensive evaluation with many reported metrics was carried out here. For further research, it is recommended to proceed in the same way and to search for evaluation metrics better suited for the small and blurry delineated new MS lesions.

Finally, it was shown qualitatively that the modeled lesions are well suited to adapt the baseline images to look like the follow-up images. When visualizing the masked appearance offsets, different intensity patterns could be observed within lesions. Further investigations are needed to assess the actual usefulness of these patterns for the classification of MS lesions. In particular, it would be interesting to extend the proposed method to several imaging modalities in which MS lesions can be better differentiated.

Overall, the detection and segmentation of new MS lesions remains a very difficult task. No method proofed able to outperform other methods substantially on all metrics, but some, including the proposed methodologies, already achieve performances close to the medical experts. Both ANCR-Net and the combination of NCR- and refinement U-Net offer reliable methods for the separation of stable and progressing patients and for the estimation of the new lesion load. While the NCR-Net extension needs two deep CNNs to be trained, ANCR-Net is more lightweight with only one CNN and offers the additional possibility to model the intensity profile of new lesions, enabling morphology analyses of individual lesions.

Chapter 6

Pathology-aware Image Registration in Retinal OCT Images

Chronic eye diseases such as AMD or CSCR are characterized by pathological fluid deposits below and inside the retina that lead to a severe loss of vision. It is known that the disease state can change very quickly, which can lead to large amounts of new fluid being detected at a visit, even though no fluid was present at the previous examination. There are various treatment options to reduce or even completely dissolve existing fluids. The dynamic nature of retinal fluid growth leads to images that vary significantly between visits, not only in appearance but also in shape. While the longitudinal registration of MR images of MS brains, investigated in the previous chapter, requires rather small deformations, the longitudinal registration of retinal OCT images requires much larger displacements in order to capture the dynamics of fluid development. However, even very large deformations cannot depict the differences between time points in cases where structures disappear completely or are newly evolving. New as well as dissolving fluids lead to image areas for which there are no corresponding pixel values in the image of the previous visit.

The approaches presented in Chapters 4 and 5 could be used for this scenario and would provide segmentations of the non-corresponding areas. However, the resulting deformation fields would not reflect the real behavior of the fluids: Retinal layers would be stretched or compressed to cover emerging and disappearing fluids. In reality, the retinal layers are displaced by fluids and hardly expand or shrink. Therefore, a method that

- shifts the retinal layers without significant volume change, while keeping the possibility for a strong volume change in the region of pathological fluids
- offers the possibility for new fluids to appear between layers

is needed.

In this chapter, another deep learning-based registration framework is presented that explicitly addresses the problem of newly forming retinal fluids by a sparse appearance seed approach. In a fully unsupervised manner, the CNN learns to insert new fluids artificially into the moving image as small areas of changed intensities, which are then stretched to cover the fluid area in the fixed image by the deformation field. Dissolving fluids are also covered within the framework by masking out the deformation field regularizer in fluid regions. As a trade-off between the deformation field regularity, the image similarity and the sparsity of the appearance offset map, the CNN is trained to generate not only the registration deformation field but also the fluid seeds. No manual

fluid annotations are required for training, but only rough fluid segmentations that are automatically generated from the OCT images with a simple thresholding approach. It is shown that the proposed fluid registration network, FluidRegNet, leads to more realistic deformation fields with reduced volume change of normal retinal tissue. Furthermore, the appearance offset maps are used together with the deformation fields to detect and segment new fluids in an unsupervised manner. FluidRegNet was published in [Andresen et al., 2024] and used there for the chronological segmentation of OCT time series data, where manual segmentations are only required for the first examination as a starting point for the follow-up segmentations. In the second part of this chapter, FluidRegNet is additionally trained for inter-patient registration of healthy and pathological images. This extension is used for the detection of anomalies in the pathological images, which enables the segmentation of all pathologies rather than only the new ones. The anomaly detection is used to segment fluids IRF, SRF and PED as well as deviations from the normal reflectivity of the photoreceptor ellipsoid zone (PEZ).

6.1 Related Literature

While image registration techniques are often used for motion correction [Ricco et al., 2009; Kraus et al., 2012; Li et al., 2020; Makita et al., 2020; Ntatsis et al., 2022] and denoising [Alonso-Caneiro et al., 2011; Baghaie et al., 2015; Zhang et al., 2015; Cheng et al., 2016; Ma et al., 2018; Shi et al., 2019; Wang et al., 2021a; Cheng et al., 2021] of OCT images, and multimodal registration of en-face retinal images is also widely studied [Chen et al., 2010; Ghassabi et al., 2013; Wang et al., 2015; Hossein-Nejad and Nasri, 2018; Lee et al., 2019; Santarossa et al., 2022a], the longitudinal registration of OCT images has been addressed less often and mostly for non-pathological images [Niemeijer et al., 2009; Lang et al., 2016; Lee et al., 2017; Gong et al., 2019]. Probably due to the missing correspondences introduced by the dynamic changes in fluid presence, size and extent, there appears to be no image registration method in the literature to date that can generate realistic deformation fields for pathology cases without the need for tedious manual annotations of retinal layers [Wei et al., 2017; Pan et al., 2019, 2020]. Additionally, they fail to realistically capture the dynamics introduced by new (and dissolving) fluids, since the deformations do not explicitly take missing correspondences into account. In [Uzunova et al., 2022], generative adversarial networks (GANs) are used to generate pathological OCT images. Relying on segmentations of retinal layers of healthy images and fluid segmentations of pathological images, first a GAN is trained to predict the pathology-induced deformations and the deformed retinal layer and pathology segmentations then serve as input for a label to image translation network. However, this approach models the fluid-induced deformations only implicitly and does not guarantee them to be realistic.

Metamorphosis models have been developed to tackle the problem of differing appearances between image pairs to be registered. Here, appearance and shape differences are jointly estimated in the same process, which leads to more exact and realistic registration

results [Niethammer et al., 2011; Meng et al., 2022]. One idea to model the deformations introduced by pathologies with metamorphosis models is to use appearance offsets in pathological image regions only. Most existing methods that do this, require given manual annotations of the pathologies [Niethammer et al., 2011; Maillard et al., 2022; François et al., 2022; Joshi and Hong, 2023; Wang et al., 2023]. In [Uzunova et al., 2023], low-rank and sparse metamorphic autoencoders are used to separate pathology and general appearance offset maps. However, the disentanglement of pathology-related and inter-patient variations between images is not yet accurate enough, resulting in "normal" non-correspondences between patients being mapped in the pathology offset map and vice versa. Inspired by metamorphosis models, the here presented FluidRegNet explicitly models newly developing fluids as sparse appearance changes, which are then extended with the deformation field. The resulting deformations can better reflect the changes actually occurring in the eye than previous registration methods, and are used for unsupervised pathology segmentation [Andresen et al., 2024].

Another way of enabling the unsupervised segmentation of pathologies, is the usage of anomaly detection. Anomaly detection approaches aim to recognize features that deviate from the normal data distribution, where normal is defined as healthy in the medical field. Instead of training the CNNs to segment a specific type of pathology, which would require extensive labeling of the images, the networks are trained to recognize anything that looks other than healthy. The most common approaches for DL-based anomaly detection are based on autoencoders that are trained to reconstruct healthy images, e.g. [Uzunova et al., 2019; Venkataramanan et al., 2020; Shvetsova et al., 2021; Tian et al., 2023]. These models are expected to fail to reconstruct image parts that contain out-of-distribution pathologies and to replace them with in-distribution, i.e. healthy-looking, anatomy. The residual error between the original and the reconstructed image is then used to locate pathologies. In addition to the reconstruction error, patch-based methods allow the localization of anomalies through the analysis of the encoded patch representations in the latent space, e.g. [Uzunova et al., 2019]. While autoencoder-based approaches provide good results for anomalies whose intensities are far from those of healthy tissue, these methods have been shown to have a blind spot for abnormalities whose intensity range is approximately equal to that of healthy tissue and to ignore structural features of the tissue [Meissen et al., 2022].

Self-supervised anomaly segmentation networks are trained with synthetic anomalies, which are segmented in a supervised manner. This enables the CNNs to directly predict an anomaly score rather than relying on the reconstruction error, e.g. [Li et al., 2021; Schlüter et al., 2022; Baugh et al., 2023]. [Baugh et al., 2023] use five types of synthetic anomalies, i.e. intra-dataset and inter-dataset blending, sink and source deformations as well as masking with smooth blending at the borders of the masked area. Using these tasks, they are able to segment real pathologies in brain MR images and won the Medical Out-of-Distribution Analysis Challenge 2024 (MOOD, [Zimmerer et al., 2022]).

Anomaly detection methods developed for OCT images can be found in [Schlegl et al., 2019; Seeböck et al., 2020; Mou et al., 2022; Li et al., 2023]. Schlegl et al. use a GAN, where the generator is trained to reconstruct healthy images and the discriminator dif-

ferentiates real from fake images. An anomaly score is calculated based on the reconstruction error and the difference between features generated within the discriminator network for real and reconstructed images. Seeböck et al. and Mou et al. train Bayesian U-Nets to segment retinal layers on healthy images and use Monte Carlo dropout to estimate epistemic uncertainty of the segmentations. The uncertainty maps are post-processed to give segmentation of pathologies [Seeböck et al., 2020; Mou et al., 2022]. In [Li et al., 2023], SSL is used to enhance the training of an autoencoder and the reconstructed images are further used to stage anomalies and to segment retinal pathologies. They report excellent classification and moderate segmentation results. This reflects the results observed for most state-of-the-art anomaly detection methods that achieve very good detection performances but struggle with the exact localization and delimitation of anomalous regions.

6.2 Methods

The goal of this chapter is the longitudinal intra-patient registration of pathological OCT images with realistic, fluid-induced deformations. Since for newly emerging fluids, there are no corresponding pathology pixels in the baseline image, it is not sufficient to use deformation fields to match baseline and follow-up images. Instead, it is proposed to imitate the start of fluid formation by changing the intensity of small areas of the baseline moving image: So-called fluid seed points are implicitly learned and placed into the moving image. The moving image with adapted appearance is then registered with the reference scan and the fluid seed is extended to the area of fluid observed in the reference space.

6.2.1 Sparse Appearance Seed Approach

As in metamorphic approaches, it is assumed that the fixed image F can be represented as a deformed and appearance adapted version of the moving image M, i.e.

$$F \approx (M + A) \circ \varphi. \tag{6.1}$$

Here, A is the appearance offset and φ the deformation field. For network training, the following objective is used:

$$\mathcal{L} = \mathcal{L}_{\text{Dist}}\big(F, (M + A) \circ \varphi\big) + \mathcal{L}_{\text{Reg}}(\varphi) + \mathcal{L}_{\text{Sparse}}(A), \tag{6.2}$$

where $\mathcal{L}_{\text{Dist}}$ is an image distance loss, $\mathcal{L}_{\text{Reg}}$ regularizes the deformation field and $\mathcal{L}_{\text{Sparse}}$ favors sparse appearance offsets. The normalized cross-correlation image distance proved to be well suited for OCT registration in the experiments carried out for this thesis and is used here for $\mathcal{L}_{\text{Dist}}$. By requiring sparsity of the appearance offset map A, it is used to model the starting point of newly emerging fluids. The appearance adapted moving

image $M + A$ is then registered onto the fixed image such that the fluid seed is extended to match the volume of the pathology in the reference image.

To achieve realistic deformations, a multi-component loss function is used to regularize the deformation field:

$$\mathcal{L}_{\text{Reg}}(\varphi) = \alpha \mathcal{L}_{\text{Dice}}\left(\mathrm{S}^F_{\text{retina}}, \mathrm{S}^M_{\text{retina}} \circ \varphi\right) + \beta \mathcal{L}_{\text{Jac}}(\varphi) + \gamma \mathcal{L}^{\text{masked}}_{\text{Diff}}(\varphi). \tag{6.3}$$

As in the previous chapters, the Dice loss $\mathcal{L}_{\text{Dice}}$ between the retina segmentation of the follow-up image $\mathrm{S}^F_{\text{retina}}$ and the warped retina segmentation of the moving image $\mathrm{S}^M_{\text{retina}}$ gives weak guidance to the registration task by ensuring good overlap of the entire imaged tissue.

A masked diffusion loss $\mathcal{L}^{\text{masked}}_{\text{Diff}}$ is used to represent the expected behavior of the deformation: The normal tissue should be smoothly displaced and not changed in volume, while no such assumption can be made for the deformation in fluid regions. The fluid volume can change drastically between visits and to varying degrees for each patient and image pair. Therefore, arbitrary deformations within the fluids are allowed, and the regularizer is only applied to the healthy tissue. For this purpose, rough segmentations of pathology regions are generated as follows: The 70th percentile of the B-scans' intensities is used for binary thresholding of moving and fixed images. The background is removed from the masks using the manual retina segmentations. The resulting fluid segmentations are dilated with an approximately circular structuring element of 11×11 pixels. Finally, the fluid region $\mathcal{O}$ that is not considered in the evaluation of the diffusion loss is defined as the combined fluid regions of fixed and moving images. The masked diffusion loss is given by

$$\mathcal{L}^{\text{masked}}_{\text{Diff}}(\varphi) = \frac{1}{|\Omega \setminus \mathcal{O}|} \sum_{\boldsymbol{x} \in \Omega \setminus \mathcal{O}} \sum_{i=1}^{2} \|\nabla \varphi^{(i)}(\boldsymbol{x})\|_2^2, \tag{6.4}$$

where Ω is the set of all image pixels, and i indicates the spatial dimension. The masked diffusion loss implicitly also tackles the problem of dissolving fluids, which, as the new pathologies, lead to non-correspondences between time points. Inside fluids, the deformation is only very slightly restricted by avoiding inversions in the deformation field. To do so, negative values of the Jacobian determinant of the deformation field J_φ are penalized on the entire image domain using

$$\mathcal{L}_{\text{Jac}}(\varphi) = \frac{1}{|\Omega|} \sum_{\boldsymbol{x} \in \Omega} \max\left(0, -|J_\varphi(\boldsymbol{x})|\right)^2. \tag{6.5}$$

The final term in loss function (6.2) is the appearance sparsity loss

$$\mathcal{L}_{\text{Sparse}}(A) = \delta \sum_{\boldsymbol{x} \in \Omega} \left(|A(\boldsymbol{x})| + \eta \max\left(0, A(\boldsymbol{x})\right)^2\right) \tag{6.6}$$

that enforces sparse appearance offsets. The second term in Eq. (6.6) additionally encourages negative values of A, which reflects the low intensities of fluid-filled regions that need to be inpainted onto the light retinal tissue.

6.2.2 Network Architecture and Training Details

The architecture of the proposed network for fluid-aware image registration (FluidRegNet) is shown in detail in Fig. 6.1. A U-Net-like architecture is used with two separate input blocks, as in the previous chapters. Different from (A)NCR-Net, FluidRegNet does not have two separate decoders but only two separate output blocks. The first input block receives a two-channel image consisting of corresponding B-scans of fixed and moving images, and the second input block receives the subtraction image of fixed and moving scan. Each input block consists of two convolutional layers with kernel size 3×3 followed by batch normalization and leaky ReLU. The eight feature maps resulting from each input block are concatenated and then passed through the next layers of the network together. The U-Net core of FluidRegNet comprises four levels of resolution, on each of which two convolutions are performed, followed by batch normalization and leaky ReLU activation. The decoding part of the network mirrors the encoder. Bilinear interpolation is used for upsampling of the feature maps. The two output blocks use 1×1 convolutions and generate the appearance offset map and the deformation field, respectively.

Given suitable weighting parameters, the training objective (6.2) enables the network to automatically learn to compromise between appearance sparsity and deformation regularity to achieve similar looking images F and $(M + A) \circ \varphi$. Practically, it has proven favorable to support the registration task with a three-step training scheme:

1. Pre-training of network backbone and deformation output block without appearance offsets using loss function $\mathcal{L} = \mathcal{L}_{\text{Dist}}(F, M \circ \varphi) + \mathcal{L}_{\text{Reg}}(\varphi)$.

2. Warm-up step: The pre-trained network parts are kept frozen and only the appearance output block is trained.

3. Fine-tuning of the entire network.

Steps 1 to 3 are performed for 200, 50 and 450 epochs, respectively, to ensure full convergence. Network training is performed with a batch size of 10, Adam optimization, an initial learning rate of 1e-4 and a learning rate decay of 0.8 after every 50 epochs. The weight parameters α, β, γ, δ and η are found empirically and set to 10, 1000, 1, 3e-5, and 6.4e-3.

Network training is performed with a subset of the longitudinal CSCR OCT dataset described in Sec. 2.4.1. The training dataset consists of 369 images from 61 eyes of 33 CSCR patients, all of which have a field of view of 2×6×6 mm^3 and an image size of 496× 512×25 voxels. For all images, automatically generated retina segmentations provided by the OCT device are available and for 19 patients, the automatic segmentations were manually revised and supplemented by annotations of IRF, SRF and PED. The CSCR dataset is described in detail in Tabs. A1 - A3. The tables include for each patient the number of available OCT and segmentation images, as well as the information whether the images were used for the training of FluidRegNet. For FluidRegNet training, the images are pre-processed with a flattening at the Bruch's membrane, denoising with

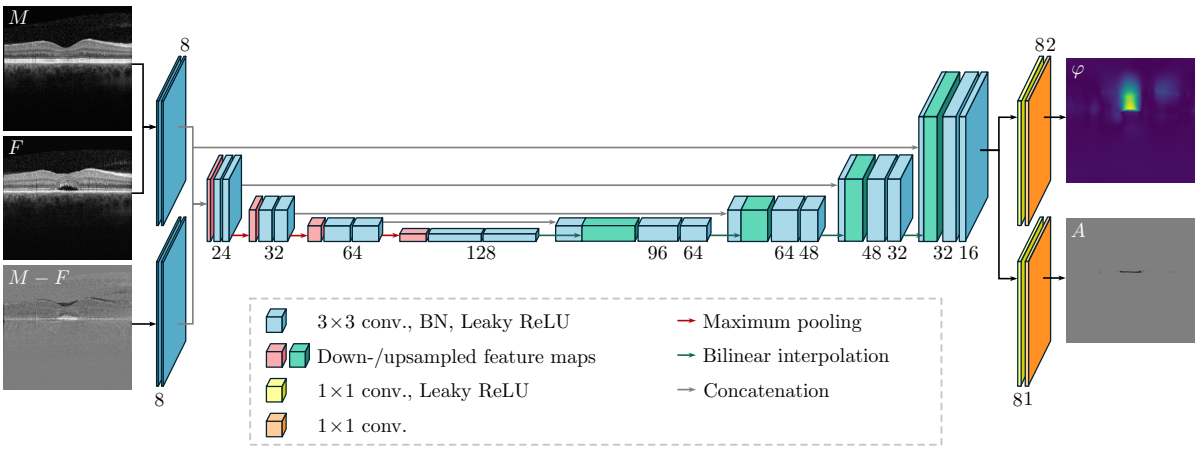


Figure 6.1: The architecture of FluidRegNet. Two separate input blocks process the OCT images to be registered and their difference image. Two output heads generate the deformation field φ and the appearance offset map A. The numbers below the boxes indicate the number of feature maps resulting from the respective convolutions.

a guided filter with radius 1 [He and Sun, 2015] and intensity normalization to $[0, 1]$. All experiments are performed with a five-fold cross-validation, splitting the data at patient-level and using approximately 80% of the images for training.

6.3 Experiments and Results

In this section, several experiments are performed to evaluate the registration accuracy of FluidRegNet and the realism of the generated deformation fields. Additionally, it is shown that the proposed framework can be used for the unsupervised detection of new fluids and an automatic evaluation of disease progression.

6.3.1 Registration Accuracy

For the given data, no anatomical labels apart from the ILM and the BM are given. Since the images are flattened at the BM, the registration accuracy can only be assessed with regard to the alignment of the ILM. Deformation realism, however, can be evaluated for the images with given pathological annotations by measuring volume change inside and outside of fluids.

Registration results are compared to two other image registration methods, VoxelMorph [Balakrishnan et al., 2019] and symmetric image normalization (ANTs SyN) [Avants et al., 2008]. Additionally, several ablation studies are performed to assess the influence of different components of the framework on the registration performance of FluidRegNet. In ablation 1, FluidRegNet is trained without the three-step training

scheme, adapting all network parameters throughout 700 epochs. For ablation 2, the appearance offset map is not used. In ablation 3, again, the appearance offset map is not used, but additionally the masked diffusion regularizer is replaced with regularization on the entire image domain. Finally, to assess the influence of the difference image input, in ablation 4, the difference image is not given to the network. To do so, the network architecture is slightly adapted by removing the respective input block. VoxelMorph is trained with the same loss function and for the same number of epochs as ablation 3. Since flattened images are used for registration, ANTs SyN is performed without initial affine transformation. Here, for comparability with the other methods, the cross correlation image similarity metric is used. In contrast to the default settings of ANTs SyN, four resolution levels are used, as three did not give satisfactory results for the CSCR images. Otherwise, the default settings are retained.

For evaluation, only those images for which manual annotations are available are used. In contrast to the training data, images with various numbers of B-scans are used for validation, provided that the images from consecutive visits were taken with the same resolution. Additionally, the CSCR dataset and the manual annotations are continuously being expanded, such that a further 15 annotated images of three patients (for two of which segmentations were already available at the time of network development, one was newly added) are available for evaluation at the time of writing. In total, 177 3D OCT images from 20 patients (3403 2D B-scans) are used for the evaluation. Network training was performed with five-fold cross-validation, and the evaluation results are averaged over all test images. For new patients, all five networks are used and the results of the individual networks are averaged.

Table 6.1: Registration results for VoxelMorph (VXM), symmetric image normalization (SyN), FluidRegNet (FRN) and several ablation studies: First, FRN is trained without the three-step training (Abl. 1). For ablations 2 and 3, FRN is trained without the appearance offset map, using masked (Abl. 2) and full (Abl. 3) diffusion regularization. Finally, the difference image between moving and fixed scan is not given as additional input (Abl. 4). Reported are ILM surface distances in μm before and after registration, the amount of inversions introduced by the deformation ($|J_\varphi| \leq 0$) given as the average percentage of pixels with negative Jacobian and the volume change of healthy tissue and fluids. Best and worst results are highlighted in bold and red, respectively.

Method	ASSD ILM ↓	HD ILM ↓	$\lvert J_\varphi\rvert \leq 0$ [%] ↓	$\lvert 1-\lvert J_\varphi\rvert\rvert_{\text{healthy}}$ ↓	$\lvert 1-\lvert J_\varphi\rvert\rvert_{\text{fluid}}$ ↑
Before	14.93 (23.40)	37.95 (56.56)	-	-	-
Abl. 1	8.41 (16.93)	23.50 (43.61)	**0.0109** (0.0259)	0.0854 (0.0538)	**1.00** (0.77)
Abl. 2	7.78 (15.32)	21.98 (40.46)	0.0258 (0.0692)	0.0914 (0.0612)	0.93 (0.73)
Abl. 3	8.39 (16.26)	24.17 (44.74)	0.0160 (0.0411)	0.0863 (0.0565)	0.36 (0.17)
Abl. 4	8.23 (16.55)	23.11 (42.91)	0.0615 (0.1698)	0.0862 (0.0541)	0.98 (0.76)
VXM	9.89 (19.42)	27.96 (51.08)	0.0289 (0.0572)	**0.0836** (0.0474)	0.30 (0.13)
SyN	**5.54** (9.86)	**16.56** (37.53)	0.2498 (0.4016)	0.1327 (0.0665)	0.46 (0.21)
FRN	8.01 (15.93)	22.55 (41.70)	0.0209 (0.0546)	0.0873 (0.0550)	0.98 (0.78)

Results are reported in Tab. 6.1, giving the ASSD and the HD of the ILM as well as several evaluations of the deformation fields. To assess the regularity of the deformation, the average number of voxels with negative Jacobian of the deformation field is specified. Finally, the realism of the deformations is evaluated by assessing the volume change separately for the healthy tissue and the fluid regions. Tab. 6.1 shows that VoxelMorph generates very uniform deformation fields. These, however, have the disadvantage that the inner limiting membrane is not sufficiently aligned. Conversely, SyN can align the ILM very well with an ASSD of 5.54 μm, but the deformation fields contain many inversions (3292 out of 3403 image registrations affected) and change not only the volume of the fluids, but also the healthy tissue. FluidRegNet is the only method that generates deformation fields capable of both concentrating volume changes in the pathological regions and generating few inversions, with average volume changes of 0.98 in pathology and 0.09 in non-pathology regions, and 0.02% foldings on average. The generated deformation fields are, therefore, far more realistic than those of the other registration methods, with ILM matching accuracy between that of ANTs SyN and VoxelMorph. This can also be observed in Fig. 6.2 that shows exemplar registration results for the three frameworks. In particular, it shows how the volume changes introduced by FluidRegNet are concentrated in fluid regions, whereas the healthy tissue is displaced according to the fluid changes (Subfigs. (a) - (e)). Furthermore, FluidRegNet can deal with imaging artifacts such as empty image areas due to eye movements, as shown in Subfig. (b). Finally, an example of a stable eye is shown in Subfig. (f), demonstrating that the learning-based methods are better able to generate uniform deformation fields when no changes occur.

From the ablation results in Tab. 6.1, it can be seen that the three-step training scheme slightly improves the alignment of the inner limiting membrane (ablation 1, ASSD of 8.41 μm vs. 8.01 μm). Training the proposed network as a "classical" image registration framework, i.e. performing spatial regularization on the entire images and using no appearance offsets (ablation 3), leads to overly smooth deformation fields similar to VoxelMorph results. Still, registration performance is better for the proposed network architecture than for the VoxelMorph network, with an ASSD of 9.89 μm for VoxelMorph and 8.39 μm for ablation 3. Introducing masked regularization (without appearance offsets) already improves registration performance and also leads to volume changes focusing on fluid regions (ablation 2, ASSD of 7.78 μm and volume changes of 0.09 in healthy tissue and 0.93 inside pathologies). Still, the appearance offset maps reduce inversions in the deformation fields and further improve the volume preservation of the normal tissue. With ablation study 4, it can be seen that the additional information provided by the difference image input improves the registration accuracy and reduces the size of inversions introduced by the deformations, since the network trained without the difference image input introduces 0.06% foldings and an ASSD of the ILM of 8.23 μm.

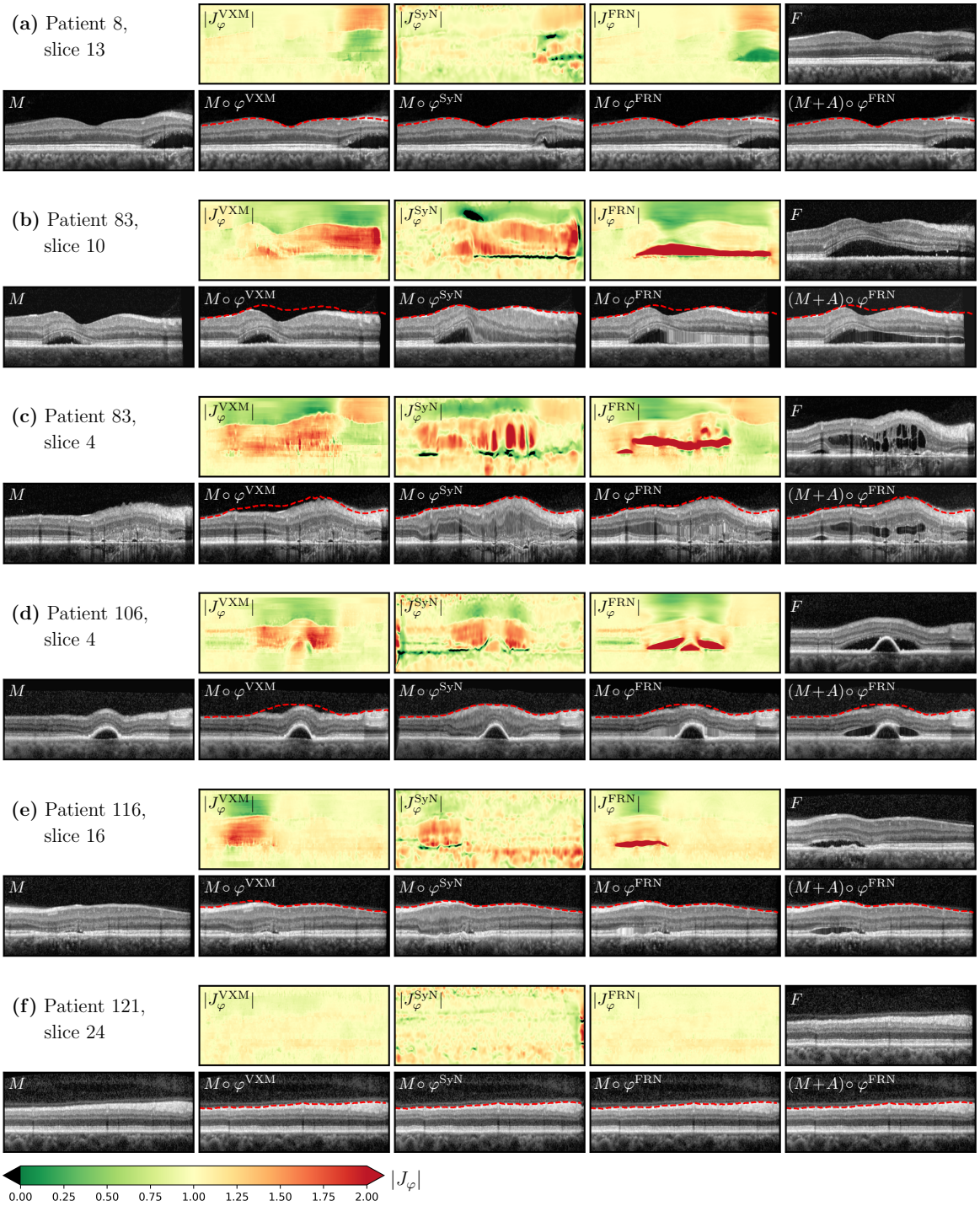


Figure 6.2: Exemplar registration results for VoxelMorph, ANTs SyN and FluidRegNet. Shown are the fixed image, the original and the warped moving image, as well as the respective Jacobian determinant of the deformation field. For the warped moving images, the ILM of the fixed image is overlaid in red. Subfig. (a) shows an example of a shrinking fluid, Subfigs. (b) - (e) show examples of new or growing fluids and in Subfig. (f) an unchanged eye is shown.

6.3.2 Detection and Segmentation of New Pathologies

The core idea of FluidRegNet is to capture new fluids in the registration process by introducing sparse intensity offsets and increasing the volume of the inpainted regions with the deformation field. Based on this, the trained network may be used to detect and segment newly developed pathologies by evaluating the network outputs A and φ:

$$\mathcal{S} = \{\boldsymbol{x} \in \Omega \mid A(\varphi(\boldsymbol{x})) < \tau_A \vee |J_\varphi(\boldsymbol{x})| > \tau_\varphi\}, \tag{6.7}$$

where $\mathcal{S}$ defines the set of pixels segmented to belong to new fluids. The thresholds τ_A and τ_φ are found empirically using grid search in order to maximize the detection rate of new fluids at B-scan level (true positive rate, TPR). For parameters yielding the same TPR, those that give the lowest false positive rate (FPR) are selected. The thresholds are set to τ_A = -0.028 and $\tau_\varphi = 2.9$. For evaluation, ground truth segmentations of new fluids are generated per B-scan from the manual annotations as described in Sec. 4.3.1.2.

In Tabs. 6.2 and 6.3, new lesion segmentation and detection results are reported on B-scan- and lesion-level and compared to anomaly detection methods f-AnoGAN [Schlegl et al., 2019] and natural synthetic anomalies (NSA) [Schlüter et al., 2022] as well as NCR-Net, the non-correspondence detection and image registration network proposed in Chapter 4. The fast anomaly detection based on GAN training (f-AnoGAN) proposed by Schlegl et al. uses a cohort of healthy images to train a Wasserstein GAN for the reconstruction of normal image appearance. During inference, pathology images are entered into the generator and an anomaly score is generated from the image reconstruction error and the residual error between the feature vectors of the discriminator for the original and reconstructed image. The f-AnoGAN approach was developed for OCT images, and the only modification applied here is the usage of larger image patches for network training to better cover different retinal thicknesses.

The NSA approach by Schlüter et al. uses self-supervised anomaly detection. Patches that serve as anomalies are cut out of images and seamlessly inserted into other images from the same cohort using Poisson image editing. Supervised training is carried out for the segmentation of the inserted patches. An extension of this approach [Baugh et al., 2023] recently won the MOOD challenge, showing state-of-the-art performance in the generalization to unknown real anomalies. Both NSA and f-AnoGAN deliver an anomaly score rather than a binary segmentation of anomalies. For comparison to FluidRegNet, the scores are binarized using a threshold of 0.08 for f-AnoGAN and 0.75 for NSA. Thresholds are chosen to best compromise TPR and FPR.

While the anomaly detection methods are designed to segment any anomaly in the images, i.e. any pathology that is present in the OCT scans, the segmentations resulting from FluidRegNet are specific to new fluids. For a fair comparison, therefore, the following metrics are considered: The detection rate of new fluids and the Dice score of the fluids that are detected. A B-scan containing new fluids is considered detected if there is an overlap between the ground truth of the new fluids and the predicted segmentation. Results are first reported for the comparison of the predicted segmentations and the binarized ground truth segmentation of new fluids containing all new fluids, including

new IRF, new SRF and new PED. Further, results per fluid type are reported. Neither the new fluid segmentations generated by FluidRegNet nor the anomaly segmentations can differentiate between different types of pathologies. The results per fluid type are created by splitting the binary segmentations: Each predicted fluid that overlaps with the ground truth of the new fluids is assigned to the fluid type given by the ground truth. Additionally, for images containing no fluid, the average number and volume of false positive fluids is reported.

Results show that the proposed algorithm for new fluid segmentation delivers high detection rates both for IRF and for SRF with 91.89% and 84.43% detected fluids on B-scan level. Newly emerging PEDs are less reliably detected (30.00%). Compared to the previously proposed NCR-Net, detection rates have improved and compared to the state-of-the-art anomaly detection methods, both NCR-Net and FluidRegNet deliver lower detection rates than NSA but better than f-AnoGAN. PED detection shows to be challenging for all considered methods. PEDs can have very different morphologies, leading to a highly variable appearance in OCT images. Since FluidRegNet is designed to primarily find pathologies with a lower intensity than the surrounding healthy tissue, this could explain why FluidRegNet underperforms in the identification of new PEDs. Additionally, newly emerging PEDs are rarely included in the CSCR dataset, typically

Table 6.2: New fluids detection and segmentation results on B-scan-level for NSA, f-AnoGAN (GAN), NCR-Net (NCR) and FluidRegNet (FRN). Reported are the number of correctly identified new fluid containing B-scans (sens^{img}, noted as sens for brevity) and the Dice score averaged over all B-scans that are detected as containing new fluids (DSC_{det}, given in percent and noted as DSC for brevity). For binary segmentations, the number and volume (in pixels) of false positive fluids, averaged over all B-scans containing no fluid, is reported additionally.

