UNIVERSITÄT ZU LÜBECK

From the Institute of Mathematics and Image Computing
of the University of Lübeck
Director: Prof. Dr. rer. nat. Jan Modersitzki

SqN: A Novel Distance Measure for Multiple Images and Applications

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

from the Department of Computer Sciences

Submitted by

Kai Brehmer
from Preetz

Lübeck, 2020

First referee: Prof. Dr. Jan Modersitzki
Second referee: Prof. Dr. Heinz Handels
Date of oral examination: March 24, 2021
Approved for printing: Lübeck, March 25, 2021

Abstract

In this thesis, a framework for measuring the similarity between several images in the form of a distance measure, namely SqN, is presented and discussed. The design of SqN naturally fits into the standard formulation of variational image registration and generalizes the concept of a distance measure to more than two images.
Quantifying the similarity of images is an important task in modern image processing. In the literature, a variety of approaches to measure the similarity of two images are discussed. Applications such as dynamic imaging or histological serial sectioning require the registration of a large number of images. For these cases an extension to a method for similarity measurement of multiple images is a logical next step. The literature is still rather modest on this subject.
Especially for the registration of multiple images, a novel distance measure SqN was developed in this work. The basic idea is to evaluate a feature matrix by its singular values to align the images. For this reason, SqN is based on the Schatten-q-Norms, which are computed using the Singular Value Decomposition (SVD). We derive the distance measure SqN by mainly using three different geometric ideas to achieve image alignment. These ideas include rank minimization of the feature matrix, volume minimization, and correlation maximization. This results in several special cases, including a close relationship to the Normalized Gradient Fields distance measure when normalized intensity gradients are used as features within the feature matrix. Finally, a correlation-based approach generalizes the introduced ideas and provides a general framework for distance measurement with SqN. This framework comprises three main points, which we verify by practical application to medical image registration problems.
First, SqN can register more than two images with comparable quality as is possible with standard two-image based methods such as the Normalized Gradient Fields distance measure or the sum of squared differences. Secondly, registration of comparable quality can be achieved in less time than with the standard methods mentioned above, which are used in a pairwise registration scenario. Furthermore, the order of the images has no influence on the registration results due to the use of the SVD. Finally, the novel distance measure SqN can be applied to real-world medical image registration problems in practice. Our numerical experiments have verified all key points so that SqN is a promising distance measure for actual application to solve real medical image registration problems.

Zusammenfassung

In dieser Arbeit wird ein Rahmen für die Ähnlichkeitsmessung zwischen mehreren Bildern in Form eines Distanzmaßes, namentlich SqN, vorgestellt und diskutiert. Das Design von SqN gliedert sich natürlicherweise in die Standardformulierung der variationellen Bildregistrierung ein und verallgemeinert so das Konzept eines Distanzmaßes für mehr als zwei Bilder.
Die Quantifizierung der Ähnlichkeit von Bildern ist eine wichtige Aufgabe in der modernen Bildverarbeitung. In der Literatur werden eine Vielzahl von Ansätzen zur Ähnlichkeitsmessung zweier Bilder diskutiert. Anwendungen wie z.B. die dynamische Bildgebung oder histologische Serienschnitte erfordern die Registrierung einer teilweise großen Anzahl von Bildern. Für diese Fälle ist eine Erweiterung auf eine Methode zur Ähnlichkeitsmessung mehrerer Bilder ein logischer nächster Schritt. Hierfür ist die Literatur noch vergleichsweise überschaubar.
Speziell für die Registrierung mehrerer Bilder wurde in dieser Arbeit ein neues Distanzmaß SqN entwickelt. Grundlegende Idee ist die Bewertung einer Feature-Matrix über ihre Singulärwerte, um die Bilder aneinander auszurichten. Aus diesem Grund basiert SqN auf den Schatten-q-Normen, die mit Hilfe der Singulärwertzerlegung (SVD) berechnet werden. Wir leiten das Distanzmaß SqN her, indem wir hauptsächlich drei verschiedene geometrische Ideen verwenden, um die Ausrichtung der Bilder zu erzielen. Diese Ideen umfassen die Rangminimierung der Feature-Matrix, eine Volumenminimierung sowie eine Korrelationsmaximierung. Daraus ergeben sich verschiedene Sonderfälle, wie unter anderem eine enge Beziehung zum Normalized Gradient Fields Distanzmaß, wenn normalisierte Intensitätsgradienten als Features innerhalb der Feature-Matrix verwendet werden. Letztlich verallgemeinert ein korrelationsbasierter Ansatz die eingeführten Ideen und bildet einen allgemeinen Rahmen für die Distanzmessung mit SqN. Dieser Rahmen umfasst drei Hauptpunkte, die wir durch praktische Anwendung auf medizinische Bildregistrierungsprobleme verifizieren.
Erstens ist SqN in der Lage, mehr als zwei Bilder in vergleichbarer Qualität zu registrieren, wie dies mit standardmäßigen zwei-Bild basierten Methoden wie dem Normalized Gradient Fields Distanzmaß oder der Summe der quadrierten Differenzen möglich ist. Zweitens kann die Registrierung in vergleichbarer Qualität in kürzerer Zeit erreicht werden als mit den erwähnten Standardmethoden, welche in einem paarweisen Registrierungsszenario verwendet werden. Außerdem hat die Reihenfolge der Bilder aufgrund der Verwendung der SVD keinen Einfluss auf die Registrierungsergebnisse. Schließlich kann das neuartige Distanzmaß SqN auf reale medizinische Bildregistrierungsprobleme in der Praxis angewendet werden. Alle Schlüsselpunkte konnten mithilfe der numerischen Experimente nachgewiesen werden, sodass SqN ein vielversprechendes Distanzmaß für die tatsächliche Anwendung zur Lösung realer Bildregistrierungsprobleme ist.

Acknowledgements

I want to thank everyone who supported me during my time at the MIC.
First and foremost, I'd like to thank my supervisor Jan for giving me the opportunity of writing this thesis. I am grateful for his guidance and advice, from which I learned a lot, both scientifically and personally.
Of course, I would also like to thank all the colleagues who gave me a fantastic time at the Institute. Professional as well as private conversations and activities were so much fun with you.
I also thank all the colleagues involved in the MED4D project from Münster, Bochum, and Erlangen for many fruitful discussions and a great time at all our meetings.
Furthermore, I like to thank Prof. Prestin for chairing and Prof. Handels for taking the effort of being a referee.
I am infinitely grateful to my family, who always supported me. Without their unconditional support, I may not have written this thesis.

My special thanks go to

my family	Christina, Gabi, Manni, Marc and Jasmin
my friends and colleagues	Alessa, Alex, Annkristin, Benjamin, Bernhard, Caterina, Constantin, Dana, Daniel, Danielle, Frederic, Hari, Herbert, Jan L., Jan M., Jan R., Janine, Johannes, Judith, Kerstin, Lars K., Lars R., Marc, Marian, Mark, Maximilian, Nadine, Nick, Nils, Roland, Sonja, Stefan, Stepahnie, Sven, Thomas P., Thomas V., Yury

CONTENTS

LIST OF FIGURES

1 Introduction to Image Registration of Multiple Images

Image registration is an inverse problem that is typically phrased as an optimization problem, where the objective function is based on a regularizer and a data fidelity term [32], [99]. In summary, image registration is about the comparison of images of different scenes, time points, or modalities. In this thesis, we suggest a novel data fidelity term called SqN, which is particularly suited for the registration of multiple images. This novel data fidelity term naturally fits into the class of distance measures used for variational image registration. In this first introduction chapter, we motivate the use of SqN for the registration of multiple images over standard two-image based measures. For this purpose, we briefly introduce the medical applications we use to demonstrate the image registration potential of SqN. Furthermore, we embed the new distance measure into existing, related literature. This chapter is concluded by an overview of the contributions and the outline of this work.

1.1 Registration of Multiple Images

Image registration is an essential component of image processing and is studied extensively [81], [90], [128]. Especially for the alignment of two images, image registration is a well-studied problem, see e.g. [81], [90], [99], [100], [128], [150]. Typically, the images show the same object or scenery, where one of the images usually is fixed. Thus, the moving image is to be aligned to the fixed image. Following Modersitzki [100], the registration problem is about

> "[finding] a reasonable transformation such that a transformed version of a template image is similar to a reference image."

This general definition allows for more than two images and is not restricted to a pixel-basis for matching. Because of its generality, we use this definition as a basis for this work.
An essential question of the variational construct is, what similarity of images is or how it is measured. The choice of a suitable distance measure is vital for the registration problem because it directly affects the alignment of the images [55]. A wide variety of distance measures for two images have been proposed and explored to the fullest extent possible; see, e.g., [21], [47], [143]. More or less, every measure known from the relevant literature uses the image's gray values for either direct comparison or to compute certain features that can be compared. The idea is to have a data fidelity term that can measure the similarity or dissimilarity of images or its features and fulfills specific properties that are needed for optimization.

According to [115], standard measures can be hierarchically organized concerning their assumed relationship between the images. These categories include – but are not limited to – functional relationships, affine relationships, as also statistical relationships. Also, distance measures are categorized as monomodal or multimodal. Concerning the latter distinction, some familiar examples are the sum of squared differences (SSD) or the normalized cross-correlation (NCC) on the monomodal side. These have proven to be very well-suited in a wide variety of applications, in which only the same modality is used. Medical applications comprise follow-up examination [112] or histological tissue reconstruction [86], [135], to name a few. On the multimodal side, there are measures like mutual information (MI) or normalized gradient fields (NGF). As these are not only based on the intensity information of the image data, they are almost always used for multimodal applications [80], [142], [144]. However, the use-cases of these measures are not limited to multimodal cases. They have also proven to be very useful for monomodal cases as well as for the registration of dynamic imaging [76], [86], [112]. Mostly, dynamic imaging comprises multiple timeframes of the same modality while consisting of some change over time. In medical dynamic imaging, most of the time, this is due to contrast agent uptake.

The standard measures are designed to compare a reference and a template image. In this thesis, we formulate the novel singular value based distance measure called SqN following the problem definition and nomenclature of Modersitzki [100]. The acronym SqN originates from the used Schatten-q-Norms. The design of SqN naturally fits into the standard variational image registration problem formulation of Modersitzki [99], [100], and generalizes the concept of distance measurement. For this purpose, we take advantage of the generality of the image registration problem formulation by Modersitzki [100]. The formulation does neither specify if the reference and template images need to be explicit or implicit nor specifies if any of the images is fixed. With SqN we take advantage of these freedoms and generalize the distance measure concept to more than two images. For SqN we don't need to choose a specific reference image as all images are intended to be transformed. Furthermore, the formulation maintains the opportunity to choose for various image features for distance measurement. With such a generalization of distance measurement, we can take advantage of image-relationships over more than just two neighboring images. The versatile design allows to solve medical image registration problems comprising more than two images, overcoming the need to register in a pairwise manner.

1.2 Medical Applications

In the field of medical image processing, many applications cover multiple images. Among the vast field of medical applications we choose two exciting applications to demonstrate the potential of SqNẆe show results for motion correction in dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) as well as for histological serial sectioning.

Motion Correction in DCE-MRI

Dynamic contrast-enhanced magnetic resonance imaging is an evolving factor for examination of kidney malfunction [62], [63]. It is an in vivo imaging technique for the measurement of several clinically relevant parameters of the kidneys function. The potential of this imaging method is to replace methods for which patients need to be exposed to radiation as in, e.g., nuclear imaging methods like Positron Emission Tomography (PET) [63]. Still, without reproducible results and high accuracy of concentration measurement, the clinical use is not reasonable [97]. Figure 1 illustrates a promising registration result of a DCE-MRI data set of a human kidney for 2D images over time.

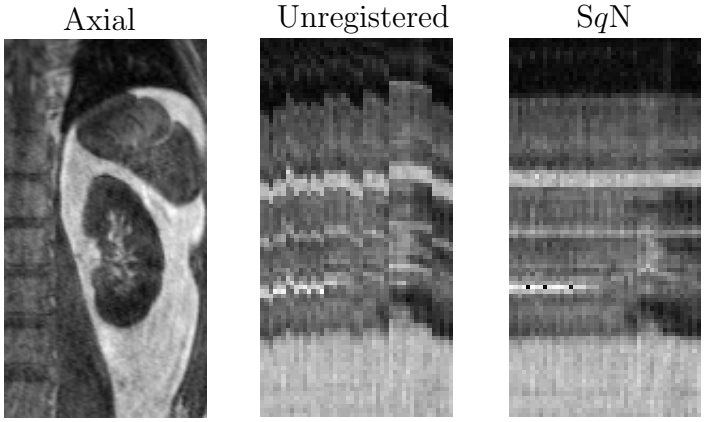


Figure 1: Visualization of a part of 4D (x,y,z,t) DCE-MRI data. A representative axial slice (x,y) at a representative time t is shown on the left. The profile (x,t) is shown for unregistered (center) and for registered data using SqN (right). Details are outlined in Sec 4. Data courtesy of Jarle Rørvik, Haukeland University Hospital, Bergen, Norway.

DCE-MRI sequences are four-dimensional datasets. The data is acquired as a time series in three dimensions to capture the uptake of the non-radioactive, Gadolinium-based contrast agent [63]. For an accurate analysis, however, an exact point-to-point correspondence is crucial. In order to achieve such correspondences,

motion correction is inevitable [62], [63], [120]. Motion correction of DCE-MRI certainly is a challenging task as there are intensity changes due to contrast agent uptake, noise, and various movements of tissue. Our results in Sec. 4 show that SqN is well suited to correct for motion in a multi-modal scenario, which is why we use DCE-MRI sequences to demonstrate the potential of SqN. We show in Sec. 4, that SqN is able to correct for motion while maintaining the tracer dynamics. Nevertheless, once achieved a proper motion correction, DCE-MRI is superior to different diagnostics due to its high sensitivity [82]. When corrected for motion, DCE-MRI has the potential to become an essential tool in clinical use for kidney malfunction examination [62].
The current trend in image processing is clearly towards the use of neural networks [36], [59], [67], [78]. Not only are neural network based methods for image processing very advanced by now, concerning computation times they often outperform conventional methods [36], [78]. Recently proposed methods for image registration have also shown that the image registration results are of comparable quality; see e.g. [57]–[59], [78]. All critical mathematical and numerical methods can be used, like e.g. variational energy-functionals comprising distance measure and regularization and even multi-level methods. Eventhough the learning process is slow and may take up to several days, the evaluation is super-fast [36]. Once the network is trained, a full 3D image registration is done within fractions of a second, which is a clear advantage, compared to conventional methods. Registration of multiple images using neural networks may be interesting for time-critical applications or for processing large data sets in less time than it would be possible with conventional registration methods. Because of that, we show numerical results for a proof-of-concept experiment on motion-correction using SqN for registration of multiple images within a recently proposed approach by Kuckertz et al. [77] at the end of Section 4.

Registration and Fusion of Histological Serial Sectioning

The registration of histological serial sectioning is about the reconstruction of three-dimensional tissue and the possibility of improved examination across sections. Following [85], [86], the digital pathology extends the examination possibilities of traditional histology. Three-dimensional reconstruction is needed for the virtual examination of tissue across slices. In order to reconstruct a series of histological sections to its original volume, image registration is needed [85]. Furthermore, a series of histological sections is likely to be examined in various stainings and modalities [83], [114], [125], [135]. For a detailed analysis of biopsy sections, a fusion of three-dimensional volumes of different modalities is essential to correlate molecular and microscopic levels of tissue [83], [114], [125], [135].
The potential of SqN lies in the multi-modality. Because of that, we choose the

reconstruction and fusion of histological serial sectioning. In Sec. 4 we show the potential of SqN by reconstruction and fusion of multiple modalities. Just aligning pairs of images may lead to cumulated errors like the banana-effect or z-drift [34], [123], [130], [131], [138]. We aditionally show, that the use of SqN for image registration practically avoids such errors. Figure 2 illustrates a promising profile view of a registration of histological serial sectioning of a monkey brain using our novel distance measure SqN.

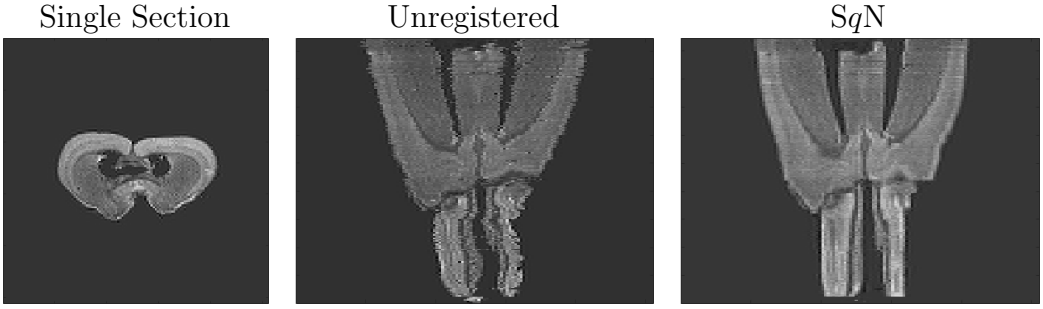


Figure 2: Visualization of a stack of histological serial sectioning (x,y,z). Illustrated on the left is a representative histological section (x,y) of a monkey brain. The profile (x,z) is shown for unregistered data (center) and registered data using SqN (right). Details on the numerical results are outlined in Sec 4. Data courtesy of O. Schmitt, University of Rostock, Germany; see [123] for experimental details.

All measures mentioned above are defined for registration of two images, but applications like dynamic imaging or serial sectioning involve the analysis of sequences of many images. The single images of these applications have an intrinsic relation that needs special treatment. A two-image distance measure only captures the relation between two single frames at a time or may be used for comparison of single frames to an average image. Especially in dynamic imaging, intensity fluctuations occur due to contrast agent uptake. Sometimes, data is also corrupted, or there are gaps between particular time windows [62], [63], [85], [130]. In histological serial sectioning, there are, for example, drift-away effects during 3D tissue reconstruction due to an accumulation of small errors resulting in an odd, unwanted shape; e.g., z-drift or banana-effects [130], [138]. All these problems benefit from a distance measurement over more than two images, as can be seen in Section 4. Intensity fluctuations can be compensated, and the 3D volume for reconstruction cannot deform in odd, unexpected ways. Furthermore, most approaches select a reference frame for the use of two-image distances; see, e.g., [85], [101], [142]. This determines the ordering of the images and favors accumulation errors like a z-drift or unwanted changes of intensities in dynamic imaging.
In Section 4, we demonstrate that our novel data fidelity term SqN can resolve

the mentioned issues due to the use of intrinsic structure with help of singular values. By design, information is spread over all considered images in an automatic, global way. This unique design allows for registration without choosing a specific reference image. Moreover, it allows registration without any specific order of images. It is possible to register just two images or even a large number of images. Due to its use of singular values of specific, deterministic features, special cases can be built. The well-studied normalized gradient fields (NGF) [47] distance measure is amongst these special cases. Therefore, it can also be understood as an extension or generalization of two-image distance measures like NGF. This allows flexible applicability and straightforward geometrical interpretability due to the connections to well-studied distance measures.

1.3 Related Literature

There is a wide variety of different approaches and also a sheer amount of applications of image registration. Consequently, we focus on an overview of the literature that is related to our work. To find general access to the topic of image registration, we refer to [37], [38], [99], [100]. A non-representative selection of survey articles, that reflect the evolution of image registration over the last decades, are [11], [61], [90], [91], [95], [104], [122], [128], [150].

Related Frameworks and Data Fidelity Terms

Since this work deals with our novel distance measure SqN, we give an overview of related registration frameworks and distance measures, respectively. A pinnacle in the development of image registration for images from different sceneries or modalities is the discovery of distance measures capable of comparing images not only by their pixel-wise intensity value equality, like, for example, the sum of squared differences (SSD) does. An example is mutual information (MI) [21], [143]. Mutual information originates in information theory. It minimizes the entropy of the joint distribution of the sequence of intensity values of both images. Another example is the normalized gradient fields (NGF) distance measure [47], [100]. This approach is based on intensity gradients to align corresponding edges. This is achieved by aligning the directions of locally normalized intensity gradients. Both MI and NGF are measures designed for multi-modal image registration. Our novel distance measure SqN is based on ideas similar to those of NGF, but generalizes the capability of comparing two images. Furthermore it is straightforward to interpret as it is deterministic and based on geometric ideas comparable to the idea of deterministic measures like NGF or the normalized cross correlation (NCC).
Measures like SSD, MI or NGF and its variants are designed to register two images, but many applications comprise more than two images. Time-dependent

data, like in dynamic imaging scenarios or several tissue slices in histological serial sectioning, are compelling examples amongst many. In a pairwise manner, such measures are also used for registration of multiple images in practice [63], [76], [85]. Nevertheless, they are accompanied by the necessity to select a reference image all other images are aligned to. The particular choice of a reference has been shown to produce inconsistent results, depending on the chosen reference [70], [98]. With SqN we overcome this necessity. Besides SqN, several developments comprise extensions of SSD, MI as well as NGF [4], [7]–[9], [70], [98], [111]. Many of these extensions have emerged from different fields of applications. An exemplary application is the generation of anatomical atlases, which is a very well-studied field. Many authors understand the process of finding anatomical atlases as an averaging over many different shapes. Shape analysis is distantly related to distance measures for registration of multiple images. The shape analysis of the human brain is a notable example in this context. Not only have landmark-based registration approaches been developed for an examination of the brain anatomy; see, e.g., [28], [39], [49], [110], [119]. Physical principles are also well-studied, e.g., with elasticity concepts from continuum mechanics [119], which are also closely related to elastic regularization techniques; see, e.g., [13], [109], [118]. Alongside this physical point of view, methods have evolved that use the well-known principal component analysis (PCA) [17], [28], [127], which includes both the analysis of geometric transformations [17] as well as using eigenanalysis for displacement computation [127]. Besides, the analysis of principal components is the examination of whole shape spaces using the correlation of specific shapes within these spaces as well as the use of entropy minimizing methods related to MI [18], [24], [72], [96], [132], which is mostly done on a statistical basis. Altogether, these methods have one crucial point in common. They need a distance measure capable of comparing multiple data sets and use the information across these. Summarizing, our novel distance measure combines ideas from well-studied fields, which makes it very versatile for application in different areas. Nevertheless, not only shape analysis and atlas generation is related to the construction of our novel distance measure for multiple images.