	All fluids				IRF		SRF		PED	
Method	sens ↑	DSC ↑	n_{FP} ↓	V_{FP} ↓	sens ↑	DSC ↑	sens ↑	DSC ↑	sens ↑	DSC ↑
NSA	**273/293**	19.93	0.75	828.34	**36/37**	19.01	**210/212**	20.14	**42/60**	14.97
GAN	191/293	27.02	**0.45**	744.52	30/37	21.35	147/212	29.86	25/60	21.11
NCR	213/293	45.79	2.07	335.25	29/37	47.77	174/212	57.44	19/60	40.60
FRN	222/293	**59.49**	1.09	**91.48**	34/37	**55.95**	179/212	**66.57**	18/60	**50.95**

Table 6.3: New fluids detection and segmentation results on lesion-level for NSA, f-AnoGAN (GAN), NCR-Net (NCR) and FluidRegNet (FRN). Reported are the number of correctly identified new fluids ($\text{sens}^{\text{fluid}}$) and the Dice score averaged over all pathologies that are detected.

	IRF		SRF		PED	
Method	$\text{sens}^{\text{fluid}}$ ↑	DSC_{det} ↑	$\text{sens}^{\text{fluid}}$ ↑	DSC_{det} ↑	$\text{sens}^{\text{fluid}}$ ↑	DSC_{det} ↑
NSA	**392/415**	0.0186	**441/448**	0.0578	**61/92**	0.0508
GAN	225/415	0.0324	198/448	0.1545	24/92	0.2003
NCR	203/415	**0.3362**	269/448	0.4633	22/92	0.3676
FRN	201/415	0.2296	295/448	**0.4802**	20/92	**0.4821**

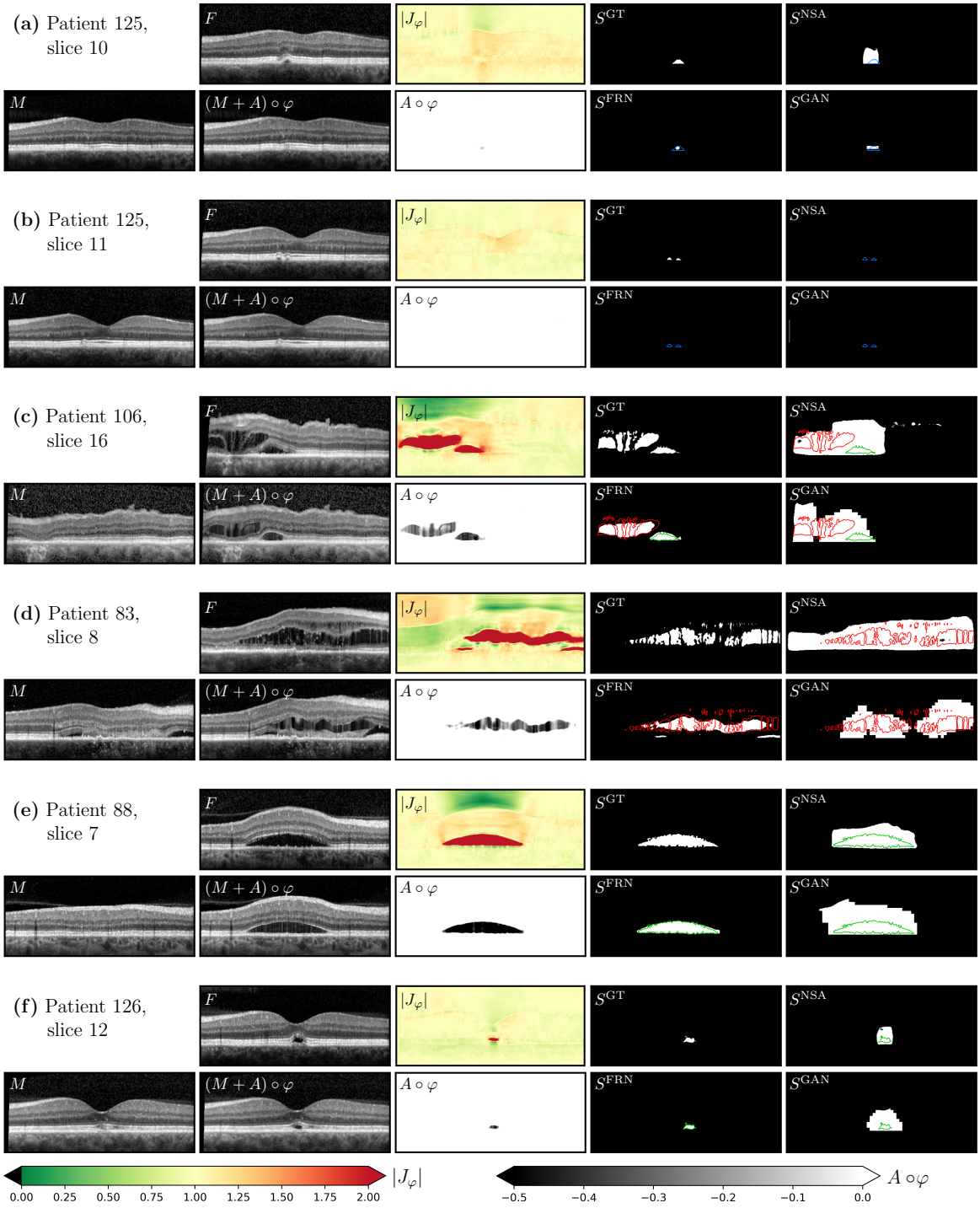


Figure 6.3: Segmentation and registration results of FluidRegNet for several examples of new fluids. For each registration pair, the moving and fixed images, the appearance adapted warped moving image, the Jacobian determinant of the deformation and the deformed appearance map are shown. The ground truth segmentation of new fluids S^{GT} is shown in the top row and fourth column for each patient, and its contours are overlaid onto the segmentation results from FluidRegNet, NSA and f-AnoGAN (S^{FRN}, S^{NSA} and S^{GAN}). The IRF is shown in red, SRF in green and PED in blue.

are small and occur in the same ocular region as SRF. All these factors could explain why the detection of new PEDs is so difficult with all methods.

Regarding segmentation performance, FluidRegNet delivers a much better delineation of new fluids than all competitive methods, except for IRF segmentation on lesion-level where NCR-Net performs better. NCR-Net was trained on images from the longitudinal AMD dataset, in which the IRF class is much better represented, which could explain this result. Overall, anomaly detection methods NSA and f-AnoGAN tend to be oversensitive and over-segment pathologies, whereas the segmentations resulting from FluidRegNet are much more precise and informative but tend to be undersensitive. This is further illustrated in Fig. 6.3 showing exemplary results for new fluid segmentation based on FluidRegNet. For comparison, anomaly detection results of NSA and f-AnoGAN are also shown.

6.4 Unsupervised Anomaly Detection with FluidRegNet

Whereas in longitudinal intra-patient image registration, non-correspondences are solely due to disease progression (except for imaging artefacts), in inter-patient registration of healthy and pathological images, non-correspondences may arise due to several reasons. Normal inter-individual variations such as the morphology of blood vessels may be one reason, but more importantly, all types of disease-related tissue changes cause non-correspondences. This is exploited in this section to use FluidRegNet for anomaly detection, i.e. to detect all types of pathological changes in the images. Firstly, the methodological extension of FluidRegNet for the detection of anomalies is described. This method is then used in Sec. 6.4.2 for the unsupervised segmentation of IRF, SRF and PED and in Sec. 6.4.3 for the investigation of pathological changes in the photoreceptor layer.

6.4.1 Methods

For anomaly detection, FluidRegNet is retrained for the registration of healthy to pathological images. Four OCT datasets are used for this purpose, one healthy dataset and three pathological datasets. The healthy dataset is the IMI dataset (c.f. Sec. 2.4.1) comprising 100 Spectralis OCT images from 50 healthy volunteers. Each image consists of 49 B-scans with a resolution of 496×512 pixels. There are no manual annotations for these images, only the automatic segmentation of the retinal layers provided by the Spectralis device. During a manual review of these segmentations, 14 images were excluded for further processing. The healthy dataset used for network training, thus, consists of 86 OCT images from 47 subjects, with images from both eyes for 39 subjects and images from only one eye for 8 subjects.

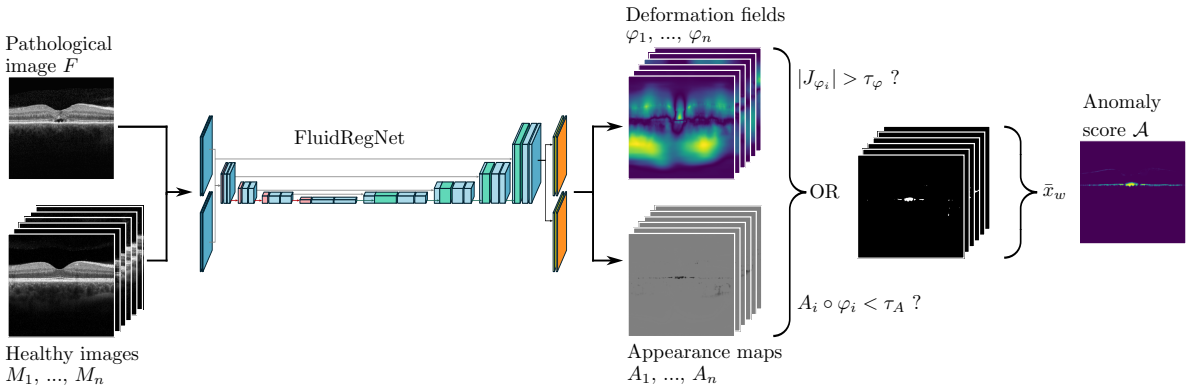


Figure 6.4: Overview of the proposed FluidRegNet-based anomaly detection method, FRN-Ano. The healthy OCT images M_1, ..., M_n are registered to the pathological image F for which the anomaly detection is to be performed. The resulting deformation fields and appearance offset maps are processed to produce n binary anomaly segmentations. These are weighted based on the image similarity between the pathological and the matched healthy image and summed to calculate the anomaly score ($\bar{x}_w$: weighted averaging with weights $w_i = \text{MSE}[F, (M_i + A_i) \circ \varphi_i]^{-1}$).

In order to better cover the wide variety of pathologies in retinal diseases, the pathological data in this section comes from different sources. Firstly, the CSCR dataset used in the previous section is also used here. This time, only the manually labeled images are used for training to ensure a good flattening of all the images. Still, the manual fluid segmentations are not used for network training, but only the automatically generated fluid segmentations as described before. The CSCR training dataset consists of 165 OCT images from 35 eyes of 19 patients (one to ten time points per eye). The second pathology dataset is the publicly available training data of the RETOUCH challenge [Bogunović et al., 2019]. For anomaly detection with FluidRegNet, only the 24 images taken with a Spectralis device are used. The third pathology dataset is the AMD dataset described in Sec. 2.4.1, consisting of 60 OCT images from 60 eyes of 36 AMD patients. For both the RETOUCH and the AMD data, manual annotations of IRF, SRF and PED are available.

Since greater deviations between images are expected for the inter-patient compared to the intra-patient registration, the different loss components are weighted slightly differently compared to the previous section to account for the larger deformations and the inter-individual differences. The weighting parameters α, β, γ, δ and η are set to 3000, 1000, 0.3, $1.6e^{-5}$ and $5e^{-3}$, respectively.

Once FluidRegNet is readily trained, anomaly detection is done as follows: Each image in the normal cohort is registered onto the pathology image to be analyzed. For each registration, the deformation field's Jacobian determinant and appearance map are thresholded similar to the previous section to generate segmentations of non-correspondent image regions. Again, the best suited thresholds are found via grid search and set to τ_J = 2.0 and τ_A = -0.13 giving the highest average DSC when comparing anomaly

segmentations with manual fluid segmentations. A weighted average of the resulting 86 non-correspondence segmentations is calculated to give an anomaly score. The segmentations are weighted by calculating the mean squared error between the respective warped healthy image and the pathological reference image and using the reciprocal of the error as the weight. An overview diagram of the FluidRegNet-based method for detecting anomalies (FRN-Ano) is shown in Fig. 6.4. A binary segmentation of anomalies can be achieved with thresholding on the anomaly score.

6.4.2 Fluid Segmentation

To evaluate the anomaly detection performance, in this section, the resulting segmentations are compared to manual segmentations of the different fluid types. Evaluation is performed for the CSCR, AMD and RETOUCH datasets separately. Since it is not necessary to have at least two segmented time points per patient to perform the proposed anomaly detection, the CSCR test data used in this experiment differs from the test data in the previous section in that it contains four additional images from four eyes, resulting in 181 images from 21 patients (4429 B-scans).

As already mentioned, all types of pathologies cause mismatches in image registration between healthy subjects and diseased patients. The most notable deviations from normal anatomy are the IRF, SRF and PED fluids, all of which are segmented unsupervised in this section, in contrast to the previous section where only the new fluids could be detected fully unsupervised. For quantitative evaluation, the FRN-Ano segmentations are compared to binarized anomaly maps of NSA and f-AnoGAN for the CSCR, the AMD and the RETOUCH dataset. The thresholds for binarization of the continuous anomaly maps are found empirically and chosen to maximize Dice scores on image-level (NSA, f-AnoGAN) or fluid-level (FRN-Ano) for the biggest dataset, the CSCR data. The thresholds are set to $\tau_{\text{FRN}} = 0.8$, $\tau_{\text{NSA}} = 0.95$ and $\tau_{\text{GAN}} = 0.09$. Results are given in Tab. 6.4. For each dataset, the proportion of correctly recognized pathological and non-pathological B-scans is reported. For pathological B-scans, the detection rate of the individual fluids is also reported, as well as the DSCs at image- and fluid-level. For B-scans falsely classified as pathological, the mean number and volume of false positive fluids is reported.

From the results, it can be seen that the proposed anomaly detection method achieves very good detection rates at image-level, close to the state-of-the-art method NSA. At the fluid-level, detection rates are lower, which can be explained by the difficulty of the method to detect very large deformations. Heavily swollen retinae often have numerous small IRF inclusions that are missed if the deformation is insufficient, as can be seen in the last example in Fig. 6.5. Regarding the segmentation performance, the results from the previous section can be confirmed: The detected fluids are delineated more precisely compared to other anomaly detection methods. While NSA and f-AnoGAN tend to strongly over-segment pathologies, FRN-Ano, in contrast, tends to under-segment

Table 6.4: Anomaly detection and segmentation results for FRN-Ano compared to NSA and f-AnoGAN (GAN). For B-scans containing pathologies, the detection rate (sens) is reported on image- and fluid-level. DSCs of detected B-scans/fluids are also reported, assessing segmentation accuracy of the anomaly detection methods. For images without pathologies, the amount of correctly identified negative samples is reported (spec). Additionally, for false positive images, the average number and volume (in pixels) of predicted pathologies is given.

		B-Scans with fluids				Fluid-free B-scans		
Dataset	Method	$\text{sens}^{\text{img}} \uparrow$	$\text{DSC}_{\text{det}}^{\text{img}} \uparrow$	$\text{sens}^{\text{fluid}} \uparrow$	$\text{DSC}_{\text{det}}^{\text{fluid}} \uparrow$	spec $\uparrow$	$\text{n}_{\text{FP}} \downarrow$	$\text{V}_{\text{FP}} \downarrow$
AMD	GAN	0.8507	**0.3283**	0.7303	0.2624	**0.6024**	**1.48**	3075.28
	NSA	**0.9746**	0.2048	**0.9489**	0.1087	0.4373	1.96	7499.61
	FRN-Ano	0.8851	0.1877	0.4606	**0.3180**	0.4482	3.21	**158.22**
CSCR	GAN	0.7777	0.2834	0.6296	0.1879	0.7383	**1.17**	1915.26
	NSA	**0.9413**	0.2639	**0.8929**	0.1184	**0.8668**	1.84	4175.96
	FRN-Ano	0.9035	**0.5138**	0.6001	**0.4437**	0.6800	2.07	**118.56**
RETOUCH	GAN	0.5905	0.1182	0.1556	0.1325	**0.9378**	**1.00**	5152.00
	NSA	**1.0000**	0.2040	**0.9465**	0.0255	0.0041	1.80	22543.64
	FRN-Ano	0.9721	**0.3067**	0.3363	**0.2108**	0.4315	3.64	**532.20**

pathological regions. This is also reflected in the analysis of false positive fluids, the volume of which is significantly greater with NSA and f-AnoGAN than with FRN-Ano.

The different methods' performances vary between datasets. Although all images were acquired with a Heidelberg Spectralis device, the image contrast differs between the datasets, as can be seen in Fig. 6.5. In particular, the RETOUCH data generally shows lower intensity values inside the retina than the images from the other two datasets. NSA and f-AnoGAN were trained on healthy images with a similar contrast to the CSCR data. Consequently, both methods show severe performance drops for the RETOUCH data. For NSA, the amount of false positives increases strongly, as oftentimes the entire retina is marked as anomalous. In contrast, f-AnoGAN tends to miss a lot of pathologies, probably due to the reduced contrast between retinal tissue and fluids. Although FRN-Ano was trained with all three datasets, it shows a higher false positive rate and a reduced segmentation performance for the AMD and RETOUCH data compared to the CSCR data. Since the anomaly detection of FRN-Ano also depends on the healthy images, in that they are registered onto the pathological images, this behavior can again be explained by the differences between the datasets. In addition, the CSCR images make up a large part of the training data, so pathologies specific to other diseases are rarely represented in the data (see arrows in Fig. 6.5). The highly heterogeneous appearance of AMD, thus, poses problems to FRN-Ano. However, in contrast to NSA and f-AnoGAN, which either mark the entire tissue or nothing as an anomaly at low image contrast, FRN-Ano never fails completely. The susceptibility to domain shifts between datasets appears to be lower than with the other methods, so that the method can be used directly as a guide for observing pathologies in different datasets.

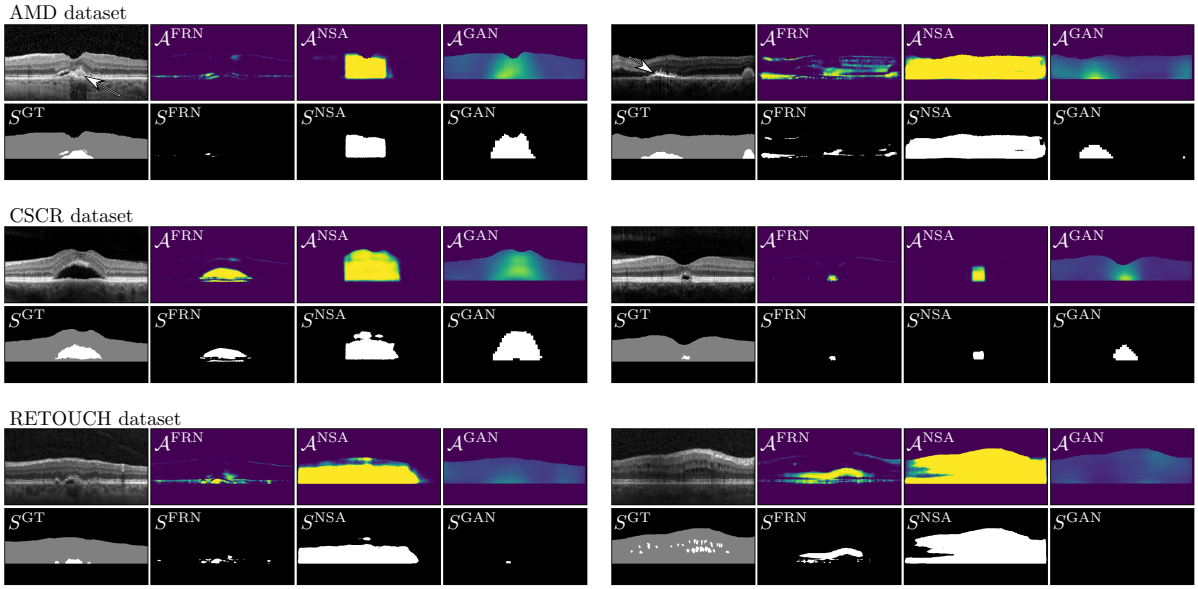


Figure 6.5: Exemplar anomaly detection results for NSA, f-AnoGAN and FRN-Ano on three different datasets. For each example, the continuous anomaly map $\mathcal{A}$ as well as the binarized anomaly segmentation S is shown. Arrows mark AMD-specific pathologies that FRN-Ano has difficulty recognizing. Image contrast in the RETOUCH data is low, but FRN-Ano still provides meaningful results.

In the following, an additional 3D evaluation is performed for all anomaly detection methods on the CSCR data. Given the 2D anomaly detection methods, the fluid extension in 3D can be estimated both volumetrically and in the en-face plane. For volumetric measurement, the predicted pathologies are totaled for all B-scans in a scan to measure the volume of pathologies. For the analysis in the en-face plane, qualitative results are shown in Fig. 6.6, where the segmented pixels per A-scan are summed up to project the 3D segmentations into the en-face plane. Quantitative evaluation is done by thresholding the projected segmentations and comparing with the projected ground truth segmentation. To do so, each A-scan that contains more than two segmented pixels is segmented in the en-face plane. On average, this results in a DSC_{proj} of 0.5614 for FRN-Ano, while NSA achieves a DSC_{proj} of 0.5202 and f-AnoGAN 0.4713. FRN-Ano, thus, proves to be the most suitable method for estimating the spread of pathological fluids in the retina. Also, the volumetric assessment of fluids is the most accurate using FRN-Ano. Despite tending to underestimate fluid volume, the average absolute deviation from the ground truth volume is 5171.81 pixels (corresponding to 0.06 mm^3) whereas NSA and f-AnoGAN tend to highly overestimate fluid volume with average absolute deviations of 42446.94 and 26516.37 pixels (0.48 and 0.31 mm^3, respectively). This can also be observed in the examples in Fig. 6.6, where it comes apparent that fluid expansion predicted by FRN-Ano closely reflects real pathologies, while the oversized anomaly segmentations of the other methods cannot represent the actual shape of the fluid inclusions. For more examples, please refer to the appendix.

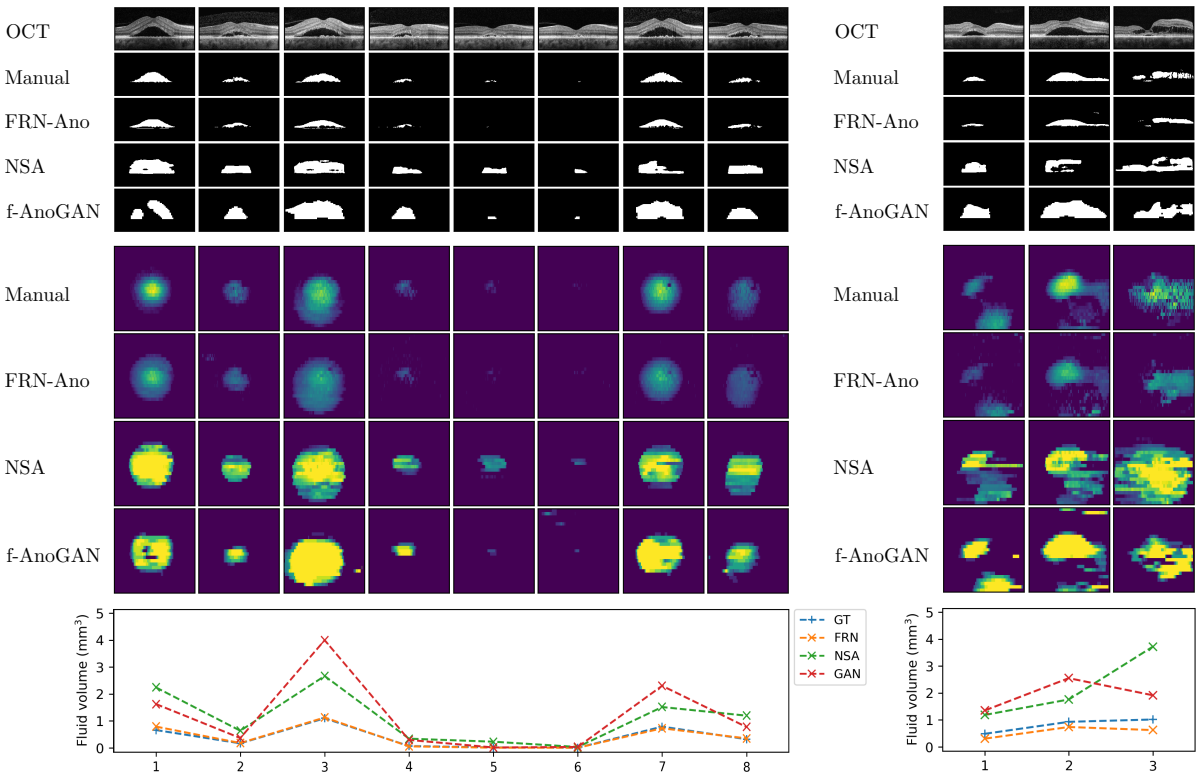


Figure 6.6: Anomaly detection results for NSA, f-AnoGAN and FRN-Ano in the en-face plane. Timelines of two CSCR patients are shown (left and right): The top two rows show the central OCT-B scans of each visit and the corresponding manual segmentations. Below, the corresponding anomaly segmentations of FRN-Ano, NSA and f-AnoGAN are shown from top to bottom, followed by the segmentations of the entire volume projected into the en face plane. At the bottom, the fluid volume in mm^3 is shown for all methods and the ground truth.

To summarize, FRN-Ano delivers good detection rates and segmentation performances compared to other anomaly detection methods, but has problems with severely deformed retinae, which leads to fluids located in the upper part of the retina being missed. The ability to capture large deformations would, therefore, potentially significantly improve FluidRegNet-based anomaly detection. In addition, susceptibility to image contrast proved to be superior to other anomaly detection methods, but could probably be further improved if more balanced pathological and more diverse healthy images were used for training and registration-based anomaly detection. For images taken with the same imaging protocol, FRN-Ano allows accurate estimations of pathological fluid volume and spread in the eye, as shown for the CSCR dataset.

The choice of binarization thresholds has a major impact on both anomaly detection and segmentation performance of all methods. Therefore, the use of continuous anomaly maps seems to be more useful in clinical practice, allowing medical experts to set the thresholds individually for each image. Another general problem in the evaluation of

the anomaly detection results is that fluids may not be the only anomalies present in the images. Instead, there can be other types of pathologies, such as macular holes or hyperreflective foci. Additionally, technical sources for anomalies cannot be covered by comparing with fluid segmentations: Overall or regional intensity variations, noise level fluctuations, mis-registration in the eye-tracking system are just a few examples of sources for "non-normal" looking images. A frequently observed pathology in CSCR are changes in the reflectivity of the photoreceptor ellipsoid zone (PEZ). In the following, FRN-Ano is shown to be capable to also detect this type of pathology, which is much more subtle than most fluids.

6.4.3 Detection of Pathological Changes in the Photoreceptor Ellipsoid Zone

In retinal OCT images of healthy eyes, the PEZ appears as a bright band above the RPE and the photoreceptor outer segments, as highlighted in yellow in Fig. 3.5 (a). Clinically, the PEZ is a widely used OCT-based biomarker of photoreceptor structure. While PEZ integrity, PEZ lesion size and width of preserved PEZ are already established metrics that have been shown to correlate with visual acuity, PEZ band reflectivity or intensity is a less explored biomarker [Lee et al., 2021]. The reflectivity of the PEZ could be of great clinical benefit, as changes in reflectivity have been shown to precede more severe PEZ pathologies in neovascular AMD and idiopathic epimacular membrane disease [Toprak et al., 2014, 2017; Pasricha et al., 2021].

Changes in photoreceptor reflectivity are visible as hypointensity within the normally bright PEZ band, some examples are shown in Fig. 6.7 for four CSCR patients. However, the changes in brightness observed are often only slightly pronounced and no clear boundaries of these regions are visible, which makes an objective evaluation of PEZ pathologies difficult and prevents the creation of annotated datasets that would be necessary for automatic segmentation methods requiring supervised training. FRN-Ano is capable of not only highlighting comparably obvious deviations from the normal anatomy such as pathological fluids, but also more subtle anatomical transformations in the PEZ region (see e.g. Fig. 6.5 the examples top left and bottom left). The goal in this section, therefore, is the usage of the FluidRegNet-based anomaly detection for the automatic delineation of photoreceptor changes. To do so, the anomaly maps generated in the previous section are refined as follows: Two U-Net-based CNNs are used to segment SRF and the photoreceptor layer and subsequently, all A-scans segmented as containing SRF are excluded from the anomaly map. In consultation with ophthalmologists, the remaining anomalous regions are masked to the PEZ region in two ways. First, the anomaly map is multiplied directly with the binary segmentations of the photoreceptor layer. Second, the PEZ segmentation is widened downwards to include the interdigitation zone (IZ, the bright band below the PEZ band), as abnormalities in the PEZ and the IZ often occur together. In the following, the first method is called FRN-Ano$_{\text{PEZ}}$ and the second FRN-Ano$_{\text{PEZ+IZ}}$.

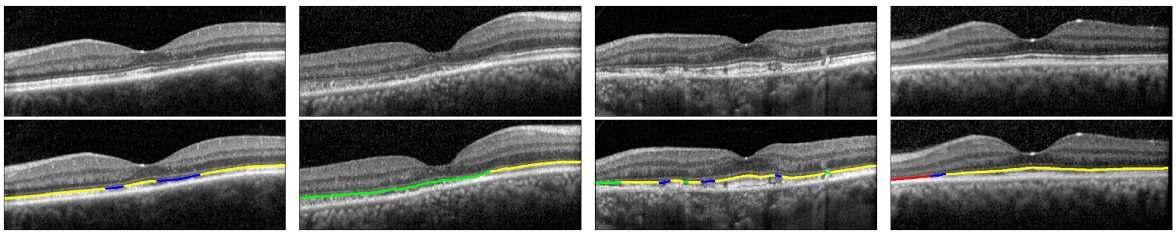


Figure 6.7: Four examples of PEZ pathologies. The top row shows the central OCT B-scans of CSCR patients 35, 7, 30 and 5. In the bottom row, the respective manual PEZ segmentations are superimposed. The normal PEZ is shown in yellow, inconsistent PEZ in blue, non-localizable PEZ in green and atrophy in red.

The architecture of the network for SRF segmentation is shown in Fig. A4. Two separate convolutional blocks at the beginning of the network take the OCT B-scan to be segmented and the corresponding retina segmentation as input. The feature maps generated from the two input blocks are concatenated after the first maximum pooling and processed together afterward. Nearest neighbor interpolation is used in the decoding path, and the network finally outputs segmentations of SRF and PED. Here, only the SRF segmentations are used. The SRF segmentation network developed and trained as part of this work was successfully used in [Santarossa et al., 2022b; von der Burchard et al., 2024; Scharf et al., 2024] for various medical analyses. The CNN for segmentation of the PEZ was developed by M. Santarossa, has a U-Net-like architecture and uses ResNet-32 as backbone [He et al., 2016]. The network is pretrained on ImageNet [Russakovsky et al., 2015] and is trained for the OCT data with a combination of Dice loss and binary focal loss. Since PEZ atrophy is characterized by the absence of the PEZ and not by reduced reflectivity, the FRN-Ano-based anomaly detection is expected to miss atrophic eye regions. In a final experiment, FRN-Ano is, therefore, combined with the atrophy detection from Sec. 3.4.

The PEZ anomaly detection results are evaluated by comparing them to manual annotations of PEZ pathologies. Such manual labels were generated for 167 OCT images of 66 CSCR patients, differentiating normal PEZ, atrophy, inconsistent and non-localizable PEZ. Manual PEZ annotation was carried out in two rounds. In the first round, annotations were done for 143 images of 46 patients and only the five central B-scans were labeled. PEZ atrophy was only very rarely observed in these images (11 OCT images of 6 patients), which is why 20 additional patients with atrophy were selected in the second round, for which one or, in four cases, two images were completely segmented. To quantitatively evaluate the PEZ anomaly detection results, the manual labels are divided into normal and pathological regions and compared to binarized PEZ anomaly maps. For this purpose, the pathology labels atrophy, non-localizable and inconsistent are combined into one label for the manual annotations. The PEZ anomaly maps are subdivided into normal and pathological, with a threshold value of 0.3. The binary results are compared at A-scan level, measuring the overlap of anomaly and pathology (projected Dice score $\mathrm{DSC}_{\mathrm{proj}}$) as well as sensitivity and specificity of the segmentations.

To avoid biases introduced by differing numbers of segmented B-scans per image, evaluation is performed on the five central slices of each segmented volume and the metrics are then averaged over all available B-scans.

Results are reported in Tab. 6.5, showing that the detection of PEZ abnormalities with FluidRegNet is moving in the right direction, as on average 49% of abnormal A-scans and 93% of normal A-scans are correctly classified when masking the anomaly maps to the PEZ. When the IZ is included, 65% of pathological A-scans and 82% of normal A-scans are identified correctly. Nevertheless, the delineation accuracy with a value of 0.4255 DSC_{proj} for FRN-Ano_{PEZ} and 0.4536 for FRN-Ano_{PEZ+IZ} does not yet seem satisfactory. While the manual annotations show abnormalities in the PEZ only, inclusion of the IZ for the generation of the PEZ anomaly maps greatly improves the method's sensitivity. At the same time, the number of false positive A-scans is increasing, which can be partly explained by pathologies that only occur in the IZ and not in the PEZ and are, therefore, not recorded in the manual annotations. Here, it should also be noted that the manual ground truth is not always completely reliable due to the high degree of difficulty of manual PEZ assessment. The annotations were made in the knowledge that not all abnormalities in the PEZ could be manually delineated. Although the aim was to provide the best possible training data for DL segmentation methods, it was also hoped that the trained networks would recognize more areas as pathological than could be manually annotated. Thus, further investigation of the predicted anomalies by medical experts is required to ensure that the predictions are meaningful and may reveal anomalies that are not captured by the manual segmentations.

Exemplary results for PEZ anomaly detection are shown in Fig. 6.8 for the better performing method FRN-Ano_{PEZ+IZ}. If available, manual annotations of PEZ pathologies are shown for comparison. It comes apparent that FRN-Ano is able to detect small areas of PEZ abnormalities as well as more extensive pathologies. Although there is good agreement with the manual ground truth for many cases, there are examples where the manual segmentation and the FRN-Ano results completely contradict each other. As expected, the detection of atrophic eye regions poses problems to the anomaly detection. An example is given in the second row of Fig. 6.8, where FRN-Ano entirely misses the atrophy region. The areas that were marked as anomalous though appear plausible despite not being marked in the ground truth.

To alleviate the problem of missed atrophic regions, the FRN-Ano results are combined with the atrophy segmentation results from Chapter 3.4. To do so, the binary atrophy and anomaly segmentations are fused with logical OR. This results in an improvement

Table 6.5: PEZ anomaly detection results for FRN-Ano and FRN-Ano combined with the atrophy detection (AD) from Chapter 3.4.

Method	DSC^{all}_{proj} ↑	sens ↑	spec ↑
FRN-Ano_{PEZ}	0.4255 (0.3033)	0.4901 (0.3770)	**0.9276** (0.1245)
FRN-Ano_{PEZ+IZ}	0.4536 (0.2813)	0.6470 (0.3517)	0.8170 (0.2113)
FRN-Ano_{PEZ} + AD	**0.5542** (0.2795)	0.6441 (0.3263)	0.9142 (0.1383)
FRN-Ano_{PEZ+IZ} + AD	0.5449 (0.2692)	**0.7776** (0.2692)	0.8050 (0.2179)

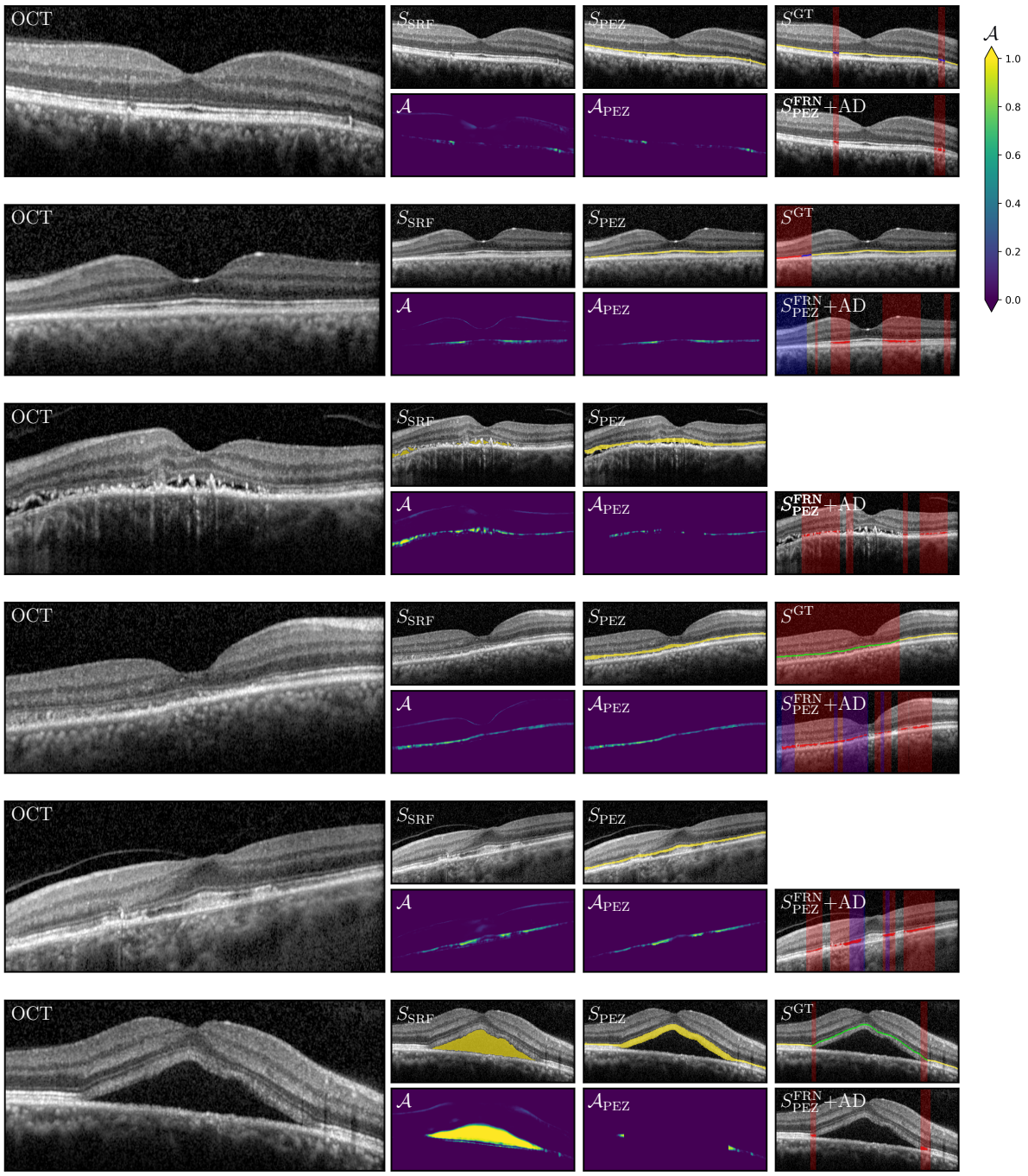


Figure 6.8: Photoreceptor anomaly detection results for six CSCR patients. For each example, the OCT-B scan is shown (large image on the left) as well as the SRF and PEZ segmentation results (S_{SRF} and S_{PEZ}, second and third column above). The anomaly score maps before and after processing with the segmentations are shown below ($\mathcal{A}$ and $\mathcal{A}_{\mathrm{PEZ}}$). In the rightmost column, the binary PEZ anomaly segmentations (S^{FRN}, red) are shown together with the atrophy detection results (AD, blue overlay) and, if applicable, the manual segmentation of the PEZ anomalies is shown for comparison (S^{GT}).

of the average DSC_{proj} to 0.5449 using FRN-Ano_{PEZ+IZ}. The sensitivity increases to 0.7776 while the specificity remains stable with 0.8050, which shows that more anomalous regions are segmented while only few false positives are introduced. Overall, the combination of anomaly detection with FluidRegNet and partially supervised atrophy clustering leads to a good differentiation between healthy and damaged PEZ, although the accuracy of the spatial delineation could still be improved.

To summarize, FRN-Ano is a first approach towards automatic PEZ anomaly detection. Since manual annotations of PEZ pathologies are scarce and the existing annotations are potentially incomplete, the quantitative analyses can only provide an initial indication of the actual performance of the procedure. Still, the results show that FRN-Ano is suitable as a supportive system for the detection of PEZ pathologies, potentially identifying abnormalities that are difficult to be assessed manually. Not part of this work, but a potential possibility, is also the use of the continuous anomaly scores rather than binarized segmentations, which allow not only the detection but also the quantification of anomalies.

6.5 Discussion

In this chapter, FluidRegNet, an application-driven deep learning-based image registration framework capable of aligning retinal OCT images with severe pathological changes was developed. Newly emerging fluids are incorporated into the registration process by letting the network learn to place an appearance seed in the moving image and to extend this fluid seed to cover the volume of the fluid observed in the fixed image. This results in much more realistic deformation fields than with previous registration methods, even those taking non-correspondences into account explicitly. The deformation fields can, thus, be used for volume change assessments and to highlight areas of disease progression. Together with the appearance offsets, the fluid-aware deformation fields offer the possibility to perform unsupervised segmentation of new pathologies, as shown for the longitudinal CSCR data.

Retraining FluidRegNet for the registration of healthy to pathological OCT images enables the segmentation of all, rather than only new, pathologies and the usage of FluidRegNet as anomaly detection method. Compared to state-of-the-art anomaly detection methods, FRN-Ano achieves lower detection rates, but delineates the detected pathologies much more precisely. Also, FRN-Ano proofed more stable to domain shifts between different OCT datasets. The precise anomaly localization was used to segment prominent pathologies IRF, SRF and PED as well as less obvious changes in the PEZ, which are generally hard to localize even manually. The segmentation of such subtle changes is not possible with the competitor anomaly detection algorithms. FRN-Ano showed promising results here, but often failed to delineate atrophic areas. The combination with the atrophy segmentation from Chapter 3.4 substantially improved the results, so that overall a good delineation of pathological PEZ areas is possible.

A limitation of FluidRegNet is that non-correspondences due to dissolving fluids are only implicitly covered with the masked regularization. The resulting deformation fields usually only depict such changes incompletely, which is generally a problem of deformation-based approaches that cannot make voxels disappear. Nevertheless, a strong reduction in volume can be observed with dissolving or shrinking fluids (cf. Fig. 6.2, first example). The deformation fields can, therefore, still be used to determine not only deteriorating but also improving eye regions. Furthermore, like most registration methods that use some kind of deformation regularization, FluidRegNet still struggles with very large deformations that are to be expected in retinal disease OCT images. Since extreme cases are rather seldom in the CSCR data, a more balanced data selection might help circumvent this problem. From an architectural point of view, an improvement might be achieved using transformer networks, iterative registration or pyramid registration. Transformers differ from CNNs in that they employ a self-attention mechanism, comparing each image patch to all other patches to guide the network focus, resulting in a superior ability to capture long-range dependencies [Vaswani et al., 2017]. Compared to CNNs, transformers proofed more difficult to train and like iterative and pyramid registration have a higher computational complexity. In contrast, FluidRegNet's single-stream architecture is relatively lightweight.

FluidRegNet provides a valuable tool for the longitudinal registration of OCT images, capturing the dynamics of retinal fluid growth more realistically than other image registration methods. The realistic deformation fields enable a detailed visualization of regressive and progressive areas in the eye for time-series images. Needing only segmentations of the entire retina rather than pathology labels, FluidRegNet can be used for the unsupervised segmentation of newly emerging fluids. Additionally, the extension of FluidRegNet for anomaly detection, FRN-Ano, enables the detection and exact localization of a lot of retinal pathologies, including fluid deposits as well as subtle changes in the photoreceptor ellipsoid zone. Therefore, FRN-Ano potentially is a valuable guidance tool to help clinicians find areas of disease activity and might significantly reduce the cost of the laborios manual annotation of retinal OCT volumes. The methods presented here, thus, make an important contribution to solving the problem of data scarcity in the field of ophthalmology.

Chapter 7

Implicit Neural Representations for OCT Interpolation and Registration

Real-world signals are continuous, yet, traditionally, they are explicitly discretized to store and process them. For example, shapes are commonly represented as point clouds or meshes and images as a grid of pixels or voxels. However, discretization comes with a number of drawbacks, especially for high-dimensional data: Its computational cost rises with increasing dimensionality, and the memory requirements for discrete data are very large [Essakine et al., 2024]. Furthermore, discretized data only contains information per segment (voxel, point, mesh, ...), binding it to the selected resolution and preventing effective representation of irregularly sampled or sparse data. Recently, implicit neural representations (INRs) have emerged as a new paradigm to represent data. Rather than directly storing the signal values, INRs are neural networks, typically MLPs, that map input coordinates to the signal values and, thus, implicitly encode the continuous signal in the weights of the network (cf. Sec. 2.2.3). They are built on the idea that neural networks can estimate complex functions after being trained on discretely represented data samples. Once an INR has been adapted to a certain signal, any input coordinates may be inserted to represent the signal at arbitrary resolutions. Different from discrete representations, the storage requirement of INRs is, thus, independent of the signal's resolution, scaling only with the complexity of the signal.

INRs have gained substantial interest in the past years, leading to diverse medical image processing applications such as super-resolution [Wu et al., 2023a; McGinnis et al., 2023; Fang et al., 2024], image reconstruction [Reed et al., 2021; Wu et al., 2021; Shen et al., 2024] and image registration [Wolterink et al., 2022; Byra et al., 2023a; Sideri-Lampretsa et al., 2024]. Whereas CNNs are limited to densely sampled input images, INRs have been shown to be especially powerful for the processing of highly anisotropic and sparse data. Retinal OCT images are sampled with high inter-slice distances in clinical routine, which is why many methods proposed for OCT processing rely on 2D CNN architectures. Here, INRs offer a promising tool to overcome this trend and to process full OCT image volumes.

However, an INR is typically adapted to a single scene by training its weights on only one image (volume). While this leads to an exceptional representation of the respective image, the INR does not generalize to other cases. Instead, a new MLP needs to be trained for each image. To enable the representation and simultaneous segmentation of more than one image, generalizable INRs have been proposed recently [Amiranashvili et al., 2022; Kim et al., 2023; Stolt-Ansó et al., 2023]. Generalizable INRs work by separating instance-specific and instance-agnostic shared parts of the network. In [Amiranashvili et al., 2022] and [Stolt-Ansó et al., 2023], one MLP is trained

on the entire training dataset and instance-specific priors are learned for each image individually. The priors are passed on to the network alongside the spatial coordinates to condition the MLP on the respective instance. This type of generalizable INR offers the possibility to be applied to images unseen during training by adapting new priors for the test images.

In this chapter, INRs are used for two OCT applications: First, INRs are used for B-scan interpolation to reduce the anisotropy of clinical OCT data. A generalizable INR is used for this approach, enabling not only the interpolation of OCT images but also the segmentation of retinal layers. To improve the performance for interpolated B-scans and to enable the integration of small structures in between slices, it is proposed to condition the MLP on multimodal image information. Second, INRs are used for the registration of retinal OCT images with pathologies. As in traditional image registration methods, optimization is performed for individual image pairs using standard INRs. To cope with the structural differences introduced by pathologies, two MLPs are trained per registration task. The first INR predicts the deformation field to map the moving to the fixed image, whereas the other one produces the so-called residual image, containing the pathological differences between moving and fixed images. The following two sections address the two OCT applications, i.e. B-scan interpolation and registration, in more detail. Each section begins with a short motivation of the respective use case and the associated literature. Afterward, the INR methodologies are described in detail, which are then applied to OCT images and evaluated in comparison to competitive methods. Each section closes with a discussion of the presented approach.

7.1 OCT Interpolation with SLO Integration Using Generalizable INRs

Despite bearing the risk of missing small anatomical or pathological structures, OCT images of the retina are usually acquired with large B-scan distance in clinical practice to reduce acquisition time. Modern CNN-based segmentation frameworks are typically developed for isotropic images, which is why the segmentation of the highly anisotropic OCT images is most often performed with 2D approaches. However, 2D segmentations inherently lead to irregular surfaces in 3D, necessitating expensive post-processing such as graphcut approaches, e.g. [Li et al., 2005]. Also, 2D approaches do not take into account the information of neighboring image slices, potentially leading to inconsistencies between segmentations of individual B-scans. Moreover, volumetric measurements of retinal structures or pathologies based on such 2D segmentations are imprecise due to the high inter-slice distances.

Interpolation methods may be used to generate densely sampled OCT volumes, but suffer from interpolation artifacts due to severe shape and localization differences between neighboring B-scans. Furthermore, they cannot infer information in between B-scans, while GAN-based methods for OCT super-resolution require large datasets

of either paired [Huang et al., 2019] or unpaired [Das et al., 2020] low- and high-resolution images. INRs, in turn, are resolution-agnostic and can successfully be trained on anisotropic or sparse data [Amiranashvili et al., 2022]. Given their success in generating plausible shapes for irregular training data, this section explores their usage for OCT representation and interpolation. A generalizable INR framework is proposed that allows not only the reconstruction of the OCT intensities but also the segmentation of retinal layers. The backbone of the INR is an MLP that is shared between all training images and that receives spatial coordinates as well as patient-specific priors as input. These priors are learned during the training process and can also be adapted to new patients once training is completed. Therefore, the proposed INR can be used to segment retinal layers in images unseen during training. Furthermore, it is proposed to combine multimodal images in the INR to enable the introduction of small structures that are located in between the B-scans.

7.1.1 Related Literature

OCT super-resolution has been done with intensity-based interpolation methods [Lindberg et al., 2018; Chen et al., 2024] and GAN-based approaches [Huang et al., 2019; Das et al., 2020; Yuan et al., 2023]. For example, [Lindberg et al., 2018] propose a weighted combination of linear and transfinite interpolation to improve upon linear or transfinite interpolation alone. However, this method requires two OCT images of the same scene taken with orthogonal scan directions. In general, intensity-based interpolation methods have difficulties to continue the retinal shape between B-scans, leading to severe interpolation artifacts. Registration-based interpolation, as presented in [Ehrhardt et al., 2007], manages to better compensate for shape differences between neighboring slices but still cannot introduce small structures that are located in between the known image slices.

The GAN-based approaches typically perform super-resolution only per B-scan but not in between B-scans [Huang et al., 2019; Das et al., 2020; Yuan et al., 2023] and often require paired low- and high-resolution datasets [Huang et al., 2019; Yuan et al., 2023]. Some approaches try to interpolate the segmentations generated on sparsely sampled B-scans rather than the OCT intensities by using morphological operations [Szeskin et al., 2021] or shape filling algorithms [Santarossa et al., 2022b]. Both works first project the OCT segmentations into the en-face plane and then fill the space between neighboring B-scans to analyze the lateral extension of pathologies.

INRs are by design resolution-agnostic and can, therefore, directly be applied for image super-resolution, e.g. [Fang et al., 2024] for CT and [Wu et al., 2023a] for MRI. [Amiranashvili et al., 2022] propose a generalizable INR for the representation of anatomical shapes. A single MLP, which is trained for all training samples, receives not only spatial coordinates as input but also shape-specific latent vectors. These vectors are learnable parameters and adapted together with the MLP. This setup allows the separation of features shared between all training samples and individual features specific to the training samples. The common features are represented in the backbone MLP,

whereas the shape-specific features are stored in the latent vectors. The authors show excellent shape representations despite training their framework with highly anisotropic and sparse input data. Furthermore, their method allows adaptation to unseen cases by freezing the MLP and learning a new latent prior for the new shape.

[Stolt-Ansó et al., 2023] build on this idea and propose neural implicit segmentation functions (NISF), a generalizable INR with two outputs that reconstruct and segment the input images based on the input coordinates and patient-specific latent codes. Despite using low-resolution inputs, they are able to generate smooth segmentation masks. Segmentation of new cases is possible by adapting new latent codes while freezing the MLP, as in [Amiranashvili et al., 2022]. To do so, no manual segmentation labels are required, since the latent code is learned by concentrating on the reconstruction of image intensities and omitting the segmentation task used during the population-based training.

To date, only very few works use implicit neural representations for OCT applications. OCT image reconstruction with INRs is done in [Li et al., 2025], and [Ziemann et al., 2024] propose time-varying INRs for OCT denoising. The goal of this section is to explore the usability of generalizable INRs for OCT interpolation and segmentation. This work was published in [Kepp et al., 2025] and still seems to be the only work using INRs for this application. The method extends NISF for multimodal inputs to allow the integration of small structures in between B-scans. The proposed INR is shown to outperform other interpolation methods, in terms of both shape interpolation and representation of the interpolated B-scans. At the same time, the segmentation of retinal layers is possible for unseen cases, which is not the case for pure interpolation methods.

7.1.2 Methods

To enable the simultaneous interpolation and segmentation of retinal OCT images, a generalizable INR (GenINR) extending the approaches of [Amiranashvili et al., 2022] and [Stolt-Ansó et al., 2023] is proposed in this section. The core idea of the GenINR is to use separate population- and instance-based network parts to isolate common and patient-specific image features. For this purpose, a backbone MLP, whose parameters are shared throughout the entire training set, and image-specific learnable priors are used, serving as conditioning inputs for the MLP. More specifically, the proposed INR $f_\theta = (f_\theta^{\text{recon}}, f_\theta^{\text{seg}})$ with learnable parameters θ consists of six linear layers of size 128 that use wavelet activation functions (WIRE, cf. activation function (2.6)), followed by two separate output heads, both consisting of a single linear layer. The first output head predicts the OCT image intensities (f_θ^{recon}) whereas the second outputs the segmentation labels of twelve retinal layers (f_θ^{seg}). The reconstruction head uses sigmoid activation, whereas the segmentation head uses a softmax activation.

Despite an INR's capability to represent data at arbitrary resolutions, it can only derive the information that lies between the coordinates used for training from the

samples that lie at those coordinates. The OCT images alone are, therefore, not sufficient to represent small structures, e.g. blood vessels, that are not captured by the sparsely sampled B-scans. To enable the INR to integrate inter-slice information, it is proposed here to use multimodal data, where the additional imaging modality serves to provide guidance about the presence and the exact location of structures in between B-scans. As a first proof of concept, the scanning laser ophthalmoscopy (SLO) modality is used, which is routinely acquired alongside the OCT images in Heidelberg Engineering Spectralis scanners. The proposed INR receives three inputs: A spatial coordinate $(x, y, z) \in \mathbb{R}^3$ representing a position in OCT_i, the OCT image of patient i, the patient-specific latent prior $\boldsymbol{p}_i \in \mathbb{R}^L$ and the intensity of the corresponding SLO image SLO_i at en-face position $(x, z) \in \mathbb{R}^2$. The relationship between the intensities and localization in OCT and SLO images is depicted in Fig. 7.1. The length L of the latent prior is set to 128 empirically. For each training step, the image coordinates (x, y, z) of an entire B-scan are used, and the loss function is calculated via

$$\begin{aligned}\mathcal{L} = \sum_{i=1}^{n} \sum_{(x,y,z)} [&\mathcal{L}_{\mathrm{recon}}(f_\theta^{\mathrm{recon}}(x, y, z, \mathrm{SLO}_i(x, z), \boldsymbol{p}_i), \mathrm{OCT}_i(x, y, z)) \\ &+\alpha \mathcal{L}_{\mathrm{seg}}(f_\theta^{\mathrm{seg}}(x, y, z, \mathrm{SLO}_i(x, z), \boldsymbol{p}_i), \mathrm{S}_i^{\mathrm{GT}}(x, y, z)) + \beta \mathcal{L}_{\mathrm{reg}}(\boldsymbol{p}_i)]\end{aligned} \tag{7.1}$$

for training images $i = 1, ..., n$. Here, $\mathrm{OCT}_i(x, y, z)$ and $\mathrm{S}_i^{\mathrm{GT}}(x, y, z)$ are the ground truth intensity value and segmentation label of the respective OCT image at position (x, y, z). The reconstruction loss $\mathcal{L}_{\mathrm{recon}}$ uses a combination of MSE and SSIM loss (cf. Sec. 2.3) and is calculated as $\mathcal{L}_{\mathrm{recon}} = \mathcal{L}_{\mathrm{MSE}} + 0.1 \cdot \mathcal{L}_{\mathrm{SSIM}}$. The binary cross entropy loss is used for the segmentation loss $\mathcal{L}_{\mathrm{seg}}$. Additionally, a regularization term $\mathcal{L}_{\mathrm{reg}} = \sum_{i=1}^{n} \|\boldsymbol{p}_i\|_2$ is used that favors small values in the latent priors. The weighting parameters are set to $\alpha = \beta = 0.2$.

After training, the INR is fit to a new patient by keeping the network parameters fixed and learning a new latent code $\boldsymbol{p}_{n+1}$. While the intensity values of the new image are known, there are typically no segmentation labels given for this image. The latent prior is adapted based on the following reduced loss function

$$\mathcal{L} = \sum_{(x,y,z)} \mathcal{L}_{\mathrm{recon}}(f_\theta^{\mathrm{recon}}(x, y, z, \mathrm{SLO}_{n+1}(x, z), \boldsymbol{p}_{n+1}), \mathrm{OCT}_{n+1}(x, y, z)) + \beta \mathcal{L}_{\mathrm{reg}}(\boldsymbol{p}_{n+1}), \tag{7.2}$$

which has been shown to provide meaningful segmentation predictions in [Stolt-Ansó et al., 2023] for cardiac MR images.

The architecture of the proposed INR, as well as the multimodal inputs and the principle of the population-based training and instance-wise optimization for inference, are shown in Fig. 7.1. The backbone MLP (highlighted in blue) uses residual connections between hidden layers and to further improve focus on the network inputs, the spatial coordinates combined with the SLO information are passed to each of these layers. The MLP and the latent priors of the training images are adapted for 1500 epochs, using Adam optimization, an initial learning rate of 1e - 4 and an exponential learning rate decay with factor 0.99 in each epoch. The adjustment of latent priors for test images

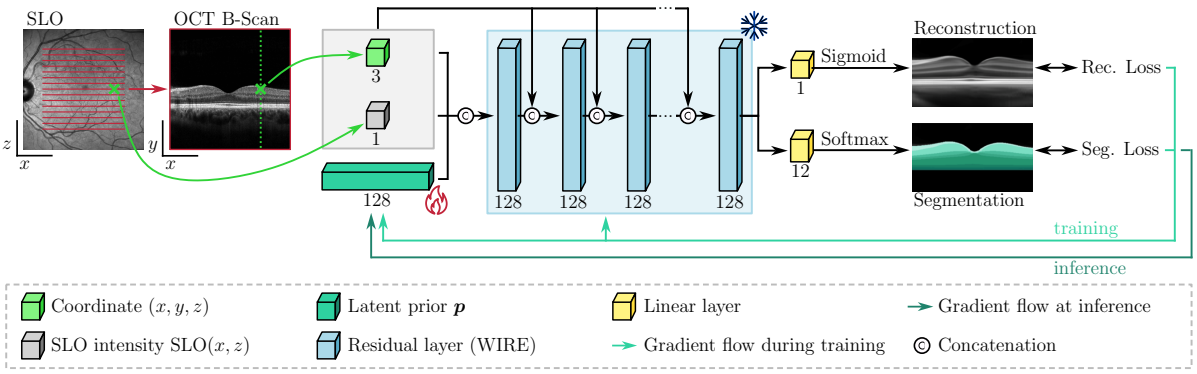


Figure 7.1: The architecture of the generalizable INR for OCT interpolation and retinal layer segmentation. On the left, the relationship between the SLO and the OCT image is shown. Red lines on the SLO image indicate the positions of the OCT B-scans. The position marked with a green cross in the SLO corresponds to the A-scan highlighted with a green dotted line in the OCT. Input to the INR are patient-specific latent priors, 3D coordinates in the OCT image and SLO intensities at the respective en-face positions. Output are the OCT reconstruction and segmentation. During training, the network and the latent priors are adapted, whereas at inference time, only the latent priors of the new images are updated, as highlighted with snowflake and flame symbols.

is run for 200 iterations per image, again using Adam optimization, but with an initial learning rate of 1e-3.

7.1.3 Experiments and Results

The proposed generalizable INR is trained on 80 OCT angiography images with manual segmentations of retinal layers and corresponding SLO images from 40 volunteers of the IMI dataset described in Sec. 2.4.1. After convergence, the framework is adapted to the remaining 20 cases of the IMI dataset by freezing the backbone network and learning new latent codes. All images are preprocessed by flattening at the Bruch's membrane, and cropping of B-scans to a size of 230×512 pixels, centered at the retina. Additionally, images of right eyes are flipped, such that all images have the same orientation. To train the network and to adapt the latent priors, only 16 B-scans are used per image. These are sampled equidistantly from the entire depth of the images, resulting in inter-B-scan distances even larger than in typical clinical settings. Evaluation is then performed on the intermediate B-scans not seen for training or inference optimization.

For comparison, non-generalizable INRs, simple linear interpolation and a more sophisticated registration-based interpolation method [Ehrhardt et al., 2007] are used. For linear interpolation, the segmentation labels are transferred to intermediate, i.e. the interpolated, B-scan positions by performing linear interpolation of the segmentation labels. The resulting continuous values are rounded to integers to receive segmentation labels again. For the registration-based interpolation, the given segmentation labels are

transferred with the deformation fields calculated on the OCT B-scans by using nearest neighbor interpolation. The non-generalizable INRs (SingleINRs) are trained for each OCT image individually using SIREN networks with three linear layers of size 512. Sine activation functions are used here instead of wavelet activation functions, as SIREN performed better than WIRE in the experiments conducted for this work. Both SingleINR and GenINR are trained once with and once without the additional SLO input to be able to quantify the benefit from the multimodal guidance.

Evaluation results are shown in Tabs. 7.1 and 7.2 for the interpolated B-scans of the 80 training images, and for GenINR, the performance on the 20 test cases is also reported. Since the other methods do not generalize to unseen images, they cannot be applied to unlabeled test images. In Tab. 7.1, image similarity metrics MAE, PSNR, SSIM and LPIPS are reported to assess the interpolation quality, whereas in Tab. 7.2, segmentation results are given for the interpolated B-scans, evaluated by DSC, ASSD and HD, averaged over all retinal layer classes (cf. Sec. 2.3 for definitions of the metrics). Finally, as an upper bound for segmentation performance, an nnU-Net [Isensee et al., 2021a] is trained on the same training data as the GenINR and its performance for the test images is reported in Tab. 7.2.

Regarding interpolation performance, the proposed GenINR and $\text{GenINR}_{\text{SLO}}$ deliver best results for three out of four image similarity metrics. Only for LPIPS, the generalizable INRs underperform. Here, the linear and registration-based interpolation achieve the best values, which might be attributed to the fact that simple interpolation hardly changes the contrast and noise level, so that a visually similar impression is created. Compared to that, the images produced by the GenINRs show an overly smooth appearance, since they lack the high-frequency components, as can be seen exemplarily in Fig. 7.2.

The instance-based methods (linear, registration, SingleINR, $\text{SingleINR}_{\text{SLO}}$), in turn, are susceptible to interpolation artifacts. Especially in the foveal region, these methods

Table 7.1: OCT interpolation results for linear and registration-based interpolation, and for INR-based methods. GenINR stands for the proposed generalizable INR, whereas SingleINR indicates networks trained on individual images. The subscript SLO indicates that the corresponding SLO image intensities have been used as additional input to the INR. Results are reported for the interpolated B-scans on train and test images separately. The best results are shown in bold font.

	Method	MAE [%] ↓	PSNR ↑	SSIM [%] ↑	LPIPS ↓
train	linear	5.92 ± 0.55	21.7 ± 0.5	39.7 ± 5.6	**0.12** ± 0.01
	registration	5.76 ± 0.55	22.0 ± 0.5	41.3 ± 5.6	**0.12** ± 0.01
	SingleINR	6.65 ± 0.64	20.6 ± 0.5	36.7 ± 5.9	0.26 ± 0.02
	$\text{SingleINR}_{\text{SLO}}$	9.42 ± 1.51	17.5 ± 1.0	26.8 ± 6.4	0.45 ± 0.07
	GenINR	**5.24** ± 0.51	**22.6** ± 0.5	**48.9** ± 5.7	0.53 ± 0.03
	$\text{GenINR}_{\text{SLO}}$	**5.24** ± 0.51	**22.6** ± 0.5	**48.9** ± 5.7	0.52 ± 0.03
test	GenINR	**6.04** ± 0.57	**21.4** ± 0.5	**44.3** ± 6.0	0.55 ± 0.03
	$\text{GenINR}_{\text{SLO}}$	6.08 ± 0.59	**21.4** ± 0.5	44.2 ± 6.0	**0.54** ± 0.03

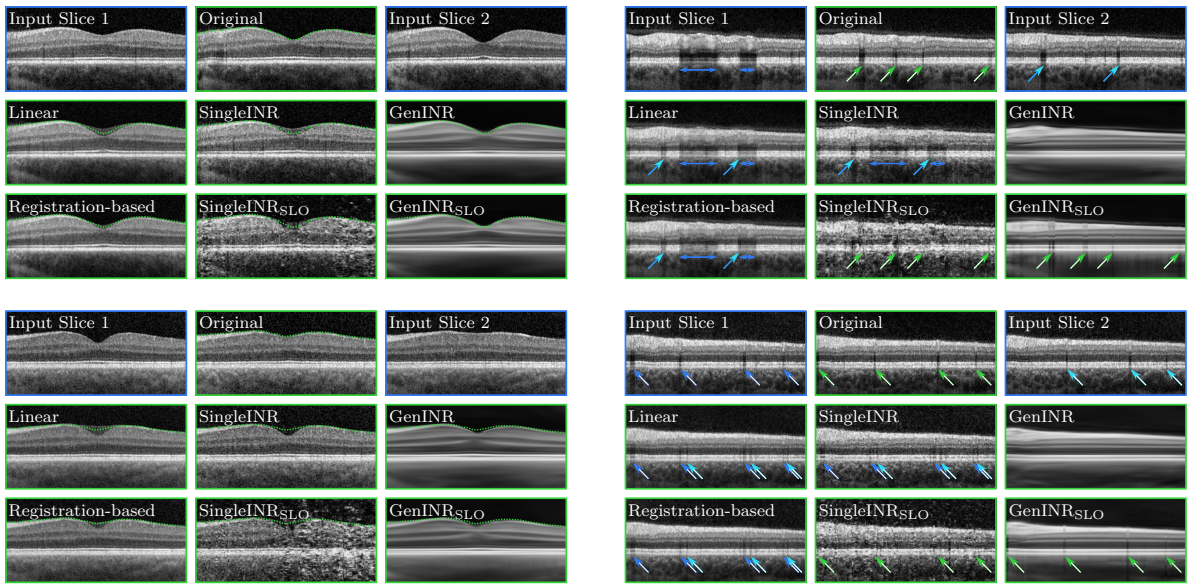


Figure 7.2: Representative OCT interpolation results using linear and registration-based interpolation, SingleINR and GenINR with and without additional SLO input. For each example, two input layers are shown in the top row as well as the original B-scan, which lies in the middle between these two layers. The interpolation results are shown in the two rows below. For examples from the center of the retina, the inner limiting membrane is indicated with a green dashed line (left side), and for examples from the border of the retina, vessel locations are indicated with arrows (right side). Blue arrows mark vessel positions from B-scans used as input to the interpolation methods. These vessels lead to interpolation artifacts for instance-based methods. Green arrows, in turn, highlight the correct locations as seen on the original intermediate slice.

fail to produce a continuous shape (examples on the left in Fig. 7.2) and blood vessels are mislocated since they are propagated from the known B-scans (examples on the right in Fig. 7.2). In contrast, the population-based training enables the generalizable INR to meaningfully continue the shape of the retina on the interpolated B-scans and, with additional SLO information, it also manages to correctly position the blood vessels, as can also be seen in Fig. 7.2. Despite the visually better representation of blood vessels, the quantitative results for GenINR and GenINR$_{SLO}$ are almost the same. Unlike the GenINR, the SingleINR does not profit from the additional SLO information. Since it is adapted to single images, it highly overfits to the small amount of samples given during training, which hinders the adaptability to intermediate image slices.

The segmentation results in Tab. 7.2 confirm the observation that the GenINRs interpolate the retinal shape best, as they show the highest segmentation performance, achieving an average DSC of 0.919 on the interpolated B-scans of the training images. A qualitative segmentation example is shown in Fig. 7.3, highlighting the smoothness of the retinal layer segmentations produced by the GenINRs. For test cases, an average DSC of 0.865 can be observed, which still shows room for improvement compared to the performance of the nnU-Net that receives an average DSC of 0.919 on the test images.

Table 7.2: Segmentation performance for B-scans interpolated with linear and registration-based interpolation, and INR-based methods. SingleINR indicates networks adapted to individual images, whereas GenINR stands for generalizable INR. The INRs are trained once with (SingleINR$_{\text{SLO}}$, GenINR$_{\text{SLO}}$) and once without (SingleINR, GenINR) SLO guidance. Again, results are reported separately for train and test images and best results are given in bold font.

	Method	DSC [%] ↑	ASSD [μm] ↓	HD [μm] ↓
train	linear	88.8 ± 1.1	9.3 ± 0.7	37.2 ± 3.3
	registration	89.1 ± 1.2	7.5 ± 0.8	39.2 ± 4.4
	SingleINR	90.8 ± 1.1	7.7 ± 0.7	37.6 ± 3.6
	SingleINR$_{\text{SLO}}$	83.6 ± 5.3	18.7 ± 12.0	603.4 ± 179.8
	GenINR	**91.9** ± 1.0	**7.0** ± 0.7	**32.5** ± 3.7
	GenINR$_{\text{SLO}}$	**91.9** ± 1.0	**7.0** ± 0.6	**32.5** ± 3.7
test	nnU-Net	**91.9** ± 4.0	**5.0** ± 1.8	**25.1** ± 4.1
	GenINR	86.5 ± 3.7	11.4 ± 2.	45.1 ± 4.9
	GenINR$_{\text{SLO}}$	86.2 ± 3.8	11.7 ± 2.2	45.4 ± 5.0

However, nnU-Net is incapable of interpolation and, thus, receives the original intermediate B-scans as input for segmentation, which gives the method an advantage over the others. Out of all considered methods, the GenINR is the only one capable of generating dense OCT images, while being also able to provide segmentations for unseen cases.