Application-driven Techniques for Image Registration

Application-driven approaches have evolved a high potential for the use of information across many data sets and in particular images. A recurring term in the context of multiple-image registration is *groupwise*; see, e.g., [23], [41], [42], [68]–[70], [73], [93], [106], [129], [139], [140], [147], [148], [151]. Note that the term *group* may not always refer to the common algebraic term. Sometimes it is derived from the literal groups of images that are registered. Although, some authors build a group structure using multiple images, which is a very exciting field of mathemat-

ical research. Many of the groupwise approaches are motivated by registration of medical image data, such as functional magnetic resonance imaging (fMRI), dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), dynamic sequences from computer tomography (CT) or even positron emission tomography (PET), to name a few. The recent groupwise approaches comprise the use of extensions of mutual information [129], [132], [151], correlation-based approaches [41], [42] and even principal component analysis-based dissimilarity measures [50], [68]–[70]. Especially noteworthy is, that many of the groupwise approaches have a very similar construction of a data matrix which is used for computing distance energies. In order to bring all images together into one mathematical structure, many authors make use of an array structure in which the images are organized in columns. Although it has different names, it basically is the same structure from which principal component analysis [43], [50], [68]–[70], singular value decomposition [111] and also correlation matrices can be derived [41], [42]. For SqN we use a similar structure but generalize it for a greater versatility.
The use of intrinsic information using such array-structures is mostly based on geometrical ideas, like on linear dependency of the images or image gradients, respectively. The hustle and bustle of the methods using neural networks and deep learning has become greater and greater within the last decade. These methods are also able to learn intrinsic structures of given data, even if this might be not understandable for humans. Many novel approaches have shown, that a combination of traditional measures and regularizers with neural networks are superior in specific tasks, when having enough data for learning and evaluation [57]–[59], [78]. Furthermore, groupwise approaches are developed for use with neural networks [19], [126]. Beyond medical imaging, applications of non-medical nature like e.g. stereo vision [89], optical flow and tracking for e.g. autonomous driving [12], [65], [108] or even satellite image registration [74] that use image registration, which may profit from an approach like SqN, are worth mentioning.

1.4 Contributions and Outline

This work introduces the novel distance measurement framework SqN for image registration of multiple images. First, we introduce the general image registration problem and a variational framework for solving it. We give a pertinent overview of state-of-the-art distance measures and regularizers for image registration of two images. This is accompanied by a comprehensive review of relevant optimization methods and numerical methods.
The central part of this work is the introduction of the distance measurement framework SqN for the registration of multiple images. This central part starts by extending state-of-the-art methods in a straightforward, sequential way and continues with a brief discussion of reasons to consider different methods. Subsequently,

we start the novelty part of this work by introducing the distance measurement framework with concomitant analysis of properties and numerical behavior. This comprises different, geometrically motivated approaches that we merge into one general framework. Here, we analyze the versatility and potential of the novel framework. Furthermore, brief error analysis and pertinent analysis of the numerical derivative follow.
With SqN we introduce a distance measurement framework that is capable of registering multiple images in a quality comparable to standard methods in less time. Furthermore, we deploy a versatility that allows application of SqN in medical real-world applications seamlessly. Finally, we demonstrate the potential of the framework and validate the theses. For this purpose, we present numerical results for medical image registration applications. We focus on the different, above introduced applications. We address these applications with our specially developed Python-based toolbox *Groupwise Registration Algorithms and Tools (GReAT)* as well as with an extension of the freely available MATLAB-based toolbox *Flexible Algorithms for Image Registration (FAIR)* [100].

2 Variational Image Registration for Two Images

For the formulation of the image registration problem, we follow Modersitzki [100], using a general, variational framework. This framework is built upon an optimization problem consisting of a data fidelity and a regularization term, which are combined are the objective function. The objective function is to be minimized, as it describes the energy of the problem. The distance measure is small when the images are similar. An optimization problem only based on a distance measure is ill-posed in the sense of Hadamard [48]. Small changes of the input data may lead to significant changes in the output data, the problem is not convex and thus the solution is not unique, and the deformation can be non-continuous [99]. This is resolved by using regularization [32]. In Section 2.3 and Section 2.4 we summarize different choices for both parts.

We choose the variational construct because of its generality. The versatile description of the registration problem as a combination of image similarity and regularity serves as a basis to fit our novel distance measurement approach into a well-studied standard approach. For the computation of a solution, we follow the discretize-then-optimize approach proposed by Haber and Modersitzki [45], as it is successfully used for different medical image registration applications [76], [85], [112]. The discretize-then-optimize approach is summarized in Section 2.2. We focus on using differentiable distances and regularizers, such that we can use standard optimization methods. For the computation of analytic derivatives, we discretize the used terms and use iterative optimization schemes like Gauß-Newton and L-BFGS; see Section 3.6 and Section 2.7.

In this section, we outline the variational framework following [100]. For this purpose, we summarize the mathematical preliminaries and define the image registration problem for two images, followed by an outline of discretization methods and a summary of optimization strategies. The chapter is mainly based on textbooks like [99], [100], [103] and references therein.

2.1 The Image Registration Problem

For the definition of the image registration problem (IRP), we first define images and transformations to have a proper notation. In this work, we understand images as mappings $\Omega \subset \mathbb{R}^d \to G$ with domain Ω and $d \in \mathbb{N}$ the dimension of the space. Typically, $d \in \{2, 3\}$ for spatial dimensions. Here, we also use $d = 4$ for an additional time dimension. The target domain G can represent binary- or gray-

values and can also be a tensor for more abstract constructs. As we mainly focus on the registration of gray-valued images, $G \subset \mathbb{R}$ is sufficient as a target domain. In Section 5, we briefly discuss the other cases.

Definition 1 (Image (cf. [99, p. 14]))**.** Let $d \in \mathbb{N}$. Then

$$\mathcal{I} : \Omega \subset \mathbb{R}^d \to \mathbb{R}$$

is called *image*, if

1. $\mathcal{I}$ is compactly supported,
2. $\mathcal{I}(x) < \infty$ for all $x \in \mathbb{R}^d$,
3. $\int_{\mathbb{R}^d} \mathcal{I}^k(x) < \infty$ for $k > 0$.

Ω is called *image domain*. We denote the set of all images by

$$\text{Img} := \{\mathcal{I} : \mathbb{R}^d \to \mathbb{R} \mid \mathcal{I} \text{ is an image}\}.$$

In this work, we focus on connected, rectangular domains for images. With this definition, we can evaluate an image at every point in the image domain. Knowing how an image is defined, we can define how to transform images in order to be able to compare them. This we can accomplish this with a transformation.

Definition 2 (Transformation (cf. [99, p. 77]))**.** Let $d \in \mathbb{N}$. A vector field

$$y : \mathbb{R}^d \to \mathbb{R}^d$$

is called *transformation*. For a given transformation y, the mapping $u : \mathbb{R}^d \to \mathbb{R}^d$ is called the displacement.

We follow [99] and split the transformation y into an identity part and the displacement u as

$$y(x) = x + u(x).$$

The deformations can be subdivided into parametric and non-parametric deformations. We use parametric transformations for pre-registration, as these transform each point of an image in the same way [99], [100]. In many cases, this does not allow for a point-to-point correspondence, as often non-linear deformations occur. This phenomenon we also explore in our medical applications in Section 4.

Parametric and Non-Parametric Deformations

Generally, a parametric approach restricts the number of parameters for determining of the deformations y to a finite number. A special case is to determine y by solving an affine linear equation system

$$y(x) \coloneqq Ax + b.$$

Here, a point $x \in \mathbb{R}^d$ with $d \in \mathbb{N}$ is affine linearly transformed by application of a linear transformation matrix $A \in \mathbb{R}^{d \times d}$ and a translation vector $b \in \mathbb{R}^d$. This model can be parameterized by the d^2 parameters determined by the entries of A and the d parameters determined by the entries of b, which gives overall $d^2 + d$ parameters. Therefore, there is a finite number of parameters to describe the transformation y. In order to generate physically explainable systems, A is required to fulfill $\det(A) \neq 0$, such that A is invertible. If A is not invertible, the singularities of A cause problems in the transformed image. The translation vector b contains real numbers to translate the coordinates of the points of the template image.
The parametric model allows for translation, rotation, scaling, and shearing in two and three spatial dimensions. This is especially useful to preserve point-to-point relations. A special group of these transformations are called rigid transformations. These are angle- and length-preserving, which means that $A^\top A = AA^\top = \mathrm{Id}$.

Non-parametric transformations have an unspecified number of unknowns, which is a difference compared to the parametric transformations. This is due to the fact that every point x is individually transformed by $y(x)$, as y is not parameterized by a finite number of parameters. The individual deformation allows us to model non-linear transformations, which is especially useful in medical applications, to capture changes in tissue or different kinds of motion of organs on the one hand. On the other hand, this comes with the need to optimize with respect to all image-points rather than with respect to a finite number of parameters.

We can now use these two definitions to define the classical image registration problem [99].

Problem 1 (Image Registration Problem (IRP, cf. [99, p. 78]))**.** Given a reference image $\mathcal{R}$ and a template image $\mathcal{T}$, the *image registration problem* aims to find a *reasonable transformation* y such that $\mathcal{T}$ and $\mathcal{R}$ are as similar as possible.

This linguistic description can be formalized, following [99], [100]. Let $\mathcal{T} \circ y$ denote the transformed image $\mathcal{T} \circ y : \mathbb{R}^d \to \mathbb{R}$ with $\mathcal{T} \circ y(x) = \mathcal{T}(y(x))$. Further, let $\mathcal{D} : \mathrm{Img} \times \mathrm{Img} \to \mathbb{R}$ be a similarity measure that quantifies the dissimilarity of the

images $\mathcal{T} \circ y$ and $\mathcal{R}$. The goal is to minimize $\mathcal{D}$ with respect to y. As outlined before, this problem is ill-posed, and we add a regularization $\mathcal{S} : \Omega_y \to \mathbb{R}$, which assures smoothness of the solution. Here, Ω_y is the space of transformations. The influence of the regularizer is controlled with a regularization parameter $\alpha \in \mathbb{R}_{>0}$. With this, the image registration model is to determine a minimizer of the functional $\mathcal{J} : \text{Img} \times \text{Img} \times \Omega_y \times \mathbb{R}_{>0} \to \mathbb{R}$ as [99]

$$\mathcal{J}(\mathcal{T}, \mathcal{R}, y, \alpha) = \mathcal{D}(\mathcal{T} \circ y, \mathcal{R}) + \alpha \mathcal{S}(y) \longrightarrow \inf \tag{1}$$

This joint functional describing the image registration problem was introduced for image registration by Fischer and Modersitzki [30] and is the basis for our generalization in Sec. 3. Note that the functional has a minimizing element due to the regularization, which in general has no explicit analytic expression. We therefore use numerical optimization methods. Following, we discuss two fundamental directions.

2.2 Optimization Strategies

Since we have defined the image registration problem we take as a basis, we shortly summarize two possible ways of minimizing a joint functional like Eq. (1). The literature indicates two paths, namely *Optimize-then-Discretize* as well as *Discretize-then-Optimize* [22], [46], [99].
The Optimize-then-Discretize approach makes use of the tools of the calculus of variations. The registration problem is modeled as a system of partial differential equations. A vanishing first variation, also known as Euler-Lagrange-Equations, is used to compute of a minimizing element of (1). For this approach, the system is discretized after solving for optimality, typically using finite elements or finite differences; see e.g. [20], [56], [99]. The approach has the advantage that well-known techniques from solving PDEs can be used [99]. A drawback is that only small steps towards a solution can be made, such that the convergence is rather slow [99]. We do not use this way, as we cannot use standard optimization methods for solving the registration problem [46].

The second way to a solution – which is used in this work – is to use the Discretize-then-Optimize approach. With this approach, we first discretize the registration problem and optimize the discrete approximation afterwards; see e.g. [45], [46], [100]. This has the following advantages [46]:

- Multi-level approaches are straightforward to use due to discretization before optimization.
- We can use Newton-type methods for solving the IRP, which converge with up to quadratic rates.

- Stopping criteria for numerical optimization can be used.

For this work, the only notable disadvantage is that the approximation's quality to the continuous solution depends on the used discretization scheme [44]. Depending on the used grid-type for discretization, different boundary conditions need to be handled; see e.g. [46], [100]. When needed, we mention these in the specific sections of this work. A detailed discussion on the implications by using this approach can be found in [46], [100]. Therefore, we follow this approach.
In Section 2 and Section 2.1 we discuss the need for two-image registration. Subsequently, in Section 2.7 we review the optimization methods we use for computation of our numerical results for the medical applications presented in Section 4.

2.3 Distance Measures

In this section, we start to enlighten the topic of measuring the similarity of images, starting for two images in order to generalize to an arbitrary number of images. The summarized data fidelity terms give geometric ideas of what image similarity means. We use these ideas for generalization in order to introduce our novel distance measure SqN. Accompanying, in the following section, we elaborate regularizers for two images as we use them to regularize our deformation fields. For the following sections, we follow original literature like [10], [29], [37], [47], [55], [81], [99], [100], [128] and references therein.
We focus on distance measures that are at least one time continuously differentiable since we use optimization methods that use derivatives.

Sum of Squared Differences

A straightforward way of comparing images is the so-called Sum of Squared Differences (SSD). Although the time of the first use may not be determined, a linearized version was already used as a standard in the early days of digital image processing in [65], [89] for optical flow estimation. It compares the images gray values in a point-wise manner. The idea is, that the compared images are similar, if their difference vanishes [99]:

$$\mathcal{D}^{\mathrm{SSD}}(R,T;y) \coloneqq \frac{1}{2}\int_{\Omega} \left(T(y(x)) - R(x)\right)^2 \,\mathrm{d}x. \tag{2}$$

The advantages of this measure are the straightforward evaluation and its robustness to noise. The main disadvantage of SSD is that it directly compares the gray values. Therefore, it is not suited for comparing images showing corresponding structures with different gray values, e.g., MRI compared to CT.

Normalized Cross Correlation

The normalized cross correlation is another standard measure in image registration [94]. It originates from comparing the reference image with a translated version of the template image [100]. The cross correlation is derived from expanding the square-form $(T(x) - R(x))^2$ and ignoring the squares [100]:

$$\mathcal{D}^{\mathrm{CC}}(R, T; y) := \int_\Omega T(y(x))\, R(x)\, \mathrm{d}x = \langle T(y(x)), R(x) \rangle. \tag{3}$$

In order to achieve scaling-invariance, the cross correlation is typically normalized and then used for registration [100]:

$$\mathcal{D}^{\mathrm{NCC}}(R, T; y) := \frac{\langle T(y(x)), R(x) \rangle}{\|T(y(x))\|\, \|R(x)\|} \tag{4}$$

with the induced norms $\|x\| = \sqrt{\langle x, x \rangle}$. Again, following the later introduced discretization method, this measure is straightforward to evaluate. Furthermore, fluctuations in gray value intensities are compensated by normalizing the images and is robust against noise. However, this measure cannot – just as SSD – compare images from modalities that show corresponding structures with different gray values. In Section 3 we show, that this measure is a special case of SqN.

Normalized Gradient Fields

The so-called Normalized Gradient Fields (NGF) distance measure is able to compare images from different modalities with differences in gray values for corresponding structures. The assumption for NGF is, that intensity changes appear at corresponding locations in the images [100]. Therefore, the spatial image gradients can be used to compare the intensity changes. However, the simple comparison of the intensity changes also takes the amount of the changes itself into account, which can differ for different modalities. In order to leave this out, a normalization of the gradients can be done. This leads to the formulation of Haber and Modersitzki [47], [100]:

$$\mathcal{D}^{\mathrm{NGF}}(R, T; y) := \int_\Omega 1 - \left(\frac{\langle \nabla T(y(x)), \nabla R(x) \rangle}{\|\nabla T(y(x))\|_\eta\, \|\nabla R(x)\|_\eta} \right)^2 \mathrm{d}x \tag{5}$$

with $\|T\|_\eta := \sqrt{\|\nabla T\|^2 + \eta^2}$ a regularized, induced norm. Here, the so-called edge parameter $\eta \in \mathbb{R}$ discriminates edge information from background noise [100] and therefore plays an important role in parameter tuning for the optimization process. We can also notice this in the numerical results of this work in Sec. 4.

2.4 Regularization

A crucial point of regularization is the ill-posedness – defined by Hadamard [48] – of the registration problem. Using a regularizer should resolve the ill-posed problem – in a best-case scenario – to have a well-posed problem. Regularization is needed to define a unique and numerically stable solution of the registration problem [99]. Our work focuses on the presentation of a distance measuring method. Therefore, we restrict the presented regularization methods to well-known methods that have proven to be well-suited for solving image registration problems in the past. In this section, we briefly summarize two regularizers from standard image registration literature; e.g., [31], [47], [55], [99], [100] we also use for our numerical experiments in Section 4. We focus on curvature regularization [29] as well as elastic regularization [10], [99] as these provide a sufficient degree of smoothness and are available in standard toolboxes like FAIR [100]. Furthermore, past results have shown, that these are well-suited for image registration; see, e.g., [7], [59], [76], [78], [100], [101], [112]. Using the regularization parameter $\alpha \in \mathbb{R}_{>0}$, we can balance how strongly the prior knowledge of the regularizer can affect the solution of the registration problem. This is a critical parameter for the optimization process since it determines the regularization energy weighting. Thus it has a direct impact on the ill-posedness. We explicitly do not claim a more in-depth analysis, as the summarized concepts have been investigated in detail in the literature referred to above and the references therein.

Curvature Regularization

Mainly, we use the curvature regularizer, introduced for image registration by Fischer and Modersitzki [29], since the deformations are typically smoother compared to the registration using other regularizers [29], [100]. In their work [29], they state that two key points are supporting the use of this approach. Firstly, it can be seen as an approximation to the curvature of the components of the deformation grid, which favors non-oscillating solutions. Secondly, the non-trivial kernel contains affine linear transformations. Concerning the non-trivial kernel, they further state that there is no need for affine linear pre-registration to be successful. Note that the curvature regularizer, by definition, does not penalize affine linear transformations.
This regularization term is a sum of the spatial second derivatives of the displacement fields components [29], [100]:

$$\mathcal{S}^{\text{curv}}(u) \coloneqq \frac{1}{2} \int_{\Omega} \sum_{j=1}^{d} (\Delta u_j)^2 \, \mathrm{d}x \tag{6}$$

where $\mathcal{S}^{\text{curv}} : \mathbb{R}^d \to \mathbb{R}$ with displacement field $u : \mathbb{R}^d \to \mathbb{R}^d$ and spatial dimension

$d \in \mathbb{N}$. Here, Δ is the Laplace operator, which is defined for a twice-differentiable, real-valued function f as $\Delta f = \partial_1^2 f_1 + \partial_2^2 f_2 + \ldots + \partial_d^2 f_d$. The displacement field u is typically characterized as the identity-free transformation grid $u = y - \mathrm{Id}$ [100].

Elastic Regularization

The motivation of the elastic regularizer is somewhat different than for the curvature regularizer. For this approach, a tangible, physical motivation exists and was already introduced in an early version by Fischler and Elschläger in 1973 [33] and finally by Broit for image registration in 1981 [10]. It measures the energy of deformations of an elastic material [100]. For an in-depth introduction to this topic and the physical motivation, see e.g., [40], [99], [100]. The regularizer is nothing more than a linearized elastic potential of the displacement field [10], [99]:

$$\mathcal{S}^{\text{elas}}(u) \coloneqq \int_\Omega \frac{\mu}{4} \sum_{i,j=1}^{d} \left(\partial_{x_i} u_j + \partial_{x_j} u_i\right)^2 + \frac{\lambda}{2} \left(\operatorname{div} u\right)^2 \, \mathrm{d}x \tag{7}$$

where $\mathcal{S}^{\text{elas}} : \mathbb{R}^d \to \mathbb{R}$ with displacement field $u : \mathbb{R}^d \to \mathbb{R}^d$ and spatial dimension $d \in \mathbb{N}$. The divergence div is given by $\operatorname{div} f = \partial_1 f + \partial_2 f + \ldots + \partial_d f$ for a differentiable, real-valued function f. Further, $\mu \geq 0$ and $\lambda \geq 0$ are the so-called Lamé constants, which are material dependent. This offers a variety of options for mimicing different materials within the registration process.