7.1.4 Discussion

INRs offer a powerful tool to process highly anisotropic and sparse data, enabling their representation at arbitrary resolutions. In this section, INRs have been proposed for the joint representation, interpolation and segmentation of OCT images that are acquired with large slice distances in clinical routine. It was shown that the population-trained INR improves over instance-based interpolation methods in terms of shape matching performance, and at the same time, is not prone to typical interpolation artifacts. As a first proof of concept, the correct positioning of blood vessels in between B-scans was enabled via the incorporation of additional SLO information.

The SLO integration was done by passing the SLO intensity at the respective en-face position alongside the spatial coordinate to the INR. This approach easily allows the integration of further imaging modalities, offering the possibility of correctly placing pathologies in the interpolated B-scans that can otherwise only be precisely examined in en-face images. For example, geographic atrophy, the advanced form of AMD, is well observable on OCT, but its extension is typically measured with higher precision on fundus autofluorescence images.

Visually, the OCT image representation of the GenINRs is biased to low frequencies, which leads to small variations within retinal layers being smoothed and small blood vessels in the center of the retina being missed. Here, further research into alternative

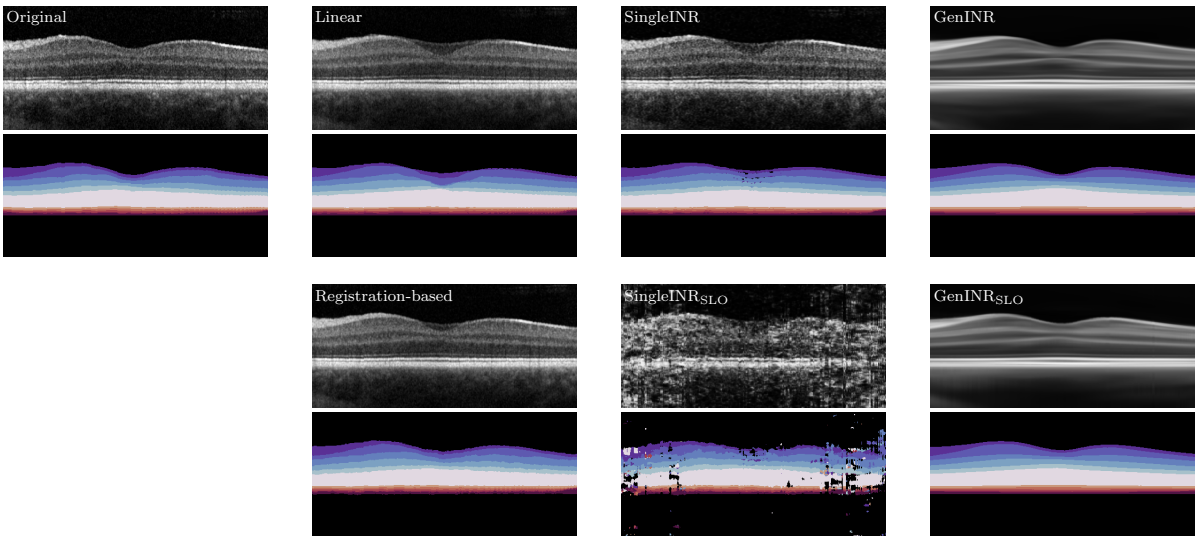


Figure 7.3: Exemplar segmentation results for an interpolated B-scan. Interpolation is performed with standard linear and registration-based interpolation, as well as INR-based methods. For each method, the interpolation result is shown at the top and the respective segmentation result at the bottom. The generalizable INRs deliver exceptionally smooth segmentation results.

network architectures or input encodings needs to be done. It would also be desirable to improve the performance for the test images, that still shows a gap to the results for the training images. Yet, the proposed GenINR method is the only one of the considered methods that is able to segment new images without given annotations. All other interpolation methods require some of the B-scans to be segmented. This is achieved by training on only 80 OCT and SLO images, for which a small number of 16 annotated B-scans is used, indicating the usability of the proposed framework for small medical datasets.

In summary, the presented method simultaneously allows the interpolation and segmentation of retinal OCT images. Based on population-trained INRs, it is adaptable to unseen images and resolution-agnostic. It can, thus, be used to provide dense OCT volumes, enabling precise volumetric measurements. Furthermore, the proposed GenINR manages to combine multimodal inputs with different resolutions, view points and spatial dimensions, offering the possibility for 3D image analysis instead of the previously often used 2D approaches.

7.2 Joint Decomposition and Registration of Pathological OCT Image Data

Image registration with INRs is a very young field of research but has shown promising registration accuracy, while combining several advantages of both classical iterative and DL-based registration algorithms. INR-based registration methods do not rely on large training datasets, but can be applied directly to individual image pairs like classical algorithms. At the same time, building on the automated differentiation techniques of modern DL libraries, they allow the usage of very complex objective functions such as CNN-based methods. Finally, INRs do not require densely sampled input images, but can be applied to any resolution, allowing the alignment of images with different spatial resolutions [Wolterink et al., 2022; Byra et al., 2023a].

However, in the case of medical images with pathologies, INRs face the same problems as other registration approaches. Non-correspondences between patients, modalities or time points can lead to registration errors, hindering the comparability of registered images and reliable volume measurements. Additionally, for newly emerging pathologies, the resulting deformation fields cannot reflect the actual movement caused by these pathologies, as deformation alone cannot insert new structures.

To cope with severe appearance differences between images, ImpRegDec [Byra et al., 2023b] proposes to use three INRs for a single image registration task, that produce the deformation field and decompose the moving image into two parts, the support image and the residual image. Support and residual image together are trained to reproduce the moving image. In addition, the deformed support image is compared to the reference image. As a result, the support contains those parts of the moving image that positively influence the registration. Primarily, these are features that the moving and the fixed image have in common [Byra et al., 2023b]. However, this approach has only been applied to images showing no pathologies and offers only limited control over which image parts are packed into the residual and which into the support image.

Inspired by ImpRegDec but closer to the idea of FluidRegNet, the CNN-based, pathology-aware image registration from the previous chapter, this section proposes to use two INRs for image registration. The first INR generates the registration deformation field and the second produces an image, that, similar to FluidRegNet's appearance offset map and the residual image of ImpRegDec, contains the structures that differ between fixed and moving images. An additional sparsity loss is introduced allowing better control of the content of the residual image, that reflects the pathological deviations between moving and fixed images. The proposed fluid registration INR (FRINR) is shown to increase plausibility of the deformation fields compared to a deformation-only INR, and to be usable for the unsupervised segmentation of non-correspondences caused by disease progression.

7.2.1 Related Literature

The first work using INRs for the representation of a transformation function for image registration was IDIR [Wolterink et al., 2022]. The authors propose to use an INR to map spatial coordinates to displacement vectors and train the network in an unsupervised fashion based on image similarity and deformation regularization. The framework is applied for the registration of 4D chest CT data and shown to surpass CNN-based methods. Furthermore, the authors investigate ReLU and sinusoidal activation functions for their INR and find SIREN to perform best.

Byra et al. extend the investigation of different activation functions for brain MRI registration [Byra et al., 2023a]. Again, SIREN is found to achieve best results. Furthermore, this work proposes a new cycle consistency loss function to promote diffeomorphic deformations in INR-based image registration and explores the usage of INR ensembles and cascades. The experiments indicate that the cascaded SIREN outperforms not only a single SIREN but also CNN-based and classical image registration frameworks. The same authors also proposed the previously mentioned ImpRegDec, which uses three INRs for joint image registration and decomposition [Byra et al., 2023b]. The resulting support and residual images are trained to contain those image features of the moving image that positively and negatively influence the registration task. The authors show improved registration performance compared to a registration-only INR, but apart from comparing the support with the moving image, the method offers no way of controlling which image parts are packed into the residual and which into the support image. In inter-patient registration, this may cause anatomical structures that both images have in common to be represented in the residual rather than in the support image. This, in turn, might cause the resulting deformation field to be overly smooth and to only match the global tissue shape instead of performing local displacements. Also, in longitudinal registration, problems could arise if growing or newly occurring pathologies are displayed excessively large in the support image (and are compensated in the residual), so that volume adjustment with the deformation is no longer necessary.

Several works investigate how to speed up INR registration [Sideri-Lampretsa et al., 2024; Tian et al., 2024a] and how to get (approximative) diffeomorphic displacements [Han et al., 2023; Sideri-Lampretsa et al., 2024; Sun et al., 2024; Tian et al., 2024a; van Harten et al., 2024]. SINR, for example, proposes to use only a subset of the images' coordinates as input and to subsequently perform B-spline interpolation to receive a dense deformation field [Sideri-Lampretsa et al., 2024]. This sparse coordinate sampling not only speeds up the INR training but also reduces the number of foldings in the deformation field.

Also based on B-splines is the method proposed in [Großbröhmer et al., 2024], that uses a generalizable INR to learn an implicit atlas image used for chest X-ray registration. During training, the INR models the atlas by mapping spatial coordinates to intensity values, which are compared to the respective patient images in the loss function. At the same time, individual deformation parameters are learned for each patient in the training set. These parameters are used to warp the input coordinates before passing them to

the INR to enable patient-specific reconstructions. During inference, the atlas image is generated by passing a regular identity grid to the INR. This work is one of the few that apply INR registration for medical images with pathologies. Explicitly considering the shape and appearance differences introduced by evolving pathologies in retinal diseases, is the goal of FRINR, the INR-based registration method proposed below.

7.2.2 Methods

To enable the pathology-aware registration of retinal OCT images, this section proposes a pairwise image registration solution based on INRs. FRINR uses two INR networks for each pair of moving and fixed images. Both networks get the same image coordinates as input to produce two separate outputs: The registration deformation field and the so-called residual image. The goal is to generate a deformation field that matches the anatomical features the moving and fixed images have in common, and a residual image that contains the pathological differences between the images. To achieve this, it is assumed that the fixed image $F : \Omega \subset [-1, 1]^3 \to \mathbb{R}$ can be represented as a combination of the moving image $M : \Omega \to \mathbb{R}$ and the residual image $R : \Omega \to \mathbb{R}$ deformed with a transformation $\varphi : \Omega \to \mathbb{R}^3$ so that

$$F \approx (M + R) \circ \varphi. \tag{7.3}$$

As first proposed by [Wolterink et al., 2022], the transformation $\varphi(\boldsymbol{x}) = \boldsymbol{x} + \Delta\boldsymbol{x}$ is represented with an implicit deformation network that maps spatial coordinates $\boldsymbol{x} \in [-1, 1]^3$ to displacement vectors $\Delta\boldsymbol{x} \in \mathbb{R}^3$ such that a coordinate $\boldsymbol{x}$ in F corresponds anatomically to the transformed coordinate $\varphi(\boldsymbol{x})$ in M.

The residual image is learned with a second implicit network that has the same network architecture as the deformation network, except for the output layer that generates grayscale values rather than displacement vectors. Both networks are trained simultaneously using loss function

$$\mathcal{L}_{\text{FRINR}} = \mathcal{L}_{\text{NCC}}(F, M \circ \varphi) + \alpha\mathcal{L}_{\text{reg}}(\varphi) + \beta\mathcal{L}_{\text{MSE}}(F, (M + R) \circ \varphi) + \gamma\mathcal{L}_{\text{sparse}}(R), \tag{7.4}$$

composed of the classical image registration objectives $\mathcal{L}_{\text{reg}}(\varphi)$ and $\mathcal{L}_{\text{NCC}}(F, M \circ \varphi)$, controlling the regularity of the deformation field φ and the image similarity between fixed image F and warped moving image $M \circ \varphi$, as well as the two additional terms $\mathcal{L}_{\text{MSE}}(F, (M + R) \circ \varphi)$ and $\mathcal{L}_{\text{sparse}}(R)$ defining the appearance of the residual image. The MSE loss ensures that the moving image, which has been modified with the residual image, looks similar to the reference image after deformation, whereas the sparsity loss favors small intensity values in the residual image. Analogously to the appearance offset of FluidRegNet (cf. Chapter 6), the residual is, thus, encouraged to contain the most dominant structural differences between fixed and moving images, which, in intra-patient registration, are primarily due to disease progression.

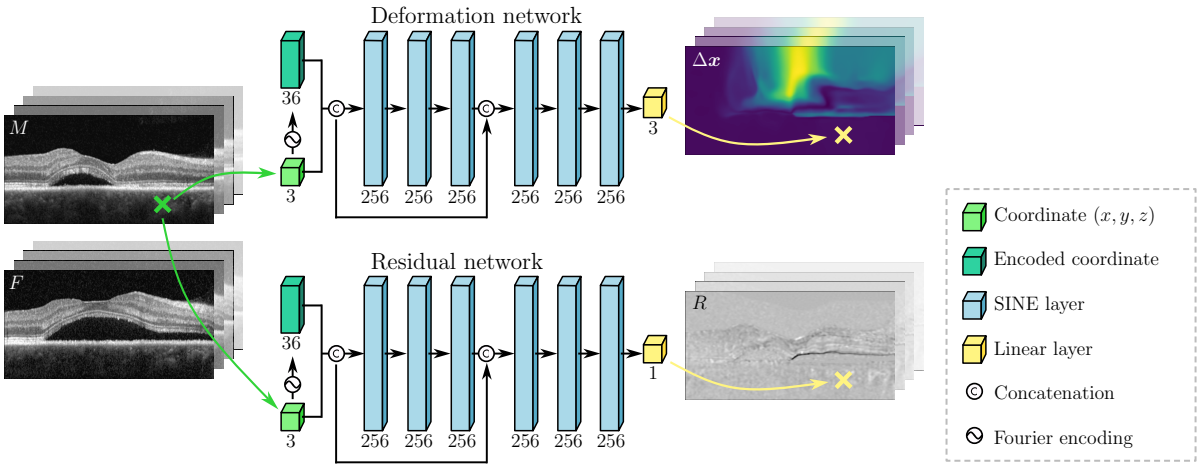


Figure 7.4: Overview of the proposed INR-based image registration method. Two INRs are used to generate residual image and deformation field. Training is performed by comparing the fixed image to the warped moving image and the warped combination of moving and residual image. Furthermore, regularization of the deformation field is used as well as sparsity of the residual image. Numbers below layers indicate the number of output features of the respective layer.

More specifically, the loss components are defined as follows. A combination of local (LNCC) and global normalized cross-correlation (NCC) is used as image similarity metric

$$\mathcal{L}_{\text{NCC}}(F, M\circ\varphi) = \frac{1}{2n}\sum_{\boldsymbol{x}\in\Omega}\Big(\text{NCC}\big(F(\boldsymbol{x}), (M\circ\varphi)(\boldsymbol{x})\big) + \text{LNCC}\big(F(\boldsymbol{x}), (M\circ\varphi)(\boldsymbol{x})\big)\Big), \quad (7.5)$$

where n is the number of coordinates in the entire discretized image domain, i.e. the number of voxels of the input images. The LNCC is calculated with a window size of $9\times9\times9$ voxels. The center coordinates of the voxels serve as input to the INRs. The deformation field regularizer consists of a penalty for negative values in the Jacobian determinant $|J_\varphi|$ and an L2 regularization of the deformation's gradient $\nabla\varphi$:

$$\mathcal{L}_{\text{reg}}(\varphi) = \frac{1}{n}\sum_{\boldsymbol{x}\in\Omega}\Big(|1 - |J_\varphi(\boldsymbol{x})|| + \lambda\|\nabla\varphi(\boldsymbol{x})\|_2\Big). \quad (7.6)$$

Based on the choice of loss functions in [Byra et al., 2023b], the image similarity between the fixed image and the deformed combination of moving and residual images is evaluated with the MSE loss function

$$\mathcal{L}_{\text{MSE}}(F, (M+R)\circ\varphi) = \frac{1}{n}\sum_{\boldsymbol{x}\in\Omega}\Big(F(\boldsymbol{x}) - \big((M+R)\circ\varphi\big)(\boldsymbol{x})\Big)^2. \quad (7.7)$$

Finally, the sparsity loss is given by

$$\mathcal{L}_{\text{sparse}}(R) = \frac{1}{n}\sum_{\boldsymbol{x}\in\Omega}\Big(R(\boldsymbol{x})\Big)^2. \quad (7.8)$$

For INR training and evaluation, the longitudinal CSCR dataset (cf. Sec. 2.4.1) is used. Image registration pairs consist of two OCT volumes from subsequent visits of the same patient. Only the manually segmented images with a B-scan resolution of 496×512 voxels are used here, resulting in 138 image volume pairs, which are registered using the proposed INR approach. The input OCT images are preprocessed with a flattening at the Bruch's membrane, and a cropping to a B-scan size of 256×496 to remove confounding background. Additionally, each B-scan is downsampled to a size of 128×248 and the images' intensities are scaled to $[-1, 1]$.

The weighting parameters α, β, γ and λ are selected empirically based on previous works and set to $\alpha = 25$, $\beta = 100$, $\gamma = 200$ and $\lambda = 0.01$. The deformation and residual networks are both SIREN networks with five FC hidden layers of size 256. Fourier encoding with six frequencies is applied to the input coordinates, which are passed to the input layer of the networks and additionally also concatenated with the output of the second hidden layer. A detailed overview of the network architectures is given in Fig. 7.4. The network weights are initialized as proposed in the original SIREN paper [Sitzmann et al., 2020], except for the last layer of the deformation network, whose weights are randomly initialized from a uniform distribution in [-0.0001, 0.0001]. This initialization ensures small deformations at the beginning of the training. Both networks are trained simultaneously with AdamW optimization, an initial learning rate of 1e-4 and an exponential learning rate scheduler with a decay rate of 0.99. Network training is performed for 1000 epochs, whereby one epoch corresponds to a run through all B-scans of the input images. In each training step, all coordinates of one B-scan are sent through the networks simultaneously.

7.2.3 Experiments and Results

In this section, first the image registration accuracy of FRINR and the plausibility of the resulting deformation fields are evaluated. Subsequently, unsupervised pathology segmentation is performed using the residual images generated by FRINR. In Tab. 7.3, OCT image registration results are given for FRINR and compared to a registration-only INR that is trained with loss function (7.4) except for the sparsity loss term, using the same weighting parameters, and calculating the MSE loss for fixed and warped moving images, since no residual is given for this method. Furthermore, results are reported for ImpRegDec as well as ANTs SyN [Avants et al., 2008] and FluidRegNet [Andresen et al., 2024], both investigated in the previous chapter (cf. Tab. 6.1). Registration accuracy and the plausibility of the resulting deformation fields are evaluated as before by measuring the ASSD and HD of the ILM, the average percentage of voxels with negative Jacobian and the volume change in pathology and non-pathology tissue regions. The settings used for ANTs SyN and FluidRegNet are the same as in the previous chapter, but results have been re-calculated on the cropped images used to train the INR-based methods. For comparison to ImpRegDec [Byra et al., 2023b], the code provided by the authors has been extended to 3D and adapted to grayscale images.

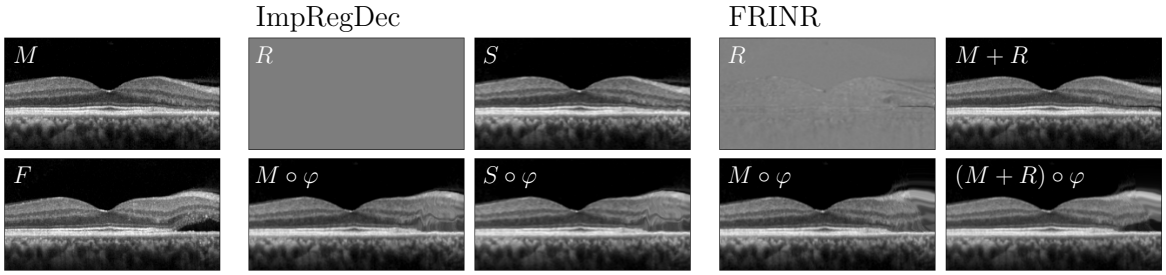
Table 7.3: Image registration results for ANTs SyN, FluidRegNet (FRN), INR-based image registration, ImpRegDec (IRD) and the proposed FRINR. The Jacobian determinant of the deformation field is calculated analytically for the INR-based methods, whereas discretized approximations have to be used for SyN and FRN.

Method	ASSD ILM ↓	HD ILM ↓	$\lvert J_\varphi\rvert \leq 0$ [%] ↓	$\lvert 1-\lvert J_\varphi\rvert\rvert_{\text{healthy}}$ ↓	$\lvert 1-\lvert J_\varphi\rvert\rvert_{\text{fluid}}$ ↑
Before	16.13 (25.05)	40.91 (59.97)	-	-	-
SyN	5.80 (10.86)	17.64 (40.26)	0.3860 (0.7952)	0.1327 (0.0684)	0.4614 (0.2123)
FRN	8.64 (17.26)	24.37 (44.63)	**0.0386** (0.0995)	0.0889 (0.0561)	**0.9826** (0.7796)
INR Reg.	5.02 (2.14)	12.43 (7.62)	0.6492 (0.7201)	0.0853 (0.1176)	0.5793 (0.6505)
IRD	5.77 (8.51)	14.82 (24.40)	0.3762 (0.4029)	0.0670 (0.0930)	0.5044 (0.5066)
FRINR	**4.95** (2.09)	**12.33** (8.73)	0.5665 (0.6285)	**0.0602** (0.1043)	0.4902 (0.5649)

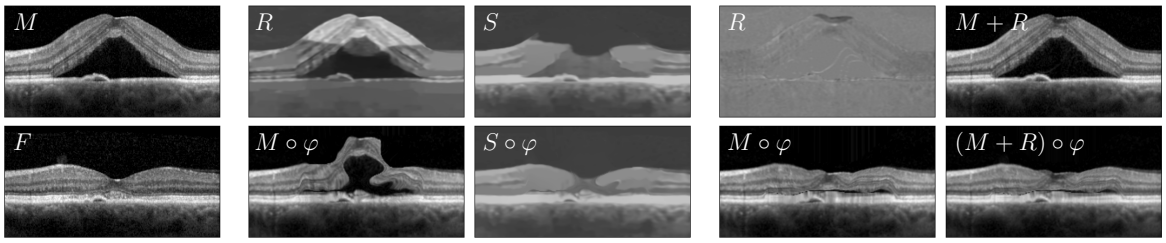
The results show that all INR-based methods outperform both the DL-based FluidRegNet and the classical iterative algorithm ANTs SyN in terms of ILM alignment. In particular, the INRs manage to map the ILM for the seldom extreme cases of the dataset, for which the other methods often fail, as indicated by the much lower standard deviation. FRINR achieves the best ILM alignment with an average ASSD of 4.95 ± 2.09 μm, compared to 5.80 ± 10.86 μm for ANTs SyN, 8.64 ± 17.26 μm for FluidRegNet, 5.02 ± 2.14 μm for the registration-only INR and 5.77 ± 8.51 μm for ImpRegDec. Two exemplary registration results for a large dissolving or forming SRF are shown in Fig. 7.5 (b) and (c) for ImpRegDec and FRINR. A known problem of image registration with INRs, however, is the increased tendency to introduce foldings into the deformation field. This can also be observed in the results here, where the INRs show an increased number of voxels with negative Jacobian.

Pathological fluids in the retina displace the surrounding tissue, but typically do not introduce strong volume changes of the tissue itself. Instead, the retinal layers show severe changes in the shape and location caused by the accumulating fluid, while the volume of the pathology may change drastically from one visit to the next. A plausible deformation should, therefore, produce stronger volume changes in the pathological regions than in the surrounding tissue. All INR-based methods manage to generate such deformations fields, introducing the largest volume changes in pathological image areas while keeping them comparably small in healthy tissue regions. In this respect, both ImpRegDec and FRINR manage to produce more realistic deformation fields compared to the registration-only method using a deformation INR and no additional intensity network. The average deviation from a unit Jacobian in healthy tissue regions is 0.0670 for ImpRegDec and 0.0602 for FRINR, compared to 0.0853 for the registration-only INR. The mean deviation in pathological areas is 6.8 times higher for the registration-only method, 7.5 times higher for ImpRegDec and 8.1 times higher for FRINR. The decomposition into common and discriminative features, as it is done in FRINR by modeling the fixed image as a combination of moving and residual image, thus, helps to improve registration performance and deformation plausibility.

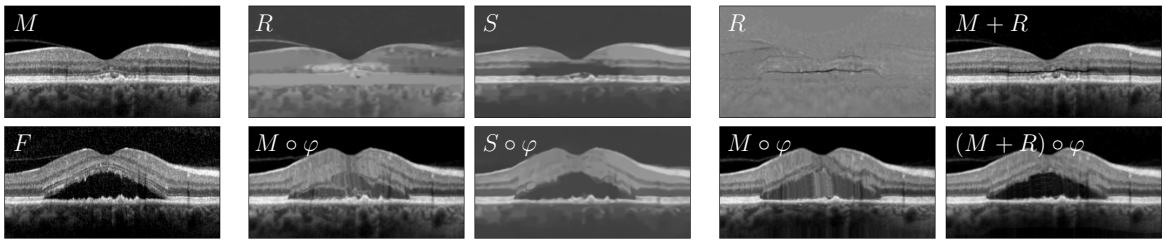
Compared to ImpRegDec, FRINR shows a higher registration accuracy. This may partially be attributed to a suboptimal choice of weighting parameters in the loss func-



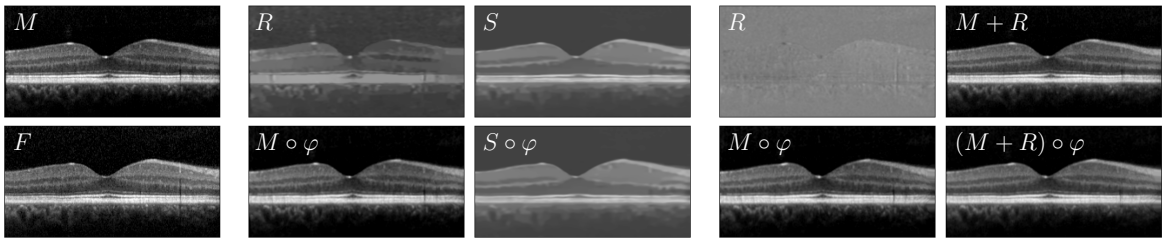
(a) Example of a newly developed SRF



(b) Example of a dissolving SRF



(c) Example of a large, newly developed SRF



(d) Example of a stable case

Figure 7.5: Image decomposition and registration results of ImpRegDec and FRINR. Shown are the moving image M and the fixed image F. For ImpRegDec, the residual image R and the support image S are depicted, as well as the registration results $M \circ \varphi$ and $S \circ \varphi$. For FRINR, the residual image R and the combination of moving and residual image are shown. Also, registration results $M \circ \varphi$ and $(M + R) \circ \varphi$ are given. The combination of residual and support images in ImpRegDec leads to an unclear separation of the tissue structures and incomplete deformations. The proposed residual-only approach, FRINR, provides more consistent results, with the residuals primarily containing the changing pathologies, allowing for more plausible registration results.

tion for ImpRegDec, since the settings originally proposed by the authors were used without further fine-tuning. However, the qualitative results in Fig. 7.5 show that the

combination of support and residual image used in [Byra et al., 2023b] oftentimes introduces unwanted behavior. A strong local optimum in ImpRegDec is the generation of either empty residual or empty support images (Subfig. (a)). Subfig. (b), in turn, shows an example of a dissolving fluid that is not fully compensated for with the deformation field. Instead, tissue is omitted in the support image and inserted into the residual. The third example (Subfig. (c) shows a new pathology that appears in the support image, causing the deformation in this area to be incomplete. In contrast, FRINR manages to produce consistent and meaningful results for all cases, showing that the proposed method offers a better control of the appearance of the residual image in longitudinal registration applications. Consequently, the more meaningful residual enables a better alignment of the ILM, explaining the higher registration accuracy of FRINR.

For the unsupervised segmentation of newly developed pathologies, the intensities given in the residual image are analyzed, similar to the experiment performed in Sec. 6.3.2 analyzing the usage of FluidRegNet's appearance offsets for pathology segmentation. New fluids are segmented with

$$\mathcal{S} = \{\boldsymbol{x} \in \Omega \mid R(\varphi(\boldsymbol{x})) < \tau_R\}, \tag{7.9}$$

where the threshold τ_R has been found via grid search and is set to -0.15. Like this, fluids are defined as image areas whose intensities have been darkened by the residual image. The deformation field extends these areas of changed intensities to the size observed in the fixed reference image, allowing direct comparison of $\mathcal{S}$ and the ground truth segmentation of new fluids in the fixed image. The generation of the ground truth segmentation is described in Sec. 4.3.1.2.

It was also analyzed to use a combination of image areas with low values in the residual and image areas with large volume increase, as done for FluidRegNet, but this resulted in reduced performance for FRINR. The unsupervised fluid detection and segmentation performances of FRINR are reported in Tab. 7.4 and compared to FluidRegNet and NCR-Net, the image registration and non-correspondence segmentation network proposed in Chapter 4.

Table 7.4: New fluids detection and segmentation results for three different registration frameworks, all proposed in this thesis: NCR-Net (NCR), FluidRegNet (FRN) and FRINR. For NCR and FRN, these are the same results as reported in Tab. 6.2. For each method, the number of correctly identified B-scans containing new fluid (sens^{img}, denoted as sens for brevity) and the DSC averaged over all B-scans that are detected as containing new fluids is reported (DSC_{det}, denoted as DSC for brevity, given in %). Additionally, the number n_{FP} and volume V_{FP} (in pixels) of generated false positive fluids is reported.

	All fluids				IRF		SRF		PED	
Method	sens ↑	DSC ↑	n_{FP} ↓	V_{FP} ↓	sens ↑	DSC ↑	sens ↑	DSC ↑	sens ↑	DSC ↑
NCR	213/293	45.79	2.07	335.25	29/37	47.77	174/212	57.44	**19/60**	40.60
FRN	**222/293**	**59.49**	**1.09**	**91.48**	34/37	**55.95**	**179/212**	**66.57**	18/60	**50.95**
FRINR	214/293	51.57	4.57	276.42	**36/37**	55.47	169/212	62.53	**19/60**	43.69

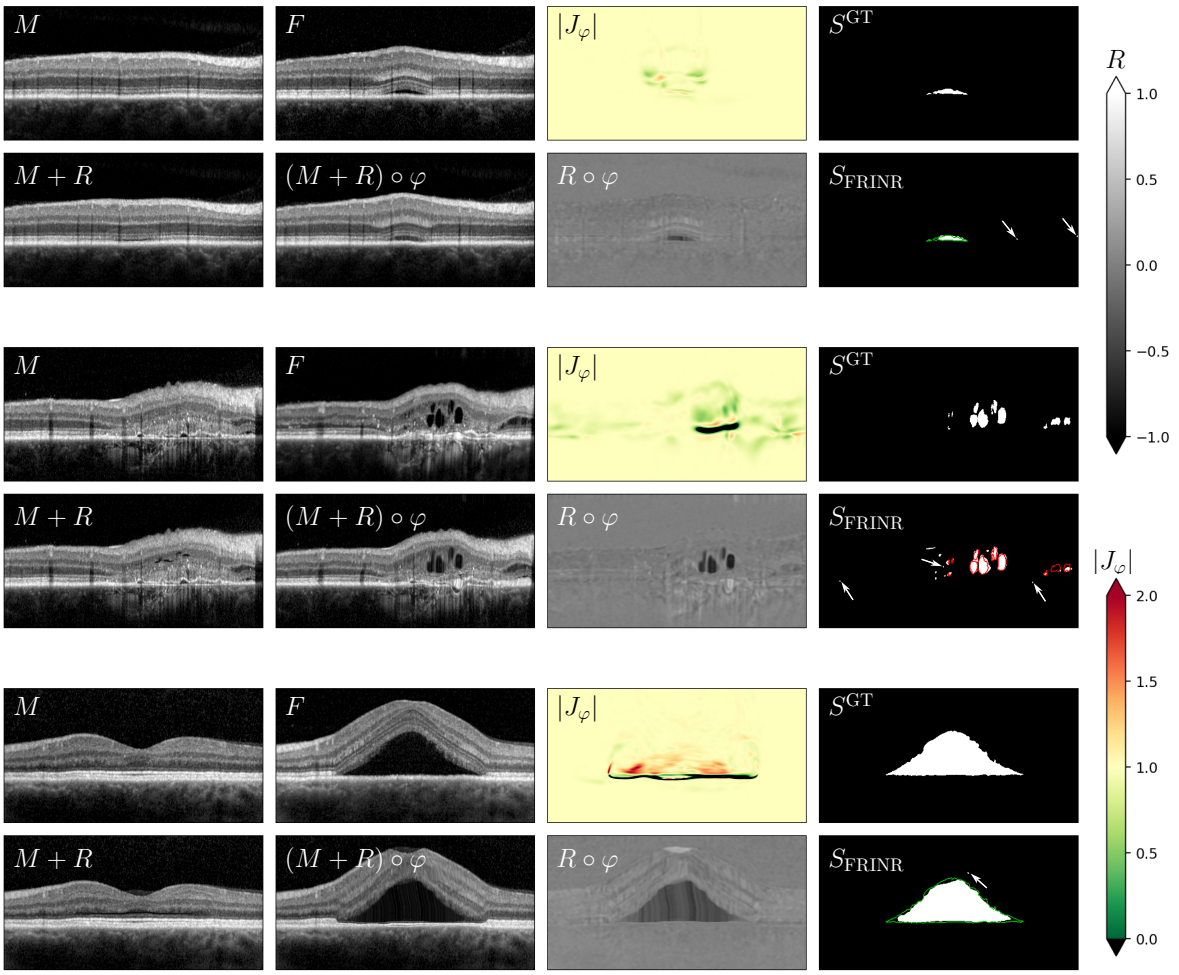


Figure 7.6: Three new fluid segmentation results of FRINR. New fluids are found based on thresholding of the warped residual image. For each registration pair, the moving and fixed images are shown, as well as the combination of moving and residual image before and after deformation. Additionally, the analytic Jacobian determinant of the deformation field, the warped residual and the ground-truth and predicted segmentation of new pathologies are shown. The ground truth segmentation borders are additionally overlaid onto the predicted segmentation (green for SRF and red for IRF). Small false positive fluids are marked with arrows.

Compared to the CNN-based methods, FRINR provides a slightly reduced detection rate for SRF, but performs comparable for PED and best for IRF, for which it misses only one B-scan. The detected fluids are segmented with an average DSC of 0.5547 for IRF, 0.6253 for SRF and 0.4369 for PED, a performance between that of NCR-Net and FluidRegNet. Despite performing on individual examples, FRINR successfully manages to decompose the moving and fixed images into common and discriminative features, with the residual containing primarily the changes caused by disease progression. However, as is to be expected, this adaptation for individual images also leads to

a higher susceptibility to false positive results compared to FluidRegNet and NCR-Net, that are both trained on large datasets. These false positive results can be caused by non-correspondences between moving and fixed images, which are not introduced by disease progression. This might include differing brightness, imaging artifacts, noise or incomplete alignment of the ILM, as can exemplarily be seen in Fig. 7.6. Here, three examples of new fluid segmentation results with FRINR are shown. Also, it can be seen that the false positive fluids produced for these examples are small (highlighted with arrows), indicating that additional post-processing of the segmentations could improve results. For a fair comparison to the other methods, it was decided to use the same processing as for NCR-Net and FluidRegNet, which consists of a masking to the retina and a removal of lesions smaller than four voxels.

7.2.4 Discussion

This section proposed a fluid-aware INR-based image registration method (FRINR), designed to allow the integration of developing pathologies into the registration with INRs. FRINR uses two separate INRs to generate the deformation field as well as the residual image, that is trained to contain structural differences between moving and fixed images. INR training is done in an unsupervised manner, based on image similarity of fixed and moving images and deformation field regularization. The registration task is supported by a second image similarity loss term that compares the fixed image with the deformed combination of moving and residual image. The residual image is, thus, encouraged to compensate for intensity differences between moving and fixed images. Together with an additional sparsity loss term, the residual image is effectively manipulated to primarily contain the main structural differences between the images to be registered. This allows the unsupervised segmentation of new pathologies, as shown exemplarily for IRF, SRF and PED.

The INR-based methods analyzed in this section all manage to outperform both ANTs SyN and FluidRegNet in terms of ILM alignment, as they are able to also align the upper retina border for cases with extreme changes between time points. Compared to an image registration INR that does not use a residual image, FRINR showed a higher registration accuracy and a reduced number of foldings in the deformation fields. In addition, the deformation fields were shown to introduce large volume changes in pathological regions while keeping the volume change small for healthy tissue.

INR registration is known to introduce more foldings than CNN-based or classical algorithms. Here, further improvement might be achieved using (approximately) diffeomorphic INRs, such as the methods presented in [Han et al., 2023; Sun et al., 2024; Tian et al., 2024a; van Harten et al., 2024]. The usage of generalizable INRs for the proposed registration method could also help to reduce the number of foldings. In addition, the population-based training of generalizable INRs could remove redundancies in the optimization processes and speed up adaptation for new cases. Finally, a generalizable INR might reduce the susceptibility to non-pathological differences of the images in the

proposed unsupervised segmentation of new fluids. Furthermore, the initial positioning of pathologies in the residual images is not anatomically grounded, but appears to be done in such a way that the deformations applied to draw the pathologies to the size of the fluids in the reference image are as small as possible (cf. Fig. 7.5 (a) and (c)). An interesting field for further research would be to exploit prior anatomical knowledge about the localization of pathologies such as SRF, IRF or PED into the optimization process. Physics-informed neural networks [Rodrigues, 2024; Zhang et al., 2025] could be a way of incorporating this knowledge.

To summarize this section, the proposed FRINR manages to perform fluid-aware longitudinal registration of 3D OCT images, similar to the previously proposed FluidRegNet. The INR-based method enables the registration of severely changed retinae, surpassing even the classical registration method ANTs SyN, while generating plausible deformation fields. The generated residual images can be used to detect and segment newly developed pathologies and might, thus, be used for automatic disease progression evaluation. FRINR requires no large training dataset but can directly be applied to any two images to be registered. Additionally, it is trained in an unsupervised manner without requiring pathology annotations, so it can be used in scenarios with scarce data and without medical expert knowledge.

Chapter 8

Summary and Outlook

Deep learning has revolutionized the image processing field, achieving or even surpassing human performance for many tasks [Kolbinger et al., 2023; Ma et al., 2024]. However, data scarcity remains a major challenge for both the development and the evaluation of deep learning-based methods for medical image analysis. Especially for longitudinal medical image data, annotations are difficult to get since expert knowledge is required to perform the segmentation on multiple 3D image volumes. In addition, many of the publicly available pre-trained networks and foundation models are not directly applicable to pathological data, as most of them were trained on healthy images. The goal of this work, therefore, was the development of deep learning-based algorithms for the automatized analysis of medical time series image data while minimizing the need for manual annotations for training. Another aim of this work was to improve image registration by automatically detecting changes between time points and by incorporating pathologies into the registration process to produce more realistic deformations. These goals were achieved with the solutions presented in Chapters 3 to 7. While Chapter 3 uses clustering to segment pathologies in individual images, the methods presented in Chapters 4 to 7 use image registration to jointly analyze images from different time points. The proposed networks are able to simultaneously learn to spatially align the input images, to segment non-correspondences between them (Chapter 4), to model the appearance of new lesions (Chapter 5) and to integrate new pathologies in the deformation process (Chapters 6 and 7).

After giving the methodological background of the proposed algorithms in Chapter 2, Chapter 3 focuses on the reduction of required annotated training samples. Contrastive invariant information clustering is proposed for the segmentation of pathologies in OCT images, relying on no or only a few annotated examples. For the comparably obvious wound pathologies analyzed in the first part of the chapter, the applied clustering loss function suffices to separate healthy from pathological tissue. Fully unsupervised training was not successful for the segmentation of photoreceptor atrophy considered in the second part of the chapter. This behavior can be explained on the one hand by the subtle appearance of photoreceptor pathologies and on the other hand by the many, far more obvious and confounding pathologies (SRF, IRF, PED) on the images. Still, the contrastive IIC allows the network to be trained partially supervised, and, thus, to exploit the information contained in the unannotated images. Using the same small number of annotated images, the proposed partially supervised approach achieves a substantially improved detection of atrophy regions compared to a fully supervised classification network featuring the same architecture as the IIC network.

NCR-Net, proposed in Chapter 4, is trained by masking outliers in the image distance measure and by enforcing the masked areas to be small and smoothly bordered. With

appropriately weighted loss components, this masking procedure prevents unwanted behavior that is otherwise frequently observed in the registration of pathology images. Instead of performing implausible deformations to compensate for the intensities observed in the pathologies, NCR-Net produces consistent deformation fields even within the pathologies. Although the resulting non-correspondence maps do not exclusively show pathologies, but all kinds of differences between the images, they are sharply delineated and can be used to segment pathologies with simple post-processing methods such as the region growing used here. To do so, the training of NCR-Net does not require any pathology annotations, instead only segmentations of the entire displayed organs were used. It was also shown that the performance hardly deteriorates if no annotations are used at all.

In Chapter 5, NCR-Net is extended for the segmentation of new multiple sclerosis lesions in FLAIR MR images. Here, new lesions are not only segmented, but also their appearance is estimated as offset from the baseline image. ANCR-Net is pre-trained in a self-supervised manner using artificial deformations and synthetically introduced lesions. The small size and seldomness of MS lesions necessitates a supervised fine-tuning of the network. The MSSEG-2 training data contains 232 new MS lesions in 32 images, which still is a very limited number of training samples. Compared to other supervised methods, ANCR-Net achieves a reduced bias towards under- or oversensitivity. ANCR-Net, thus, provides a reliable estimate of real new lesion load and showed robust against several imaging artifacts typically observed in MR imaging.

Chapter 6 addresses the problem of retinal OCT registration, where severe shape and appearance differences are regularly observed between subsequent examinations. FluidRegNet matches such images by estimating a sparse appearance offset, that reflects the onset of fluid formation, and deforming the altered baseline image to the follow-up reference space. This results in realistic deformation fields with strong volume changes in pathological image regions, displacing the surrounding unaffected tissue. In combination with the appearance offset maps, the fluid-aware deformation fields are successfully used to segment newly emerging pathologies and shown to provide visual guidance of improving and deteriorating eye regions. In the second part of the chapter, FRN-Ano is proposed as a direct extension of FluidRegNet. FRN-Ano performs inter-patient registration between healthy and pathological images. Since no fluids are present in the healthy images, each fluid observed in the reference images can be considered "new" and segmented by analyzing the areas of volume and appearance changes. The resulting anomaly detection is shown to surpass other anomaly detection methods in terms of delineation accuracy, while showing a reduced detection sensitivity. In addition to fluid segmentation, FRN-Ano is used to detect more subtle photoreceptor abnormalities and in combination with the atrophy segmentation from Chapter 3, a reliable separation of normal and abnormal PEZ areas is achieved.

In the first part of Chapter 7, implicit neural representations are used to interpolate and segment retinal OCT images, which are typically acquired with large slice distances. The INR combines two image modalities, with the SLO showing a top view of the retina and the OCT showing cross-sections. The SLO integration enables the depiction of blood

vessels at the correct locations, information that cannot be inferred from the neighboring OCT B-scans. Furthermore, the population-based training allows adapting the trained INR to unseen samples based on the reconstruction error only. A segmentation of new images is, thus, possible, which is exemplarily shown to be exceptionally smooth. FluidRegNet and FRN-Ano are based on the assumption that retinal pathologies manifest with darker intensities compared to the surrounding tissue. No such assumption is made for the INR-based registration proposed in the second part of Chapter 7. Rather, the images are decomposed into automatically learned common and discriminative features, which are represented in the residual images. To segment IRF, SRF and PED, thresholding is performed on the resulting residual images, again assuming a dark appearance of pathologies. The residual, however, also contains bright appearance differences, and can, therefore, potentially also be used to detect further pathologies such as hyper-reflective foci. Furthermore, the deformation fields generated by FRINR manage to align the ILM well, even for extremely distorted retinae. At the same time, the deformation fields introduce small volume changes in normal tissue and larger ones in pathological regions, which is consistent with the assumption that developing fluids primarily displace the retinal layers.

In summary, the proposed methods offer several solutions for the longitudinal comparison and analysis of medical images. Without the need for manual segmentations, change detection is enabled, as well as the segmentation of (new) pathologies. The registration procedures deliver more consistent and more realistic deformation fields than previous approaches, reflecting the actual behavior of pathologies that deform and displace the surrounding tissue.

While the evaluations in this thesis concentrate on the segmentation of pathologies, a problem often observed with medical datasets is that the images are annotated either for pathologies or for anatomy. However, for further analysis of the images, it may be useful to have both annotations. In the case of retinal images, for example, one might be interested in the retinal layers in which the fluids occur. Since FRN-Ano was trained for the registration of healthy to pathological images, it is possible to transfer layer segmentations of healthy images onto the pathology cases, as shown exemplarily in Fig. 8.1. The depicted images are fully annotated for pathological fluids and retinal layers. None of these labels were used for network training, despite annotations of the outer borders of the retina. The proposed methods, thus, pave the way for the transfer of labels, e.g. via atlas-based segmentation, which is usually hindered by the presence of pathologies in medical images.

Despite the advancements achieved, several further research questions arise. The most obvious one probably is how to further improve the accuracy of unsupervised segmentation, which has already been significantly advanced in this work, but still lags behind supervised methods. Recently, huge advancements have been made in general purpose (segmentation) networks. Foundation models are trained on vast amounts of data, often using partial or self-supervision, to achieve models that are able to extract rich features from a large variety of data sources. These models serve as an exceptional starting point for the fine-tuning to domain-specific tasks. Using foundation models for image regis-

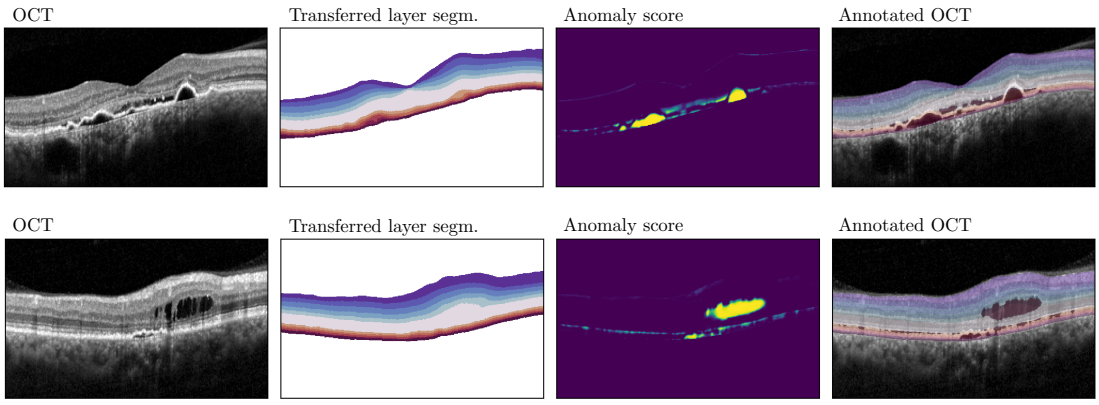


Figure 8.1: Two examples of OCT B-scans annotated with FRN-Ano for both anatomical and pathological labels. The segmentations of the retinal layers are transferred from healthy images by applying the deformations predicted by FluidRegNet to the respective segmentation masks and combining them with STAPLE [Warfield et al., 2004]. Pathology segmentations are generated by thresholding the anomaly scores derived from the deformation field's Jacobian and the inpainted pathologies.

tration, e.g. [Tian et al., 2024b; Demir et al., 2024], to train the proposed CNNs might help improve performance. However, most foundation models do not work directly for pathological data because they were trained on healthy images and, in addition, the network architectures would have to be adapted to the one of the foundation model.

The CNN-based methods still struggle with very large deformations, which might be caused by the scarcity of extreme cases in the training data. Apart from the usage of larger, potentially artificially generated, datasets, additional research in better suited network architectures would be desirable. CNNs have a rather small and local receptive field, hindering the detection of long-range dependencies. Transformer networks use self-attention to find correlations between image patches, resulting in global receptive fields, and have shown to increase registration performance [Dosovitskiy et al., 2021]. Using a transformer architecture as backbone model for the presented methods would be an interesting extension. Also, instance optimization or iterative registration could help to achieve better alignment for the seldom extreme cases. Since the INR-based methods analyzed in Chapter 7 proved to be suitable for such large deformations, these methods could also provide a tool for further processing and improvement of the CNN results.

Significant progress has been achieved regarding the plausibility of the deformation fields. However, further research is necessary to confirm their realism. To do so, a more frequent imaging of the eyes would be necessary to capture the formation of fluids over several time points. Here, new low-cost OCT systems for home use open up a lot of possibilities, e.g. [Sudkamp et al., 2016; Dahrouj and Miller, 2021; Mathai et al., 2022]. The patients are enabled to perform the imaging themselves without having to consult a physician, which makes more frequent, even daily, examinations possible. In the future, this will facilitate the close monitoring of dynamic eye diseases such as AMD or CSCR, enabling examinations of the processes involved in fluid manifestation. Finally, the

tuning of hyperparameters is a decisive component for successful and realistic DL-based registration, as is also the case with classic procedures. A self-configuring variant of the presented algorithms would be desirable here, similar to nnU-Net [Isensee et al., 2021b] or ConvexAdam [Siebert et al., 2024].

In conclusion, the present work proposed several methods that enable the automatic analysis of longitudinal medical image data. From single image clustering to registration approaches that enable change detection, the proposed algorithms are characterized by the ability to segment pathologies in an unsupervised manner and an increasing realism of the generated deformations. This allows the exact localization of altered tissue and, thus, a precise examination of the course of the considered diseases, without requiring manual pathology annotations.

Bibliography

Abbasi, S., Tavakoli, M., Boveiri, H. R., Shirazi, M. A. M., Khayami, R., Khorasani, H., Javidan, R., and Mehdizadeh, A. (2022). Medical Image Registration Using Unsupervised Deep Neural Network: A Scoping Literature Review. *Biomedical Signal Processing and Control*, 73:103444.

Abderezaei, J., Pionteck, A., Chopra, A., and Kurt, M. (2023). 3D Inception-Based TransMorph: Pre- and Post-operative Multi-contrast MRI Registration in Brain Tumors. In Bakas, S., Crimi, A., Baid, U., Malec, S., Pytlarz, M., Baheti, B., Zenk, M., and Dorent, R., editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 35–45.

Adams, R. and Bischof, L. (1994). Seeded Region Growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647.

Aljalbout, E., Golkov, V., Siddiqui, Y., Strobel, M., and Cremers, D. (2018). Clustering with Deep Learning: Taxonomy and New Methods. arXiv E-Print 1801.07648.

Almahfouz Nasser, S., Kurian, N. C., Meena, M., Shamsi, S., and Sethi, A. (2023). WSSAMNet: Weakly Supervised Semantic Attentive Medical Image Registration Network. In Bakas, S., Crimi, A., Baid, U., Malec, S., Pytlarz, M., Baheti, B., Zenk, M., and Dorent, R., editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 15–24.

Alonso-Caneiro, D., Read, S. A., and Collins, M. J. (2011). Speckle Reduction in Optical Coherence Tomography Imaging by Affine-Motion Image Registration. *Journal of Biomedical Optics*, 16(11):116027.

Amin, M., Martínez-Heras, E., Ontaneda, D., and Prados Carrasco, F. (2024). Artificial Intelligence and Multiple Sclerosis. *Current Neurology and Neuroscience Reports*, 24(8):233–243.

Amiranashvili, T., Lüdke, D., Li, H. B., Menze, B., and Zachow, S. (2022). Learning Shape Reconstruction from Sparse Measurements with Neural Implicit Functions. In Konukoglu, E., Menze, B., Venkataraman, A., Baumgartner, C., Dou, Q., and Albarqouni, S., editors, *Proceedings of The 5th International Conference on Medical Imaging with Deep Learning*, volume 172 of *Proceedings of Machine Learning Research (PMLR)*, pages 22–34.

Andresen, J., Ehrhardt, J., von der Burchard, C., Tatli, A., Roider, J., Handels, H., and Uzunova, H. (2024). FluidRegNet: Longitudinal Registration of Retinal OCT Images with New Pathological Fluids. In Burgos, N., Petitjean, C., Vakalopoulou, M., Christodoulidis, S., Coupe, P., Delingette, H., Lartizien, C., and Mateus, D., editors, *Proceedings of The 7th International Conference on Medical Imaging with Deep Learning*, volume 250 of *Proceedings of Machine Learning Research (PMLR)*, pages 48–60.

Andresen, J., Kepp, T., Ehrhardt, J., von der Burchard, C., Roider, J., and Handels, H. (2022a). Deep Learning-Based Simultaneous Registration and Unsupervised Non-Correspondence Segmentation of Medical Images with Pathologies. *International Journal of Computer Assisted Radiology and Surgery*, 17:699–710.

Andresen, J., Kepp, T., Wang-Evers, M., Ehrhardt, J., Manstein, D., and Handels, H. (2022b). Unsupervised Segmentation of Wounds in Optical Coherence Tomography Images Using Invariant Information Clustering. In Maier-Hein, K., Deserno, T. M., Handels, H., Maier, A., Palm, C., and Tolxdorff, T., editors, *Bildverarbeitung für die Medizin 2022*, pages 1–6.

Bibliography

Andresen, J., Uzunova, H., Ehrhardt, J., and Handels, H. (2021). New Multiple Sclerosis Lesion Detection with Convolutional Neural Registration Networks. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 111–114.

Andresen, J., Uzunova, H., Ehrhardt, J., Kepp, T., and Handels, H. (2022c). Image Registration and Appearance Adaptation in Non-Correspondent Image Regions for New MS Lesions Detection. *Frontiers in Neuroscience*, 16.

Ankerst, M., Breunig, M. M., Kriegel, H.-P., and Sander, J. (1999). OPTICS: Ordering Points to Identify the Clustering Structure. In *Proceedings of the ACM SIGMOD'99 International Conference on Management of Data*, pages 49–60.

Antonio, M. J., Montero, J., Yáñez, J., and Gómez, D. (2010). A Divisive Hierarchical k-Means Based Algorithm for Image Segmentation. In *2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering*, pages 300–304.

Arifin, A. Z. and Asano, A. (2006). Image Segmentation by Histogram Thresholding Using Hierarchical Cluster Analysis. *Pattern Recognition Letters*, 27(13):1515–1521.

Ashtari, P., Barile, B., Van Huffel, S., and Sappey-Marinier, D. (2021). Longitudinal Multiple Sclerosis Lesion Segmentation Using Pre-activation U-Net. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 45–51.

Ashtari, P., Barile, B., Van Huffel, S., and Sappey-Marinier, D. (2022). New Multiple Sclerosis Lesion Segmentation and Detection Using Pre-Activation U-Net. *Frontiers in Neuroscience*, 16.

Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain. *Medical Image Analysis*, 12(1):26–41.

Baghaie, A., D'souza, R. M., and Yu, Z. (2015). Sparse and Low Rank Decomposition Based Batch Image Alignment for Speckle Reduction of Retinal OCT Images. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 226–230.

Baheti, B., Chakrabarty, S., Akbari, H., Bilello, M., Wiestler, B., Schwarting, J., Calabrese, E., Rudie, J., Abidi, S., Mousa, M., Villanueva-Meyer, J., Fields, B. K. K., Kofler, F., Shinohara, R. T., Iglesias, J. E., Mok, T. C. W., Chung, A. C. S., Wodzinski, M., Jurgas, A., Marini, N., Atzori, M., Muller, H., Grobroehmer, C., Siebert, H., Hansen, L., Heinrich, M. P., Canalini, L., Klein, J., Gerken, A., Heldmann, S., Hering, A., Hahn, H. K., Meng, M., Bi, L., Feng, D., Kim, J., Zeineldin, R. A., Karar, M. E., Mathis-Ullrich, F., Burgert, O., Abderezaei, J., Pionteck, A., Chopra, A., Kurt, M., Yan, K., Yan, Y., Tang, Z., Ma, J., Nasser, S. A., Kurian, N. C., Meena, M., Shamsi, S., Sethi, A., Tustison, N. J., Avants, B. B., Cook, P., Gee, J. C., Tian, L., Greer, H., Niethammer, M., Hoopes, A., Hoffmann, M., Dalca, A. V., Christodoulidis, S., Estiene, T., Vakalopoulou, M., Paragios, N., Marcus, D. S., Davatzikos, C., Sotiras, A., Menze, B., Bakas, S., and Waldmannstetter, D. (2024). The Brain Tumor Sequence Registration (BraTS-Reg) Challenge: Establishing Correspondence Between Pre-Operative and Follow-up MRI Scans of Diffuse Glioma Patients. arXiv E-Print 2112.06979.

Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V. (2019). VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800.

Barrett, W. A. and Mortensen, E. N. (1996). Fast, Accurate, and Reproducible Live-Wire Boundary Extraction. In Höhne, K. H. and Kikinis, R., editors, *Visualization in Biomedical Computing*, pages 183–192.

Barrett, W. A. and Mortensen, E. N. (1997). Interactive Live-Wire Boundary Extraction. *Medical Image Analysis*, 1(4):331–341.

Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., and Srinivasan, P. P. (2021). Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5855–5864.

Basaran, B. D., Matthews, P. M., and Bai, W. (2021). MSSEG-2 Challenge, Team New Brain: Cascaded Networks for New MS Lesion Detection. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 77–80.

Basaran, B. D., Matthews, P. M., and Bai, W. (2022). New Lesion Segmentation for Multiple Sclerosis Brain Images with Imaging and Lesion-Aware Augmentation. *Frontiers in Neuroscience*, 16.

Basaran, B. D., Matthews, P. M., and Bai, W. (2024). SegHeD+: Segmentation of Heterogeneous Data for Multiple Sclerosis Lesions with Anatomical Constraints and Lesion-aware Augmentation. arXiv E-Print 2412.10946.

Basaran, B. D., Zhang, X., Matthews, P. M., and Bai, W. (2025). SegHeD: Segmentation of Heterogeneous Data for Multiple Sclerosis Lesions with Anatomical Constraints. In Schroder, A., Li, X., Syeda-Mahmood, T., Oxtoby, N. P., Young, A., Hering, A., Mathai, T. S., Mukherjee, P., Kuckertz, S., He, T., Llorente-Saguer, I., Maier, A., Kashyap, S., Greenspan, H., and Madabhushi, A., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024 Workshops*, pages 52–63.

Battaglini, M., Rossi, F., Grove, R. A., Stromillo, M. L., Whitcher, B., Matthews, P. M., and De Stefano, N. (2014). Automated Identification of Brain New Lesions in Multiple Sclerosis Using Subtraction Images. *Journal of Magnetic Resonance Imaging*, 39(6):1543–1549.

Baugh, M., Tan, J., Müller, J. P., Dombrowski, M., Batten, J., and Kainz, B. (2023). Many Tasks Make Light Work: Learning to Localise Medical Anomalies from Multiple Synthetic Tasks. In Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., and Taylor, R., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pages 162–172.

Beg, M. F., Miller, M. I., Trouvé, A., and Younes, L. (2005). Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. *International Journal of Computer Vision*, 61:139–157.

Bernhardt, L. (2010). *Neuronale Degeneration bei spinaler multipler Sklerose*. PhD thesis, Freie Universität Berlin, Charité – Universitätsmedizin Berlin.

Beucher, S. (1979). Use of Watersheds in Contour Detection. In *Proceedings International Workshop on Image Processing*, pages 17–21.

Beucher, S. and Meyer, F. (1992). The Morphological Approach to Segmentation: The Watershed Transformation. In *Mathematical Morphology in Image Processing*, pages 433–481. CRC Press.

Bielski, A. and Favaro, P. (2019). Emergence of Object Segmentation in Perturbed Generative Models. In *Proceedings of the 33rd Conference on Advances in Neural Information Processing Systems (NeurIPS)*, pages 7256–7266.

Bibliography

Bogunović, H., Venhuizen, F., Klimscha, S., Apostolopoulos, S., Bab-Hadiashar, A., Bagci, U., Beg, M. F., Bekalo, L., Chen, Q., Ciller, C., Gopinath, K., Gostar, A. K., Jeon, K., Ji, Z., Kang, S. H., Koozekanani, D. D., Lu, D., Morley, D., Parhi, K. K., Park, H. S., Rashno, A., Sarunic, M., Shaikh, S., Sivaswamy, J., Tennakoon, R., Yadav, S., Zanet, S. D., Waldstein, S. M., Gerendas, B. S., Klaver, C., Sánchez, C. I., and Schmidt-Erfurth, U. (2019). RETOUCH - The Retinal OCT Fluid Detection and Segmentation Benchmark and Challenge. *IEEE Transactions on Medical Imaging*, 38(8):1858–1874.

Brett, M., Leff, A. P., Rorden, C., and Ashburner, J. (2001). Spatial Normalization of Brain Images with Focal Lesions Using Cost Function Masking. *NeuroImage*, 14(2):486–500.

Bruse, J. L., Zuluaga, M. A., Khushnood, A., McLeod, K., Ntsinjana, H. N., Hsia, T.-Y., Sermesant, M., Pennec, X., Taylor, A. M., and Schievano, S. (2017). Detecting Clinically Meaningful Shape Clusters in Medical Image Data: Metrics Analysis for Hierarchical Clustering Applied to Healthy and Pathological Aortic Arches. *IEEE Transactions on Biomedical Engineering*, 64(10):2373–2383.

Burkhard, H.-D. (2006). Einführung in die KI. https://www.informatik.hu-berlin.de/de/forschung/gebiete/ki/lehre/ws0607/vlki-skript/eki0607NeuronaleNetze.pdf. Accessed: 2023-04-23.

Byra, M., Poon, C., Rachmadi, M. F., Schlachter, M., and Skibbe, H. (2023a). Exploring the Performance of Implicit Neural Representations for Brain Image Registration. *Scientific Reports*, 13(1):17334.

Byra, M., Poon, C., Shimogori, T., and Skibbe, H. (2023b). Implicit Neural Representations for Joint Decomposition and Registration of Gene Expression Images in the Marmoset Brain. In Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., and Taylor, R., editors, *International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI 2023*, pages 645–654.

Cabezas, M., Corral, J., Oliver, A., Díez, Y., Tintoré, M., Auger, C., Montalban, X., Lladó, X., Pareto, D., and Rovira, À. (2016). Improved Automatic Detection of New T2 Lesions in Multiple Sclerosis Using Deformation Fields. *American Journal of Neuroradiology*, 37(10):1816–1823.

Cabezas, M., Luo, Y., Kyle, K., Ly, L., Wang, C., and Barnett, M. (2021). Estimating Lesion Activity through Feature Similarity: A Dual Path Unet Approach for the MSSEG2 MICCAI Challenge. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 107–110.

Calderon-Delgado, M., Tjiu, J.-W., Lin, M.-Y., and Huang, S.-L. (2018). High Resolution Human Skin Image Segmentation by Means of Fully Convolutional Neural Networks. In *2018 International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD)*, pages 31–32.

Calvi, A., Haider, L., Prados, F., Tur, C., Chard, D., and Barkhof, F. (2022). In Vivo Imaging of Chronic Active Lesions in Multiple Sclerosis. *Multiple Sclerosis Journal*, 28(5):683–690.

Campbell, N. A., Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., and Jackson, R. B. (2015). *Campbell Biologie*. Pearson Studium. 10., aktualisierte Auflage.

Cao, X., Fan, J., Dong, P., Ahmad, S., Yap, P.-T., and Shen, D. (2020). Chapter 14 - Image Registration Using Machine and Deep Learning. In Zhou, S. K., Rueckert, D., and Fichtinger, G., editors, *Handbook of Medical Image Computing and Computer Assisted Intervention*, The Elsevier and MICCAI Society Book Series, pages 319–342. Elsevier.

Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., and Shen, D. (2017). Deformable Image Registration Based on Similarity-Steered CNN Regression. In Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D. L., and Duchesne, S., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, pages 300–308.

Caron, M., Bojanowski, P., Joulin, A., and Douze, M. (2018). Deep Clustering for Unsupervised Learning of Visual Features. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Chen, C.-F. R., Fan, Q., and Panda, R. (2021). CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 357–366.

Chen, C.-H., Lin, Z.-H., Lin, C.-H. D., Lu, C.-K., Wang, J.-K., and Huang, T.-L. (2024). Evaluation of Image Interpolations for AMD Detection on OCT Images Using Mobile-OCT CNN Model. In *2024 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, pages 557–558.

Chen, J., Frey, E. C., He, Y., Segars, W. P., Li, Y., and Du, Y. (2022). TransMorph: Transformer for Unsupervised Medical Image Registration. *Medical Image Analysis*, 82:102615.

Chen, J., Liu, Y., Wei, S., Bian, Z., Subramanian, S., Carass, A., Prince, J. L., and Du, Y. (2025). A Survey on Deep Learning in Medical Image Registration: New Technologies, Uncertainty, Evaluation Metrics, and Beyond. *Medical Image Analysis*, 100:103385.

Chen, J., Tian, J., Lee, N., Zheng, J., Smith, R. T., and Laine, A. F. (2010). A Partial Intensity Invariant Feature Descriptor for Multimodal Retinal Image Registration. *IEEE Transactions on Biomedical Engineering*, 57(7):1707–1718.

Chen, J. and Yang, X. (2025). Uncertainty-Guided Joint Semi-Supervised Segmentation and Registration of Cardiac Images. In Ide, I., Kompatsiaris, I., Xu, C., Yanai, K., Chu, W.-T., Nitta, N., Riegler, M., and Yamasaki, T., editors, *MultiMedia Modeling*, pages 253–267.

Chen, K., Derksen, A., Heldmann, S., Hallmann, M., and Berkels, B. (2015). Deformable Image Registration with Automatic Non-Correspondence Detection. In Aujol, J.-F., Nikolova, M., and Papadakis, N., editors, *Scale Space and Variational Methods in Computer Vision*, pages 360–371.

Chen, M., Artières, T., and Denoyer, L. (2019). Unsupervised Object Segmentation by Redrawing. In *Proceedings of the 33rd Conference on Advances in Neural Information Processing Systems (NeurIPS)*, pages 12726–12737.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations. In *ICML'20: Proceedings of the 37th International Conference on Machine Learning*, pages 1597–1607.

Cheng, J., Tao, D., Quan, Y., Wong, D. W. K., Cheung, G. C. M., Akiba, M., and Liu, J. (2016). Speckle Reduction in 3D Optical Coherence Tomography of Retina by A-Scan Reconstruction. *IEEE Transactions on Medical Imaging*, 35(10):2270–2279.

Cheng, Y., Chu, Z., and Wang, R. K. (2021). Robust Three-Dimensional Registration on Optical Coherence Tomography Angiography for Speckle Reduction and Visualization. *Quantitative Imaging in Medicine and Surgery*, 11(3):879–894.

Chitphakdithai, N. and Duncan, J. S. (2010). Non-Rigid Registration with Missing Correspondences in Preoperative and Postresection Brain Images. In *International Conference on Medical Image Computing and Computer Assisted Intervention – MICCAI 2010*, pages 367–374.

Chu, Z., Wang, L., Zhou, X., Shi, Y., Cheng, Y., Laiginhas, R., Zhou, H., Shen, M., Zhang, Q., de Sisternes, L., Lee, A. Y., Gregori, G., Rosenfeld, P. J., and Wang, R. K. (2022). Automatic Geographic Atrophy Segmentation Using Optical Attenuation in OCT Scans with Deep Learning. *Biomedical Optics Express*, 13(3):1328–1343.

Comaniciu, D. and Meer, P. (2002). Mean Shift: A Robust Approach Toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619.

Combès, B., Kerbrat, A., Pasquier, G., Commowick, O., Le Bon, B., Galassi, F., L'Hostis, P., El Graoui, N., Chouteau, R., Cordonnier, E., Edan, G., and Ferré, J.-C. (2021). A Clinically-Compatible Workflow for Computer-Aided Assessment of Brain Disease Activity in Multiple Sclerosis Patients. *Frontiers in Medicine*, 8:740248.

Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M. (2021). MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure. In *MICCAI 2021 - 24th International Conference on Medical Image Computing and Computer Assisted Intervention*.

Commowick, O., Combès, B., Cervenansky, F., and Dojat, M. (2023). Editorial: Automatic Methods for Multiple Sclerosis New Lesions Detection and Segmentation.

Commowick, O., Istace, A., Kain, M., Laurent, B., Leray, F., Simon, M., Pop, S. C., Girard, P., Améli, R., Ferré, J.-C., Kerbrat, A., Tourdias, T., Cervenansky, F., Glatard, T., Beaumont, J., Doyle, S., Forbes, F., Knight, J., Khademi, A., Mahbod, A., Wang, C., McKinley, R., Wagner, F., Muschelli, J., Sweeney, E., Roura, E., Lladó, X., Santos, M. M., Santos, W. P., Silva-Filho, A. G., Tomas-Fernandez, X., Urien, H., Bloch, I., Valverde, S., Cabezas, M., Vera-Olmos, F. J., Malpica, N., Guttmann, C., Vukusic, S., Edan, G., Dojat, M., Styner, M., Warfield, S. K., Cotton, F., and Barillot, C. (2018). Objective Evaluation of Multiple Sclerosis Lesion Segmentation Using a Data Management and Processing Infrastructure. *Scientific Reports*, 8(1):13650.

Cootes, T. F. and Taylor, C. J. (1992). Active Shape Models – 'Smart Snakes'. In Hogg, D. and Boyle, R., editors, *British Machine Vision Conference (BMVC92)*, pages 266–275.

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1992). Training Models of Shape from Sets of Examples. In Hogg, D. and Boyle, R., editors, *British Machine Vision Conference (BMVC92)*, pages 9–18.

Couprie, C., Grady, L., Najman, L., and Talbot, H. (2011). Power Watershed: A Unifying Graph-Based Optimization Framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7):1384–1399.

Dahrouj, M. and Miller, J. B. (2021). Artificial Intelligence (AI) and Retinal Optical Coherence Tomography (OCT). *Seminars in Ophthalmology*, 36(4):341–345.

Dalbis, T., Fritz, T., Grilo, J., Hitziger, S., and Ling, W. X. (2021). Triplanar U-Net with Orientation Aggregation for New Lesions Segmentation. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 57–60.

Dalca, A. V., Balakrishnan, G., Guttag, J., and Sabuncu, M. R. (2019). Unsupervised Learning of Probabilistic Diffeomorphic Registration for Images and Surfaces. *Medical Image Analysis*, 57:226–236.