2.5 Singular Values and Schatten-Norms

In this work, the main contribution is the use of singular values and Schatten-Norms for distance measurement between images. Therefore, we first recall the Singular Value Decomposition, show a relevant property and finally define the Schatten-Norms, named after American mathematician Robert Schatten.

Theorem 1 (Singular Value Decomposition; cf. [136, Thm. 2.4.1])**.** Let $A \in \mathbb{R}^{n \times K}$ be a matrix. Then there exist orthogonal matrices

$$U = [u_1 | u_2 | \ldots | u_n] \in \mathbb{R}^{n \times n} \quad \text{and} \quad V = [v_1 | v_2 | \ldots | v_K] \in \mathbb{R}^{K \times K}$$

such that

$$A = U \Sigma V^\top .$$

Here, $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)$ with $p = \min\{n, K\}$. Note, that the so-called singular values σ_i are in descending order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_K \geq 0$.

Proof. see [136, Thm. 2.4.1] □

We use the following Corollary to show properties of our novel distance measure SqN in Section 3.

Corollary 1 (cf. [136, Cor. 2.4.3])**.** Let $A \in \mathbb{R}^{n\times K}$ be a matrix. Then,

$$\|A\|_2 = \sigma_1 \quad \text{and} \quad \|A\|_F^2 = \sigma_1^2 + \cdots + \sigma_p^2,$$

with $p = \min\{n, K\}$.

Proof. see [136, Cor. 2.4.3] □

Using the SVD, we define the Schatten-q-Norms in Definition 3, which are highly relevant for defining our novel distance measure SqN.

Definition 3 (Schatten-q-Norm [6])**.** Let $A \in \mathbb{R}^{n\times K}$ be a matrix and $S \in \mathbb{R}^K$ the vector of singular values σ_i, $i = 1, 2, \ldots, K$. Here, n, $K \in \mathbb{N}$. Further, let q be in $\mathbb{R}_{\geq 0} \cup \{+\infty\}$.
Then, the *Schatten-q-Norm* is defined as the q-Norm of the vector of singular values σ_i:

$$\|A\|_{S,q}^q \coloneqq \|S\|_q^q = \sum_i \sigma_i^q.$$

For the particular choice of $0 \leq q < 1$, the triangle inequality is no longer fulfilled; hence, the Schatten-q-Norms are just quasinorms. A quasinorm fulfills all other norm axioms except the triangle inequality. The triangle inequality is weakened and thus needs an additional scaling parameter $c > 0$ such that $\|x+y\| \leq c(\|x\| + \|y\|)$. Note, that we show in our medical applications in Section 4, that $q < 1$ is no problem in practice. Besides, in Section 3 we figure out, that there are smarter choices of q for application in image registration.
Two exciting properties of the Schatten-Norms — which we use in Sec. 3.4 — are the equality to the Frobenius Norm for $q = 2$ and for that case the notation using the trace.

Lemma 1 (Equality of Frobenius- and Schatten-Norm [66])**.** For $q = 2$ and arbitrary $A \in \mathbb{R}^{n\times m}$ it holds

$$\|A\|_{S,2}^2 = \|A\|_F^2 = \operatorname{tr}(A^\top A).$$

Proof. First, we rewrite the Forbenius norm as trace of A and use that A is real-valued:

$$\|A\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} |a_{ij}|^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}\cdot a_{ij} = \operatorname{tr}(A^\top A)$$

To resolve this trace expression, we now use the singular value decomposition $A = U\Sigma V^\top$:

$$\operatorname{tr}(A^\top A) = \operatorname{tr}\left((U\Sigma V^\top)^\top (U\Sigma V^\top)\right) = \operatorname{tr}\left(V\Sigma^\top \Sigma V^\top\right) = \operatorname{tr}\left(\Sigma^\top \Sigma\right) = \sum_k \sigma_k^2.$$

The second last step is possible because of the invariance to cyclic permutations of the trace. Since the last expression is the sum of singular values of A, we have the definition of the Schatten-2-Norm $\sum_k \sigma_k^2(A) = \|A\|_{S,2}^2$. See also Corollary 1 and its proof. □

2.6 Discretization and Numerical Preliminaries

As outlined in Section 2.7, we follow the discretize-then-optimize approach and now present discretization details. We first discretize the distance measures and regularizers and then optimize numerically. For this purpose, we discuss discretization preliminaries. We first outline what grids and interpolation methods we use. Second, we investigate discretization strategies for the distance and regularization equations. The specific notations we use in this chapter help us formulate our novel distance measure SqN straightforwardly. This chapter is mainly based on standard literature like [51], [99], [100], and references therein.

Grids and Interpolation

Following [100], we start by defining the image domain, which is to discretize using a grid. The grids we work with within this thesis, consist of equidistant points on a domain Ω [100]

$$\Omega \coloneqq (\omega_1, \omega_2) \times \ldots, \times (\omega_{2d-1}, \omega_{2d}) \subset \mathbb{R}^d.$$

Here, $d \in \mathbb{N}$ denotes the spatial dimension. For $d = 2$ the domain is rectangular, for $d = 3$ and higher it is cuboid, respectively. In the literature, different types of grids are used for discretization. Primarily, these grids differ in the definition of the used cells, meaning the grid point locations within the discretization domain differ. We focus on cell-centered grids since these are well-suited for our use-case, see [100] for a discussion on different grids. The domain Ω is divided into $\bar{m} \coloneqq (m_1, m_2, \ldots, m_d)$ cells, we denote the total number of cells by $m = \prod_k m_k$. The edge lengths of the cells are denoted by $h \coloneqq (h_1, h_2, \ldots, h_d)$. A cell-centered grid is defined as a collection of points [100]

$$x_j \coloneqq \left(\omega_1 + h_1\left(i_1 - \frac{1}{2}\right), \ldots, \omega_{2d-1} + h_d\left(i_d - \frac{1}{2}\right)\right), \tag{8}$$

where i is a set of index vectors $i = (i_1, \dots, i_d)$ with $i_l = 1, \dots, m_l$ for $l = 1, \dots, d$. We arrange the grid points as a vector using a lexicographic ordering with indices

$$k \coloneqq i_1 + \sum_{l=2}^{d} (i_l - 1) \prod_{r=1}^{l-1} m_r$$

for each grid point. For $d = 3$, we then have $k = i_1 + m_1(i_2 - 1) + m_1 m_2(i_3 - 1)$, respectively. In practice, this means, we start collecting all points corresponding to the first spatial dimension and continue with the remaining dimensions in the same manner. It is now straightforward to recover a single grid point by

$$x_k = \left(x_k \dots, x_{k+(d-1)m}\right) \in \mathbb{R}^d .$$

Using these definitions we can discretize images in a vectorized manner

$$I(x) \coloneqq \left(\mathcal{I}(x_k)\right)_{k=1,\dots,m} \in \mathbb{R}^m, \tag{9}$$

where $I : \mathbb{R}^{dm} \to \mathbb{R}^m$ maps the set of coordinates to a vector of discrete intensities. Furthermore, we use the cell-centered grids as a starting guess for the deformation grids y in optimization. Here, it is not important if we use cell-centered grids or other forms. It is possible to convert the single grid forms into the other forms via grid-transformation-operators, which make them equivalent and only numerical details as e.g., the handling of boundary conditions may be more straightforward with a specific type of grid, but can be handled with all types of grids [100]. The here employed vectorized notations are used for defining the feature matrices in Sec. 3.

In order to compute deformed versions of the images, we need to be able to evaluate the images on any point of the given continuous domain. It is not sufficient to only know the values at the grid points since these are moved through transformation. For transferring the discrete data to a continuous representation, we use image interpolation. Since interpolation is not a crucial part of this work, we restrict ourselves to standard interpolation methods. Due to its differentiability properties, we mainly use cubic spline interpolation, as outlined in [25], [100], [121]. This interpolation model is well-suited for variational image registration since it is twice continuously differentiable and straightforward to compute. Furthermore, we can compute image intensity gradients using the derivatives of the polynomial basis functions. Although the computation of spline-based interpolants is more complicated than with linear- or next-neighbor-interpolation methods, for natural images, the interpolation quality and the differentiability is superior to methods like linear- or next-neighbor-interpolation [134]. In order to achieve the superior quality of spline interpolation, we have to handle a few drawbacks. The differentiability comes with a ringing-behavior [25], [134]. Due to the twice-continuously-differentiable transitions between the single splines, an issue is overshooting at

inflexion points or non-smooth points [25], [134]. As interpolation is not the main issue of this work, we stay with natural boundary conditions. Thus, the numerical handling is straightforward, since no explicit conditions need to be handled. For a detailed discussion, we refer to appropriate literature; see e.g., [25], [100], [121], [134].

Discretization of Regularization- and Distanceterms

As reviewed above, we use finite difference schemes for the discretization of differential operators on a grid in order to discretize the (IRP) as well as Problem 2, we introduce in Sec. 3. Furthermore, we discretize integrals using the midpoint rule. The midpoint rule gives us a second-order accuracy, which we need for application with second-order optimization methods [51], [100], [105]. Furthermore, it is straightforward to handle and to implement. Assuming a domain $\Omega \subset \mathbb{R}^d$ and $\mathrm{hd} = h_1 \cdot \dots \cdot h_d$ with $h = (h_1, \dots, h_d)$, the midpoint rules equation is [51]

$$\int_\Omega f(x)\,\mathrm{d}x = \mathrm{hd} \sum_k^n f(x_k) + \mathcal{O}(\|h\|_2^2),$$

Here, x_k is a cell-centered grid point with $k = 1, \dots, n$ and n is the number of grid-cells. Furthermore, f is a smooth function and Ω is bounded by definition. The error-behavior is described by $\mathcal{O}(\|h\|_2^2)$, meaning, $h \sum_k^n f(x_k)$ converges to $\int_\Omega f(x)\,\mathrm{d}x$ for $h \to 0$. We take this error-behavior as a basis for our discretization and a multi-level-approach we use in optimization, where we refine the resolution of our discretization on each level. In this section, we outline the needs for discretizing the distance measures and regularizers.
The idea of the finite-difference-schemes is to approximate a derivative for a function $y : \mathbb{R} \to \mathbb{R}$ on an equidistant grid using e.g. central differences [100]:

$$\partial y(x_i) = \frac{y\left(x_i + \frac{h}{2}\right) - y\left(x_i - \frac{h}{2}\right)}{h} + \mathcal{O}(h^2), \tag{10}$$

where x_i is a cell-centered point and h again denotes the cells edge lenghts, as introduced above. Also y is a smooth function, such that the approximation exists. This second-order approximation of the derivative on a cell-centered grid has a drawback, however. We discretize on a nodal grid, in order to obtain these differences, following the approach outline in [100]. It holds $\partial y(x_i) \approx \frac{1}{h}(y(x_i^n) - y(x_{i-1}^n))$ [100], from which we derive the $1D$ derivative operator [100]

$$\partial_m^h \coloneqq \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{m \times m+1}.$$

Using the abbreviation $y_i^n := y\left(x_i + \frac{h}{2}\right)$, we can summarize (10) to

$$\partial y = \partial_m^h y^n + \mathcal{O}(h^2).$$

It is worth noting that the long stencil $\partial y(x_i) \approx \frac{1}{2h}\left(y(x_{i+1}) - y(x_{i-1})\right)$ returns zero derivatives for high oscillatory functions [100]. This means, high frequency components are in the kernel, which is why we use the short stencil, like suggested in [100]. It is straightforward to extend the $1D$-differential operator to higher dimensions by utilizing the well-known Kronecker product as can be seen in e.g., [100].
Since both regularizers, we are interested in, are L_2-norm-based, we follow [100] to give a general discretization rule using the midpoint quadrature rule to discretize the integrals

$$\mathcal{S}(u) = \int_\Omega \langle \mathcal{B}u, \mathcal{B}u\rangle \,\mathrm{d}x = \mathrm{hd}\|Bu^h\|^2 + \mathcal{O}\left(\|h\|^2\right), \tag{11}$$

where $\mathrm{hd} = \prod_{j=1}^d h_j$ and u^h is the discretization of $u = y - \mathrm{Id}$. Furthermore, B the discretization of $\mathcal{B}$ and is either the curvature operator or the elastic operator, respectively. Using the first-order stencil from (10), it is straightforward to build the discretized elastic operator on a staggered grid as introduced in [100]. In order to build the discretized curvature operator, we set the second-order stencils, regardless of the boundary conditions, to

$$\partial^2 u(x_i) = \frac{u\left(x_{i+1}\right) - 2u(x_i) + u\left(x_{i-1}\right)}{h^2} + \mathcal{O}(h^2),$$

which is dependent on the direction of derivation. For details on the derivation of the discretized operators and the boundary conditions, we refer to [100].

We are now ready to discretize the distance measures introduced in Sections 2.1 and 3. For this part, we rely on [99], [100] for discretization of $\mathcal{D}^{\mathrm{NCC}}$ and $\mathcal{D}^{\mathrm{NGF}}$. We leave out the discretization of SSD since this is straightforward and can be found in [100]. Here, we summarize the discretization of NCC and NGF – that can also be found in [100] – to familiarize the notation of our used discretization strategy. Furthermore, the discretization of NCC and NGF plays a role in the discretization and error estimation of SqN. Using the midpoint quadrature rule and the cell-centered grid, reviewed above, we discretize the gradients using the discussed finite difference scheme.
We discretize the integrals for the Normalized Cross-Correlation (4) using the mid-point quadrature rule. The denominator is a product of the norms of the discretized images, while the nominator is an inner product [100]

$$D^{\mathrm{NCC}}(T(y(x^c)), R(x^c)) = \frac{\langle T(y(x^c)), R(x^c)\rangle}{\|T(y(x^c))\|_2 \|R(x^c)\|_2},$$

where it is assumed, that T, $T(y)$ and R are not completely zero [100]. Otherwise, the norms used in the denominator need to be regularized. Due to the discretization of integrals in the nominator as well as in the denominator, the factor of the mid-point rule hd cancels out.
For the Normalized Gradient Fields from (5), we utilize the finite-differences from (10) to approximate the gradients. Just as shown in [100, Sec. 7.4.2] the Kronecker product can be used for higher dimensionality of the finite difference matrices. We then abbreviate the discretized gradient images by $\partial T(y(x^c))$ and $\partial R(x^c)$. These are again vectorized because of the approximation on the cell-centered grid x^c, such that we obtain $\partial T, \partial R \in \mathbb{R}^{d \cdot n}$. Here, the components of the derivatives of the spatial directions are stacked, such that $\partial I = (\partial_{d_1} I \in \mathbb{R}^n, \partial_{d_2} I \in \mathbb{R}^n, \partial_{d_3} I \in \mathbb{R}^n, \dots)^\top \in \mathbb{R}^{d \cdot n}$. The computations are considered pointwise. Therefore, exemplarily for 3D, we obtain $(\partial I)_k = ((\partial_{d_1} I)_k, (\partial_{d_2} I)_k, (\partial_{d_3} I)_k)^\top \in \mathbb{R}^3$ at each grid point and [100]

$$D^{\mathrm{NGF}}(T(y(x^c)), R(x^c)) = \mathrm{hd} \sum_{k=1}^{n} (1 - r_k^2),$$

where

$$r_k := \frac{\langle (\partial T(y(x^c)))_k, (\partial R(x^c))_k \rangle}{\|(\partial T(y(x^c)))_k\|_\epsilon, \|(\partial R(x^c))_k\|_\epsilon}$$

and $\|x\|_\epsilon = \sqrt{\|x\|^2 + \epsilon^2}$, just as in (5). Note that the computations in NGF differ from those in NCC, where the normalization is done globally instead of pointwise [47], [100]. The difference between global and pointwise normalization of gradients is discussed in Section 3.2.

2.7 Numerical Optimization

In this section, we shortly outline the numerical optimization schemes we use in this thesis. First, we summarize Newton's Method. Second, building on Newton's Method, we summarize the two Newton-type methods L-BFGS and Gauß-Newton, which we use for our medical applications in Section 4. These give us the advantage of avoiding an explicit computation of a full Hessian of the objective function by rather iteratively approximate it. The explicit computation of a Hessian might be infeasible, as we point out with practical examples throughout this section. This short summarizing section is the basis for the specific extension of the examined methods at the end of Section 3 for the optimization of our novel method. This section is mainly based on original literature, see, e.g., [3], [103], [116].

Newton's-Method

Following e.g. [100], in this thesis, we use Newton-type methods for optimization due to specific convergence behavior [103]. We now summarize the idea of

Newton's-Method in order to derive the specific schemes we use. For the specific optimization schemes, we elaborate specialties concerning the Hessian we use, e.g., for implementation. To find a descent direction, we use of the first and second derivative of a two times continuous differentiable function $f : \mathbb{R}^d \to \mathbb{R}$. Using a Taylor approximation of $f(x_k + p_k)$ we obtain

$$f(x_k + p_k) = f(x_k) + \nabla f(x_k)^\top p_k + \mathcal{O}(x^2), \tag{12}$$

with $x_k \in \mathbb{R}^d$ and descent direction $p_k \in \mathbb{R}^d$ which is to be determined. Applying the first-order necessary condition to (12) while ignoring all higher order terms, we obtain a system of linear equations, called Newton equation [103]

$$\nabla^2 f(x_k) p_k = -\nabla f(x_k) \tag{13}$$

which we solve for the descent direction p_k. Here, the Hessian of f needs to be positive definite in order to obtain a descent direction by solving the system of linear equations Eq. (13) [103]. This is the case for convex optimization problems. If the Hessian is not positive definite, other methods may be used where the Hessian is not used or the Hessian needs to be pre-conditioned [100], [103]. The reason we use methods from the Newton-family is locally quadratic convergence we may achieve [100] if the initial guess is close to a minimizer or the Gauß-Newton matrix is well-posed [103, Thm. 3.5.]. Furthermore, the Hessian does not need to be explicitly inverted as it is iteratively approximated. The severe drawback of the particular approach in Eq. (13) is that we need to compute the Hessian of the objective function in each step in order to solve the linear system of equations. Concerning the fact that we handle multiple images for registration, where each image is transformed, the computation of the Hessian will get quickly very expensive in terms of computation time and memory because of a very high number of unknowns. Following [55], a simple computation illustrates the issue: Computing the Hessian $\nabla^2 f(x_k) \in \mathbb{R}^{n \cdot d \times n \cdot d}$, where $n \in \mathbb{N}$ is the number of pixels and $d \in \mathbb{N}$ is the image dimension. Then, for a 2D image with $256 \times 256 = 2^{16}$ pixels, the Hessian of f has $2^{17} \times 2^{17}$ entries. Although most of the entries are zero, the non-zero entries of the matrix are still $2n$. This is the case concerning only one moving image. Now that we have to handle an arbitrary number $T \in \mathbb{N}$ of moving images, the number of unknowns grows. The Hessian for our specific problem containing multiple images is $\nabla^2 f(x_k) \in \mathbb{R}^{n \cdot d \cdot T \times n \cdot d \cdot T}$. Already for $T = 3$ images of size 256×256 pixels, we obtain a Hessian of size $2^{51} \times 2^{51} = 2^{102}$, which is currently infeasible to store and recompute in each iteration!

Newton-type Methods

A trade-off for computation of the inverse of the Hessian is to approximate it iteratively. Here, we classify the Newton-type methods into two sub-classes [27]. There

are so-called Quasi-Newton methods and there are least-squares Newton methods [103]. Both classes have several advantages over the plain Newton method. First, we do not require to compute the Hessian explicitly. Second, we won't need as much memory for saving the approximate Hessian as we would for the exact one. The only notable drawback is that we won't obtain the theoretical, local quadratic convergence of Newton's-Method in practice. However, at least a super-linear convergence behaviour can be achieved [103, Thm. 6.6.], which is a trade-off between convergence speed, computation time and storage memory. Since both classes have successfully been applied for medical imaging problems, for example, in [63], [76], [101], [112], we specifically focus on the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method as a Quasi-Newton approach, which approximates the Hessian in each step by just using gradient information. As an alternative, we discuss the Gauss-Newton method as a least-squares Newton approach, which can exploit specific structures of the given optimization problem.

BFGS-Method

The Broyden-Fletcher-Goldfarb-Shanno-Method (BFGS-Method) uses an iteratively approximated Hessian within Newton's method and therefore is a quasi-newton method. The L-BFGS-Method is a limited-memory version of the BFGS-Method, which is especially well-suited for our problem as the number of unknowns for registration of multiple images is rather high. Therefore, we outline the idea at the end of this section, based on original literature like [27], [55], [103].