Das, V., Dandapat, S., and Bora, P. K. (2020). Unsupervised Super-Resolution of OCT Images Using Generative Adversarial Network for Improved Age-Related Macular Degeneration Diagnosis. *IEEE Sensors Journal*, 20(15):8746–8756.

de Sousa, I. M. and de Oliveira, M. (2021). Convolutional Neural Network for MS New Lesions Segmentation Challenge: Using a Data Management and Processing Infrastructure (MSSEG-2). In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 101–105.

de Vos, B. D., Berendsen, F. F., Viergever, M. A., Sokooti, H., Staring, M., and Išgum, I. (2019). A Deep Learning Framework for Unsupervised Affine and Deformable Image Registration. *Medical Image Analysis*, 52:128–143.

Deegan, A. J., Mandell, S. P., and Wang, R. K. (2019). Optical Coherence Tomography Correlates Multiple Measures of Tissue Damage Following Acute Burn Injury. *Quantitative Imaging in Medicine and Surgery*, 9(5):731–741.

Deegan, A. J., Wang, W., Men, S., Li, Y., Song, S., Xu, J., and Wang, R. K. (2018). Optical Coherence Tomography Angiography Monitors Human Cutaneous Wound Healing Over Time. *Quantitative Imaging in Medicine and Surgery*, 8(2):135–150.

Del Amor, R., Morales, S., Colomer, A., Mogensen, M., Jensen, M., Israelsen, N. M., Bang, O., and Naranjo, V. (2020). Automatic Segmentation of Epidermis and Hair Follicles in Optical Coherence Tomography Images of Normal Skin by Convolutional Neural Networks. *Frontiers in Medicine*, 7:220.

Demir, B., Tian, L., Greer, H., Kwitt, R., Vialard, F.-X., Estépar, R. S. J., Bouix, S., Rushmore, R., Ebrahim, E., and Niethammer, M. (2024). MultiGradICON: A Foundation Model for Multimodal Medical Image Registration. In Modat, M., Simpson, I., Špiclin, Ž., Bastiaansen, W., Hering, A., and Mok, T. C. W., editors, *Biomedical Image Registration*, pages 3–18.

Derradji, Y., Mosinska, A., Apostolopoulos, S., Ciller, C., De Zanet, S., and Mantel, I. (2021). Fully-Automated Atrophy Segmentation in Dry Age-Related Macular Degeneration in Optical Coherence Tomography. *Scientific Reports*, 11:21893.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations, ICLR 2021*.

Dufresne, E., Fortun, D., Kremer, S., and Noblet, V. (2022). A Unified Framework for Focal Intensity Change Detection and Deformable Image Registration. Application to the Monitoring of Multiple Sclerosis Lesions in Longitudinal 3D Brain MRI. *Frontiers in Neuroimaging*, 1.

Efird, C., Miller, D., and Cobzas, D. (2021). A UNet Pipeline for Segmentation of New MS Lesions. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 53–56.

Ehrhardt, J., Säring, D., and Handels, H. (2007). Structure-Preserving Interpolation of Temporal and Spatial Image Sequences Using an Optical Flow-Based Method. *Methods of Information in Medicine*, 46(3):300–307.

Ehrhardt, J., Schmidt-Richberg, A., Werner, R., and Handels, H. (2015). Variational Registration: A Flexible Open-Source ITK Toolbox for Nonrigid Image Registration. In Handels, H., Deserno, T. M., Meinzer, H.-P., and Tolxdorff, T., editors, *Bildverarbeitung für die Medizin (BVM) 2015*, pages 209–214.

Bibliography

Essakine, A., Cheng, Y., Cheng, C.-W., Zhang, L., Deng, Z., Zhu, L., Schönlieb, C.-B., and Aviles-Rivero, A. I. (2024). Where Do We Stand with Implicit Neural Representations? A Technical and Performance Survey. arXiv E-Print 2411.03688.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Simoudis, E., Han, J., and Fayyad, U., editors, *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 226–231.

Eurostat (2024). Eurostat: Healthcare Resource Statistics - Technical Resources and Medical Technology. https://ec.europa.eu/eurostat/databrowser/view/hlth_co_exam/default/table?lang=en. Accessed: 2025-02-25.

Fang, W., Tang, Y., Guo, H., Yuan, M., Mok, T. C., Yan, K., Yao, J., Chen, X., Liu, Z., Lu, L., Zhang, L., and Xu, M. (2024). CycleINR: Cycle Implicit Neural Representation for Arbitrary-Scale Volumetric Super-Resolution of Medical Data. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11631–11641.

Fartaria, M. J., Kober, T., Granziera, C., and Cuadra, M. B. (2019). Longitudinal Analysis of White Matter and Cortical Lesions in Multiple Sclerosis. *NeuroImage: Clinical*, 23:101938.

Felzenszwalb, P. F. and Huttenlocher, D. P. (2004). Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, 59(2):167–181.

Fenneteau, A., Bourdon, P., Helbert, D., Fernandez-Maloigne, C., Habas, C., and Guillevin, R. (2021). Siamese Convolutional Neural Network for New Multiple Sclerosis Lesion Segmentation. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 13–16.

Ferreira, A., Solak, N., Li, J., Dammann, P., Kleesiek, J., Alves, V., and Egger, J. (2024). How we won brats 2023 Adult Glioma challenge? Just faking it! Enhanced Synthetic Data Augmentation and Model Ensemble for Brain Tumour Segmentation. arXiv E-Print 2402.17317.

François, A., Maillard, M., Oppenheim, C., Pallud, J., Bloch, I., Gori, P., and Glaunès, J. (2022). Weighted Metamorphosis for Registration of Images with Different Topologies. In Hering, A., Schnabel, J., Zhang, M., Ferrante, E., Heinrich, M., and Rueckert, D., editors, *Biomedical Image Registration*, pages 8–17.

Fukushima, K. (1980). Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, 36:193–202.

Ganiler, O., Oliver, A., Diez, Y., Freixenet, J., Vilanova, J. C., Beltran, B., Ramió-Torrentà, L., Rovira, À., and Lladó, X. (2014). A Subtraction Pipeline for Automatic Detection of New Appearing Multiple Sclerosis Lesions in Longitudinal Studies. *Neuroradiology*, 56(5):363–374.

Geurts, J. J., Bö, L., Pouwels, P. J., Castelijns, J. A., Polman, C. H., and Barkhof, F. (2005). Cortical Lesions in Multiple Sclerosis: Combined Postmortem MR Imaging and Histopathology. *American Journal of Neuroradiology*, 26(3):572–577.

Ghassabi, Z., Shanbehzadeh, J., Sedaghat, A., and Fatemizadeh, E. (2013). An Efficient Approach for Robust Multimodal Retinal Image Registration Based on UR-SIFT Features and PIIFD Descriptors. *EURASIP Journal on Image and Video Processing*, 2013:25.

Ghosh, B., Mandal, M., Mitra, P., and Chatterjee, J. (2020). Attenuation Corrected-Optical Coherence Tomography for Quantitative Assessment of Skin Wound Healing and Scar Morphology. *Journal of Biophotonics*, 14(4):e202000357.

Gibicar, A., Mitha, S., and Khademi, A. (2021). Segmentation of New Multiple Sclerosis Lesions Using an Ensemble of SC U-Nets with Multi Channel Patch-Based Inputs. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 17–20.

Gong, L., Zhang, C., Duan, L., Du, X., Liu, H., Chen, X., and Zheng, J. (2019). Nonrigid Image Registration Using Spatially Region-Weighted Correlation Ratio and GPU-Acceleration. *IEEE Journal of Biomedical and Health Informatics*, 23(2):766–778.

Gong, X., Khaidem, L., Zhu, W., Zhang, B., and Doermann, D. (2022). Uncertainty Learning Towards Unsupervised Deformable Medical Image Registration. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1555–1564.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative Adversarial Networks. *Communications of the ACM*, 63(11):139–144.

Grady, L. (2006). Random Walks for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1768–1783.

Greaves, N. S., Benatar, B., Whiteside, S., Alonso-Rasgado, T., Baguneid, M., and Bayat, A. (2014). Optical Coherence Tomography: A Reliable Alternative to Invasive Histological Assessment of Acute Wound Healing in Human Skin? *British Journal of Dermatology*, 170(4):840–850.

Greaves, N. S., Iqbal, S. A., Hodgkinson, T., Morris, J., Benatar, B., Alonso-Rasgado, T., Baguneid, M., and Bayat, A. (2015). Skin Substitute-Assisted Repair Shows Reduced Dermal Fibrosis in Acute Human Wounds Validated Simultaneously by Histology and Optical Coherence Tomography. *Wound Repair and Regeneration*, 23(4):483–494.

Großbröhmer, C., Hemidi, Z. A.-H., Falta, F., and Heinrich, M. P. (2024). SINA: Sharp Implicit Neural Atlases by Joint Optimisation of Representation and Deformation. In Modat, M., Simpson, I., Špiclin, Ž., Bastiaansen, W., Hering, A., and Mok, T. C. W., editors, *International Workshop on Biomedical Image Registration (WBIR)*, pages 165–180.

Großbröhmer, C., Siebert, H., Hansen, L., and Heinrich, M. P. (2023). Employing ConvexAdam for BraTS-Reg. In Bakas, S., Crimi, A., Baid, U., Malec, S., Pytlarz, M., Baheti, B., Zenk, M., and Dorent, R., editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 252–261.

Hamzaoui, M., Soulier, T., Yazdan-Panah, A., Schmidt-Mengin, M., Colliot, O., Ayache, N., and Stankoff, B. (2021). Intensity Based Regions Of Interest (ROIs) Preselection Followed by Convolutional Neuronal Network (CNN) based Segmentation for New Lesions Detection in Multiple Sclerosis. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 93–96.

Han, K., Sun, S., Yan, X., You, C., Tang, H., Naushad, J., Ma, H., Kong, D., and Xie, X. (2023). Diffeomorphic Image Registration with Neural Velocity Field. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1869–1879.

Han, X., Zhong, Y., Cao, L., and Zhang, L. (2017). Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification. *Remote Sensing*, 9(8):848.

Handels, H. (2000). *Medizinische Bildverarbeitung*. Vieweg+Teubner Verlag, Springer Fachmedien Wiesbaden.

Hashemi, M., Akhbari, M., and Jutten, C. (2022). Delve into Multiple Sclerosis (MS) Lesion Exploration: A Modified Attention U-Net for MS Lesion Segmentation in Brain MRI. *Computers in Biology and Medicine*, 145:105402.

He, K. and Sun, J. (2015). Fast Guided Filter. arXiv E-Print 1505.00996.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Hering, A., Kuckertz, S., Heldmann, S., and Heinrich, M. P. (2019). Enhancing Label-Driven Deep Deformable Image Registration with Local Distance Metrics for State-of-the-Art Cardiac Motion Tracking. In Handels, H., Deserno, T. M., Maier, A., Maier-Hein, K. H., Palm, C., and Tolxdorff, T., editors, *Bildverarbeitung für die Medizin (BVM) 2019*, pages 309–314.

Hinneburg, A. and Keim, D. A. (1998). An Efficient Approach to Clustering in Large Multimedia Databases with Noise. In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, pages 58–65.

Hitziger, S., Ling, W. X., Fritz, T., D'Albis, T., Lemke, A., and Grilo, J. (2022). Triplanar U-Net with Lesion-Wise Voting for the Segmentation of New Lesions on Longitudinal MRI Studies. *Frontiers in Neuroscience*, 16.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. In *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS)*, pages 6840–6851.

Hori, Y., Yasuno, Y., Sakai, S., Matsumoto, M., Sugawara, T., Madjarova, V. D., Yamanari, M., Makita, S., Yasui, T., Araki, T., Itoh, M., and Yatagai, T. (2006). Automatic Characterization and Segmentation of Human Skin Using Three-Dimensional Optical Coherence Tomography. *Optics Express*, 14(5):1862–1877.

Hossein-Nejad, Z. and Nasri, M. (2018). A-RANSAC: Adaptive Random Sample Consensus Method in Multimodal Retinal Image Registration. *Biomedical Signal Processing and Control*, 45:325–338.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269.

Huang, P., Huang, Y., Wang, W., and Wang, L. (2014). Deep Embedding Network for Clustering. In *2014 22nd International Conference on Pattern Recognition*, pages 1532–1537.

Huang, Y., Lu, Z., Shao, Z., Ran, M., Zhou, J., Fang, L., and Zhang, Y. (2019). Simultaneous Denoising and Super-Resolution of Optical Coherence Tomography Images Based on Generative Adversarial Network. *Optics Express*, 27(9):12289–12307.

Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H. (2021a). nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. *Nature Methods*, 18:203 – 211.

Isensee, F., Jäger, P. F., Full, P. M., Vollmuth, P., and Maier-Hein, K. H. (2021b). nnU-Net for Brain Tumor Segmentation. In Crimi, A. and Bakas, S., editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 118–132.

Jaderberg, M., Simonyan, K., Zisserman, A., and kavukcuoglu, k. (2015). Spatial Transformer Networks. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R., editors, *Proceedings of the 29th Conference on Advances in Neural Information Processing Systems*.

Jain, S., Ribbens, A., Sima, D. M., Cambron, M., De Keyser, J., Wang, C., Barnett, M. H., Van Huffel, S., Maes, F., and Smeets, D. (2016). Two Time Point MS Lesion Segmentation in Brain MRI: An Expectation-Maximization Framework. *Frontiers in Neuroscience*, 10:576.

Ji, X., Henriques, J. F., and Vedaldi, A. (2019). Invariant Information Clustering for Unsupervised Image Classification and Segmentation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9865–9874.

Ji, Y., Yang, S., Zhou, K., Rocliffe, H. R., Pellicoro, A., Cash, J. L., Wang, R., Li, C., and Huang, Z. (2022). Deep-Learning Approach for Automated Thickness Measurement of Epithelial Tissue and Scab Using Optical Coherence Tomography. *Journal of Biomedical Optics*, 27(1):015002.

Joshi, A. and Hong, Y. (2023). MetaRegNet: Metamorphic Image Registration Using Flow-Driven Residual Networks. In Qin, W., Zaki, N., Zhang, F., Wu, J., Yang, F., and Li, C., editors, *Computational Mathematics Modeling in Cancer Analysis*, pages 160–170.

Kamraoui, R. A., Mansencal, B., Manjon, J. V., and Coupé, P. (2022a). Longitudinal Detection of New MS Lesions Using Deep Learning. *Frontiers in Neuroimaging*, 1.

Kamraoui, R. A., Ta, V.-T., Manjon, J. V., and Coupé, P. (2021a). Draw and Erase to Learn Better. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 33–36.

Kamraoui, R. A., Ta, V.-T., Manjon, J. V., and Coupé, P. (2021b). Image Quality Data Augmentation for New MS Lesion Segmentation. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 37–40.

Kamraoui, R. A., Ta, V.-T., Manjon, J. V., and Coupé, P. (2021c). New MS lesion Segmentation with Lesion-wise Metrics Learning. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 29–32.

Kamraoui, R. A., Ta, V.-T., Tourdias, T., Mansencal, B., Manjon, J. V., and Coupé, P. (2022b). DeepLesionBrain: Towards a Broader Deep-Learning Generalization for Multiple Sclerosis Lesion Segmentation. *Medical Image Analysis*, 76:102312.

Kanezaki, A. (2018). Unsupervised Image Segmentation by Backpropagation. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1543–1547.

Kang, J., Jung, S., Yim, J., Lee, H., Jang, J., and Nam, Y. (2021). Segmentation of New Multiple Sclerosis Lesions in Longitudinal MRI Analysis Using a Multi-Stage 3D Patch-Wise Deep Learning Algorithm. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 9–11.

Kepp, T., Andresen, J., Falta, F., and Handels, H. (2025). Bridging Gaps in Retinal Imaging - Fusing OCT and SLO Information with Implicit Neural Representations for Improved Interpolation and Segmentation. In Palm, C., Breininger, K., Deserno, T., Handels, H., Maier, A., Maier-Hein, K. H., and Tolxdorff, T. M., editors, *Bildverarbeitung für die Medizin 2025*, pages 107–112.

Kepp, T., Droigk, C., Casper, M., Evers, M., Hüttmann, G., Salma, N., Manstein, D., Heinrich, M. P., and Handels, H. (2019). Segmentation of Mouse Skin Layers in Optical Coherence Tomography Image Data Using Deep Convolutional Neural Networks. *Biomedical Optics Express*, 10(7):3484–3496.

Kim, C., Lee, D., Kim, S., Cho, M., and Han, W.-S. (2023). Generalizable Implicit Neural Representations via Instance Pattern Composers. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11808–11817.

Kingma, D. P. and Ba, J. (2015). Adam: A Method for Stochastic Optimization. In Bengio, Y. and LeCun, Y., editors, *International Conference on Learning Representations, ICLR 2015, Conference Track Proceedings*.

Klein, S., Staring, M., Murphy, K., Viergever, M. A., and Pluim, J. P. (2009). Elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205.

Köhler, C., Wahl, H., Ziemssen, T., Linn, J., and Kitzler, H. H. (2019). Exploring Individual Multiple Sclerosis Lesion Volume Change Over Time: Development of an Algorithm for the Analyses of Longitudinal Quantitative MRI Measures. *NeuroImage: Clinical*, 21:101623.

Kolb, H., Al-Louzi, O., Beck, E. S., Sati, P., Absinta, M., and Reich, D. S. (2022). From Pathology to MRI and Back: Clinically Relevant Biomarkers of Multiple Sclerosis Lesions. *NeuroImage: Clinical*, 36:103194.

Kolbinger, F. R., Rinner, F. M., Jenke, A. C., Carstens, M., Krell, S., Leger, S., Distler, M., Weitz, J., Speidel, S., and Bodenstedt, S. (2023). Anatomy Segmentation in Laparoscopic Surgery: Comparison of Machine Learning and Human Expertise - An Experimental Study. *International Journal of Surgery*, 109(10):2962–2974.

Kraus, M. F., Potsaid, B., Mayer, M. A., Bock, R., Baumann, B., Liu, J. J., Hornegger, J., and Fujimoto, J. G. (2012). Motion Correction in Optical Coherence Tomography Volumes on a Per A-Scan Basis Using Orthogonal Scan Patterns. *Biomedical Optics Express*, 3(6):1182–1199.

Krebs, J., Mansi, T., Mailhé, B., Ayache, N., and Delingette, H. (2018). Unsupervised Probabilistic Deformation Modeling for Robust Diffeomorphic Registration. In Stoyanov, D., Taylor, Z., Carneiro, G., Syeda-Mahmood, T., Martel, A., Maier-Hein, L., Tavares, J. M. R., Bradley, A., Papa, J. P., Belagiannis, V., Nascimento, J. C., Lu, Z., Conjeti, S., Moradi, M., Greenspan, H., and Madabhushi, A., editors, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 101–109.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, *Proceedings of the 26th Conference on Advances in Neural Information Processing Systems (NeurIPS)*.

Krüger, J., Ehrhardt, J., Schultz, S., and Handels, H. (2019). A Probabilistic Approach for the Registration of Images with Missing Correspondences. In Angelini, E. D. and Landman, B. A., editors, *Medical Imaging 2019: Image Processing*, volume 10949, page 1094925.

Krüger, J., Opfer, R., Gessert, N., Ostwaldt, A.-C., Manogaran, P., Kitzler, H. H., Schlaefer, A., and Schippling, S. (2020). Fully Automated Longitudinal Segmentation of New or Enlarged Multiple Sclerosis Lesions Using 3D Convolutional Neural Networks. *NeuroImage: Clinical*, 28:102445.

Krüger, J., Schultz, S., Handels, H., and Ehrhardt, J. (2020). Registration with Probabilistic Correspondences — Accurate and Robust Registration for Pathological and Inhomogeneous Medical Data. *Computer Vision and Image Understanding*, 190:102839.

Kuck, M., Strese, H., Alawi, S. A., Meinke, M. C., Fluhr, J. W., Burbach, G. J., Krah, M., Sterry, W., and Lademann, J. (2013). Evaluation of Optical Coherence Tomography as a Non-Invasive Diagnostic Tool in Cutaneous Wound Healing. *Skin Research and Technology*, 20(1):1–7.

Kumar, P., Dhara, S., Gope, A., Chatterjee, J., and Mandal, S. (2023). Deep Learning Based Skin-layer Segmentation for Characterizing Cutaneous Wounds from Optical Coherence Tomography Images. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1–4.

Lachinov, D., Seeböck, P., Mai, J., Goldbach, F., Schmidt-Erfurth, U., and Bogunović, H. (2021). Projective Skip-Connections for Segmentation Along a Subset of Dimensions in Retinal OCT. In de Bruijne, M., Cattin, P. C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., and Essert, C., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 431–441.

Lang, A., Carass, A., Al-Louzi, O., Bhargava, P., Solomon, S. D., Calabresi, P. A., and Prince, J. L. (2016). Combined Registration and Motion Correction of Longitudinal Retinal OCT Data. In Styner, M. A. and Angelini, E. D., editors, *Medical Imaging 2016: Image Processing*, volume 9784, page 97840X.

Lee, J. A., Liu, P., Cheng, J., and Fu, H. (2019). A Deep Step Pattern Representation for Multimodal Retinal Image Registration. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5076–5085.

Lee, K. E., Heitkotter, H., and Carroll, J. (2021). Challenges Associated With Ellipsoid Zone Intensity Measurements Using Optical Coherence Tomography. *Translational Vision Science & Technology*, 10(12).

Lee, S., Heisler, M., Mackenzie, P. J., Sarunic, M. V., and Beg, M. F. (2017). Quantifying Variability in Longitudinal Peripapillary RNFL and Choroidal Layer Thickness Using Surface Based Registration of OCT Images. *Translational Vision Science & Technology*, 6(1):11.

Li, C.-L., Sohn, K., Yoon, J., and Pfister, T. (2021). CutPaste: Self-Supervised Learning for Anomaly Detection and Localization. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9659–9669.

Li, K., Wu, X., Chen, D. Z., and Sonka, M. (2005). Optimal Surface Segmentation in Volumetric Images - A Graph-Theoretic Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):119–134.

Li, Y., Lao, Q., Kang, Q., Jiang, Z., Du, S., Zhang, S., and Li, K. (2023). Self-Supervised Anomaly Detection, Staging and Segmentation for Retinal Images. *Medical Image Analysis*, 87:102805.

Li, Y., Zhang, C., Huang, T., Fan, Y., Ning, G., and Liao, H. (2025). Computational Multi-Angle Optical Coherence Tomography Using Implicit Neural Representation. *Optics & Laser Technology*, 184:112551.

Li, Z., Pandiyan, V. P., Maloney-Bertelli, A., Jiang, X., Li, X., and Sabesan, R. (2020). Correcting Intra-Volume Distortion for AO-OCT Using 3D Correlation Based Registration. *Optics Express*, 28(25):38390–38409.

Lin, W., Hasenstab, K., Moura Cunha, G., and Schwartzman, A. (2020). Comparison of Handcrafted Features and Convolutional Neural Networks for Liver MR Image Adequacy Assessment. *Scientific Reports*, 10(1):20336.

Lindberg, A.-S. W., Jørgensen, T. M., and Dahl, V. A. (2018). Linear, Transfinite and Weighted Method for Interpolation from Grid Lines Applied to OCT Images. *Applied Soft Computing*, 68:293–302.

Liu, Z., Zhu, H., Zhang, Q., Fu, J., Deng, W., Ma, Z., Guo, Y., and Cao, X. (2024). FINER: Flexible Spectral-Bias Tuning in Implicit NEural Representation by Variable-Periodic Activation Functions. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2713–2722.

Lladó, X., Ganiler, O., Oliver, A., Martí, R., Freixenet, J., Valls, L., Vilanova, J. C., Ramió-Torrentà, L., and Rovira, À. (2012). Automated Detection of Multiple Sclerosis Lesions in Serial Brain MRI. *Neuroradiology*, 54:787–807.

Lloyd, S. (1982). Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137.

Löhr, T., Paetzold, J. C., Sekobouyina, A., Shit, S., Ezhov, I., Wiestler, B., and Menze, B. H. (2021). MSSEG-2 New MS Lesions Detection and Segmentation Challenge Using a Data Management and Processing Infrastructure. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 21–24.

Lu, J., Jin, R., and Song, E. (2024). Pyramid Convolutional Recurrent Network for Serial Medical Image Registration With Adaptive Motion Regularizations. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 8(7):800–813.

Ma, J., He, Y., Li, F., Han, L., You, C., and Wang, B. (2024). Segment Anything in Medical Images. *Nature Communications*, 15(1):654.

Ma, X., Ji, Z., Niu, S., Leng, T., Rubin, D. L., and Chen, Q. (2020). MS-CAM: Multi-Scale Class Activation Maps for Weakly-Supervised Segmentation of Geographic Atrophy Lesions in SD-OCT Images. *IEEE Journal of Biomedical and Health Informatics*, 24(12):3443–3455.

Ma, Y., Chen, X., Zhu, W., Cheng, X., Xiang, D., and Shi, F. (2018). Speckle Noise Reduction in Optical Coherence Tomography Images Based on Edge-Sensitive cGAN. *Biomedical Optics Express*, 9(11):5129–5146.

Macar, U., Karthik, E. N., Gros, C., Lemay, A., and Cohen-Adad, J. (2021). Team NeuroPoly: Description of the Pipelines for the MICCAI 2021 MS New Lesions Segmentation Challenge. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 69–75.

MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. In Le Cam, L. M. and Neyman, J., editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 5, pages 281–297.

Mai, J., Lachinov, D., Riedl, S., Reiter, G. S., Vogl, W.-D., Bogunović, H., and Schmidt-Erfurth, U. (2023). Clinical Validation for Automated Geographic Atrophy Monitoring on OCT Under Complement Inhibitory Treatment. *Scientific Reports*, 13:7028.

Maier, O., Menze, B. H., von der Gablentz, J., Häni, L., Heinrich, M. P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., Christiaens, D., Dutil, F., Egger, K., Feng, C., Glocker, B., Götz, M., Haeck, T., Halme, H.-L., Havaei, M., Iftekharuddin, K. M., Jodoin, P.-M., Kamnitsas, K., Kellner, E., Korvenoja, A., Larochelle, H., Ledig, C., Lee, J.-H., Maes, F., Mahmood, Q., Maier-Hein, K. H., McKinley, R., Muschelli, J., Pal, C., Pei, L., Rangarajan, J. R., Reza, S. M., Robben, D., Rueckert, D., Salli, E., Suetens, P., Wang, C.-W., Wilms, M., Kirschke, J. S., Krämer, U. M., Münte, T. F., Schramm, P., Wiest, R., Handels, H., and Reyes, M. (2017). ISLES 2015 - A Public Evaluation Benchmark for Ischemic Stroke Lesion Segmentation from Multispectral MRI. *Medical Image Analysis*, 35:250–269.

Maillard, M., François, A., Glaunès, J. A., Bloch, I., and Gori, P. (2022). A Deep Residual Learning Implementation of Metamorphosis. *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*.

Makita, S., Miura, M., Azuma, S., Mino, T., Yamaguchi, T., and Yasuno, Y. (2020). Accurately Motion-Corrected Lissajous OCT with Multi-Type Image Registration. *Biomedical Optics Express*, 12(1):637–653.

Martínez-Heras, E., Vicente-Gomez, A., Vivó, F., Diaz-Hurtado, M., Kanber, B., Casas-Roma, J., Llufriu, S., and Prados, F. (2023). Longitudinal Segmentation of Multiple Sclerosis Lesions Using nnU-Net Architecture. In *Artificial Intelligence Research and Development*, pages 163–172. IOS Press.

Masson, A., Le Bon, B., Kerbrat, A., Edan, G., Galassi, F., and Combes, B. (2021). A nnUnet Implementation of New Lesions Segmentation from Serial FLAIR Images of MS Patients. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 5–7.

Mathai, M., Reddy, S., Elman, M. J., Garfinkel, R. A., Ladd, B., Wagner, A. L., Sanborn, G. E., Jacobs, J. H., Busquets, M. A., Chew, E. Y., and ALOFT Study Group (2022). Analysis of the Long-Term Visual Outcomes of ForeseeHome Remote Telemonitoring: The ALOFT Study. *Ophthalmology Retina*, 6(10):922–929.

McGinnis, J., Shit, S., Li, H. B., Sideri-Lampretsa, V., Graf, R., Dannecker, M., Pan, J., Stolt-Ansó, N., Mühlau, M., Kirschke, J. S., Rueckert, D., and Wiestler, B. (2023). Single-Subject Multi-Contrast MRI Super-Resolution via Implicit Neural Representations. In Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., and Taylor, R., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pages 173–183.

McKinley, R., Wagner, F., and Wiest, R. (2021). Detection of Lesion Change in Multiple Sclerosis Using a Cascade of 3D-to-2D Networks. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 97–100.

McKinley, R., Wepfer, R., Grunder, L., Aschwanden, F., Fischer, T., Friedli, C., Muri, R., Rummel, C., Verma, R., Weisstanner, C., Wiestler, B., Berger, C., Eichinger, P., Muhlau, M., Reyes, M., Salmen, A., Chan, A., Wiest, R., and Wagner, F. (2020). Automatic Detection of Lesion Load Change in Multiple Sclerosis Using Convolutional Neural Networks with Segmentation Confidence. *NeuroImage: Clinical*, 25:102104.

Meissen, F., Wiestler, B., Kaissis, G., and Rueckert, D. (2022). On the Pitfalls of Using the Residual Error as Anomaly Score. In Konukoglu, E., Menze, B., Venkataraman, A., Baumgartner, C., Dou, Q., and Albarqouni, S., editors, *Proceedings of The 5th International Conference on Medical Imaging with Deep Learning*, volume 172 of *Proceedings of Machine Learning Research (PMLR)*, pages 914–928.

Melas-Kyriazi, L., Rupprecht, C., Laina, I., and Vedaldi, A. (2022). Deep Spectral Methods: A Surprisingly Strong Baseline for Unsupervised Semantic Segmentation and Localization. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8364–8375.

Meng, M., Bi, L., Feng, D., and Kim, J. (2023). Brain Tumor Sequence Registration with Non-Iterative Coarse-To-Fine Networks and Dual Deep Supervision. In Bakas, S., Crimi, A., Baid, U., Malec, S., Pytlarz, M., Baheti, B., Zenk, M., and Dorent, R., editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 273–282.

Meng, M., Bi, L., Fulham, M., Feng, D. D., and Kim, J. (2022). Enhancing Medical Image Registration via Appearance Adjustment Networks. *NeuroImage*, 259:119444.

Meng, Z., Hu, Y., and Ancey, C. (2020). Using a Data Driven Approach to Predict Waves Generated by Gravity Driven Mass Flows. *Water*, 12(2).

Meyer, F. (1994). Topographic Distance and Watershed Lines. *Signal Processing*, 38(1):113–125.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2021). NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In Vedaldi, A., Bischof, H., Brox, T., and Frahm, J.-M., editors, *Computer Vision – ECCV 2020*, pages 405–421.

Modersitzki, J. (2009). *FAIR: Flexible Algorithms for Image Registration*. Society for Industrial and Applied Mathematics.

Mok, T. C. W. and Chung, A. C. S. (2020). Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks. In Martel, A. L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M. A., Zhou, S. K., Racoceanu, D., and Joskowicz, L., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pages 211–221.

Mok, T. C. W. and Chung, A. C. S. (2023). Robust Image Registration with Absent Correspondences in Pre-operative and Follow-Up Brain MRI Scans of Diffuse Glioma Patients. In Bakas, S., Crimi, A., Baid, U., Malec, S., Pytlarz, M., Baheti, B., Zenk, M., and Dorent, R., editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 231–240.

Moraal, B., Wattjes, M. P., Geurts, J. J., Knol, D. L., van Schijndel, R. A., Pouwels, P. J., Vrenken, H., and Barkhof, F. (2010). Improved Detection of Active Multiple Sclerosis Lesions: 3D Subtraction Imaging. *Radiology*, 255(1):154–163.

Mortensen, E., Morse, B., Barrett, W., and Udupa, J. (1992). Adaptive Boundary Detection Using 'Live-Wire' Two-Dimensional Dynamic Programming. In *Proceedings Computers in Cardiology*, pages 635–638.

Mou, L., Liang, L., Gao, Z., and Wang, X. (2022). A Multi-Scale Anomaly Detection Framework for Retinal OCT Images Based on the Bayesian Neural Network. *Biomedical Signal Processing and Control*, 75:103619.

Müller, T., Evans, A., Schied, C., and Keller, A. (2022). Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Transactions on Graphics*, 41(4).

Nachev, P., Coulthard, E., Jäger, H. R., Kennard, C., and Husain, M. (2008). Enantiomorphic Normalization of Focally Lesioned Brains. *NeuroImage*, 39(3):1215–1226.

Najman, L. and Schmitt, M. (1996). Geodesic Saliency of Watershed Contours and Hierarchical Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1163–1173.

Nanni, L., Ghidoni, S., and Brahnam, S. (2017). Handcrafted vs. Non-Handcrafted Features for Computer Vision Classification. *Pattern Recognition*, 71:158–172.

Narayanan, A. and Bergen, K. J. (2024). Prototype-Based Methods in Explainable AI and Emerging Opportunities in the Geosciences. arXiv E-Print 2410.19856.

Nasheeda, V. P. and Rajangam, V. (2024). Deep Attention V-Net Architecture for Enhanced Multiple Sclerosis Segmentation. *IEEE Access*, 12:110550–110562.

Nichyporuk, B., Vasilevski, K., Hu, A., Myers-Colet, C., Cardinell, J., Szeto, J., Falet, J.-P., Zimmermann, E., Schroeter, J., Arnold, D. L., and Arbel, T. (2021). Consensus Learning with Multi-Rater Labels for Segmenting and Detecting New Lesions. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 85–88.

Niemeijer, M., Garvin, M. K., Lee, K., van Ginneken, B., M.D., M. D. A., and Sonka, M. (2009). Registration of 3D Spectral OCT Volumes Using 3D SIFT Feature Point Matching. In Pluim, J. P. W. and Dawant, B. M., editors, *Medical Imaging 2009: Image Processing*, volume 7259, page 72591I.

Niethammer, M., Hart, G. L., Pace, D. F., Vespa, P. M., Irimia, A., Van Horn, J. D., and Aylward, S. R. (2011). Geometric Metamorphosis. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pages 639–646.

Ntatsis, K., Brea, L. S., De Jesus, D. A., Barbosa-Breda, J. a., van Walsum, T., Bennink, E., and Klein, S. (2022). Motion Correction in Retinal Optical Coherence Tomography Imaging Using Deep Learning Registration. In Colliot, O. and Išgum, I., editors, *Medical Imaging 2022: Image Processing*, page 1203219.

Otte, R., Wippermann, B., Schade, S., and Otte, V. (2020). *Von Data Mining bis Big Data, Handbuch für die industrielle Praxis*. Carl Hanser Verlag GmbH & Co. KG.

Ou, Y., Sotiras, A., Paragios, N., and Davatzikos, C. (2011). DRAMMS: Deformable Registration via Attribute Matching and Mutual-Saliency Weighting. *Medical Image Analysis*, 15(4):622–639.

Ouali, Y., Hudelot, C., and Tami, M. (2020). Autoregressive Unsupervised Image Segmentation. In Vedaldi, A., Bischof, H., Brox, T., and Frahm, J.-M., editors, *Computer Vision – ECCV 2020*, pages 142–158.

Pan, L. and Chen, X. (2023). Retinal OCT Image Registration: Methods and Applications. *IEEE Reviews in Biomedical Engineering*, 16:307–318.

Pan, L., Guan, L., and Chen, X. (2019). Segmentation Guided Registration for 3D Spectral-Domain Optical Coherence Tomography Images. *IEEE Access*, 7:138833–138845.

Pan, L., Shi, F., Xiang, D., Yu, K., Duan, L., Zheng, J., and Chen, X. (2020). OCTRexpert: A Feature-Based 3D Registration Method for Retinal OCT Images. *IEEE Transactions on Image Processing*, 29:3885–3897.

Park, B. H., Saxer, C., Srinivas, S. M., Nelson, J. S., and de Boer, J. F. (2001). In Vivo Burn Depth Determination by High-Speed Fiber-Based Polarization Sensitive Optical Coherence Tomography. *Journal of Biomedical Optics*, 6(4):474–479.

Pasricha, M. V., Tai, V., Sleiman, K., Winter, K., Chiu, S. J., Farsiu, S., Stinnett, S. S., Lad, E. M., Wong, W. T., Chew, E. Y., Toth, C. A., and Toth, C. A. (2021). Local Anatomic Precursors to New-Onset Geographic Atrophy in Age-Related Macular Degeneration as Defined on OCT. *Ophthalmology Retina*, 5(5):396–408.

Periaswamy, S. and Farid, H. (2006). Medical Image Registration with Partial Data. *Medical Image Analysis*, 10(3):452–464.

Petersen, G., Wittmann, R., Arndt, V., and Göpffarth, D. (2014). Epidemiologie der Multiplen Sklerose in Deutschland: Regionale Unterschiede und Versorgungsstruktur in Abrechnungsdaten der gesetzlichen Krankenversicherung. *Der Nervenarzt*, 8:990–998.

Pilehvari, S., Morgan, Y., and Peng, W. (2024). An Analytical Review on the Use of Artificial Intelligence and Machine Learning in Diagnosis, Prediction, and Risk Factor Analysis of Multiple Sclerosis. *Multiple Sclerosis and Related Disorders*, 89:105761.

Prados, F. and Kanber, B. (2021). Detecting New Multiple Sclerosis Lesions Using a Mixed Approach. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 1–4.

Preloznik, D. and Žiga Špiclin (2021). Double Pathway Method For MSSEG-2 Challenge. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 41–44.

Rahimi, A. and Recht, B. (2007). Random Features for Large-Scale Kernel Machines. In Platt, J., Koller, D., Singer, Y., and Roweis, S., editors, *Proceedings of the 21st Conference on Advances in Neural Information Processing Systems (NeurIPS)*, volume 20.

Ramasinghe, S. and Lucey, S. (2022). Beyond Periodicity: Towards a Unifying Framework for Activations in Coordinate-MLPs. In Avidan, S., Brostow, G., Cissé, M., Farinella, G. M., and Hassner, T., editors, *Computer Vision – ECCV 2022*, pages 142–158.

Ramesh, K. K. D., Kumar, G. K., Swapna, K., Datta, D., and Rajest, S. S. (2021). A Review of Medical Image Segmentation Algorithms. *EAI Endorsed Transactions on Pervasive Health and Technology*, 7(27):e6.

Reed, A. W., Kim, H., Anirudh, R., Mohan, K. A., Champley, K., Kang, J., and Jayasuriya, S. (2021). Dynamic CT Reconstruction from Limited Views with Implicit Neural Representations and Parametric Motion Fields. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2238–2248.

Rey, D., Subsol, G., Delingette, H., and Ayache, N. (2002). Automatic Detection and Segmentation of Evolving Processes in 3D Medical Images: Application to Multiple Sclerosis. *Medical Image Analysis*, 6(2):163–179.

Ricco, S., Chen, M., Ishikawa, H., Wollstein, G., and Schuman, J. (2009). Correcting Motion Artifacts in Retinal Spectral Domain Optical Coherence Tomography via Image Registration. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, pages 100–107.

Rodrigues, C. N., Nunes, I. M., Pereira, M. B., Oliveira, H., and dos Santos, J. A. (2024). From Superpixels to Foundational Models: An Overview of Unsupervised and Generalizable Image Segmentation. *Computers & Graphics*, 123:104014.

Rodrigues, J. A. (2024). Using Physics-Informed Neural Networks (PINNs) for Tumor Cell Growth Modeling. *Mathematics*, 12(8):1195.

Rondinella, A., Crispino, E., Guarnera, F., Giudice, O., Ortis, A., Russo, G., Di Lorenzo, C., Maimone, D., Pappalardo, F., and Battiato, S. (2023). Boosting Multiple Sclerosis Lesion Segmentation through Attention Mechanism. *Computers in Biology and Medicine*, 161:107021.

Rondinella, A., Guarnera, F., Crispino, E., Russo, G., Di Lorenzo, C., Maimone, D., Pappalardo, F., and Battiato, S. (2025). ICPR 2024 Competition on Multiple Sclerosis Lesion Segmentation—Methods and Results. In Antonacopoulos, A., Chaudhuri, S., Chellappa, R., Liu, C.-L., Bhattacharya, S., and Pal, U., editors, *Pattern Recognition. Competitions*, pages 1–16.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241.

Rosa, F. L., Thiran, J.-P., and Cuadra, M. B. (2021). A Subtraction Image-Based Method to Detect New Appearing Multiple Sclerosis Lesions on Single-Contrast FLAIR MRI. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 65–68.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252.

Sadda, S. R., Guymer, R., Holz, F. G., Schmitz-Valckenberg, S., Curcio, C. A., Bird, A. C., Blodi, B. A., Bottoni, F., Chakravarthy, U., Chew, E. Y., Csaky, K., Danis, R. P., Fleckenstein, M., Freund, K. B., Grunwald, J., Hoyng, C. B., Jaffe, G. J., Liakopoulos, S., Monés, J. M., Pauleikhoff, D., Rosenfeld, P. J., Sarraf, D., Spaide, R. F., Tadayoni, R., Tufail, A., Wolf, S., and Staurenghi, G. (2018). Consensus Definition for Atrophy Associated with Age-Related Macular Degeneration on OCT: Classification of Atrophy Report 3. *Ophthalmology*, 125(4):537–548.

Salem, M., Cabezas, M., Valverde, S., Pareto, D., Oliver, A., Salvi, J., Rovira, À., and Lladó, X. (2018). A Supervised Framework with Intensity Subtraction and Deformation Field Features for the Detection of New T2-w Lesions in Multiple Sclerosis. *NeuroImage: Clinical*, 17:607–615.

Salem, M., Ryan, M. A., Oliver, A., Hussain, K. F., and Lladó, X. (2022). Improving the Detection of New Lesions in Multiple Sclerosis with a Cascaded 3D Fully Convolutional Neural Network Approach. *Frontiers in Neuroscience*, 16.

Salem, M., Valverde, S., Cabezas, M., Pareto, D., Oliver, A., Salvi, J., Rovira, À., and Lladó, X. (2020). A Fully Convolutional Neural Network for New T2-w Lesion Detection in Multiple Sclerosis. *NeuroImage: Clinical*, 25:102149.

Sandkühler, R., Jud, C., Andermatt, S., and Cattin, P. C. (2018). AirLab: Autograd Image Registration Laboratory. arXiv E-Print 1806.09907.

Santarossa, M., Kilic, A., von der Burchard, C., Schmarje, L., Zelenka, C., Reinhold, S., Koch, R., and Roider, J. (2022a). MedRegNet: Unsupervised Multimodal Retinal-Image Registration with GANs and Ranking Loss. In Colliot, O. and Išgum, I., editors, *Medical Imaging 2022: Image Processing*, volume 12032, page 1203218.

Santarossa, M., Tatli, A., von der Burchard, C., Andresen, J., Roider, J., Handels, H., and Koch, R. (2022b). Chronological Registration of OCT and Autofluorescence Findings in CSCR: Two Distinct Patterns in Disease Course. *Diagnostics*, 12(8).

Saragadam, V., LeJeune, D., Tan, J., Balakrishnan, G., Veeraraghavan, A., and Baraniuk, R. G. (2023). WIRE: Wavelet Implicit Neural Representations. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18507–18516.

Sarica, B. and Seker, D. Z. (2021). New MS Lesion Segmentation Using Deep Residual Attention Gate U-Net Using 2D slices of 3D MR Images. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 25–28.

Sarica, B. and Seker, D. Z. (2022). New MS Lesion Segmentation with Deep Residual Attention Gate U-Net Utilizing 2D Slices of 3D MR Images. *Frontiers in Neuroscience*, 16.

Sarica, B., Seker, D. Z., and Bayram, B. (2023). A Dense Residual U-Net for Multiple Sclerosis Lesions Segmentation from Multi-Sequence 3D MR Images. *International Journal of Medical Informatics*, 170:104965.

Schaffland, A. and Schöning, J. (2024). High Accurate Rephotographic Image Registration by Attention Masks: Enabling Intention-Driven Rephotographic Image Registration With Interactive Areas of Interest Masks. *IEEE Access*, 12:7519–7530.

Scharf, A., von der Burchard, C., Tatli, A., Santarossa, M., Andresen, J., Koch, R., Handels, H., Kepp, T., and Roider, J. (2024). Linking AI-Based Biomarker Analysis to Visual Acuity Changes in Central Serous Chorioretinopathy (CSCR). *Investigative Ophthalmology & Visual Science*, 65(7).

Schlegl, T., Seeböck, P., Waldstein, S. M., Langs, G., and Schmidt-Erfurth, U. (2019). f-AnoGAN: Fast Unsupervised Anomaly Detection with Generative Adversarial Networks. *Medical Image Analysis*, 54:30–44.

Schlüter, H. M., Tan, J., Hou, B., and Kainz, B. (2022). Natural Synthetic Anomalies for Self-Supervised Anomaly Detection and Localization. In Avidan, S., Brostow, G., Cissé, M., Farinella, G. M., and Hassner, T., editors, *Computer Vision – ECCV 2022*, pages 474–489.

Schmidt-Mengin, M., Soulier, T., Hamzaoui, M., Yazdan-Panah, A., Bodini, B., Ayache, N., Stankoff, B., and Colliot, O. (2022). Online Hard Example Mining vs. Fixed Oversampling Strategy for Segmentation of New Multiple Sclerosis Lesions from Longitudinal FLAIR MRI. *Frontiers in Neuroscience*, 16.

Schmidt-Mengin, M., Yazdan-Panah, A., Soulier, T., Hamzaoui, M., Stankoff, B., Ayache, N., and Colliot, O. (2021). Segmentation of New Multiple Sclerosis Lesions on FLAIR MRI Using Online Hard Example Mining. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 89–92.

Sedghi, A., Kapur, T., Luo, J., Mousavi, P., and Wells, W. M. (2019). Probabilistic Image Registration via Deep Multi-class Classification: Characterizing Uncertainty. In Greenspan, H., Tanno, R., Erdt, M., Arbel, T., Baumgartner, C., Dalca, A., Sudre, C. H., Wells, W. M., Drechsler, K., Linguraru, M. G., Oyarzun Laura, C., Shekhar, R., Wesarg, S., and González Ballester, M. Á., editors, *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging and Clinical Image-Based Procedures*, pages 12–22.

Seeböck, P., Orlando, J. I., Schlegl, T., Waldstein, S. M., Bogunović, H., Klimscha, S., Langs, G., and Schmidt-Erfurth, U. (2020). Exploiting Epistemic Uncertainty of Anatomy Segmentation for Anomaly Detection in Retinal OCT. *IEEE Transactions on Medical Imaging*, 39(1):87–98.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 618–626.

Semmlow, J. L. (2008). *Biosignal and Medical Image Processing*. CRC Press.

Sentker, T., Madesta, F., and Werner, R. (2018). GDL-FIRE^{4D}: Deep Learning-Based Fast 4D CT Image Registration. In Frangi, A. F., Schnabel, J. A., Davatzikos, C., Alberola-López, C., and Fichtinger, G., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 765–773.

Shamshad, F., Khan, S., Zamir, S. W., Khan, M. H., Hayat, M., Khan, F. S., and Fu, H. (2023). Transformers in Medical Imaging: A Survey. *Medical Image Analysis*, 88:102802.

Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., Poldrack, R. A., Bilder, R. M., and Toga, A. W. (2008). Construction of a 3D Probabilistic Atlas of Human Cortical Structures. *Neuroimage*, 39(3):1064–1080.

Shen, L., Pauly, J., and Xing, L. (2024). NeRP: Implicit Neural Representation Learning With Prior Embedding for Sparsely Sampled Image Reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, 35(1):770–782.

Shi, F., Cai, N., Gu, Y., Hu, D., Ma, Y., Chen, Y., and Chen, X. (2019). DeSpecNet: A CNN-Based Method for Speckle Reduction in Retinal Optical Coherence Tomography Images. *Physics in Medicine and Biology*, 64(17):175010.

Shi, J. and Malik, J. (2000). Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905.

Shvetsova, N., Bakker, B., Fedulova, I., Schulz, H., and Dylov, D. V. (2021). Anomaly Detection in Medical Imaging With Deep Perceptual Autoencoders. *IEEE Access*, 9:118571–118583.

Siddiquee, M. R. and Myronenko, A. (2021). Robust 3D MRI Segmentation of Multiple Sclerosis Lesions. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 81–84.

Sideri-Lampretsa, V., McGinnis, J., Qiu, H., Paschali, M., Simson, W., and Rueckert, D. (2024). SINR: Spline-Enhanced Implicit Neural Representation for Multi-Modal Registration. In Burgos, N., Petitjean, C., Vakalopoulou, M., Christodoulidis, S., Coupe, P., Delingette, H., Lartizien, C., and Mateus, D., editors, *Proceedings of The 7th International Conference on Medical Imaging with Deep Learning*, volume 250 of *Proceedings of Machine Learning Research (PMLR)*, pages 1462–1474.

Siebert, H., Großbröhmer, C., Hansen, L., and Heinrich, M. P. (2024). ConvexAdam: Self-Configuring Dual-Optimisation-Based 3D Multitask Medical Image Registration. *IEEE Transactions on Medical Imaging*, 44(2):738–748.

Simonyan, K., Vedaldi, A., and Zisserman, A. (2014). Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. In *Workshop at International Conference on Learning Representations*.

Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations*.

Sitzmann, V., Martel, J., Bergman, A., Lindell, D., and Wetzstein, G. (2020). Implicit Neural Representations with Periodic Activation Functions. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*.

Srivastava, R., Yow, A. P., Cheng, J., Wong, D. W., and Tey, H. L. (2018). Three-Dimensional Graph-Based Skin Layer Segmentation in Optical Coherence Tomography Images for Roughness Estimation. *Biomedical Optics Express*, 9(8):3590–3606.

Stolt-Ansó, N., McGinnis, J., Pan, J., Hammernik, K., and Rueckert, D. (2023). NISF: Neural Implicit Segmentation Functions. In Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., and Taylor, R., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pages 734–744.

Sudkamp, H., Koch, P., Spahr, H., Hillmann, D., Franke, G., Münst, M., Reinholz, F., Birngruber, R., and Hüttmann, G. (2016). In-Vivo Retinal Imaging with Off-Axis Full-Field Time-Domain Optical Coherence Tomography. *Optics Letters*, 41(21):4987–4990.

Sullivan, S. R., Underwood, R. A., Gibran, N. S., Sigle, R. O., Usui, M. L., Carter, W. G., and Olerud, J. E. (2004). Validation of a Model for the Study of Multiple Wounds in the Diabetic Mouse (db/db). *Plastic and Reconstructive Surgery*, 113(3):953–960.

Sun, S., Han, K., You, C., Tang, H., Kong, D., Naushad, J., Yan, X., Ma, H., Khosravi, P., Duncan, J. S., and Xie, X. (2024). Medical Image Registration via Neural Fields. *Medical Image Analysis*, 97:103249.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going Deeper with Convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9.

Szeskin, A., Yehuda, R., Shmueli, O., Levy, J., and Joskowicz, L. (2021). A Column-Based Deep Learning Method for the Detection and Quantification of Atrophy Associated with AMD in OCT Scans. *Medical Image Analysis*, 72:102130.

Tahghighi, P., Zhang, Y., Souza, R., and Komeili, A. (2024). Enhancing New Multiple Sclerosis Lesion Segmentation via Self-Supervised Pre-training and Synthetic Lesion Integration. In Linguraru, M. G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., and Schnabel, J. A., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, pages 263–272.

Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J., and Ng, R. (2020). Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. In *Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS)*.

Tian, L., Greer, H., José Estépar, R. S., Sengupta, R., and Niethammer, M. (2024a). NePhi: Neural Deformation Fields for Approximately Diffeomorphic Medical Image Registration. In Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., and Varol, G., editors, *Computer Vision – ECCV 2024*, pages 213–237.

Tian, L., Greer, H., Kwitt, R., Vialard, F.-X., San José Estépar, R., Bouix, S., Rushmore, R., and Niethammer, M. (2024b). uniGradICON: A Foundation Model for Medical Image Registration. In Linguraru, M. G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., and Schnabel, J. A., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*.

Tian, Y., Pang, G., Liu, Y., Wang, C., Chen, Y., Liu, F., Singh, R., Verjans, J. W., Wang, M., and Carneiro, G. (2023). Unsupervised Anomaly Detection in Medical Images with a Memory-Augmented Multi-Level Cross-Attentional Masked Autoencoder. In Cao, X., Xu, X., Rekik, I., Cui, Z., and Ouyang, X., editors, *Machine Learning in Medical Imaging: 14th International Workshop, MLMI 2023, Held in Conjunction with MICCAI 2023, Proceedings*, pages 11–21.

Toprak, I., Yaylalı, V., and Yildirim, C. (2014). Decreased Photoreceptor Inner Segment/Outer Segment Junction Reflectivity in Patients with Idiopathic Epimacular Membrane. *Eye*, 28:1126–1130.

Toprak, I., Yaylalı, V., and Yildirim, C. (2017). Early Deterioration in Ellipsoid Zone in Eyes with Non-Neovascular Age-Related Macular Degeneration. *International Ophthalmology*, 37(4):801–806.

Tran, P., Thoprakarn, U., Gourieux, E., dos Santos, C. L., Cavedo, E., Guizard, N., Cotton, F., Krolak-Salmon, P., Delmaire, C., Heidelberg, D., Pyatigorskaya, N., Ströer, S., Dormont, D., Martini, J.-B., and Chupin, M. (2022). Automatic Segmentation of White Matter Hyperintensities: Validation and Comparison with State-of-the-Art Methods on Both Multiple Sclerosis and Elderly Subjects. *NeuroImage: Clinical*, 33:102940.

Uzunova, H., Basso, L., Ehrhardt, J., and Handels, H. (2022). Synthesis of Annotated Pathological Retinal OCT Data with Pathology-Induced Deformations. In Drukker, K. and Iftekharuddin, K. M., editors, *Medical Imaging 2022: Computer-Aided Diagnosis*, volume 12033, page 120333K.

Uzunova, H., Handels, H., and Ehrhardt, J. (2021). Guided Filter Regularization for Improved Disentanglement of Shape and Appearance in Diffeomorphic Autoencoders. In Heinrich, M., Dou, Q., de Bruijne, M., Lellmann, J., Schläfer, A., and Ernst, F., editors, *Proceedings of the Fourth Conference on Medical Imaging with Deep Learning*, volume 143 of *Proceedings of Machine Learning Research (PMLR)*, pages 774–786.

Uzunova, H., Handels, H., and Ehrhardt, J. (2023). Low-Rank and Sparse Metamorphic Autoencoders for Unsupervised Pathology Disentanglement. In Fragemann, J., Li, J., Liu, X., Tsaftaris, S. A., Egger, J., and Kleesiek, J., editors, *Medical Applications with Disentanglements*, pages 59–69.

Uzunova, H., Schultz, S., Handels, H., and Ehrhardt, J. (2019). Unsupervised Pathology Detection in Medical Images Using Conditional Variational Autoencoders. *International Journal of Computer Assisted Radiology and Surgery*, 14(3):451–461.

Vadapalli, P. (2021). Biological Neural Network: Importance, Components & Comparison. https://www.upgrad.com/blog/biological-neural-network/. Accessed: 2023-04-23.

Van Gansbeke, W., Vandenhende, S., Georgoulis, S., and Van Gool, L. (2021). Unsupervised Semantic Segmentation by Contrasting Object Mask Proposals. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10032–10042.

van Harten, L. D., Stoker, J., and Išgum, I. (2024). Robust Deformable Image Registration Using Cycle-Consistent Implicit Representations. *IEEE Transactions on Medical Imaging*, 43(2):784–793.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is All you Need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Proceedings of the 31st Conference on Advances in Neural Information Processing Systems (NeurIPS)*, volume 30.

Vázquez-Marrufo, M., Sarrias-Arrabal, E., García-Torres, M., Martín-Clemente, R., and Izquierdo, G. (2023). A Systematic Review of the Application of Machine-Learning Algorithms in Multiple Sclerosis. *Neurología (English Edition)*, 38(8):577–590.

Vedaldi, A. and Soatto, S. (2008). Quick Shift and Kernel Methods for Mode Seeking. In Forsyth, D., Torr, P., and Zisserman, A., editors, *Computer Vision – ECCV 2008*, pages 705–718.

Venkataramanan, S., Peng, K.-C., Singh, R. V., and Mahalanobis, A. (2020). Attention Guided Anomaly Localization in Images. In Vedaldi, A., Bischof, H., Brox, T., and Frahm, J.-M., editors, *Computer Vision – ECCV 2020: 16th European Conference, Proceedings*, pages 485–503.

Vincent, L. and Soille, P. (1991). Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(06):583–598.

von der Burchard, C., Scharf, A., Santarossa, M., Tatli, A., Andresen, J., Koch, R., Handels, H., Kepp, T., and Roider, J. (2024). AI-Based Photoreceptor Length Segmentation and Influence on Disease Progression and Visual Acuity Changes in Central Serous Chorioretinopathy (CSCR). *Investigative Ophthalmology & Visual Science*, 65(7).

Voon, C. C., Wiltgen, T., Wiestler, B., Schlaeger, S., and Mühlau, M. (2024). Quantitative Susceptibility Mapping in Multiple Sclerosis: A Systematic Review and Meta-Analysis. *NeuroImage: Clinical*, 42:103598.

Walton, C., King, R., Rechtman, L., Kaye, W., Leray, E., Marrie, R. A., Robertson, N., La Rocca, N., Uitdehaag, B., van Der Mei, I., Wallin, M., Helme, A., Napier, C. A., Rijke, N., and Baneke, P. (2020). Rising Prevalence of Multiple Sclerosis Worldwide: Insights from the Atlas of MS. *Multiple Sclerosis Journal*, 26(14):1816–1821.

Wan, B., Ganier, C., Du-Harpur, X., Harun, N., Watt, F., Patalay, R., and Lynch, M. (2021). Applications and Future Directions for Optical Coherence Tomography in Dermatology. *British Journal of Dermatology*, 184(6):1014–1022.

Wang, G., Wang, Z., Chen, Y., and Zhao, W. (2015). Robust Point Matching Method for Multimodal Retinal Image Registration. *Biomedical Signal Processing and Control*, 19:68–76.

Wang, J., Xing, J., Druzgal, J., Wells, W. M., and Zhang, M. (2023). MetaMorph: Learning Metamorphic Image Transformation with Appearance Changes. In Frangi, A., de Bruijne, M., Wassermann, D., and Navab, N., editors, *Information Processing in Medical Imaging*, pages 576–587.

Wang, M., Zhu, W., Yu, K., Chen, Z., Shi, F., Zhou, Y., Ma, Y., Peng, Y., Bao, D., Feng, S., Ye, L., Xiang, D., and Chen, X. (2021a). Semi-Supervised Capsule cGAN for Speckle Noise Reduction in Retinal OCT Images. *IEEE Transactions on Medical Imaging*, 40(4):1168–1183.

Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Huang, J., Yang, W., and Han, X. (2021b). TransPath: Transformer-Based Self-Supervised Learning for Histopathological Image Classification. In de Bruijne, M., Cattin, P. C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., and Essert, C., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 186–195.

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

Warfield, S. K., Zou, K. H., and Wells, W. M. (2004). Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921.

Wei, Q., Shi, F., Zhu, W., Xiang, D., Chen, H., and Chen, X. (2017). Nonrigid Registration of 3D Longitudinal Optical Coherence Tomography Volumes with Choroidal Neovascularization. In Styner, M. A. and Angelini, E. D., editors, *Medical Imaging 2017: Image Processing*, volume 10133, page 101330X.

Welzel, J. (2001). Optical Coherence Tomography in Dermatology: A Review. *Skin Research and Technology*, 7(1):1–9.

Werner, R., Wilms, M., Cheng, B., and Forkert, N. D. (2016). Beyond Cost Function Masking: RPCA-Based Non-Linear Registration in the Context of VLSM. In *2016 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, pages 1–4.

WHO (2008). *Atlas: Multiple Sclerosis Resources in the World 2008*. World Health Organization & Multiple Sclerosis International Federation.

Wodzinski, M., Jurgas, A., Marini, N., Atzori, M., and Müller, H. (2023). Unsupervised Method for Intra-Patient Registration of Brain Magnetic Resonance Images Based on Objective Function Weighting by Inverse Consistency: Contribution to the BraTS-Reg Challenge. In Bakas, S., Crimi, A., Baid, U., Malec, S., Pytlarz, M., Baheti, B., Zenk, M., and Dorent, R., editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 241–251.

Wolterink, J. M., Zwienenberg, J. C., and Brune, C. (2022). Implicit Neural Representations for Deformable Image Registration. In Konukoglu, E., Menze, B., Venkataraman, A., Baumgartner, C., Dou, Q., and Albarqouni, S., editors, *Proceedings of The 5th International Conference on Medical Imaging with Deep Learning*, volume 172 of *Proceedings of Machine Learning Research (PMLR)*, pages 1349–1359.

Wu, J., Zhou, S., Lin, L., Wang, X., and Tan, W. (2024). Fast Diffeomorphic Image Registration Using Patch Based Fully Convolutional Networks. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1–4.

Wu, M., Cai, X., Chen, Q., Ji, Z., Niu, S., Leng, T., Rubin, D. L., and Park, H. (2019). Geographic Atrophy Segmentation in SD-OCT Images Using Synthesized Fundus Autofluorescence Imaging. *Computer Methods and Programs in Biomedicine*, 182:105101.

Wu, Q., Li, Y., Sun, Y., Zhou, Y., Wei, H., Yu, J., and Zhang, Y. (2023a). An Arbitrary Scale Super-Resolution Approach for 3D MR Images via Implicit Neural Representation. *IEEE Journal of Biomedical and Health Informatics*, 27(2):1004–1015.

Wu, Q., Li, Y., Xu, L., Feng, R., Wei, H., Yang, Q., Yu, B., Liu, X., Yu, J., and Zhang, Y. (2021). IREM: High-Resolution Magnetic Resonance Image Reconstruction via Implicit Neural Representation. In de Bruijne, M., Cattin, P. C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., and Essert, C., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 65–74.

Wu, Y., Wu, Z., Shi, H., Picker, B., Chong, W., and Cai, J. (2023b). CoactSeg: Learning from Heterogeneous Data for New Multiple Sclerosis Lesion Segmentation. In Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., and Taylor, R., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pages 3–13.

Wysocki, A. B. (1996). Wound Measurement. *International Journal of Dermatology*, 35(2):82–91.

Xia, X. and Kulis, B. (2017). W-Net: A Deep Model for Fully Unsupervised Image Segmentation.

Xie, X., Song, Y., Ye, F., Yan, H., Wang, S., Zhao, X., and Dai, J. (2021). Improving Deformable Image Registration with Point Metric and Masking Technique for Postoperative Breast Cancer Radiotherapy. *Quantitative Imaging in Medicine and Surgery*, 11(4):1196–1208.

Xie, Y., Rizzi, S. C., Dawson, R., Lynam, E., Richards, S., Leavesley, D. I., and Upton, Z. (2010). Development of a Three-Dimensional Human Skin Equivalent Wound Model for Investigating Novel Wound Healing Therapies. *Tissue Engineering Part C: Methods*, 16(5):1111–1123.

Xu, Z., Luo, J., Lu, D., Yan, J., Frisken, S., Jagadeesan, J., Wells, W. M., Li, X., Zheng, Y., and Tong, R. K.-y. (2022). Double-Uncertainty Guided Spatial and Temporal Consistency Regularization Weighting for Learning-Based Abdominal Registration. In Wang, L., Dou, Q., Fletcher, P. T., Speidel, S., and Li, S., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 14–24.

Yang, S., Zhu, F., Ling, X., Liu, Q., and Zhao, P. (2021). Intelligent Health Care: Applications of Deep Learning in Computational Medicine. *Frontiers in Genetics*, 12:607471.

Yang, X., Kwitt, R., Styner, M., and Niethammer, M. (2017). Quicksilver: Fast Predictive Image Registration - A Deep Learning Approach. *NeuroImage*, 158:378–396.

Yeh, A. T., Kao, B., Jung, W. G., Chen, Z., Nelson, J. S., and Tromberg, B. J. (2004). Imaging Wound Healing Using Optical Coherence Tomography and Multiphoton Microscopy in an In Vitro Skin-Equivalent Tissue Model. *Journal of Biomedical Optics*, 9(2):248–253.

Yousefi, S., Qin, J., Dziennis, S., and Wang, R. K. (2014). Assessment of Microcirculation Dynamics During Cutaneous Wound Healing Phases In Vivo Using Optical Microangiography. *Journal of Biomedical Optics*, 19(7):076015.

Yuan, Z., Yang, D., Wang, W., Zhao, J., and Liang, Y. (2023). Self Super-Resolution of Optical Coherence Tomography Images Based on Deep Learning. *Optics Express*, 31(17):27566–27581.

Yuan, Z., Zakhaleva, J., Ren, H., Liu, J., Chen, W., and Pan, Y. (2010). Noninvasive and High-Resolution Optical Monitoring of Healing of Diabetic Dermal Excisional Wounds Implanted with Biodegradable In Situ Gelable Hydrogels. *Tissue Engineering Part C: Methods*, 16(2):237–247.

Zeineldin, R. A., Karar, M. E., Mathis-Ullrich, F., and Burgert, O. (2023). Self-Supervised iRegNet for the Registration of Longitudinal Brain MRI of Diffuse Glioma Patients. In Bakas, S., Crimi, A., Baid, U., Malec, S., Pytlarz, M., Baheti, B., Zenk, M., and Dorent, R., editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 25–34.

Zhang, H., Li, H., and Oguz, I. (2021). Segmentation of New MS Lesions with Tiramisu and 2.5D Stacked Slices. In Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M., editors, *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure*, pages 61–64.

Zhang, H., Li, Z., Wang, X., and Zhang, X. (2015). Speckle Reduction in Optical Coherence Tomography by Two-Step Image Registration. *Journal of Biomedical Optics*, 20(3):036013.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. (2018). The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 586–595.

Zhang, R. Z., Ezhov, I., Balcerak, M., Zhu, A., Wiestler, B., Menze, B., and Lowengrub, J. S. (2025). Personalized Predictions of Glioblastoma Infiltration: Mathematical Models, Physics-Informed Neural Networks and Multimodal Scans. *Medical Image Analysis*, 101:103423.

Zhang, Y., Li, Z., Nan, N., and Wang, X. (2023). TranSegNet: Hybrid CNN-Vision Transformers Encoder for Retina Segmentation of Optical Coherence Tomography. *Life*, 13(4):976.

Zhou, C. and Paffenroth, R. C. (2017). Anomaly Detection with Robust Deep Autoencoders. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 665–674.

Ziemann, M. R., Joshi, R. R., and Metzler, C. A. (2024). Time-Varying Implicit Neural Representations for Unsupervised Speckle Denoising in Dynamic Scenes. In *Unconventional Imaging, Sensing, and Adaptive Optics 2024*, volume 13149, page 1314907.

Zimmerer, D., Full, P. M., Isensee, F., Jager, P., Adler, T., Petersen, J., Kohler, G., Ross, T., Reinke, A., Kascenas, A., Jensen, B. S., O'Neil, A. Q., Tan, J., Hou, B., Batten, J., Qiu, H., Kainz, B., Nina Shvetsova, I. F., Dylov, D. V., Yu, B., Zhai, J., Hu, J., Si, R., Zhou, S., Wang, S., Li, X., Chen, X., Zhao, Y., Marimont, S. N., Tarroni, G., Saase, V., Maier-Hein, L., and Maier-Hein, K. (2022). MOOD 2020: A Public Benchmark for Out-of-Distribution Detection and Localization on Medical Images. *IEEE Transactions on Medical Imaging*, 41(10):2728–2738.

Appendix

A1 CSCR Dataset Characteristics

In this section, the CSCR dataset used in Chapters 3, 6 and 7 is described in detail. The dataset is partially annotated for PEZ pathologies and/or fluid accumulations. Tab. A1 lists the patients who were segmented for PEZ pathologies, indicating the number of images available per eye (#L and #R), the number of segmented images (#S_{PEZ}), the pathologies observed (Patho.), the follow-up time (FUT), and for which network trainings the patients' images were used. In Tab. A2, the patients segmented for fluids are listed, again giving the number of (segmented) images per eye, the observed pathologies, the follow-up time and the methods the images were used for. Finally, in Tab. A3, for patients without given annotations, the number of available OCT images, the follow-up times and an indication whether the images were used for the training of IIC and FluidRegNet are given. Patients 35, 36, 104, 106 and 122 were segmented for both PEZ and fluid pathologies and, thus, are listed twice. Follow-up times are given in months.

Table A1: Part 1 of the longitudinal CSCR dataset. For patients listed here, manual segmentations of the PEZ are available. For each patient, the number of available OCT images, the number of segmented OCT images and the observed pathologies are given per eye (1: Inconsistent PEZ, 2: Non-localizable PEZ, 3: Atrophy and - indicates no pathology). Also, the follow-up time is given, and it is indicated whether the patient's images were used for the training of the atrophy segmentation network in Sec. 3.4 (IIC), FluidRegNet in Chapter 6 (FRN), and FRN-Ano in Sec. 6.4.