The underlying idea is to solve the Newton equation (13), but using an approximation H instead of the Hessian of the objective f

$$H(x_k)p_k = -\nabla f(x_k).$$

When using such an update, we want it to mimic the behaviour of the exact Hessian in such a way that our approximation is a symmetric positive definite matrix. We can obtain this behaviour by fulfilling the so-called curvature condition as well as the so-called secant equation [103]. The latter one is derived by constructing a quadratic model for the new iterate x_{k+1} for the line search update, like in (12)

$$f(x_{k+1} + p) = f(x_{k+1}) + \nabla f(x_{k+1})^\top p + \frac{1}{2} p^\top H(x_{k+1})p.$$

Following [103], applying the first-order necessary condition for both sides of the equation and requiring this new system to match the gradients of the two latest steps x_{k+1} and x_k, it is

$$\nabla f(x_{k+1}) - \alpha_k H(x_{k+1})p_k = \nabla f(x_k)$$

obtained, what is rearranged to

$$H(x_{k+1})\alpha_k p_k = \nabla f(x_{k+1}) - \nabla f(x_k). \tag{14}$$

Shortly, (14) is the so-called secant equation [103]

$$H(x_{k+1})s_k = g_k \tag{15}$$

where $s_k = x_{k+1} - x_k = \alpha_k p_k$ and $g_k = \nabla f(x_{k+1}) - \nabla f(x_k)$. The secant equation requires that the symmetric, positive definite Hessian approximation H maps s_k on g_k. Indeed, this is only possible if further conditions are imposed. For strongly convex f, we can multiply (15) with $s_k^\top$ from the left and see, that the so-called curvature condition is fulfilled

$$s_k^\top g_k > 0. \tag{16}$$

To always fulfill (16) and have $H(x_{k+1})$ unique, $H(x_{k+1})$ it is chosen in such a way, that it is the closest solution to the current $H(x_k)$ in sense of a weighted Frobenius norm $\|H(x_{k+1}) - H(x_k)\|$ [103, Chpt. 6.1]. Using these prerequisites, the literature – e.g. [27], [103] – gives a direct update formula for the Hessian approximation H with

$$H(x_{k+1}) = H(x_k) - \frac{H(x_k)s_k s_k^\top H(x_k)}{s_k^\top H(x_k) s_k} + \frac{g_k g_k^\top}{g_k^\top s_k}. \tag{17}$$

One advantage of this update formula is that the inverse of $H(x_{k+1})$ can be computed. With this formula, it can be updated in each step to solve (13), instead of solving it anew in each iteration. An explicit formula for the computation of the inverse can be found in standard optimization literature as e.g. [27], [103]. Another advantage is that we don't need to store the matrix H. Instead, the current update can directly be computed by just using the last s_i and g_i with $i = 0, \ldots, k$ and an initial value H_0. A disadvantage compared to Gauß-Newton is that we may need more iterations to converge using L-BFGS. This is due to more information that is used to compute the Gauß-Newton approximation.
In Section 3.9, we discuss details on how we handle the extension for multiple images. There we extend the approximation for H_0, or more precisely, the approximation for H_0^{-1}.

Gauß-Newton

Like for the BFGS method, the Gauß-Newton method avoids the expensive computation of the second-order derivative of the objective function. Different than the BFGS method, Gauß-Newton can be understood as a least-squares Newton method [27]. Its Hessian approximation formula is derived via a nonlinear least-squares problem. The underlying least-squares-problem is

$$f(x) = r(x)^\top r(x) = \|r(x)\|_2^2 \longrightarrow \min. \tag{18}$$

Here, f again is a sufficiently smooth function as stated before for Newton's method. Furthermore, the residual vector $r : \mathbb{R}^n \to \mathbb{R}^m$ is a concatenation of $m \in \mathbb{N}$ nonlinear residual functions $r_j : \mathbb{R}^n \to \mathbb{R}$ [103, Chpt. 10]. Moreover, f is a combination of squared residuals

$$f(x) = \frac{1}{2} \sum_{j=1}^{m} r_j^2(x).$$

Instead of solving Problem (18) by using simple gradient based methods, like steepest descent [103, Chpt. 2], Newton's method is applied. We summarize the Method, following [103]. First, the problem locally linearized using a Taylor approximation

$$\|r(x_k) + \mathrm{d}r(x_k)(x - x_k)\|_2^2 \longrightarrow \min . \tag{19}$$

Here, dr is the Jacobian with

$$\mathrm{d}r = \left(\frac{\partial r_j}{\partial x_i} \right)_{\substack{j=1,\dots,m \\ i=1,\dots,n}} .$$

With $m \geq n$ and f sufficiently smooth, the gradient of f is expressed as

$$\nabla f(x) = \mathrm{d}r(x)^\top r(x)$$

and analogously the Hessian as

$$\nabla^2 f(x) = \mathrm{d}r(x)^\top \mathrm{d}r(x) + \sum_{j=1}^{m} r_j(x) \nabla^2 r_j(x).$$

If the residuals are small, close to the solution, the second term is typically neglected; cf. [103, Chpt. 10]. This gives the Hessian approximation

$$\nabla^2 f(x) \approx \mathrm{d}r(x)^\top \mathrm{d}r(x). \tag{20}$$

Coming back to the linearization (19), it is checked for the first-order necessary condition in order to derive the Gauß-Newton method.
Computing the derivative and rearrangement of (19) gives

$$r(x_k) = - \mathrm{d}\, r(x_k)(x - x_k). \tag{21}$$

By multiplication of $\mathrm{d}\, r(x_k)^\top$ on each side and solving for x we obtain the Gauß-Newton update

$$x = x_k - \left(\mathrm{d}\, r(x_k)^\top \mathrm{d}\, r(x_k) \right)^{-1} \mathrm{d}\, r(x_k)^\top r(x_k), \tag{22}$$

which is nothing else than solving the Newton equation (13) using the Gauß-Newton approximation of the Hessian. The quadratic form $\mathrm{d}\, r(x_k)^\top \mathrm{d}\, r(x_k)$ is positive definite and therefore can be inverted. We see that the problem is solved by the pseudo-inverse – what is to be expected for least-squares problems. Indeed, the pseudo-inverse inherits the Gauß-Newton approximation of the Hessian. Representing our search direction here as $p_k = x - x_k$ and using the representations of the gradient and Hessian of f, especially the approximated Hessian, the Newton system of f is

$$\begin{aligned} & x = x_k - \left(\mathrm{d}\, r(x_k)^\top \mathrm{d}\, r(x_k)\right)^{-1} \mathrm{d}\, r(x_k)^\top r(x_k) \\ \Leftrightarrow\, & x - x_k = -\left(\nabla^2 f(x)\right)^{-1} \nabla f(x) \\ \Leftrightarrow\, & \nabla^2 f(x) p_k = -\nabla f(x) \end{aligned} \tag{23}$$

The Gauß-Newton method gives several advantages over the plain Newton's method [103]. With the Hessian approximation, we don't need to compute the individual Hessians of each residual function r_j. Indeed, we just need the gradients to compute the Hessian approximation. Furthermore, if $\nabla^2 f(x)$ has full rank and the gradient of f is nonzero, p_k is a descent direction and for zero gradients, x_k is a stationary point [103, Chpt. 10.3]. A drawback is the rather slow convergence behavior for ill-conditioned $\nabla^2 f(x)$, which results in an ill-conditioned system of equations.
From our derivation, we can also see that the search direction p_k is the solution of a least-squares problem, which means that we wouldn't even need to compute the Hessian approximation explicitly when using standard least-squares methods [103]. Another drawback is that rather large systems of equations need to be solved in each step. Therefore, we focus on L-BFGS for our applications in Section 4.
At the end of Section 3 we extend the Jacobian used for approximation of H.
For optimization, we rely on standard stopping criteria by Gill, Murray and Wright, derived in detail in [35]. Here, we just give a short summary of the stopping rules: These comprise five different stopping rules concerning the rate of change of the objective function, the rate of change of the respective transformation field, the rate of change of the gradient of the objective function as well as two fixed values. The fixed values are a small lower bound (typically machine precision) for the norm of the objective-gradient as well as a maximum number of iterations. Either the rate of changes must all be fulfilled or one of the fixed values needs to be fulfilled in order to stop the optimization process.

2.8 Conclusion

In this section, we have summarized the preliminaries of image registration and numerical optimization. We have summarized the variational framework consisting

of distance measure and regularizer to build a feasible optimization problem. Furthermore, we have summarized standard distance measures as well as regularizers. Especially the distance measures with its discretization and the specific notation lays a basis for our further elaboration on image registration tools for registration of multiple images. From the given measures, we take geometrical ideas to derive measures for multiple images.
We have closed this chapter with a brief summary of Newton-type optimization methods. At the end of Chapter 3, we denote remarks on slight changes for the optimization of a variational framework for multiple image registration.

3 Image Registration of Multiple Images

In this chapter, we introduce a mathematical framework for the registration of multiple images. This framework allows for deriving various specific distance measures. Here, we understand this framework as an extension of the distance measure concept as we already know it from the two-image registration setting. We will first introduce the mathematical framework and show a straightforward extension from two-images to multiple-images by using two-image tools. Second, we define an image feature array, with which we can take advantage of different interesting properties we examine. Third, we specify three central derivations of the here introduced framework. Furthermore, we give a geometrical intuition for each derived approach and perform an error analysis for the novel distance measure. This is accompanied by a numerical examination of the critical parameter q. Ultimately, we give a general distance measurement framework and draw a link to related work.

3.1 Notation and Extension of Two-Image Methods

We are now ready to extend the registration framework for two images to an arbitrary number T of images. We keep the variational approach, i.e., we minimize a joint functional consisting of a distance measure and a regularization. We extend the image and transformation sets to $I = (I_1, \dots, I_T)$ and $Y = (y_1, \dots, y_T)$, where $I_t \in \mathrm{Img}$ and $y_t : \mathbb{R}^d \to \mathbb{R}^d$ for all $t = 1, \dots, T$, respectively. We use the suggestive notation $I \circ Y := (I_1 \circ y_1, \dots, I_T \circ y_T)$.
The most interesting parts are the generalization of the distance measure and the regularization,

$$\mathcal{D} : \mathrm{Img}^T \to \mathbb{R}, \quad \mathcal{S} : \{y : \mathbb{R}^d \to \mathbb{R}^d\}^T \to \mathbb{R}\,.$$

These are discussed in Sec. 3.3 and Sec. 3.4.

Problem 2 (Variational Multiple-Images Registration)**.** Given $I \in \mathrm{Img}^T$, determine a minimizing element $Y \in \{y : \mathbb{R}^d \to \mathbb{R}^d\}^T$ of $\mathcal{J} : \mathrm{Img}^T \times \{\mathbb{R}^d\}^T \to \mathbb{R}$, with

$$\mathcal{J}(Y) \coloneqq \mathcal{D}(I \circ Y) + \mathcal{S}(Y).$$

We now show a straightforward extension utilizing the tools of the two image registration setting. After that, we introduce our novel multiple images registration framework following the notation of Problem 2.

Sequential Extension

The extension from an image pair to multiple images might be realized in a sequential manner [137]. For that, two-image tools – as described before – are used

for registration of a stack of image pairs. This sequential approach is certainly a straightforward way of registering multiple images. Using the variational framework of Problem 2, we are able to define this sequential setting. The distance measures are the standard pair-wise distances and are summed up in a sequential manner. Each pair of images is regularized individually [7]:

$$\mathcal{J}^{\text{seq}}(Y) \coloneqq \sum_{t=1}^{T-1} \mathcal{D}(I_t \circ y_t, I_{t+1} \circ y_{t+1}) + \mathcal{S}(y_{t+1}). \tag{24}$$

Typically, one of the deformation grids is fixed, e.g. $y_1 = \text{Id}$ [101], [137].
One of the advantages of this approach is that all similarity measures and regularizers from classical image registration are usable straight away. However, one drawback is that a summation like in (24) is rather memory consuming [99] to solve. Therefore, one way of solving is by either fixing the components and solve sequentially for just one of the components at a time. This is non-linear Gauss-Seidel type iteration [7], [100] which is rather slow as the computations can not be done in parallel [136]. Another way is to solve for all components simultaneously, which needs the same amount of memory but may be faster.
The approaches we introduce for multiple image registration in this Chapter, are solving for all components at once. Furthermore, the coupling of the different components within a pair-wise approach is weak – meaning that it is restricted to only two images at once – because of the sequential, pair-wise updates. Especially in applications like the 3D reconstruction of histological serial sections, error propagation can occur due to the pair-wise coupling, which deforms the actual shape of the volume, looking like a straightened banana or cause a z-shift; see banana-effect in e.g., [131]. The use of multiple-image based approaches also shows in practice that we can avoid such an error propagation; see the numerical results in Sec. 4. Nevertheless, the pair-wise sequential approaches are regularly used in practice [62], [63], [85], [86] due to the simplicity of the extension from two to multiple images.
In the following sections, we will elaborate methods to overcome the two-image sequential manner and use all available information from all given datasets. These methods comprise the use of image features as well as the singular values of the image matrix or feature matrix, respectively.

3.2 Features for Image Registration

Everything arranged, we will now combine and extend image features for use with our novel multiple-images approaches. Here, the features can be, e.g., the image intensities itself or even normalized gradient fields to fit the multi-modal

registration case. Since in this work we focus on using singular values to measure image similarity, we define a feature operator F as

$$F : \mathrm{Img}^T \to \mathbb{R}^{d \cdot k \times T},$$

where $d \in \mathbb{N}$ is the spatial dimension and $k \in \mathbb{N}$ a feature specific multiplicator. Specifying F like that, we can use intensity and gradient features within a feature array. As an example, we show how the feature array looks for gradient features of T images $I = (I_t)_{t=1}^T$:

$$F_\nabla(I(x)) = \begin{pmatrix} | & & | \\ \nabla I_1(x) & \dots & \nabla I_T(x) \\ | & & | \end{pmatrix} \in \mathbb{R}^{d \cdot k \times T}.$$

Here, the parameter k is the spatial dimension, since the gradient is taken in each spatial direction for each point $x \in \mathbb{R}^d$. For intensities, the same example would be $F_{\text{intensity}}(I(x)) \in \mathbb{R}^{1 \times T}$, since each point maps to a scalar intensity value, for gray-valued images. Typically in this work, the features are either intensities or normalized gradients, such that we obtain a feature array – as shown above. In this work, we will mainly focus on gradient-based features. This concept may be further generalized, such that other features like SIFT [87], [88] or MIND [53], [54] can also be used.

Normalization of Features

Another compelling objective of the use of image features for image registration is normalization. As indicated in the discussion in [47], it is worthwhile to investigate a normalization strategy of features. In order to overcome a difference in magnitudes of intensity gradients of different modalities, we use two different strategies. In Sec. 3.6, we will also investigate the numerical properties of the strategies. As already seen for NGF in Eq. (5), normalization is used to overcome differences in magnitudes. Our first normalization strategy is a local normalization, also used for the NGF distance measure. For this purpose, we take advantage of the different directional derivatives, which we normalize in a pointwise manner. Note that we consider the unnormalized gradients as features in this work. The corresponding, normalizing norm is denoted by $\|\cdot\|_\eta = \sqrt{\langle \cdot, \cdot \rangle + \eta^2}$ with $\eta > 0$. The second approach is to normalize over the whole feature domain in a global, non-pointwise manner. For ease of presentation, we denote this using the standard $L_2(\Omega)$-Norm $\|\cdot\|_{L_2(\Omega)}$.

Both normalizing approaches mentioned have several drawbacks. Considering intensity gradients as image features, the discussion in [47] points out a high sensitivity to small values. Furthermore, in constant areas of the image, the normalization-term is non-differentiable, which is why it needs to be regularized by $\eta > 0$. This

Local normalization	Global normalization
$\|F(I(x))\|_\eta = \sqrt{\sum_k \|\nabla_k I(x)\|^2 \,\mathrm{d}x + \eta^2}$	$\|F(I)\|_{L_2(\Omega)} \coloneqq \sqrt{\int_\Omega \|\nabla I(x)\|^2 \,\mathrm{d}x}$

Figure 3: Comparison of global and local normalization divisors for gradient-based image features $F(I) = \nabla I$.

introduces an additional parameter that needs to be tuned for optimization. Although a semi-automatic choice of η is proposed in [47], it is still highly dependent on the data.
Both approaches have in common that the magnitude of spatially corresponding features over different modalities are adjusted. The goal is that the adjustment makes the different modalities comparable at the locations of the features. In fact, with both introduced normalizations, the magnitudes are handled differently. The local normalization, like it is used for NGF, scales all local features to unit-length. Here, it is very likely to scale noise with small magnitude also to unit-length, which can be seen as a different reason why the normalization needs to be regularized. In contrast to that, the global normalization approach using the $L_2(\Omega)$-Norm does not change the local magnitude differences. Instead, it scales the magnitudes of different images to $[-1, 1]$ while maintaining the same scaling between features within one image. In this way, the noise is not a significant issue. However, details with small magnitude remain to have just a small overall influence. Because of this, we need to find a compromise. Either we want to use also fine details for exact alignment, but need to care more about the noise level. Alternatively, we do not want to care about the noise level but need to accept that small details may not influence an alignment. A comparison for different choices of η is visualized in Figure 4. The locally normalized gradient images show more noise for small values of η than the globally normalized images. This means the choice of η does not play a major role when using globally normalized gradient features, compared to the use of the local normalization strategy. However, the globally normalized gradient features are not as sensible as the locally normalized features. Edges with low amplitude may not be considered when using globally normalized gradient features, while they can be considered with a careful choice for locally normalized gradient features. Figure 4 also shows, the higher the choice of η is, the more equal both feature-images get. The choice of the normalization and the choice of η are highly dependent on the application. Strategies for choosing η, as pointed out in [47], may be proper guidance for different applications, but they may also need manual interaction. A meaningful, fully automatic choice of η is not within the scope of this work. Numerical issues and differences for the specific choice of

global or local normalization are discussed in Section 3.6.

3.3 Extension for Regularization of Multiple Images

We extend the regularization in a suggestive, straightforward way. Since each image has its respective transformation grid, we use a regularizing term for each transformation grid $y \in \mathbb{R}^d \to \mathbb{R}^d$. The respective energy for all grids $Y = (y_1, \dots, y_T)$ is a summation of all individual energies

$$S(Y) = \sum_{t=1}^{T} \mathcal{S}(y_t - \mathrm{Id}),$$

where $\mathcal{S}$ is one of the introduced two-image regularizers from Sec. 2.4 and Id the corresponding regular grid, such that just the displacement field is regularized. This extension approach is straightforward yet effective since well-known regularizing methods can be combined with multiple images distances. Furthermore, results for the choice of parameters from past studies using two-image registration approaches can be considered for use in a more convenient way. The downside of this choice is that it still has a sequential character to it. Just building the sum of individual regularization terms may regularize the individual transformation grids quite well, but there is no coupling across the images. There are different coupling strategies available – see, e.g., [1], [26] – that need to be explored in combination with distance measurement techniques for multiple images. Since, in this work, we focus on the elaboration of a distance measure framework for multiple images, we do not elaborate on the regularization part. A brief discussion of different concepts can be found in Section 5.

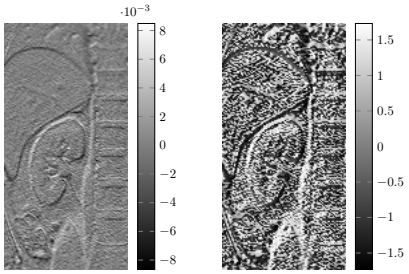
3.4 Schatten-Norms for Global Registration

This section is the heart of this work. We analyze different multiple image distance measurement approaches using Schatten-q-Norms. Throughout this chapter, we show that the different approaches originate from different geometrical ideas like rank-minimization, volume-minimization, or even correlation maximization and, therefore, also combine different fields of applications. Moreover, we analyze the properties of the framework and show compelling details for practical application. Ultimately, we show that all approaches have a common basis, and though they are very versatile.

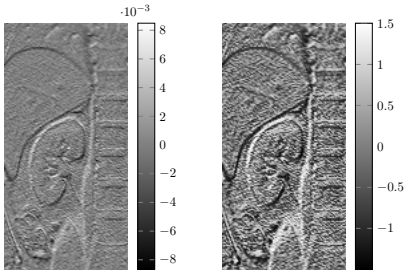
Rank Minimization by Linear Dependence of Image Features

We start analyzing the first of three presented approaches for distance measurement of multiple images by exploiting the linear dependence of images features.

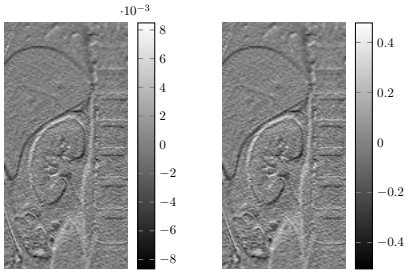
(a) Visualization of a 2D (x,y) section from a 4D (x,y,z,t) DCE-MRI dataset of a human kidney. Data courtesy of Jarle Rørvik, Haukeland University Hospital, Bergen, Norway.



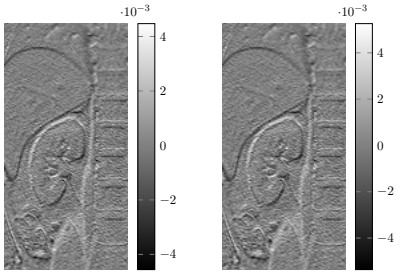
(b) Gradients for $\eta = 10^{-4}$



(c) Gradients for $\eta = 10^{1}$



(d) Gradients for $\eta = 10^{2}$



(e) Gradients for $\eta = 10^{4}$

Figure 4: Illustrated are the accumulated gradient images of a human kidney from DCE-MRI for different η. The images are illustrated as the sum of the components of the gradient. The original 2D slice from the 4D dataset is shown in Fig. 4a. Here, on the left, the globally normalized gradients are illustrated. The images on the right illustrate the respective locally normalized gradients.