ID	#L	#S_{PEZ}	Patho.	#R	#S_{PEZ}	Patho.	FUT	IIC	FRN	FRN-Ano
001	14	5	1, 2	14	3	-	65			
002	2	2	-	2	2	1, 2	1			
003	11	11	1, 2	11	11	1, 2	69			
004	5	5	1, 2	5	5	1, 2	13			
005	9	8	1, 2, 3	9	9	-	24	✓		
006	22	1	2, 3	22	1	1, 2, 3	69			
007	1	1	-	1	1	1, 2	0			
013	3	1	1, 2	3	1	-	11			
014	1	1	-	1	1	1	0			
015	5	1	-	5	1	1, 2, 3	13	✓		
016	1	1	1, 2	1	1	1, 3	0			
017	3	1	1, 2, 3	3	1	1	18			
018	1	1	1	1	1	-	0			
019	1	1	-	1	1	1, 2	0			
020	24	1	1, 2	24	1	1, 3	94	✓		
021	6	1	-	6	1	1, 2	58			

022	1	1	-	1	1	-	0			
023	6	1	-	5	1	1, 2	40			
024	6	1	-	6	1	1, 2, 3	24			
026	3	1	1, 2	3	1	1	19			
027	2	1	1, 2	2	1	1, 3	60	✓		
028	2	1	-	2	1	1	41			
029	11	1	1, 2	10	1	1, 3	34	✓		
030	2	1	-	2	1	1, 2	10			
031	17	1	-	17	1	1, 2	88		✓	
032	3	1	1	3	1	-	6			
033	2	1	-	2	1	1	2			
034	2	1	1, 2	2	1	1	2			
035	12	1	1, 2	10	1	-	143	✓	✓	✓
036	4	1	-	4	1	-	68		✓	✓
037	2	1	1, 2	1	1	-	3			
038	1	1	1, 2	1	1	-	0			
039	4	1	-	4	1	1, 2	32			
040	1	1	1, 2	1	1	-	0			
041	1	1	-	1	1	1, 2	0			
042	4	1	-	4	1	1, 2	15			
043	3	1	-	3	1	-	10			
045	1	1	-	1	1	1, 2	0			
046	7	1	1	7	1	1	19			
048	4	1	1, 2	4	1	-	17			
049	1	1	1	1	1	1, 2	0			
050	3	1	1, 2	3	1	-	4			
051	5	1	-	5	1	-	42			
052	1	1	1, 2	1	1	-	0			
053	1	1	-	1	1	1, 2	0			
054	11	1	1, 2	9	1	-	55			
055	19	1	1, 2, 3	19	0	n/a	91	✓		
061	19	0	n/a	19	1	1, 2, 3	80	✓		
064	0	0	n/a	1	1	1, 2, 3	0	✓		
070	4	0	n/a	5	1	1, 2, 3	113	✓		
086	0	0	n/a	1	1	1, 2, 3	0	✓		
087	4	1	1, 3	4	0	n/a	51	✓		
091	0	0	n/a	1	1	1, 3	0	✓		
095	1	1	1, 2, 3	0	0	n/a	0			
098	1	1	1, 2, 3	0	0	n/a	0	✓		
099	8	0	n/a	8	1	1, 3	76	✓	✓	
102	4	1	1, 2, 3	4	1	1, 3	9	✓		
104	12	1	1, 3	12	1	1, 2, 3	77	✓		
106	11	1	1, 2, 3	10	0	n/a	126		✓	
109	1	1	1, 3	1	1	1, 2, 3	0	✓		
112	0	0	n/a	17	1	1, 2, 3	155	✓		
113	1	1	1, 2, 3	1	1	1, 2, 3	0	✓		
117	3	1	1, 2, 3	2	0	n/1	52	✓	✓	
118	14	0	n/a	15	1	1, 2, 3	128	✓	✓	

122	6	1	1, 2, 3	7	0	n/1	90	✓
124	22	0	n/a	22	1	1, 2, 3	121	✓

Table A2: Part 2 of the longitudinal CSCR dataset. For patients listed here, fluid segmentations are given. For each patient, the number of available OCT images, the number of segmented OCT images and the observed pathologies are given per eye (4: IRF, 5: SRF, 6: PED and - indicates no pathology). Also, the follow-up time is given, and it is indicated whether the patient's images were used for the training of the atrophy segmentation network in Sec. 3.4 (IIC), FluidRegNet in Chapter 6 (FRN), and FRN-Ano in Sec. 6.4.

ID	#L	#S_{Fluid}	Patho.	#R	#S_{Fluid}	Patho.	FUT	IIC	FRN	FRN-Ano
008	8	8	5, 6	8	8	5, 6	31		✓	✓
009	1	0	n/a	1	1	5	0			✓
035	12	8	5, 6	10	6	-	143	✓	✓	✓
036	4	4	-	4	4	-	68		✓	✓
057	2	2	-	2	2	5	91	✓	✓	✓
082	5	5	5, 6	5	5	-	72		✓	✓
083	5	5	5, 6	5	5	4, 5, 6	40	✓	✓	✓
088	8	8	-	8	8	5	27		✓	✓
094	8	8	5	8	8	-	24		✓	✓
097	4	4	5, 6	4	4	5, 6	30		✓	✓
100	4	4	-	4	4	6	58		✓	✓
101	0	0	n/a	1	1	-	0		✓	
104	12	3	5	12	2	5	77	✓	✓	✓
106	11	10	4, 5, 6	10	10	4, 5, 6	126			✓
116	6	6	5, 6	6	6	5	77		✓	✓
121	7	5	-	7	5	4, 5, 6	135	✓	✓	✓
122	6	6	5	7	6	-	90	✓		
123	3	3	-	3	3	5, 6	75	✓	✓	✓
125	11	10	5, 6	11	9	5, 6	129		✓	✓
126	4	4	5, 6	4	4	6	18		✓	✓

Table A3: Part 3 of the longitudinal CSCR dataset. For patients listed here, no manual segmentations are available. For each patient, the number of available OCT images is given per eye. Also, it is indicated whether the patient's images were used for the training of the atrophy segmentation network in Sec. 3.4 (IIC) and FluidRegNet in Chapter 6 (FRN).

ID	#L	#R	FUT	IIC	FRN	ID	#L	#R	FUT	IIC	FRN
010	5	5	38	✓		077	1	1	0		
011	11	11	50			078	6	6	13		
012	2	2	47	✓		079	1	1	0		
025	5	5	14			080	3	3	3		
044	4	4	60			081	9	8	31		
047	15	18	158	✓	✓	084	6	6	27		✓
056	2	2	9			085	13	13	42		
058	7	6	14			089	2	2	3		✓
059	2	2	9			090	34	35	119		
060	1	1	0			092	4	4	7		
062	1	1	0			093	5	6	33		✓
063	2	2	113		✓	096	1	1	0		
065	4	4	10			103	10	10	163		
066	1	1	0			105	10	10	40	✓	✓
067	12	12	61			107	31	31	164	✓	
068	5	5	12			108	10	10	114		✓
069	1	1	0			110	21	20	168		
071	7	7	15			111	12	12	53		
072	1	1	0			114	4	4	7		
073	2	2	3			115	3	3	72		✓
074	1	1	0			119	2	2	2		✓
075	3	2	6			120	3	3	8		✓
076	5	5	16								

A2 Additional Implementation Details

In this section, three figures are given showing the exact architectures of networks used additionally to the main methods of the respective chapters. Fig. A3 illustrates the combination of NCR-Net and U-Net for new MS lesion segmentation. This method served as competitive method to ANCR-Net in Sec. 5.3.1. Fig. A3 shows the U-Net for retina segmentation used in Section 3.4, and Fig A4 shows the U-Net which is used for SRF segmentation in Sec. 6.4.3.

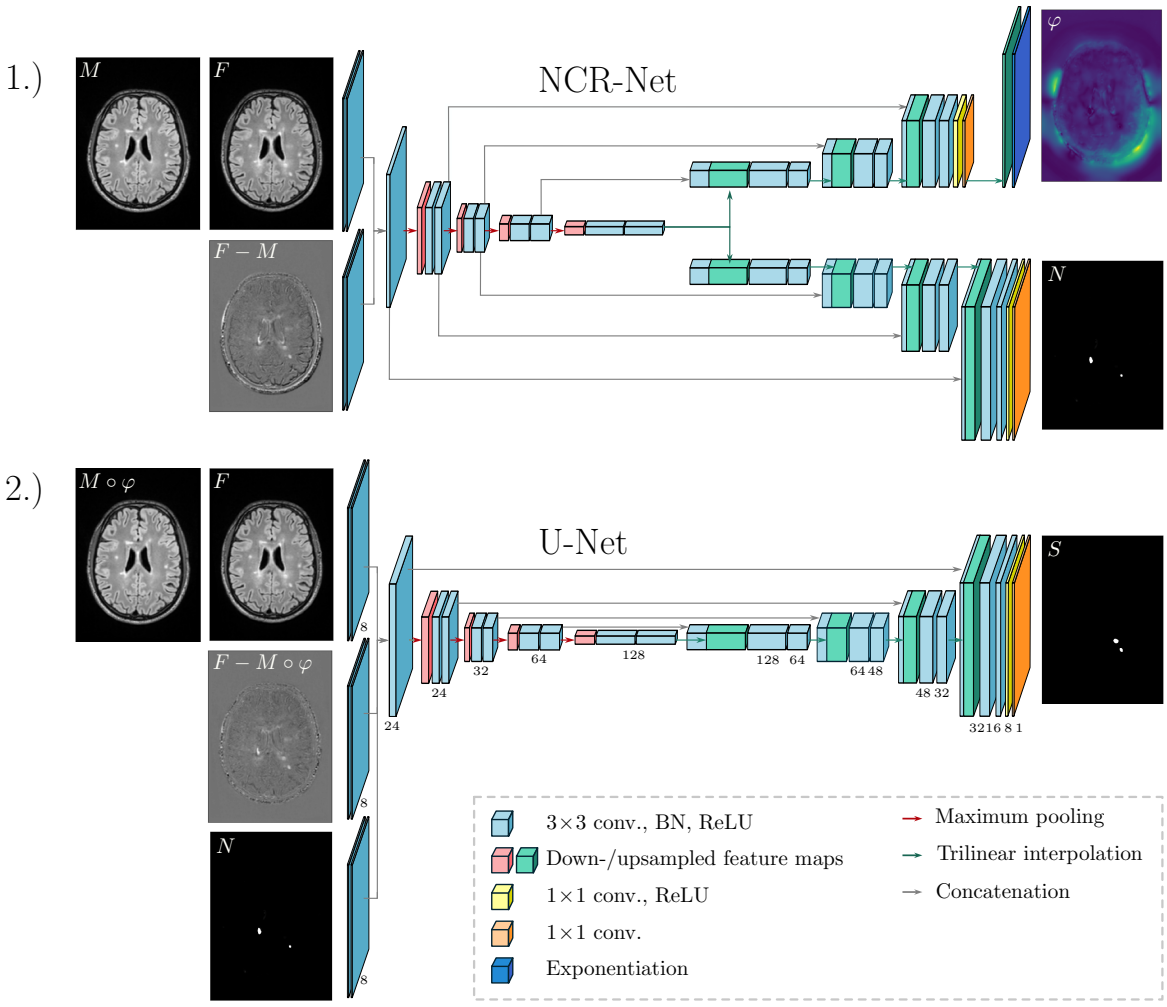


Figure A2: The proposed pipeline to use NCR-Net for new MS lesion detection and segmentation. First, NCR-Net is trained in a fully unsupervised manner. The registration results and non-correspondence maps are then used as inputs for a U-Net which is trained supervised. For NCR-Net, the outputs on the lower resolution levels have been omitted for better visualization. The complete architecture, including also the numbers of feature maps, can be seen in Fig. 4.1. For the U-Net, all implementation details are shown here, with the numbers below feature maps indicating the numbers of feature maps of the respective convolutional layers.

A Appendix

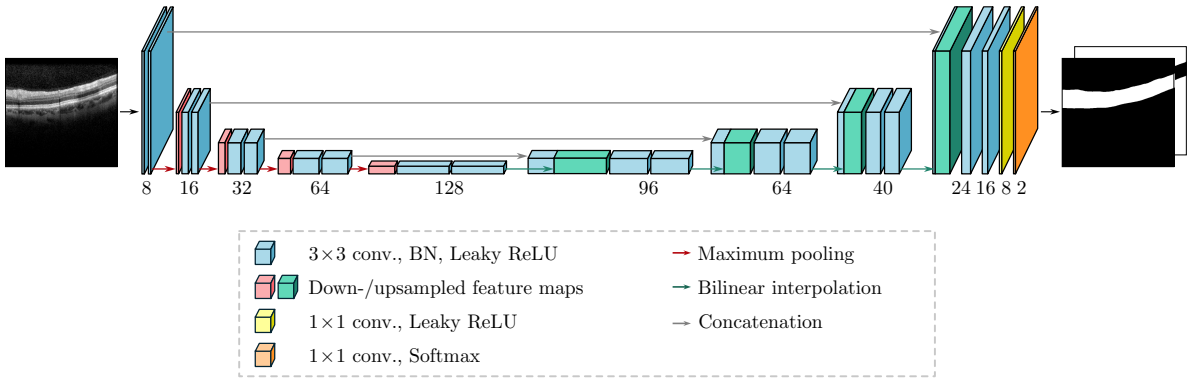


Figure A3: U-Net used for retina segmentation in Chapter 3.4.

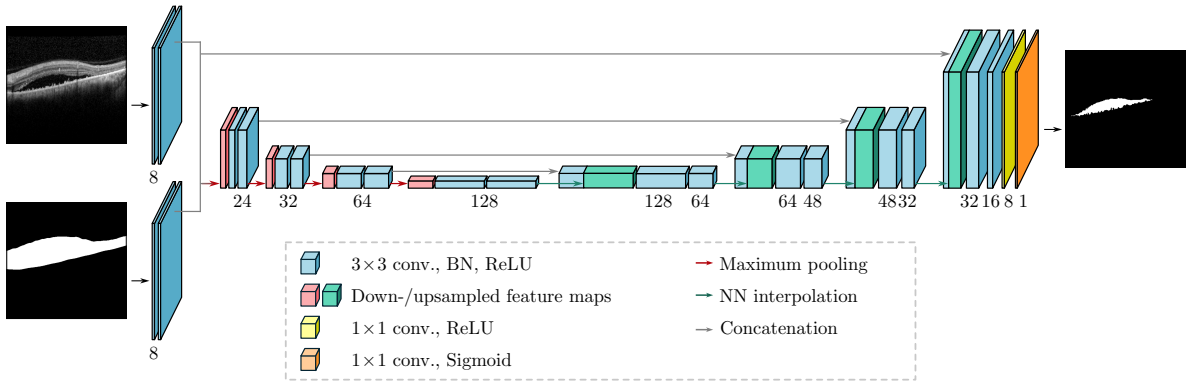


Figure A4: U-Net used in Sec. 6.4.3 for the segmentation of subretinal fluid. The generated segmentations serve to post-process the anomaly maps generated by FluidRegNet to refine them for PEZ pathology segmentation.

A3 Additional FluidRegNet Anomaly Detection Results

In this section, additional anomaly detection results are shown supplementing Chapter 6.4.2. Each image shows anomaly detection results for left and right eyes of one CSCR patient (left eye on the left, right eye on the right). For each eye, the central OCT-B scans of each visit and the corresponding manual segmentations are shown. The detection of anomalies for FRN-Ano, NSA and f-AnoGAN is shown from top to bottom, first for the central B-scans and then in the en-face plane. The color coding is scaled according to the maximum amount of fluid per A-scan observed per eye in all examinations. In the bottom row, the real and estimated fluid volumes are depicted.

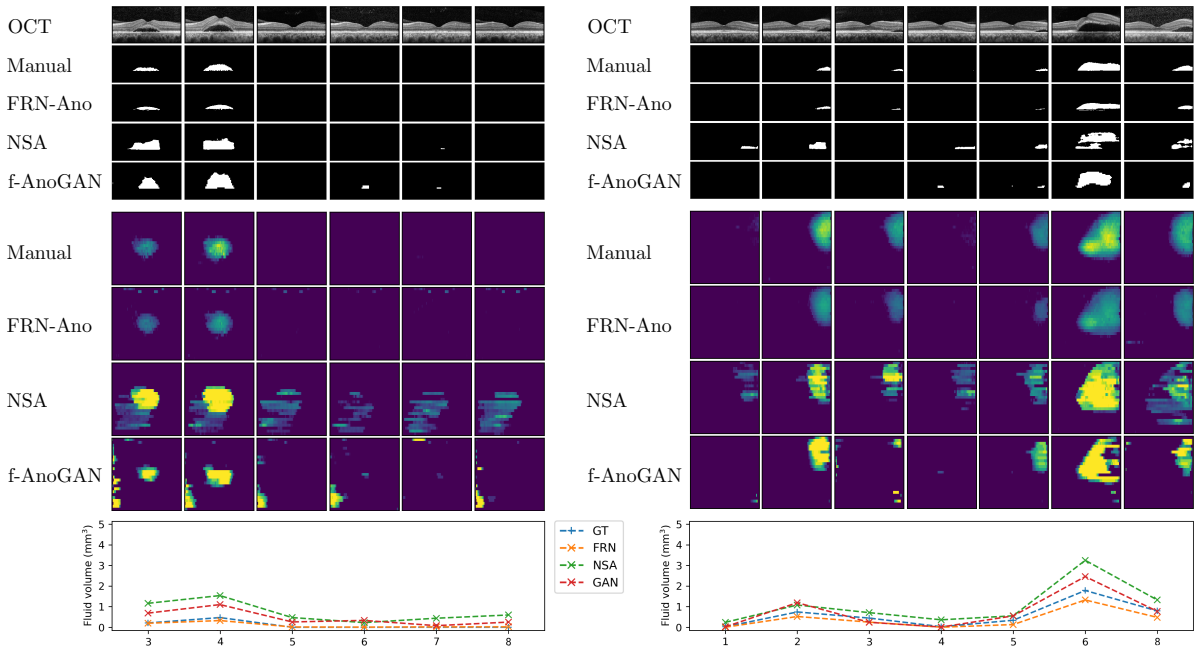


Figure A5: Anomaly detection results for CSCR patient 8. For some time points, only one eye was monitored or images had to be excluded due to low image quality. Same numbers in the bottom plots indicate same visits.

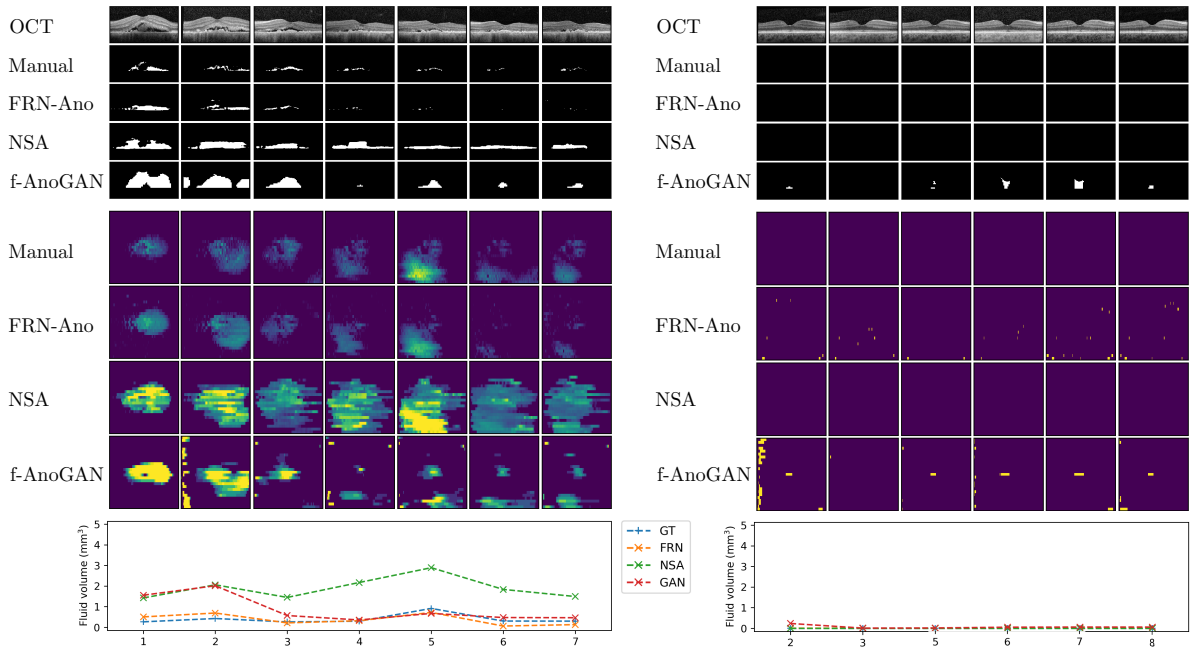


Figure A6: Anomaly detection results for CSCR patient 35. For some time points, only one eye was monitored or images had to be excluded due to low image quality. Same numbers in the bottom plots indicate same visits.

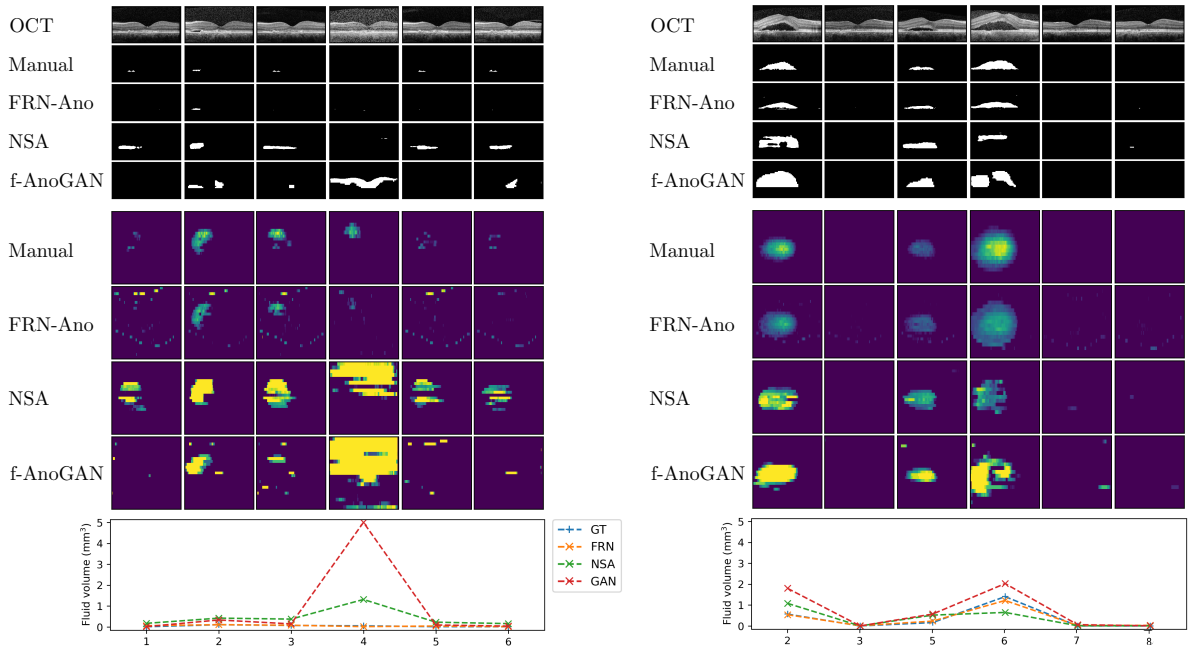


Figure A7: Anomaly detection results for CSCR patient 116. While NSA and f-AnoGAN are both susceptible to the high noise at visit 4 (left eye), FRN-Ano proves to be stable against this form of imaging artefact.

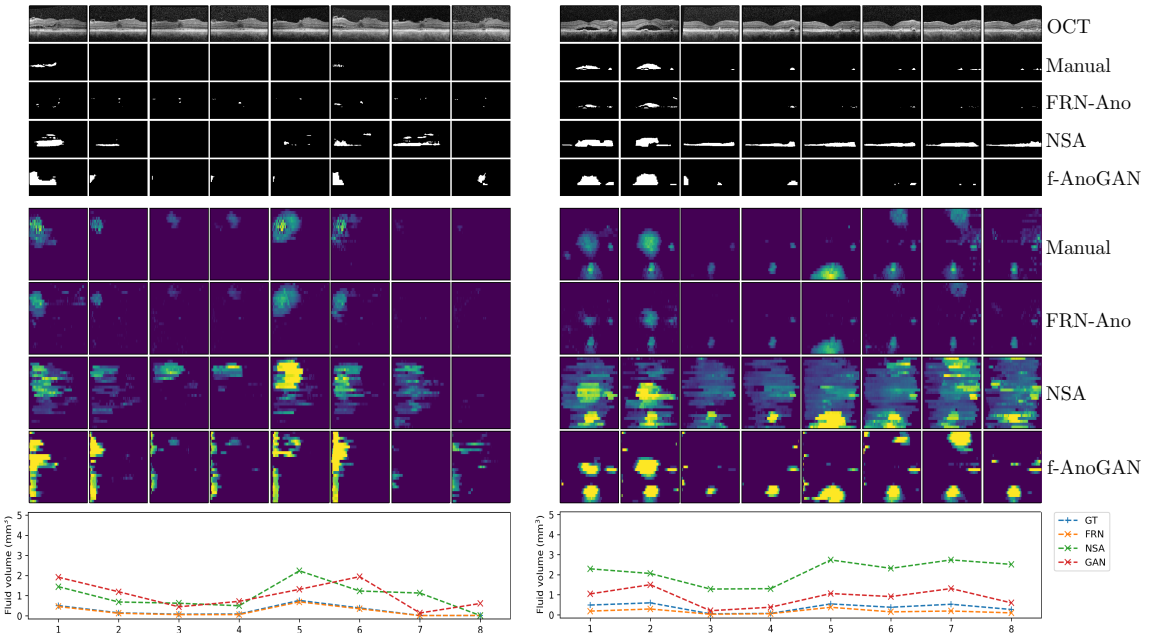


Figure A8: Anomaly detection results for CSCR patient 106. f-AnoGAN tends to produce lots of false positives if the retina does not cover the entire width of the B-scans (black border on the left side of the B-scans of the left eye).

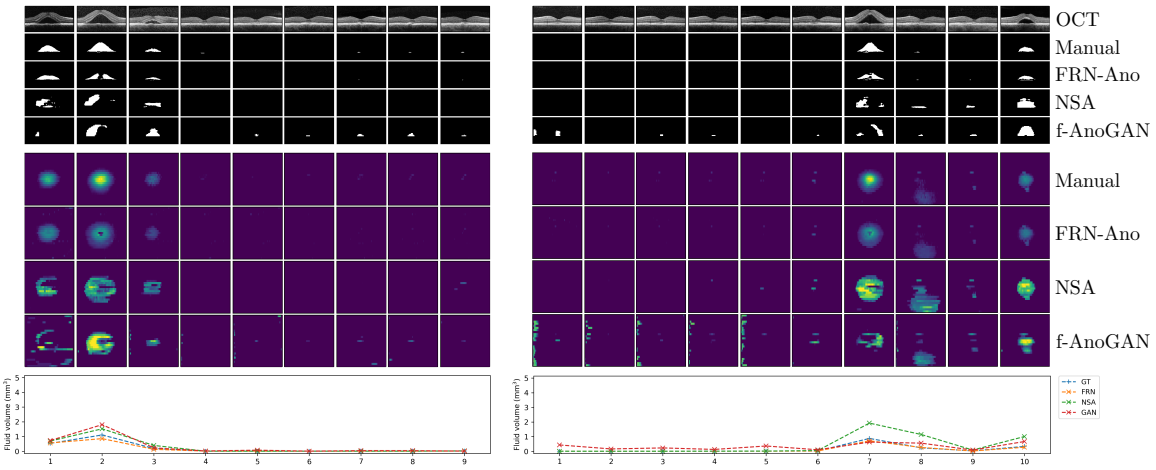


Figure A9: Anomaly detection results for CSCR patient 126. Very large fluids still cause problems for FRN-Ano, as the strong deformation cannot yet be completely covered by the framework. The other methods also have difficulties with such extreme cases (see visit 2 of the left eye and visit 7 of the right eye).

Own Publications

Journal articles as first author

- **J. Andresen**, H. Uzunova, J. Ehrhardt, T. Kepp and H. Handels: "Image Registration and Appearance Adaptation in Non-Correspondent Image Regions for New MS Lesions Detection" *Frontiers in Neuroscience* 16, Frontiers Media SA, Lausanne, Switzerland, 2022

- **J. Andresen**, T. Kepp, J. Ehrhardt, C. von der Burchard, J. Roider and H. Handels: "Deep Learning-Based Simultaneous Registration and Unsupervised Non-Correspondence Segmentation of Medical Images with Pathologies" *International Journal of Computer Assisted Radiology and Surgery* 17(4), Springer International Publishing, 2022

Journal articles as co-author

- M. Santarossa, A. Tatli, C. von der Burchard, **J. Andresen**, J. Roider, H. Handels and R. Koch: "Chronological Registration of OCT and Autofluorescence Findings in CSCR: Two Distinct Patterns in Disease Course " *Diagnostics* 12(8), MDPI, Basel, Switzerland, 2022

Conference contributions as first author

- T. Kepp, **J. Andresen**, F. Falta and H. Handels: "Bridging Gaps in Retinal Imaging: Fusing OCT and SLO Information with Implicit Neural Representations for Improved Interpolation and Segmentation" in: *Bildverarbeitung für die Medizin 2025* (editors: C. Palm, K. Breininger, T. Deserno, H. Handels, A. Maier, K. Maier-Hein, T. Tolxdorff), Springer Fachmedien Wiesbaden, Regensburg, Germany, 2025

- **J. Andresen**, J. Ehrhardt, C. von der Burchard, A. Tatli, J. Roider, H. Handels and H. Uzunova: "FluidRegNet: Longitudinal Registration of Retinal OCT Images with New Pathological Fluids" in: *Proceedings of The 7th International Conference on Medical Imaging with Deep Learning (MIDL)* volume 250 of *Proceedings of Machine Learning Research* (editors: N. Burgos, C. Petitjean, M. Vakalopoulou, S. Christodoulidis, P. Coupe, H. Delingette, C. Lartizien and D. Mateus), Paris, France, 2024

- **J. Andresen**, T. Kepp, J. Ehrhardt, C. von der Burchard, J. Roider and H. Handels: "Unsupervised Non-correspondence Detection in Medical Images Using an Image Registration Convolutional Neural Network" in: *International Workshop on Biomedical Image Registration (WBIR)* (editors: A. Hering, J. Schnabel, M. Zhang, E. Ferrante, M. Heinrich and D. Rueckert), Springer International Publishing, Munich, Germany, 2022

- **J. Andresen**, T. Kepp, M. Wang-Evers, J. Ehrhardt, D. Manstein and H. Handels: "Unsupervised Segmentation of Wounds in Optical Coherence Tomography Images Using Invariant Information Clustering" in: *Bildverarbeitung für die Medizin 2022* (editors: K. Maier-Hein, T. Deserno, H. Handels, A. Maier, C. Palm and T. Tolxdorff), Springer Fachmedien Wiesbaden, Heidelberg, Germany, 2022

- **J. Andresen**, H. Uzunova, J. Ehrhardt and H. Handels: "New Multiple Sclerosis Lesion Detection with Convolutional Neural Registration Networks" in: *MSSEG-2 Challenge Proceedings: Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure. MICCAI 2021 - 24th International Conference on Medical Image Computing and Computer Assisted Intervention* (editors: O. Commowick, F. Cervenansky, F. Cotton and M. Dojat), Strasbourg, France, 2021

Conference contributions as co-author

- C. von Dresky, C. von der Burchard, **J. Andresen**, M. Seibel, M. Rowedder, T. Kepp, J. Roider and H. Handels: "Visual Acuity Assessment from Optical Coherence Tomography Images Using the Foundation Model RETFound" in: *Medical Imaging 2025: Computer-Aided Diagnosis* 13407 (editors: S. M. Astley and A. Wismüller), International Society for Optics and Photonics, SPIE, San Diego, USA, 2025

- M. Seibel, M. Rowedder, **J. Andresen**, R. Neffin, T. Neumann, H. Sudkamp, H. Handels and T. Kepp: "Enhancing Retinal SELFF-OCT Image Quality: A Deep Learning-Based Pipeline" in: *Medical Imaging 2025: Computer-Aided Diagnosis* 13407 (editors: S. M. Astley and A. Wismüller), International Society for Optics and Photonics, SPIE, San Diego, USA, 2025

- T. Kepp, **J. Andresen**, C. von der Burchard, J. Roider, G. Hüttmann and H. Handels: "Shape-Based Segmentation of Retinal Layers and Fluids in OCT Image Data" in: *Medical Imaging 2023: Computer-Aided Diagnosis* 12465 (editors: K. M. Iftekharuddin and W. Chen), International Society for Optics and Photonics, SPIE, San Diego, USA, 2023

- T. Kepp, **J. Andresen**, H. Sudkamp, C. von der Burchard, J. Roider, G. Hüttmann, J. Ehrhardt and H. Handels: "Epistemic and Aleatoric Uncertainty Estimation for

PED Segmentation in Home OCT Images" in: *Bildverarbeitung für die Medizin 2022* (editors: K. Maier-Hein, T. Deserno, H. Handels, A. Maier, C. Palm, T. Tolxdorff), Springer Fachmedien Wiesbaden, Heidelberg, Germany, 2022

Abstracts

- C. von der Burchard, M. Santarossa, A. Scharf, T. Kepp, J. Roider, H. Handels and **J. Andresen**: "Zusammenhang zwischen Sensibilitätsverlust in der Mikroperimetrie und subretinaler Flüssigkeit sowie Photorezeptor-Inhomogenitäten bei Patienten mit Chorioretinopathia centralis serosa (CCS)" *Die Ophthalmologie – Abstractband DOG 2025*, Springer Nature, Berlin, Germany, 2025

- M. Santarossa, C. von der Burchard, A. Scharf, A. Tatli, **J. Andresen**, H. Handels and J. Roider: "Chronology of Hyperfluorescence and Subretinal Fluid Indicates Visual Acuity Development in Central Serous Chorioretinopathy" *Investigative Ophthalmology & Visual Science* 66, The Association for Research in Vision and Ophthalmology, Salt Lake City, USA, 2025

- Y. Elser, S. Engelson, J. Ehrhardt, S. Schierholz, **J. Andresen**, D. Drömann, T. Keck, J. Barkhausen, H. Handels and M. Sieren: "KI-gestützte Erkennung von metastasierenden Lymphknoten des Mediastinums und automatische Bestimmung des N-status bei Lungenkarzinomen" *RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren* 197(1), Georg Thieme Verlag KG, Stuttgart, Germany, 2025

- C. von der Burchard, A. Scharf, M. Santarossa, A. Tatli, **J. Andresen**, R. Koch, H. Handels, T. Kepp and J. Roider: "AI-Based Photoreceptor Length Segmentation and Influence on Disease Progression and Visual Acuity Changes in Central Serous Chorioretinopathy (CSCR)" *Investigative Ophthalmology & Visual Science* 65(7), The Association for Research in Vision and Ophthalmology, Seattle, USA, 2024

- A. Scharf, C. von der Burchard, A. Tatli, M. Santarossa, **J. Andresen**, R. Koch, H. Handels, T. Kepp and J. Roider: "Linking AI-Based Biomarker Analysis to Visual Acuity Changes in Central Serous Chorioretinopathy (CSCR)" *Investigative Ophthalmology & Visual Science* 65(7), The Association for Research in Vision and Ophthalmology, Seattle, USA, 2024

- T. Kepp, **J. Andresen**, C. von der Burchard, J. Roider, G. Hüttmann and H. Handels: "Shape-Based Segmentation of Retinal Layers and Fluids in OCT Image Data" in *Bildverarbeitung für die Medizin 2023* (editors: T. Deserno, H. Handels, A. Maier, K. Maier-Hein, C. Palm, T. Tolxdorff), Springer Fachmedien Wiesbaden, Braunschweig, Germany, 2023

- **J. Andresen**, T. Kepp, J. Ehrhardt, C. von der Burchard, J. Roider and H. Handels: "CNN-Based Joint Non-Correspondence Detection and Registration of Retinal Optical Coherence Tomography Images" in: *International Journal of Computer Assisted Radiology and Surgery – CARS 2021: Computer Assisted Radiology and Surgery Proceedings of the 35th International Congress and Exhibition*, Munich, Germany, 2021

Awards

- Third prize at the Best Paper Award, Bildverarbeitung für die Medizin (BVM) 2025
- Runner-up for the Audience Prize (Best Presentation), Workshop on Biomedical Image Registration (WBIR) 2022