The main idea originates in color image regularization, as first introduced by Möllenhoff et al. in [102]. There, the Schatten-q-Norms are used in a TV-like regularization scheme. Here, we build a new similarity measure using singular values of the feature matrix $F(I)$ obtained with the help of the Schatten-q-Norms from Definition 3.
The idea is to create a similarity of images by forcing a linear dependency of the columns of the feature matrix $F(I)$. We control this by manipulating the rank of $F(I)$. For example, by choosing $q = 0$ and assuming $0^0 = 0$, we obtain the rank-function itself from the Schatten-Norm:

$$\|F(I)\|_{S,0} = \sum_{i=1}^{K} \sigma_i^0 = \sum_{i=1}^{\#\{\sigma_i \neq 0\}} i = \operatorname{rank}(F(I)).$$

The number of non-zero singular values determines the rank of F. Therefore we use $\|F(I)\|_{S,q}$ as a similarity measure in order to manipulate $\operatorname{rank}(F(I))$ [7]–[9].

Definition 4 (The Image Similarity Measure $\mathrm{S}q\mathrm{N}_{\mathrm{rank}}$)**.** Let $I \in \mathrm{Img}^T$ be a tuple of images and $F(I)$ a corresponding feature array and $q > 0$. Then, the image similarity measure $\mathrm{S}q\mathrm{N}_{\mathrm{rank}}$ is defined as

$$\mathrm{S}q\mathrm{N}_{\mathrm{rank}}(F(I)) = \mathrm{S}q\mathrm{N}(F(I_1), \ldots, F(I_T)) \coloneqq \|F(I)\|_{S,q} = \left(\sum_{t=1}^{T} \sigma_t(F(I))^q \right)^{\frac{1}{q}}. \tag{25}$$

Here, σ_t, $t = 1, \ldots, T$ are the non-zero singular values of the feature array $F(I)$. Note, $\mathrm{S}q\mathrm{N}_{\mathrm{rank}}$ is the name, where the term rank is just an identifier-term.

However, one of the approaches drawbacks is $F(I) = 0$, for which all singular values become zero. For example, we obtain this by shifting all images out of the domain, such that I and, therefore, $F(I)$ is the all-zero matrix. This particular case is a non-desired global minimum. It is not desired because all image information is dismissed. Nevertheless, with a careful choice of q and the use of a pre-registration scheme, we show in Section 4 that this approach works, and for this particular measure, we instead desire a local minimum. In the following sections, we discuss approaches that resolve this problem.

Alignment by Volume Minimization

Closely related to $\mathrm{S}q\mathrm{N}_{\mathrm{rank}}$ is a volume minimization approach as studied in e.g., [7], [41]. The assumption for such volume minimizing approaches is that the columns of the feature array F span a parallelotope. The spanned volume of this parallelotope is minimal if all the edges collapse and lie on the same line. Building

a correlation matrix $C(I) = F(I)^\top F(I)$, we determine the spanned volume of the parallelotope by computing the determinant $\det(C(I))$ [41], [136]. Since $C(I)$ is quadratic, symmetric, and positive semi-definite, we take the logarithm of the determinant without changing convexity and monotonicity properties. Finally, we obtain the volume minimizing distance measure by

$$D_{\mathrm{vol}}(I) \coloneqq \log(\det(C(I))) = \log\left(\prod_{t=1}^{T} \sigma_t^2(F(I))\right) = 2\sum_{t=1}^{T} \log(\sigma_t(F(I))). \tag{26}$$

This approach has several drawbacks which are also already discussed in [47]. Nevertheless, we briefly summarize them here and discuss an issue using the logarithm. An illustration of this problem is given in [7] by considering the volume of two vectors $\mathrm{vol}(v_1, v_2) = \|v_1\|\|v_2\| \sin\alpha$, where $v_1 \neq 0$ and α is an angle. This is minimal if both vectors are linear dependent, which is the case for $v_2 = 0$. Also the zero-solution discussed for $\mathrm{S}q\mathrm{N}_{\mathrm{rank}}$ is a problem. Furthermore, if the parallelotope spanning vectors collapse, some of the singular values vanish. This leads to issues using the logarithm. Because of this, it is worthwhile to think about regularization strategies for vanishing singular values. In [47] it is suggested to overcome this problem by maximization of a correlation instead of a volume minimization. For this case the sinus-function is replaced by a $|\cos|$ and is maximal iff and only iff $v_2 = \pm v_1$ [7], [47].

Alignment using Correlation of Features

Another way of defining a similarity measure is to use the correlation of normalized features $\|F(I_k)\| = 1$, where the norm for normalizing the features is either the L_2-Norm or the η-Norm as shown in Sec. 3.2. Building a correlation matrix $C(I) \in \mathbb{R}^{T\times T}$, we obtain the following properties; cf. [7]:

$$C_{kk} = 1, \quad C_{jk} = \langle F(I_j), F(I_k)\rangle = \cos\gamma_{jk}, \tag{27}$$

where γ_{jk} denotes the angle between the j-th and the k-th feature. The idea of maximizing this correlation is well-known from the two-image setting, namely by the Normalized Cross Correlation (NCC) defined in Eq. 4. The best case scenario is a 2-by-2 matrix filled with ± 1, the worst case the 2-by-2 identity matrix. This is straightforwardly extended to multiple images, such that the best case is $C(I) \in \{\pm 1\}^{T\times T}$ and the worst case $C(I) = \mathrm{Id}$. Hence, an appropriate way of similarity measurement is to take the distance between the correlation matrix and the worst case scenario, the identity matrix using a suitable matrix norm $\|\cdot\|_M$ [7]:

$$D_{\mathrm{corr}}(I) \coloneqq \|C(I) - \mathrm{Id}\|_M. \tag{28}$$

In our case, we combine Eq. (28) with the Schatten-Norms from Definition 3 as many other norms are only special cases of the Schatten-Norms for particular choices of q

$$D_{\mathrm{S}q\mathrm{N}}(I) = \|C(I) - \mathrm{Id}\|_{S,q} = \left(\sum_{k=1}^{T} (\sigma_k^2(F(I)) - 1)^q \right)^{\frac{1}{q}} \tag{29}$$

for which we will investigate some special cases and their properties. Especially interesting for the numerical results of this work are the cases for the parameter $q = 2$ and $q = \infty$. On the one hand, SqNfor $q = 2$ can be rephrased, such that we may use C or also F. On the other hand, SqNfor $q = \infty$ has interesting properties, which we discuss in the following. Furthermore, our numerical results in Sec. 4 show better numbers for these two cases.
Alternatives are, e.g., the Frobenius Norm, which is a special case of the Schatten-Norms as indicated in Thm. 1. Besides that, the L_1-Norm and the L_∞-Norm are alternatives. Using the Schatten-Norms allows us to maintain a generality that covers many choices like the Frobenius Norm, the Nuclear Norm, or a Maximum-Norm. In Sec. 3.6, we show numerical results for different choices of q that cover some of the mentioned special cases.
The Schatten-2-Norm is the quadratic mean of the correlation among the image features [7]. In fact, for $q = 2$ we have

$$D_{\mathrm{S}q\mathrm{N}}(T) = \|C(T) - \mathrm{Id}\|_{S,2}^2 = \sum_{i \neq j} |C_{ij}|^2.$$

Theorem 2. For a correlation matrix $C = F^\top F \in \mathbb{R}^{T \times T}$ with normalized feature arrays $F \in \mathbb{R}^{n \times T}$ it holds

$$\|C - \mathrm{Id}\|_{S,2}^2 = \|F\|_{S,4}^4 - T.$$

Proof. With Corollary 1 we can rephrase:

$$\|C - \mathrm{Id}\|_{S,2}^2 = \mathrm{tr}((C - \mathrm{Id})^\top (C - \mathrm{Id}))$$

Using the linearity of the trace and the fact that the diagonal $C_{ii} = 1$ with $i \in \{1, 2, \ldots, T\}$, we have

$$\mathrm{tr}(C^\top C) - \mathrm{tr}(C^\top) - \mathrm{tr}(C) + \mathrm{tr}(\mathrm{Id}) = \mathrm{tr}(C^\top C) - T.$$

Again with Corollary 1, we obtain

$$\mathrm{tr}(C^\top C) - T = \|C\|_{S,2}^2 - T = \|F^\top F\|_{S,2}^2 - T = \|F\|_{S,4}^4 - T.$$

□

Using the formulation of Eq. (29), we have a maximization problem for achieving maximum similarity. Therefore, we rephrase (29) to obtain an equation that can be minimized in order to obtain maximum similarity of images:

$$D_{\mathrm{S}q\mathrm{N}}(T) = -\|C(T)\|_{S,q}. \tag{30}$$

From Thm. 2, we can see that T can be canceled out and does not play an essential role during optimization. Based on Eq. (30), we define a weighted version.

Weighted Correlation of Features

Not only the close relationship to the rank-minimization and volume-minimization approaches is a positive argument to use this approach. We further extend the novel approach using a weighting matrix $W \in \mathbb{R}^{T\times T}$. For example, such a weighting matrix allows controlling how many images are considered for similarity computation. This extension opens up even more possibilities for registration. It even allows registration in a pairwise manner, comparing two images and a groupwise manner more or less globally, comparing more than two images. Therefore, we can use one unified approach and only need to choose how many images should be used for comparison and features appropriate for the application.
According to Equation (29), by appending $W \in \mathbb{R}^{T\times T}$ to the equation, we obtain

$$D_{\mathrm{S}q\mathrm{N}}(W, I) = -\|W \odot C(I)\|_{S,q}. \tag{31}$$

Note, that the product indicated by $\odot$ is the Hadamard-Product multiplying the elements of both matrices element-wise. Choosing W as a matrix consisting of all ones, we obtain the SqN approach defined in Eq. (30).
Choosing e.g.

$$W = \begin{pmatrix} 1 & 1 & 0 & \dots \\ 0 & 1 & 1 & \\ \vdots & & \ddots & \\ 0 & & & 1 \end{pmatrix}$$

by elementwise multiplication with C we obtain a bi-diagonal matrix. This matrix consists of all ones on the main diagonal and $\langle F(I_j), F(I_k)\rangle$ on the first superdiagonal, where $j = 1, \dots, T-1$ and $k = j+1$. Concludingly, this gives a pairwise registration approach. The only difference here is that none of the images is fixed, as it is for the standard two-image approaches. Furthermore, the number of images considered for similarity computation may be varied for different slices or time points or even on different levels, when using a multi-level approach. It is also conceivable that different weightings like a standard distribution or a linear decay may be useful in different applications. We refer also to the discussion (Sec. 5).

3.5 General Framework and Important Properties

We are now ready to give a final, comprehensive definition for our novel framework now that we have studied different ways of defining a distance measure for image registration of multiple images.

Definition 5 (SqN Distance Measure)**.** Given a tuple of images $I \in \mathrm{Img}^T$, $T \in \mathbb{N}$ and $q > 0$, the function

$$D_{\mathrm{S}q\mathrm{N}}(I) = -\|W \odot C(I)\|_{S,q}$$

is called SqN distance measure. Here, $W \in \mathbb{R}^{T \times T}$ is a specific weight matrix and $C(I) = F(I)^\top F(I)$ is the correlation-matrix of features $F(I) \in \mathbb{R}^{n \times T}$. Note, that $\odot$ denotes the Hadamard product.

Typically, we choose $W = \mathbf{1}\mathbf{1}^\top$, where $\mathbf{1} \in \mathbb{R}^T$ is a vector of all ones. With Definition 5, we have a versatile framework for measuring image similarity of multiple images that merges all introduced ideas into one equation. Using W, we can control how many images are considered for comparison. This also comprises a specific weighting that can control the influence of specific features. Furthermore, due to a free choice of image-features, we can control how the images are compared to each other. The given examples include intensity images as well as gradient features which are differently normalized. Note that the features need to be normalized in order to fulfill the needs of the correlation matrix. A discussion on the normalization is found in Sec. 3.2. Compared to the two-images methods introduced in Sec. 2.1, we do not need to choose a specific reference image. Due to the use of the Schatten-Norms – or, more precisely, the singular values of the correlation matrix – the ordering of images does not play a role for the registration. This might be handy for applications where a specific order is not needed, e.g., for building an atlas. Other applications like dynamic imaging or histological sectioning have an intrinsic ordering. The given order can be used as-is without a problem. Although, the fact that there is a specific ordering is not be exploited as it might be the case in pairwise registration cases. There, the ordering plays a major role and the comparison is completely built upon the comparison of neighbouring images. Still, our framework is very versatile. Different choices of image-features, the specific choice of W, and the variability in formulation due to the use of singular values allow the interpretation of the framework as an independent distance measure or as a generalization of different two-image measures like NGF. For the choice of features, we suggest either image intensities or image gradients, as we discuss these within this thesis. We base the choice of q on the discussed variants of our novel measure SqN. Thus, the standard choices might be $q = 0.5$, as suggested by Möllenhoff et al. in [102]. Furthermore, as we mainly use it in this thesis, $q = 2$, $q = 4$ as

well as $q = \infty$ are choices for successful registration results, as we show in Sec. 4. This makes the framework straightforward to interpret as well as straightforward to implement. We discuss more important invariances and properties in the course of the next subsections.

Geometrical Interpretation and Related Work

In this section, we give some geometrical interpretations of the different singular value-based approaches. Exemplary, we take normalized gradients as image features. Referring to [47], the alignment of gradient-based features is achieved by aligning the gradient directions. This corresponds to a linear dependency of the set of gradient vectors, which is based on the assumption that a pair of vectors form an angle in between [47]. In the case of Definition 4, the rank minimization of the feature matrix achieves a linear dependency of the columns. In this case, the columns are the feature vectors. The use of such gradient features is well-known from NGF [47]. Thus, such features are especially useful in a multi-modal image registration scenario. This idea was first used in color image regularization; see [102]. For color image regularization, the idea is that the gradients should have the same direction in each color channel. If there is a variation between the channels, color noise occurs. Taking a matrix of gradients and minimizing its rank, e.g., using Schatten-Norms, achieves a coupling of the gradients directions. The common idea of these approaches is illustrated in Fig. 5.

One might depict the scenario of rank minimization for image registration as reducing all information to one main component. The idea is quite similar to a Principal Component Analysis, as a data set might be explained with only the principal component, e.g., for compression reasons. There are image registration approaches using variations of principal component analysis for image registration. In [43], [50], for example, the robust PCA is used to split dynamic imaging data into a part containing the motion and another part containing the anatomical information. The idea is to register the motion part while keeping the anatomical information as is. In the end, the registered motion component and the untouched anatomy component can be recombined to achieve a fully registered data set. All these singular value-based approaches have in common that a linear dependency of either feature vectors or principal components is taken as a basis. The idea of the PCA is a good starting point, as it can be interpreted as a fitting of the orthogonal principal axes of an ellipsoid to a given data set. Inspired by that, we might see the image registration based on singular value based distance measures as the approach to collapse the principal axes by transforming the data. Therefore, the images are represented by the principal axes.
Further, it is worth investigating some particular choices of q for the SqN-approach,

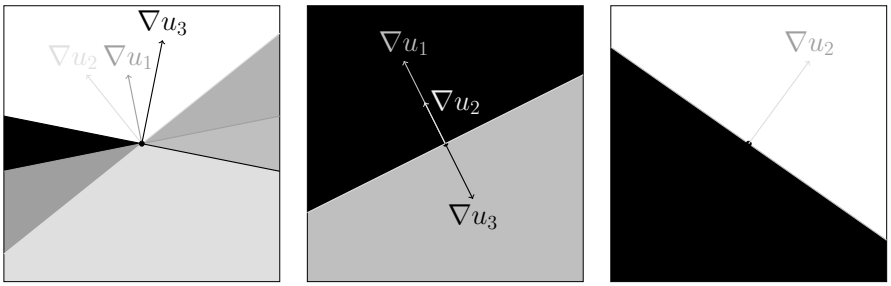


Figure 5: Illustration of a local gradient matrix $A = [\nabla u_1, \nabla u_2, \nabla u_3] \in \mathbb{R}^{2,3}$ of three color channels; illustration adapted from [102]. The rank of A is two (left) or one (center and right). The color-channels are illustrated gray-valued intentionally.

as those indicate connections to other well-studied norms used in image processing. Especially $q = 1$, which is the nuclear norm, as well as $q = 2$, which is the Frobenius norm, are commonly used in image processing. As the nuclear norm is used in a variety of regularization strategies in image processing; see e.g. [43], [102], the connection to the Frobenius norm is most interesting for us, since

$$q = 2 : \qquad \|F\|_{S,2} = \|F\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_K^2}.$$

This draws a direct connection to the NGF distance measure [47] from Eq. (5). Theorem 2 illustrates how both approaches SqNand NGF are connected and NGF can be understood as a special case of SqN.

Invariances and Energies

The behavior of the distance-energy concerning the choice of q is noteworthy, as q is an important parameter to SqN. The parameter q is chosen as $0 \leq q \leq \infty$. As suggested in Section 3.4, $q = 4$ is a noteworthy choice when using just the feature matrix F instead of the correlation matrix C for computation of SqN. In Fig. 8- 10 we show the energies for affine linear transformations of an image against itself. For this experiment, we choose a 2D section from a DCE-MRI data set of a human kidney, as illustrated in Figure 27 on page 75. Illustrated are the energies for $q \in \{0.5, 1, 4, \infty\}$. Note that the choices of q are based on SqNusing the feature matrix F instead of the correlation matrix C, such that $q = 4$ for the use with F is equal to $q = 2$ for the use with C. The choice of these specific q is based on the geometrical ideas discussed above. For comparison, the energy behavior of NGF is illustrated in Fig. 11. In particular, the translation plots make it obvious that for $q < 2$, the trivial solution is the best case. The respective Figures 7a and 8a show a behavior where the values near the boundaries of the considered domain are much lower than the perfect match at position zero. Each

time the transformed image is translated out of the respective image domain, the fixed images are compared to an image of zeros. This is not the case for $q > 2$, as we can see from Figures 9a and 10a. This is due to the change of signs. For $q > 2$ the perfect match is in a maximum situation as discussed in Sec. 3.4. Apart from the translation, each case has its global minimum at location zero, where the perfect match of both images is. Additionally, we see that the energy-plot of SqN is much smoother than the energy of, e.g., the NGF distance measure or comparable approaches for multiple images. Details on the results of the affine linear transformation experiment are found in the Example below.

A noteworthy problem of registration approaches that transform all images and do not employ a fixed reference image is a non-uniqueness of the solution. Moreover, the trivial solution might be the best case concerning the energy that is to minimize. A typical problem that can be observed is the shifting of images out of the domain, such that they become zero everywhere. Often, multiple image approaches are used with additional constraints concerning the uniqueness of a solution; see e.g., [5], [42], [43], [70], [98], [111]. Typically, a constraint is deployed that demands the mean of all deformations to be zero in all grid points [5]. A related approach is to force the sum of all deformations to be zero in each coordinate direction [43]. In theory, it still is necessary to deploy such a constraint as an additional regularizer for use together with SqN, as the regularizers presented in Sec. 2.4 do not handle this problem. The need for such an additional regularizing term is due to the absence of a specific reference image, which gives the approach a certain degree of freedom. If there is no reference image, the images may be aligned in any arbitrary position within the image domain. In practice, we did not notice any shifting-issues for SqN with $q > 2$ since we typically choose a rather high regularization parameter, such that a proper regularization is provided, as we can also see in Sec. 4. Furthermore, with an additional linear pre-registration routine – as summarized in Sec. 2.1 – the images become more similar such that the starting-guess of the respective SqN-energy for a non-linear registration routine is close enough to a minimizer, such that these issues do not play a prominent role in practice. A desirable minimizer is found using the stopping criteria by Gill, Murray and Wright [35] summarized in Sec. 2.7.
Another property of the Schatten-Norm-based distances is its permutational invariance, which allows a random ordering of the images without changing the registration result. This property can be proved by using the properties of the underlying SVD. Since the Schatten-Norms are unitarily invariant and permutation matrices are unitary, we can permute the order of the registered images or respective feature maps without changing the result. As shown in [43], a straightforward one-liner proofs this for arbitrary matrix $A \in \mathbb{R}^{n\times K}$ with SVD $A = U\Sigma V$

and permutation matrix $P \in \{0,1\}^{n\times n}$:

$$(PU)^\top (PU) = U^\top P^\top PU = U^\top U = \mathrm{Id}.$$

The singular values do not change with the permutation of the columns. Thus, the values of the Schatten-Norms do not change, which we also see in practice in section 4. This proofs the invariance to the ordering of the images for computing the SqN distance measure.
Also, the normalization of the features plays a significant role in the energy-behavior. Figure 9 illustrates the affine linear experiments for global normalization in the upper row and local normalization in the lower row. The used features are normalized gradients. The y-axis indicates that the energy levels and the smoothness of the energy plots are quite different. These differences indicate the discussed properties from Section 3.2. Moreover, it can be relevant in practical use, as shown in different application scenarios in Section 4.

Example for q under Affine Linear Transformations

This example illustrates the energy behavior of SqN for different q under affine linear transformations of a 2D DCE-MRI image section against itself. The used image is illustrated in Figure 6. We use translation, rotation, scaling as well as shearing of one of the images. The other image is fixed. The experiment is inspired by an analog experiment published in [43]. For each transformation type, we use 51 different positions which are interpolated linearly. Furthermore, we show the resulting graphs for globally as well as locally normalized gradient features. For normalization the used norm is regularized using the parameter $\eta = 10^{-4}$. We transform the images in full image resolution without any smoothing. For the transformation, the moving image moves from the bottom left corner to the upper right corner of the fixed image domain. Starting in the bottom left corner, the moving image is completely outside the fixed images domain also ending completely outside the domain in the upper right corner. This results in a comparison with all zeros. This is reflected in the graphs as an outstanding plateau at the left and right boundaries of the shown graph having value 1 or -1 in the case of globally normalized gradient features. The graphs of the locally normalized gradient features show an analog behaviour while having a smoother appearance. The graphs are illustrated in Fig. 8 - 10. This behavior is especially unfavourable in the case of $q < 2$ as the resulting value is obviously lower than the minimum for a perfect match of both images. For $q \geq 2$ this problem is fixed due to the properties outlined in Sec. 3.4. Similar results can be seen for the results using rotation around the image center as well as for the scaling and the shearing of the moving image. Note that the moving image is scaled in both x- and y-directions using the same scaling-factor. The shearing is only executed in

x-direction as a similar result is to be expected for shearing in y-direction. Most of the graphs show a rather flat course from the boundaries to the minimum. In some, the graph is even dropping towards the boundaries. For $q = 4$ and $q = \infty$ this behaviour is a bit less than for the lower q but can still be recognized. From these results, we may conclude that $q \geq 2$ may be interesting, as there is no drop towards a vanishing image. Furthermore, for all q, we seem to need a scenario in that the images are already quite similar in order to find the desired minimum. Meaning, we need to start close enough to the desired minimizer in order to find it. For the computed examples, it is to be expected, that a smoother graph can be obtained by smoothing the images on different scales or using a multi-level approach, as claimed in [46]. In comparison with the graphs for NGF shown in Fig. 11, the graphs for SqN seem to be rather smooth. For NGF, a drastically smoother behaviour is to be expected for larger η as well as for smoothed images. Besides, we can see from the graphs for NGF that there is no noticeable drop towards the boundaries.

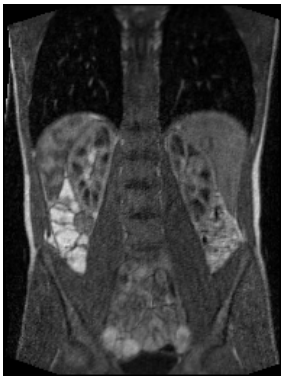
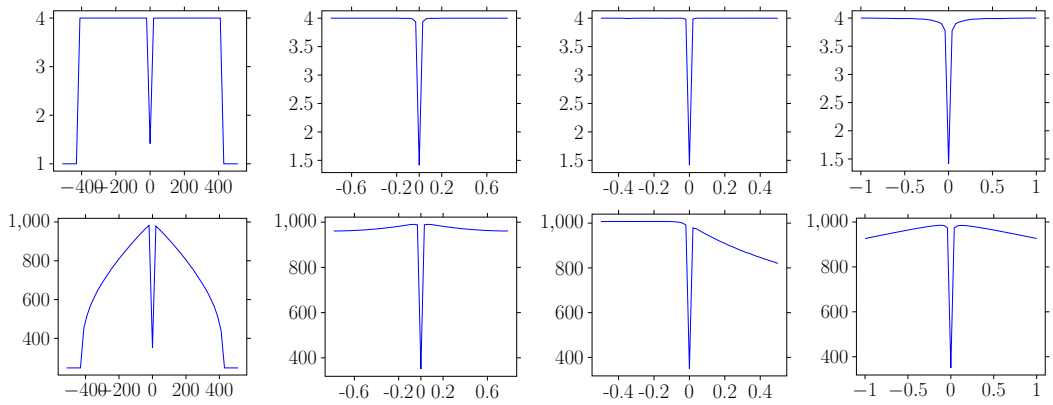
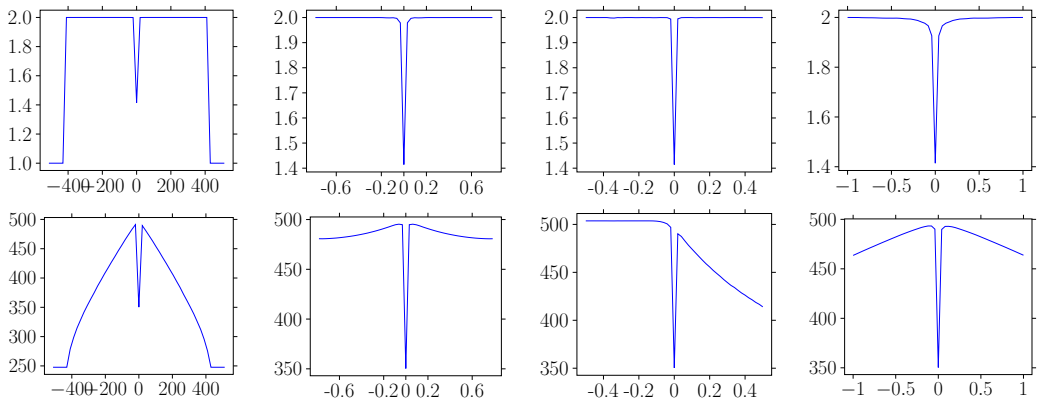


Figure 6: Visualization of 4D (x,y,z,t) DCE-MRI data. A representative axial slice (x,y) at a representative time t is shown. The image is used for the affine linear transformation experiment summarized in Example 3.5. Data courtesy of Jarle Rørvik, Haukeland University Hospital, Bergen, Norway.



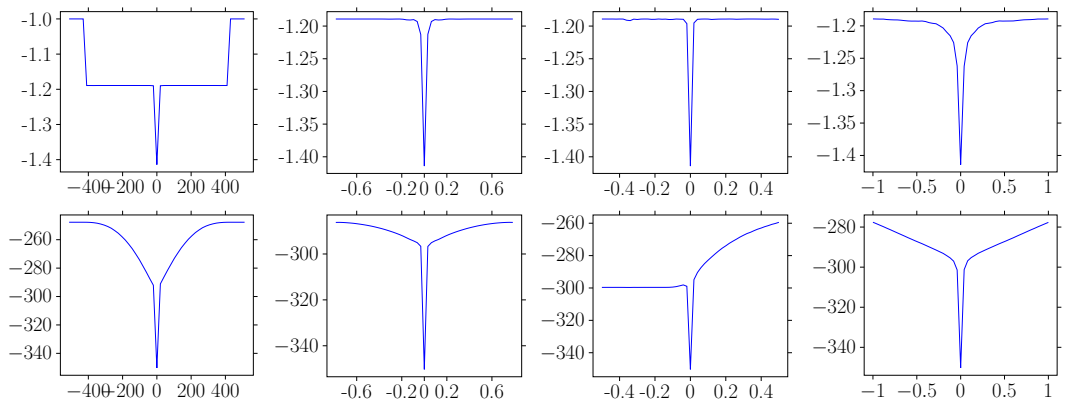
(a) Translation (b) Rotation (c) Scaling (d) Shearing

Figure 7: Illustrated are the energies of the affine linear experiments for SqN with $q = 0.5$ and normalized gradient features. The upper row shows the energies for globally normalized features. The energies in the lower row are computed using locally normalized gradient features, respectively.



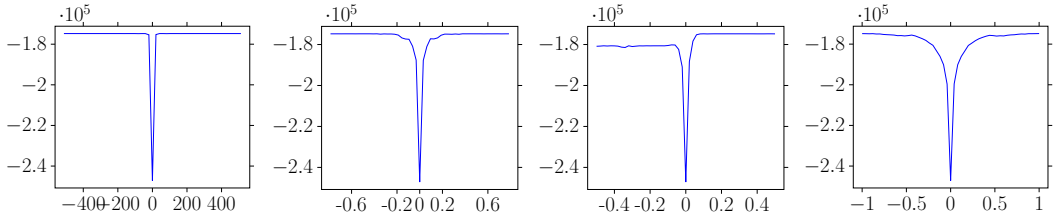
(a) Translation (b) Rotation (c) Scaling (d) Shearing

Figure 8: Illustrated are the energies of the affine linear experiments for SqN with $q = 1$ and normalized gradient features. The upper row shows the energies for globally normalized features. The energies in the lower row are computed using locally normalized gradient features, respectively.



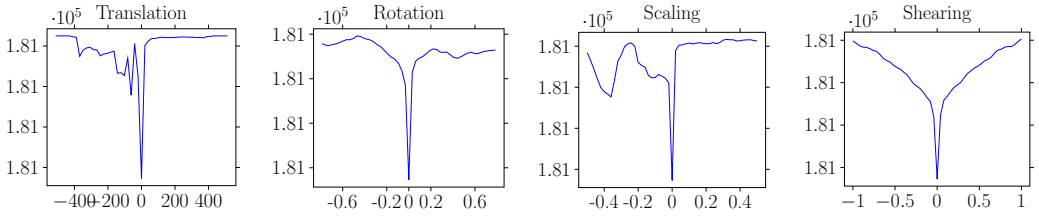
(a) Translation (b) Rotation (c) Scaling (d) Shearing

Figure 9: Illustrated are the energies of the affine linear experiments for SqN with $q = 4$ and normalized gradient features. The upper row shows the energies for globally normalized features. The energies in the lower row are computed using locally normalized gradient features, respectively.



(a) Translation (b) Rotation (c) Scaling (d) Shearing

Figure 10: Illustrated are the energies of the affine linear experiments for SqN with $q = \infty$ and normalized gradient features. The energies are computed using locally normalized gradient features.



(a) Translation (b) Rotation (c) Scaling (d) Shearing

Figure 11: Illustrated are the energies of the affine linear experiments for NGF. Details on the rough appearance of the graphs are summarized in Example 3.5. Smoother graphs are to be expected for larger η or smoothed images.

3.6 Discretization of Schatten-Norm-based Distances

We are now set for discretizing the Schatten-Norm-based distance measures. For this purpose, we take already computed parts, rephrase them as needed, and compose the discretized Schatten-Norm-based measures. We elaborate the discretization on the SqN-definition, using only the feature matrix

$$\mathrm{S}q\mathrm{N}(I, q, F) = \|F(I)\|_{S,q} = \left(\sum_{t=1}^{T} \sigma_t(F(I))^q \right)^{\frac{1}{q}},$$

since it is straightforward to refomulate to the correlation-based approach from (29) as noted in Theorem 2. For ease of presentation, we set $W = \mathbf{1}\mathbf{1}^\top$, as it just represents a factor for each element that does not change the discretization. Exemplarily, we show the discretization for NGF-based features. The used features can be interchanged with other discretized features if needed. For our case, we claim Theorem 3.

Theorem 3 (Discretization of SqN)**.** Let $D_{\mathrm{S}q\mathrm{N}}(I) = \|W \odot \left(F(I)^\top F(I)\right)\|_{S,q}$ be the introduced distance measure, defined in Definition 5. Given a tuple of images $I \in \mathrm{Img}^T, T \in \mathbb{N}$ with $q > 0$, a feature matrix consisting normalized features $F(I) \in \mathbb{R}^{n \times T}$ and a weighting matrix $W \in R^{T \times T}$. Without loss of generality, we set $W = \mathbf{1}\mathbf{1}^\top$ and use Theorem 2, such that we can use only $F(I)$ instead of $F(I)^\top F(I)$. Then, a discretization of the simpliedfied term

$$\mathrm{S}q\mathrm{N}(I, q, F) = \|F(I)\|_{S,q} = \left(\sum_{t=1}^{T} \sigma_t(F(I))^q \right)^{\frac{1}{q}}$$

may be written as

$$\mathrm{S}q\mathrm{N}_{\mathrm{discrete}}(I, q, F) = \left(\sum_{t=1}^{T} \sigma_t^q \Big(F((I \circ Y)(x^c)) \Big) \right)^{\frac{1}{q}}.$$

Proof. Beginning, each column of the feature matrix is discretized, like the template image T of the NGF-distance. In this way, we obtain

$$F(I) = \begin{pmatrix} \frac{\partial I_1(y_1(x_1^c))}{\|(\partial I_1(y_1(x_1^c)))\|_\epsilon} & & \frac{\partial I_T(y_T(x_1^c))}{\|(\partial I_T(y_T(x_1^c)))\|_\epsilon} \\ \vdots & \dots & \vdots \\ \frac{\partial I_1(y_1(x_m^c))}{\|(\partial I_1(y_1(x_m^c)))\|_\epsilon} & & \frac{\partial I_T(y_T(x_m^c))}{\|(\partial I_T(y_T(x_m^c)))\|_\epsilon} \end{pmatrix} \in \mathbb{R}^{d \cdot m \times T}.$$

Now applying the SVD while taking advantage of the unitary invariance of the Schatten-Norm, we get

$$\begin{aligned}\|F(I)\|_{S,q} &= \|U\Sigma V^\top\|_{S,q} = \|\Sigma\|_{S,q} \\ &= \left(\sum_{t=1}^{T} \sigma_t^q \left(\left(\begin{array}{ccc} \frac{\partial I_1(y_1(x_1^c))}{\|(\partial I_1(y_1(x_1^c)))\|_\epsilon} & & \frac{\partial I_T(y_T(x_1^c))}{\|(\partial I_T(y_T(x_1^c)))\|_\epsilon} \\ \vdots & \dots & \vdots \\ \frac{\partial I_1(y_1(x_m^c))}{\|(\partial I_1(y_1(x_m^c)))\|_\epsilon} & & \frac{\partial I_T(y_T(x_m^c))}{\|(\partial I_T(y_T(x_m^c)))\|_\epsilon} \end{array}\right)\right)\right)^{\frac{1}{q}}\end{aligned}$$

which gives for normalized features the discretization

$$\mathrm{S}q\mathrm{N}_{\mathrm{discrete}}(I, q, F) = \left(\sum_{t=1}^{T} \sigma_t^q\Big(F((I \circ Y)(x^c))\Big)\right)^{\frac{1}{q}}.$$

This result doesn't contain the factor hd, as seen in the other discretization formulas, which is due to the defined normalization of the features. Note that x^c is a cell-centered grid with which we use the notation for the discretization noted in Section 2.6 to obtain the vectors in each column. □

3.7 Derivatives of Normalized Features and Singular Values

As discussed in Section 2.7 we aim for second-order optimization methods. As we focus on Newton-type methods that approximate the second derivative, we only need to compute the first derivative of the objective function. We can use the results from [100] for the derivative of the normalized gradient fields to show the complete derivative of SqN with normalized features using the chain rule.
In this chapter, we use the NGF distance measure derivative results from [100] to build the numerical derivative of SqN. Furthermore, the transformations derivatives, as well as the derivative of the singular values, are needed. The latter is also summarized in this section, relying on [107]. For all results on the derivative of the normalized gradient fields, we refer to [47], [100] and references therein.
Consider the normalized gradients $g : \mathbb{R}^d \to \mathbb{R}$ with transformation fields $y : \mathbb{R}^d \to \mathbb{R}^d$ and features $f : \mathbb{R}^d \to \mathbb{R}$ as $g(y) = \frac{f(y)}{\|f(y)\|_2}$. Following we use the suggestive abbreviations $\mathrm{d}f(y)$ as substitute for $\mathrm{d}f(y) \coloneqq \frac{\partial}{\partial y} f(y)$ and $y = y(x^c) \in \mathbb{R}^{d\cdot m\times 1}$, such that $g(y(x^c)) \in \mathbb{R}^{m\times 1}$. Note, that the vectorization of y and g in the latter cases is due to the definition of the cell-centered grid, we elaborated in Eq. (8). We now derive the normalized gradients as

$$\begin{aligned}\frac{\partial}{\partial y}g(y) &= \frac{\partial}{\partial y}\frac{f(y)}{\|f(y)\|}\\ &= \frac{\partial}{\partial y}\left(f(y)\|f(y)\|^{-1}\right)\\ &= \frac{\partial}{\partial y}\left(f(y)\left(f(y)^{\top}f(y)\right)^{-\frac{1}{2}}\right)\\ &= \mathrm{d}f(y)\left(f(y)^{\top}f(y)\right)^{-\frac{1}{2}} + f(y)\left(-\frac{1}{2}\right)\left(f(y)^{\top}f(y)\right)^{-\frac{3}{2}}2f(y)^{\top}\mathrm{d}f(y)\\ &= \frac{\mathrm{d}f(y)}{\|f(y)\|} - \frac{f(y)f(y)^{\top}\mathrm{d}f(y)}{\|f(y)\|^3}\\ &= \left(\mathrm{Id} - g(y)^{\top}g(y)\right)\frac{\mathrm{d}f(y)}{\|f(y)\|}.\end{aligned} \tag{32}$$

Since the normalized gradient fields are an inner part of the composition for the SqN distance, we use the chain rule. Consider σ as the corresponding singular value for one of the feature columns f of the feature matrix F. Again, we use $\mathrm{d}\sigma$ as a suggestive abbreviation. Then we obtain

$$\begin{aligned}\frac{\partial}{\partial y}\sigma(g(y)) &= \frac{\partial}{\partial y}\sigma\left(\frac{f(y)}{\|f(y)\|}\right)\\ &= \mathrm{d}\sigma\left(\frac{f(y)}{\|f(y)\|}\right)(1 - g(y)g(y)^{\top})\frac{\mathrm{d}f(y)}{\|f(y)\|}.\end{aligned} \tag{33}$$

An essential part in order to compute the derivative of SqN is the derivative of singular values. For all the following results, we refer to [107]. We start over with the SVD of the feature matrix F, which gives $F = U\Sigma V^{\top} = \sum_{t=1}^{T}\sigma_t U_t{}^{\top}V_t$. Note, U and V are orthonormal matrices. Here, f_{ij} denotes a single entry of the feature matrix F. Now the component-wise derivative of the SVD gives [107]

$$\frac{\partial F}{\partial f_{ij}} = \frac{\partial U}{\partial f_{ij}}\Sigma V^{\top} + U\frac{\partial \Sigma}{\partial f_{ij}}V^{\top} + U\Sigma\frac{\partial V^{\top}}{\partial f_{ij}}.$$

It holds that $\frac{\partial f_{kl}}{\partial f_{ij}} = 0$ for all $(k,l) \neq (i,j)$, while $\frac{\partial f_{ij}}{\partial f_{ij}} = 1$. Following [107] a suggestive abbreviation gives

$$U^{\top}\frac{\partial F}{\partial f_{ij}}V = \Omega_U^{ij}\Sigma + \frac{\partial \Sigma}{\partial f_{ij}} + \Sigma\Omega_V^{ij} \tag{34}$$

with $\Omega_U^{ij} = U^\top \frac{\partial U}{\partial f_{ij}}$ and an analog expression for V. Here, Ω_U^{ij} is an anti-symmetric matrix. From this, the relation [107]

$$(\Omega_U^{ij})^\top + \Omega_U^{ij} = \frac{\partial U^\top}{\partial f_{ij}} U + U^\top \frac{\partial U}{\partial f_{ij}} = 0$$

is obtained because of the orthogonality of U. The same principle applies for V. From (34) and with the diagonal matrix Σ as well as the anti-symmetric matrices Ω_U^{ij} and Ω_V^{ij}, the derivatives of the singular values are obtained by [107]

$$\frac{\partial \sigma_k}{\partial f_j} = u_i v_{jk}. \tag{35}$$

We are now ready to elaborate on the derivative of the SqN based distance measures. Using (33) and (35) we compose the first derivative to

$$\begin{aligned}
&\frac{\partial}{\partial y_t} \|F(I \circ Y)\|_{S,q} \\
&= \frac{\partial}{\partial y_t} \left(\sum_{t=1}^{T} \sigma_t (F(I))^q \right)^{\frac{1}{q}} \\
&= \frac{1}{q} \left(\sum_{t=1}^{T} \sigma_t (F(I \circ Y))^q \right)^{\frac{1}{q}-1} \sum_{t=1}^{T} \frac{\partial}{\partial y_t} \sigma_t (F(I \circ Y))^q \\
&= \frac{1}{q} \left(\sum_{t=1}^{T} \sigma_t (F(I \circ Y))^q \right)^{\frac{1}{q}-1} \sum_{t=1}^{T} q \sigma_t (F(I \circ Y))^{q-1} \frac{\partial}{\partial y_t} \sigma_t (F(I \circ Y)) \\
&= \left(\sum_{t=1}^{T} \sigma_t (F(I \circ Y))^q \right)^{\frac{1}{q}-1} \sum_{t=1}^{T} \sigma_t (F(I \circ Y))^{q-1} \cdot \\
&\quad \cdot \frac{\partial}{\partial F} \sigma_t (F(I \circ Y)) \frac{\partial}{\partial f_t} F(I \circ Y) \frac{\partial}{\partial I_t} f_t \frac{\partial}{\partial y_t} I_t.
\end{aligned} \tag{36}$$

Here, $\frac{\partial}{\partial F} \sigma_k (F(I \circ Y)) \frac{\partial}{\partial f_t} F(I \circ Y) \frac{\partial}{\partial I_t} f_t$ is the part, we elaborated in (33). We write (33) for a feature matrix $F \in \mathbb{R}^{m \times T}$, discretized on a cell-centered grid, with m grid-points and T feature vectors explicitly for one feature vector. With (35), using the columns of U and V as $u_t \in \mathbb{R}^{m \times 1}, v_t \in \mathbb{R}^{T \times 1}$ and the normalized feature vectors $g = \frac{f}{\|f\|} \in \mathbb{R}^{m \times 1}$, we get

$$\underbrace{\underbrace{\frac{1}{\|f\|} (\partial_m^h)^\top}_{\mathbb{R}^{d \cdot m \times m}} \left(\mathrm{Id} - \underbrace{g^\top g}_{\mathbb{R}^{m \times m}} \right) \underbrace{u_t v_t^\top}_{\mathbb{R}^{m \times T}}}_{\mathbb{R}^{d \cdot m \times T}} =: D_m^h, \tag{37}$$

where $\mathrm{Id} \in \mathbb{R}^{m \times m}$ is the identity of dimension $m \times m$. By multiplication from the right-hand-side with the vector of singular values $\mathrm{diag}(\Sigma) = (\sigma_1, \dots, \sigma_T)^\top \in \mathbb{R}^{T \times 1}$, we can directly compute the latter sum from (36). We then obtain

$$\partial_{I_t} \|F\|_{S,q} = \left(\sum_{t=1}^{T} \sigma_t(F)^q \right)^{\frac{1}{q}-1} D_m^h \mathrm{diag}(\Sigma) \in \mathbb{R}^{d \cdot m \times 1} \,. \tag{38}$$

We did not consider the part $\frac{\mathrm{d}I}{\mathrm{d}y}$, deriving I_t with respect to y_t. This is handled by multiplying the latter result with the numerical derivative of the images I_t with respect to the corresponding grid y_t. This specific part of the full derivative is, for example, computed using finite differences; see, e.g., [100].

Differences between Derivatives for Locally and Globally Normalized Features

As summarized in Section 3.2, features can be normalized locally or globally. A local normalization requires more than just a single value at the location of one grid point. An example are spatial gradients. At one grid point, mulitple values are indicating the gradients spatial direction. If given only intensities as features, local normalization can not be applied in a meaningful way, as there is only one given value at the location of a grid point. Still, in this section, we describe the differences in the derivatives. Equations (37) and (38) can directly be used for global normalization. For local normalization, the notation needs to be changed slightly. In this case, we compute everything shown before, for each single grid point.
The main difference lies in equation (37). While (37) is straightforward to use for globally normalized features, we need to change $\frac{1}{\|f\|}\left(\mathrm{Id} - g^\top g\right)$ for locally normalized features. For local normalization, we consider the feature vector g to be normalized at each grid point x_i^c, such that $g(x_i^c) = \frac{f(x_i^c)}{\|f(x_i^c)\|}$. However, instead of writing everything for each grid point, we rather use a proper rephrasing of $\frac{1}{\|f\|}\left(\mathrm{Id} - g^\top g\right)$. For this purpose, we consider the normalized feature vector to be a matrix, such that

$$g_{\text{local}} = \begin{pmatrix} \frac{f_{x^1}(x_1^c)}{\|f(x_1^c)\|} & \cdots & \frac{f_{x^\delta}(x_1^c)}{\|f(x_1^c)\|} \\ \vdots & & \vdots \\ \frac{f_{x^1}(x_m^c)}{\|f(x_m^c)\|} & \cdots & \frac{f_{x^\delta}(x_m^c)}{\|f(x_m^c)\|} \end{pmatrix} \in \mathbb{R}^{m \times \delta},$$

where δ is the number of feature-characteristics at one grid point and f_{x^δ} is the corresponding feature-characteristic. For example, for gradient features this corresponds to the spatial dimension d, such that $\delta = d$; e.g., for two-dimensional

images this gives $\delta = d = 2$. As a formula for each grid point, we obtain

$$\mathrm{Id} - g_{\mathrm{local}}(x_i^c)^\top g_{\mathrm{local}}(x_i^c).$$

Here, $\mathrm{Id} \in \mathbb{R}^{\delta\times\delta}$ is the respective identity and $g_{\mathrm{local}}(x_i^c) \in \mathbb{R}^{1\times\delta}$ is the vector of feature-characteristics at grid-point x_i^c. We now abbreviate

$$\mathrm{Id} - g_{\mathrm{local}}(x_i^c)^\top g_{\mathrm{local}}(x_i^c) =: G_{\mathrm{local}}(x_i^c) \in \mathbb{R}^{\delta\times\delta}.$$

Furthermore, we consider $(G_{\mathrm{local}}(x_i^c))_j \in R^{\delta\times 1}$ to be a column of $G_{\mathrm{local}}(x_i^c)$. This allows us to set up a block-structured matrix that we can directly use like in (37). Taking the columns $(G_{\mathrm{local}}(x_i^c))_j$, we concatenate a block structure matrix of diagonal matrices

$$\Gamma_{\mathrm{local}} \coloneqq \begin{pmatrix} \mathrm{diag}(\eta_1(G_{\mathrm{local}}(x_1^c))_1) & \dots & \mathrm{diag}(\eta_1(G_{\mathrm{local}}(x_1^c))_\delta) \\ \vdots & & \vdots \\ \mathrm{diag}(\eta_m(G_{\mathrm{local}}(x_m^c))_1) & \dots & \mathrm{diag}(\eta_m(G_{\mathrm{local}}(x_m^c))_\delta) \end{pmatrix} \in \mathbb{R}^{\delta\cdot m\times\delta^2}, \quad (39)$$

where $\eta = (\eta_1, \dots, \eta_m)^\top \in \mathbb{R}^{m\times 1}$ is the vector of normalization factors $\eta_i = \frac{1}{\|g_{\mathrm{local}}(x_i^c)\|}$. This matrix is substituted into (37), which gives the equivalent equation for locally normalized feature vectors:

$$\underbrace{\underbrace{(\partial_m^h)^\top}_{\mathbb{R}^{\delta\cdot\hat{m}\times\hat{m}}} \Gamma_{\mathrm{local}} \underbrace{u_i v_{jk}^\top}_{\mathbb{R}^{\delta^2\times K}}}_{\mathbb{R}^{\delta\cdot\hat{m}\times K}}. \quad (40)$$

Exemplary, for $T = 3$ two-dimensional ($d = \delta = 2$) images, we obtain for locally normalized images with m pixels:

$$\underbrace{\underbrace{(\partial_m^h)^\top}_{\mathbb{R}^{2\cdot\hat{m}\times\hat{m}}} \Gamma_{\mathrm{local}} \underbrace{u_i v_{jt}^\top}_{\mathbb{R}^{2^2\times 3}}}_{\mathbb{R}^{2\cdot\hat{m}\times 3}}.$$

Here, $\Gamma_{\mathrm{local}} \in \mathbb{R}^{2m\times 2^2}$ consists of 2-by-2 blocks. More precisely, the i-th row of Γ_{local} for the t-th image I_t comprises

$$\eta_i \begin{pmatrix} 1 - \left(\frac{\partial_{x_1} I_t(x_i^c)}{\|\partial I_t(x_i^c)\|}\right)^2 & \frac{\partial_{x_1}\partial_{x_2} I_t(x_i^c)}{\|\partial I_t(x_i^c)\|} & 1 - \left(\frac{\partial_{x_2} I_t(x_i^c)}{\|\partial I_t(x_i^c)\|}\right)^2 & \frac{\partial_{x_2}\partial_{x_1} I_t(x_i^c)}{\|\partial I_t(x_i^c)\|} \end{pmatrix}.$$

In contrast to that, the globally normalized features are multiplied with the same factor at each grid point. Because of that, we can directly use the form of (37). For that case, $g^\top g$ is a vector-multiplication of the vectorized, normalized feature images. The corresponding matrix is ${g_{\mathrm{global}}}^\top g_{\mathrm{global}} = \Gamma_{\mathrm{global}} \in \mathbb{R}^{2m\times 2m}$, where $g_{\mathrm{global}} \in \mathbb{R}^{2m\times 1}$.

3.8 Error Estimates

We elaborate on the discretization of our distance- and regularization functionals by utilizing a midpoint quadrature rule. Since the discretization introduces errors, we investigate for an error bound in this section. We focus on the errors arising from estimating the interpolation of discretized images within the numerical schemes. Since we are using a multi-level scheme for optimization, we want the error to decrease with increasing data resolution. This section is based on the fact that SqN can be computed analytically for non-discrete functions. The assumption is that the given images are non-discrete, continuous functions, such that the interpolation error for interpolation of the given, discretized images decreases when the grid is refined.

We start our discussion by the introduction of some abbreviations. Let $\mathcal{D}(Y)$ be the continuous distance functional. On the one hand, we have interpolation errors, which is why we define $\mathcal{D}_m(Y)$ as approximated distance functional using interpolated images with data resolution m. On the other hand, we use the midpoint quadrature for integration approximation, which is why we introduce $\mathcal{D}_{m,h}$ as an approximation of the distance functional comprising the quadrature as well as the interpolated images. Following [120], we use the triangle inequality to obtain a splitting of the error bounds

$$|\mathcal{D}(Y_h) - \mathcal{D}_{m,h}(Y_h)| \leq |\mathcal{D}(Y_h) - \mathcal{D}_m(Y_h)| + |\mathcal{D}_m(Y_h) - \mathcal{D}_{m,h}(Y_h)|.$$

This splits up the error into an interpolation error part and a quadrature error part. For increasing data resolution, especially in the limit with $m \to \infty$, the first summand vanishes. This is based on an approximation result in [134]. We now need to focus on the second summand for which we will investigate the quadrature error of the correlation-based SqN distance functional $\mathcal{D}_{SqN}$. We restrict to the $2D$-case since the extension to higher dimensions is straightforward, as we will also see later. Furthermore, we use the suggestive abbreviations $F = F(I(Y))$ and $h = (h_1, h_2)$. Therefore, we use the discretized intensity images as columns of the feature array F for ease of presentation.

We now begin by analyzing the quadrature error of one element of the correlation matrix of features. For this purpose, we use a first-order Taylor approximation at x for

$$\sqrt{x+e} \approx \sqrt{x} + \frac{e}{2\sqrt{x}} + \mathcal{O}(e^2) = \sqrt{x} + \mathcal{O}(\|h\|^2),$$

where $e = \mathcal{O}(\|h\|^2)$ and $x > 0$. For practical reasons, we consider images not being entirely zero; otherwise the error would increase and tend to become infinity. This means the integrals are well-defined and none of the images tends to zero. We

therefore obtain

$$\begin{aligned}\left(F^\top F\right)_{i,j} &= \frac{\int I_i(y_i(x))I_j(y_j(x))\,\mathrm{d}\Omega}{\sqrt{\int I_i^2(y_i(x))\,\mathrm{d}\Omega}\sqrt{\int I_j^2(y_j(x))\,\mathrm{d}\Omega}} \\ &= \frac{\mathrm{hd}\sum_{k=1}^n I_i(x_k)I_j(x_k) + \mathcal{O}(\|h\|_2^2)}{\left(\sqrt{\mathrm{hd}\sum_{k=1}^n I_i^2(x_k)} + \mathcal{O}(\|h\|_2^2)\right)\left(\sqrt{\mathrm{hd}\sum_{k=1}^n I_j^2(x_k)} + \mathcal{O}(\|h\|_2^2)\right)} \\ &= \frac{\sum_{k=1}^n I_i(x_k)I_j(x_k) + \mathcal{O}(\|h\|_2^2)}{\sqrt{\sum_{k=1}^n I_i^2(x_k)\sum_{k=1}^n I_j^2(x_k)} + \mathcal{O}(\|h\|_2^2)},\end{aligned}$$

where x_k are cell-centered grid points and n is the number of total grid points. Now the whole fraction is

$$\begin{aligned}&\frac{\sum_{k=1}^n I_i(x_k)I_j(x_k) + \mathcal{O}(\|h\|_2^2)}{\sqrt{\sum_{k=1}^n I_i^2(x_k)\sum_{k=1}^n I_j^2(x_k)} + \mathcal{O}(\|h\|_2^2)} \\ &= \frac{\sum_{k=1}^n I_i(x_k)I_j(x_k)}{\sqrt{\sum_{k=1}^n I_i^2(x_k)\sum_{k=1}^n I_j^2(x_k)}} + \mathcal{O}(\|h\|_2^2) \\ &= \left({F_h}^\top F_h\right)_{i,j} + \mathcal{O}(\|h\|_2^2),\end{aligned}$$

which gives us an upper error bound for each element of the feature array, where F_h is the approximation of F for cell-size h. That $\frac{a+\mathcal{O}(x^2)}{b+\mathcal{O}(x^2)} = \frac{a}{b} + \mathcal{O}(x^2),\ b \neq 0$ can be seen by computing $\frac{1}{b+\mathcal{O}(x^2)} - \frac{1}{b}$, which gives $\mathcal{O}(x^2)$, such that the total error rate of the fraction is $\mathcal{O}(x^2)$. We use this error bound to investigate the behavior of the error bound for the whole array composed with the Schatten-Norm. Let us abbreviate $f_{i,j} = \left(F^\top F\right)_{i,j}$, then the feature array is

$$F^\top F = \left(f_{i,j} + \mathcal{O}(\|h\|_2^2)\right)_{i,j=1}^T = {F_h}^\top F_h + \mathcal{O}(\|h\|_2^2)\mathbf{1}\mathbf{1}^\top,$$

where T is the number of images and $\mathbf{1} \in \mathbb{R}^T$ is a vector of all ones. Now by substitution, we obtain for our distance functional

$$\left\|F^\top F\right\|_{S,q} = \left\|{F_h}^\top F_h + \mathcal{O}(\|h\|_2^2)\mathbf{1}\mathbf{1}^\top\right\|_{S,q}.$$

Again, using the triangle inequality, we can split the Schatten-Norm to

$$\left\|{F_h}^\top F_h + \mathcal{O}(\|h\|_2^2)\mathbf{1}\mathbf{1}^\top\right\|_{S,q} \leq \left\|{F_h}^\top F_h\right\|_{S,q} + \left\|\mathcal{O}(\|h\|_2^2)\mathbf{1}\mathbf{1}^\top\right\|_{S,q}.$$

Since we can factor out our error estiamte $\mathcal{O}(\|h\|_2^2)$ from the norm and the rank-1 matrix $\mathbf{1}\mathbf{1}^\top$ has the single singular value $\sigma = T$, we obtain

$$\left\|{F_h}^\top F_h\right\|_{S,q} + \mathcal{O}(\|h\|_2^2). \tag{41}$$

Utilizing the fact that $F^\top F$ is symmetric, positive semi-definite, we know that $\sigma(F)^2 = \lambda(F^\top F) = \sigma(F^\top F)$. Furthermore, $\|F^\top F\|_{S,q} = \|F\|_{S,q}^2$. With this, we deduce the total error bound e_{total} by

$$\begin{aligned} e_{\text{total}} &= |\|F\|_{S,q} - \|F_h\|_{S,q}| \\ &= \left| \left(\|{F_h}^\top F_h\|_{S,q} + \mathcal{O}(\|h\|_2^2) \right)^{\frac{1}{2}} - \|{F_h}^\top F_h\|_{S,q}^{\frac{1}{2}} \right| \\ &\leq \left| \|{F_h}^\top F_h\|_{S,q}^{\frac{1}{2}} + \mathcal{O}(\|h\|_2^2) - \|{F_h}^\top F_h\|_{S,q}^{\frac{1}{2}} \right| \\ &= \mathcal{O}(\|h\|_2^2), \end{aligned} \tag{42}$$

with the help of $\sqrt{x+h} \approx \sqrt{x} + h$.
From (41) and (42), we can see, that the quadrature error for the whole feature array vanishes for $h \to 0$, as desired. Therefore, we can conclude that for arbitrary fine resolution $|m| \to \infty$ and arbitrary fine quadrature intervals $h \to 0$, the approximated distance functional converges to the continuous functional.
In addition, this result is still valid for different normalized features like e.g., normalized gradients. When using finite differences to approximate the intensity gradients, the error of the finite difference scheme adds up to the interpolation and quadrature error. This is achieved by substituting the function $f : \Omega \to \mathbb{R}$ that is to be integrated using the midpoint quadrature by its finite difference approximation $f(x_k) = f_h(x_k) + E_{\text{fd}}$, where E_{fd} is the error of the finite difference scheme:

$$\begin{aligned} \int_\Omega f(x)\mathrm{d}x &= \mathrm{hd} \sum_{k=1}^{n} f(x_k) + \mathcal{O}(\|h\|_2^2) \\ &= \mathrm{hd} \sum_{k=1}^{n} \left(f_h(x_k) + (E_{\text{fd}})_k \right) + \mathcal{O}(\|h\|_2^2) \\ &= \mathrm{hd} \sum_{k=1}^{n} f_h(x_k) + E_{\text{fd}} + \mathcal{O}(\|h\|_2^2) \end{aligned}$$

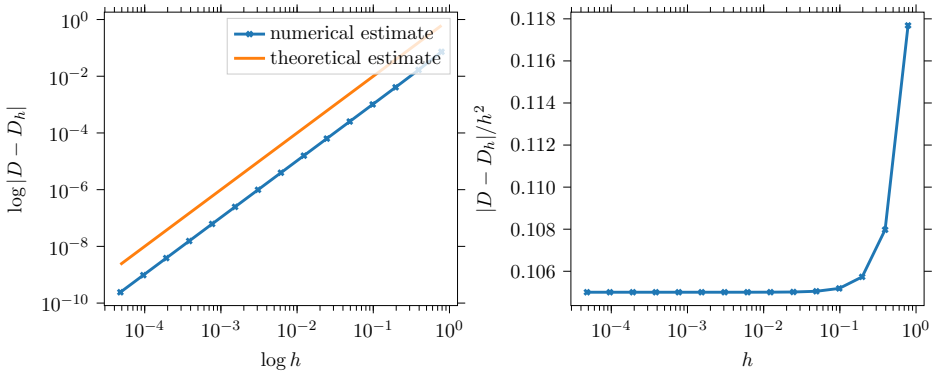
A discussion about the accuracy of numerical minimizers with respect to the continuous solutions is not within the scope of this work.

Examples

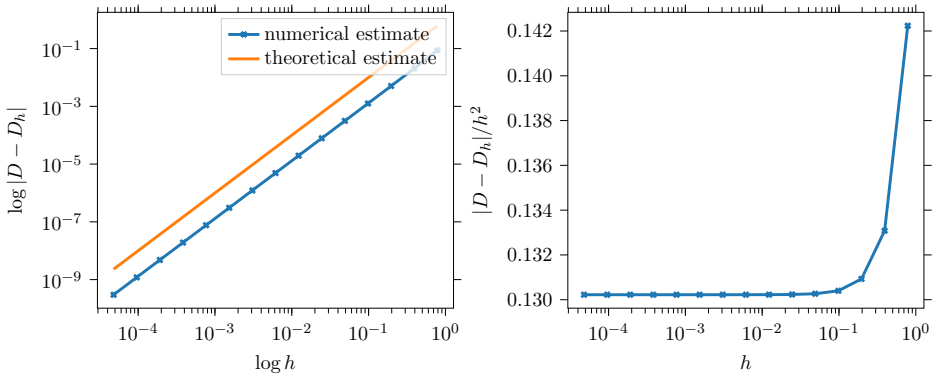
We now exemplarily verify the error estimates of our novel distance measure. First, we estimate the error behavior for $I = (\sin(x), \cos(x), \sin(x))$, assuming normalized intensities as features. With these analytical functions, there is no doubt that the expected behavior can be shown. For the experiments we further assume that $\Omega = [0, \frac{\pi}{2}]$. Therefore, with

$$F_{\text{trig.}}(I) = \begin{pmatrix} \frac{\int \sin}{\|\sin\|} & \frac{\int \cos}{\|\cos\|} & \frac{\int \sin}{\|\sin\|} \end{pmatrix},$$

we compute the theoretical values of $\mathrm{S}q\mathrm{N}(F_{\text{trig.}}(I))$ and compare them for different sampling resolutions on cell-centered grids. Figure 12 illustrates the error estimates for $q \in \{1, 2, 4\}$ and 15 different resolution levels. The resolution of the different levels is obtained by $r = 2^{\mathrm{l}}$, where $l \in \{1, \ldots, 15\}$. The left-hand side illustrates the error between the theoretical value D and the numerical value D_h as absolute value $|D - D_h|$ against the grid spacing h in a $\log - \log$ plot. Furthermore, the assumed quadratic behavior is drawn by h^2 against h. On the right-hand side, a semi-log plot shows the behavior of the error for $h \to 0$. For verification of the quadratic behavior of the error estimates, we observe the following: On the left-hand side, we see the same slope as the theoretical estimate of the numerical computation. On the right-hand side, as h decreases, we observe the actual behavior, which needs to decrease in a quadratic manner. Note, that the computation of SqN is based on the correlation matrix $C = F^{\top}F$. For a direct comparison to SqN using just F as a computation basis, we refer to Theorem 2.
Now that the quadratic behavior is shown to be correct also for analytic functions like sine and cosine for intensity features, it is especially engaging to see the behavior for actual images. We investigate the error behavior for a part of a DCE-MRI data set; data courtesy of Jarle Rørvik, Haukeland University Hospital, Bergen, Norway. Figure 13 illustrates three representative coronal slices of the DCE-MRI data set. The full data set is acquired as a time series of three-dimensional time points. Here, we use multiple two-dimensional sagittal planes cut out of the three-dimensional time points. The reference for our error estimates is the computation of SqN on the non-interpolated data set in its full resolution – this time, we use normalized gradients as features. As a basis, we compute SqN using just the feature matrix F. That means our primary focus lies on $q = 4$, due to the relation to the NGF distance measure. The single images of the data set contain 256-by-256 pixels. The domain (in millimeters) of the images is $\Omega = [0, 425] \times [0, 425]$. The error is estimated on six different grid resolutions, starting from 8-by-8 in steps of power-2 up to original resolution. Therefore, the cell size h is ranging from ~ 1.7 mm up to ~ 53 mm. For interpolation, a standard cubic-spline method is used; see e.g. [100] for details. The error behavior is illustrated in Figures 14, 15 and 16. The behavior of the error is, as expected, quadratic for all chosen q. The log-log plots of the theoretical error vs. the numerical error estimates showcases the quadratic error decrease. There, the slope of the numerical estimate needs to be the same as the slope of the theoretical estimate to fulfill the quadratic decrease. We further see that the error-decrease remains quadratic for different numbers of images, as we show plots for three as well as for 45 images. We also observe a proportional increase of error for a coarse grid for a higher number of images. Nevertheless, it always tends to zero. Very compelling is also the behavior concerning the normalizations. The error for globally normalized features is lower than for features that



(a) Error behavior for $q = 2$



(b) Error behavior for $q = 4$

Figure 12: Illustrated is the quadratically decreasing error for an increasing number of grid points. Here, the error is based on the integral computations for the trigonometric functions $(\sin, \cos, \sin)$. The functions represent three images with intensity features. The results are computed on the domain $\Omega = [0, \frac{\pi}{2}]$.
Note that for $q = 1$ with $h \to 0$, the tiny increase is due to the scaling of the y-axis.

are normalized locally. This observation is due to the overall values of the feature maps, which are much smaller per cell for global normalization. Thus, the error between the original data and the interpolation on the grid is smaller. Captivating is also the sharp decrease of the locally normalized features when turning to the same grid resolution the original data has. Here, the error of the locally normalized features is much lower than the globally normalized features. The error value is as low as machine precision. The latter two observations support the discussion from Sec. 3.2 that globally normalized features may not consider small details. In contrast – and this is what we see in the plots – the fine details considered by local normalization lead to a lower error on full resolution.

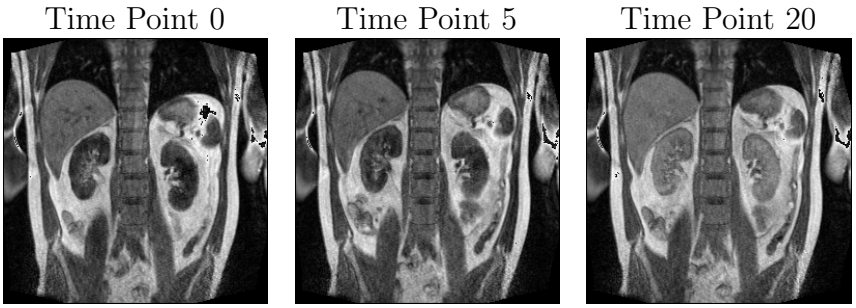
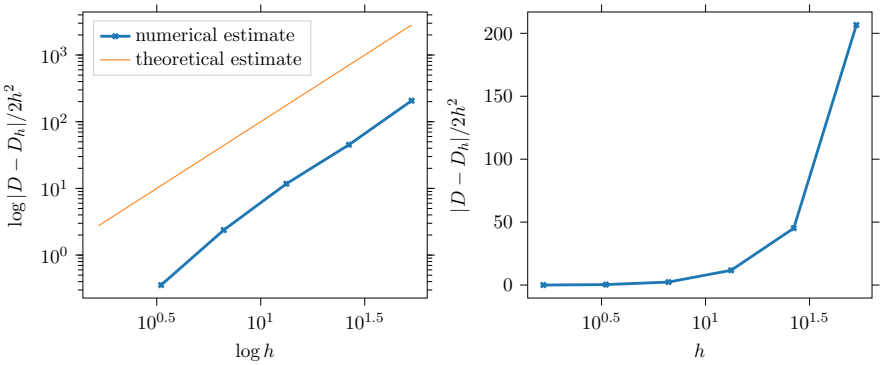
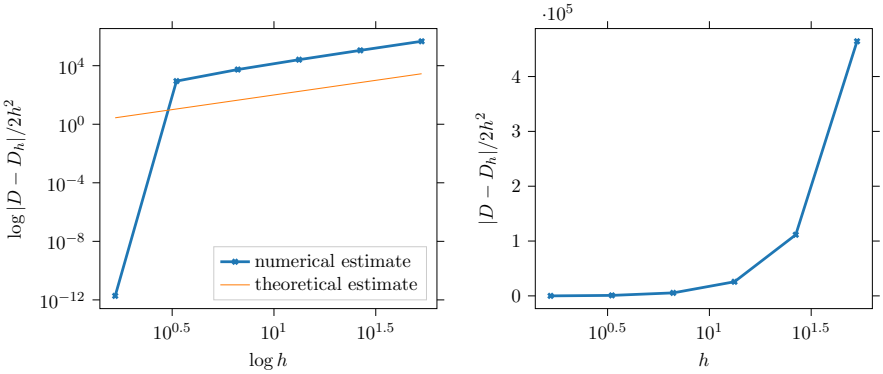


Figure 13: Illustrated are three representative coronal slices of a 4D DCE-MRI dataset of a human kidney at z-position 14. The full data set comprises 45 volumes containing 256-by-256-by-30 voxels. Hence, the sagittal slices contain 256-by-256 pixels. Data courtesy of Jarle Rørvik, Haukeland University Hospital, Bergen, Norway.

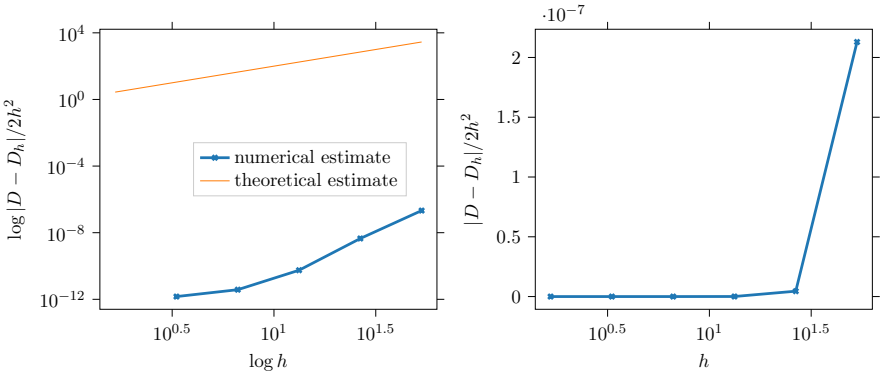


(a) Error for 3 images and $q = 4$ with globally normalized features.

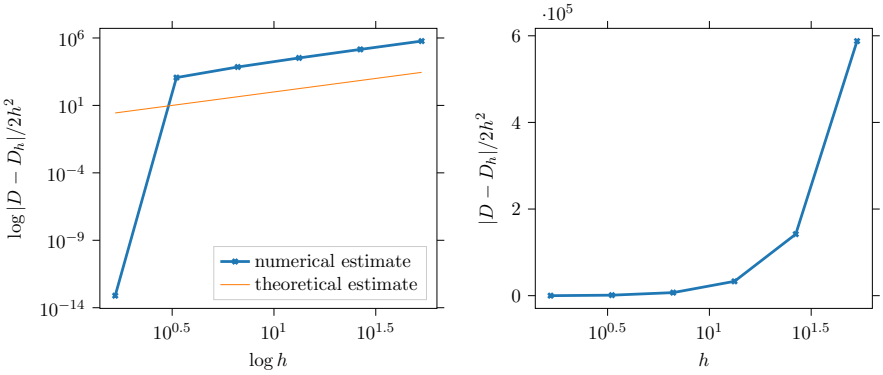


(b) Error for 3 images and $q = 4$ with locally normalized features.

Figure 14: Illustrated is the quadratically decreasing error for an increasing number of grid points, equivalent to decreasing h. Here, the error is based on the resolution of the non-interpolated data, which is 256-by-256 per image. Considered for the computation were three images. The used features are normalized gradients. The plots show the error behavior for global as well as local normalization.

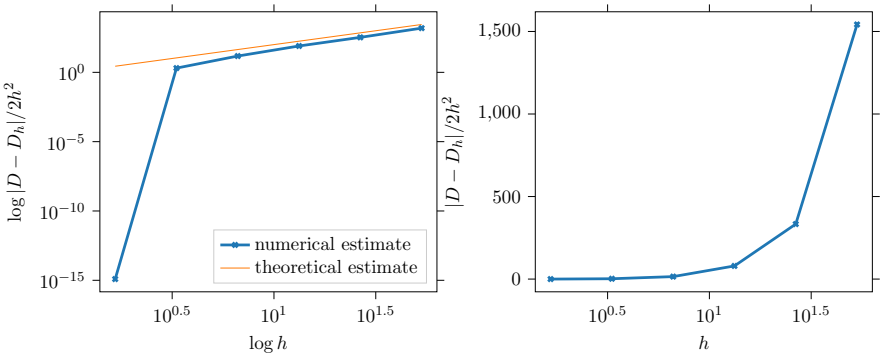


(a) Error for 3 images and $q = 2$ with globally normalized features.

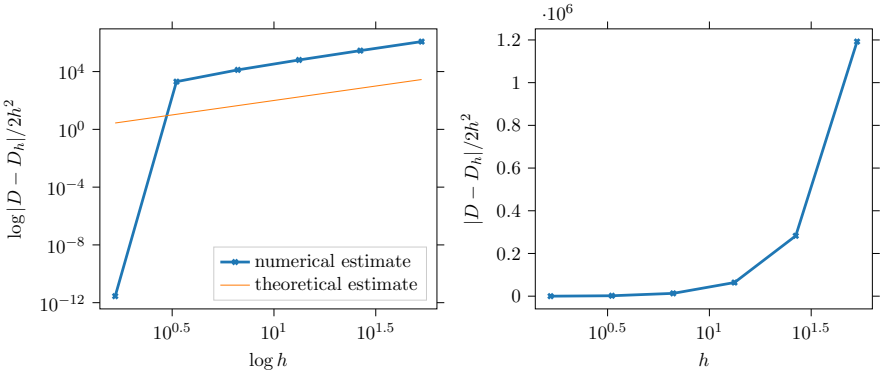


(b) Error for 3 images and $q = 2$ with locally normalized features.

Figure 15: Illustrated is the quadratically decreasing error for an increasing number of grid points, equivalent to decreasing h. Here, the error is based on the resolution of the non-interpolated data, which is 256-by-256 per image. Considered for the computation were three images. The used features are normalized gradients. The plots show the error behavior for global as well as local normalization.



(a) Error for 45 images and $q = 4$ with globally normalized features.



(b) Error for 45 images and $q = 4$ with locally normalized features.

Figure 16: Illustrated is the quadratically decreasing error for an increasing number of grid points, equivalent to decreasing h. Here, the error is based on the resolution of the non-interpolated data, which is 256-by-256 per image. Considered for the computation were 45 images. The used features are normalized gradients. The plots show the error behavior for global as well as local normalization.

3.9 Optimization and Implementation Remarks

This section is the respective extension of Sec. 2.7. The focus is rather on deploying details for handling multiple images with the outlined optimization schemes than on reviewing the full schemes for multiple images. This is accompanied by implementation remarks as we discuss the discretized setting. We mainly base this part on original literature, see, e.g., [3], [55], [103], [116].

BFGS

For the BFGS method, we start over at the choice of the starting-guess for H. Here, we follow [55] for the choice of H_0 or the inverse $B_0 = H_0^{-1}$ for an inital guess. Considering $f(x)$ to be our objective function, then the Hessian is [55]

$$\nabla^2 f(x) = \nabla^2 D(x) + \alpha \nabla^2 S(x) = \nabla^2 D(x) + \alpha h A.$$

The Matrix A is, considering Dirichlet boundary conditions [55], symmetric positive definite. Thus, we choose B_0 following [55] as

$$B_0 = (\alpha h)^{-1} A^{-1} \approx (\nabla^2 f(x))^{-1}$$

as a starting-guess that approximates the Hessian of our objective function. Numerical experiments in [55] show, that this choice outperforms a simple identity as starting-guess. Also for different boundary conditions it is possible to use this starting-guess. Adding a regularizing identity to A and using a pseudo-inverse B_0 is still positive definite for [55]

$$B_0 = (\epsilon I + \alpha h A)^{-1},$$

for $\epsilon > 0$. This formula holds for the chosen Dirichlet boundary conditions. When choosing Neumann boundary conditions, A is no longer positive definite [55]. Instead, it is positive semi-definite and a Moore-Penrose pseudo inverse may be computed [55]. We use this for deploying a starting-guess for our specific multiple-image approach. As outlined in Sec. 3.3, we use a straightforward extension of the regularization by using a summation of regularization-energies for each image. Taking into consideration that we concatenate all transformation grids into a long transformation vector $y \in \mathbb{R}^{d \cdot n \cdot T}$ we build A for the approximation of $\nabla^2 f(x)$ as a block-diagonal matrix. With this, we obtain

$$A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_T \end{pmatrix},$$

where A_t is the respective approximation for the single images. This specific construction has the advantage that each block can be determined individually as the inverse of a block-diagonal matrix is a block-diagonal of the inverse blocks. A great advantage is that this is straightforward to parallelize. When using the curvature regularizer, all A_t are equal, for example.
For the limited-memory version of the BFGS-Method, only a few, meaning five to ten, of the last s_i and g_i, are used to save memory. This means we just need to compute the approximation to the Hessian once and update it in each step. Considering a full A for approximation, we would need to save a matrix of dimension $n \cdot d \cdot T \times n \cdot d \cdot T$. With the low-memory version, we only need to save two vectors of dimension $n \cdot d \cdot T$. For our example for the 2D images of size 256×256 pixels, we just need to save $\approx 4 \cdot 10^6$ entries instead of $\approx 2 \cdot 10^{11}$ entries, which makes the recomputation in each iteration much faster.

Gauß-Newton

In our multiple image case, for vectorized Jacobian, we need to handle $\mathrm{d}\, r(x_k) \in \mathbb{R}^{d \cdot n \cdot T}$, which ends up in a Hessian approximation $H \in \mathbb{R}^{d \cdot n \cdot T \times d \cdot n \cdot T}$.
Unlike as for the BFGS method, for solving the Newton system with a Gauß-Newton approximation of the Hessian, we need to compute the approximation anew in each iteration. Although this may get infeasible for many large images, due to the better convergence, it is worthwhile to consider Gauß-Newton for optimization, at least for problems containing small images or a small number of images.

3.10 Guideline for the Choice of Parameters

Since there are a few parameters we have introduced along with SqN we now give a guideline to a choice of parameters in a practical way. We focus on the parameters for the distance measure SqN rather than on the choice of a suitable regularization parameter α. For this, we refer to appropriate image registration literature; see e.g. [29], [32], [100].
The most prominent parameter along with SqN is the q. In Section 3.4, the use of $q \geq 2$ is recommended. This is due to the behavior of the distance energy. In that case, the global maximum is the best match for the given images. Therefore, a simple switch of signs deploys a minimization problem. Furthermore, the Schatten-Norms for $q \geq 2$ are actual norms. Thus, SqN is convex. In particular, $q = 2$ (or $q = 4$, depending on the use of the correlation matrix C or the feature matrix F in the formulation of SqN) and $q = \infty$ are choices worthwhile to try. The choice $q = 4$ along with the feature matrix F in the formulation of SqN have shown to be a good choice in practice; c.f. [7]–[9], [60]. The choice of $q = \infty$

may be engaging in combination with eigenvalue or singular value approximating approaches as for $q = \infty$, only the largest singular value is relevant. A method like the so-called power iteration might be worthwhile to investigate together with SqN and $q = \infty$ eliminate the SVD computation in each iteration. We also use $q = 0.5$ which originates from color image denoising; see [102]. This deploys a non-convex quasi-norm case for the Schatten-Norms, which can be solved using the methods reviewed in this work but may rather be solved using optimization methods not utilizing the objective function's derivatives.
The choice of the weighting matrix W is shortly discussed in the respective subsection within Section 3.4. The edge parameter η for normalized gradient features may be chosen following the rule shown in [47]. This parameter is highly dependent on the data and may also be chosen manually by sight. For that, one can plot the different spatial parts of the normalized intensity gradients while adapting η, such that the noise reduces.

3.11 Summary

With Equations (25), (26), (29) and (31), SqN is a very flexible distance-measuring framework. Due to its use of the Schatten-Norms, it is possible to combine various advantages from well-studied approaches for two-images. It is possible to use different image features, such as simple intensities, intensity-gradients or even more advanced features like MIND [53], [54] or local scale-invariant features (SIFT) [87], [88]. By building a feature matrix, various geometrical ideas can be used for similarity measurement taking advantage of the specific features. This allows for the use of SqN in a wide variety of applications, not only in medicine. Furthermore, the architecture of the specific feature matrix allows the registration of multiple images in one process. Due to the use of singular values modeled by the Schatten-Norms, intrinsic properties over a whole stack of images can be considered for registration. Of course, by using Equation (31), the neighborhood-size for similarity measurement can be determined. With this, also a pair-wise image registration lies within the possibilities of the SqN-framework. Moreover, an arbitrary number of images can be considered for the measurement of similarity between features. Altogether, we have seen a straightforward discretization of the framework as well as an analysis of the first derivative, such that the framework can be used with higher-order optimization methods. Also, the framework has upper and lower distance-energy bounds and the error estimates for the discretization lies within the expectations. On the one hand, the error vanishes quadratically for refined grids. On the other hand, the error rises linearly for a higher number of images considered for registration, as expected. Finally, the elaboration of the numerical optimization schemes with details on the starting-guess for Hessian-approximation makes it straightforward to implement SqN for use in various toolboxes.

4 Numerical Results

In this chapter, we elaborate on numerical results for medical applications using the novel distance measure SqN. We start with a short overview of the datasets we use for our numerical experiments. Following, we use an academic example in which we register five x-ray images of a human hand. This dataset is annotated with landmarks, which allows us to discuss the results concerning different parameters. The discussion of the parameters gives a first impression of the behavior of SqN in comparison to the standard measures like NGF or SSD. Subsequently, we present results for motion correction of DCE-MRI datasets as well as results for 3D reconstruction of histological tissue. A fusion application, as requested by clinicians, accompanies this. Finally, this chapter presents a proof-of-concept application using a recently published convolutional neural network to show the applicability of SqN with modern neural network methods.

4.1 Overview of the used Datasets

For the different numerical experiments, we use the following, shortly introduced, datasets. This section intends to give a short overview of the datasets and to clarify data courtesy. Details on the datasets and the number of pixels, voxels, slices or time points are given in the respective section.
Our first experiment utilizes an x-ray image of a human hand, as found in the FAIR toolbox [100] or in [2]. The toolbox provides landmarks that can be seen in Figure 17 overlayed on the hand.

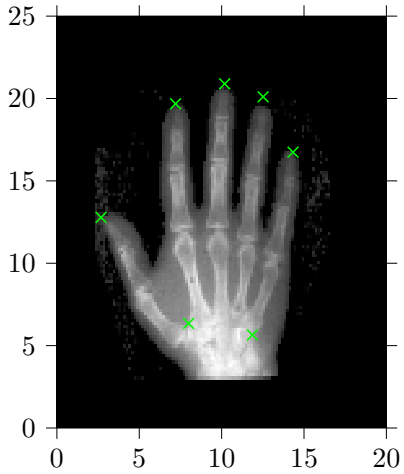


Figure 17: X-Ray image of a human hand manually annotated with landmarks. The image data and the landmarks are taken from the FAIR-Toolbox [100].

The second experiment is about motion correction in dynamic imaging. More precisely, we use multiple DCE-MRI datasets with courtesy of Jarle Rørvik from Haukeland University Hospital Bergen, Norway. Figure 18 shows two-dimensional, coronal overview slices from a representative dataset. For reasonable computation times, we also use smaller parts of the datasets, only showing one kidney, as shown in Figure 19.

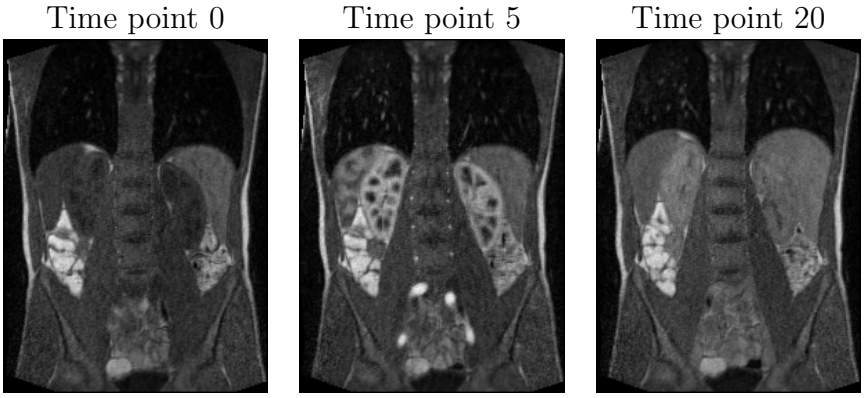


Figure 18: Illustrated are three representative 2D coronal slices of two different 4D DCE-MRI datasets of human kidneys; data courtesy of Jarle Rørvik, Haukeland University Hospital, Bergen, Norway.

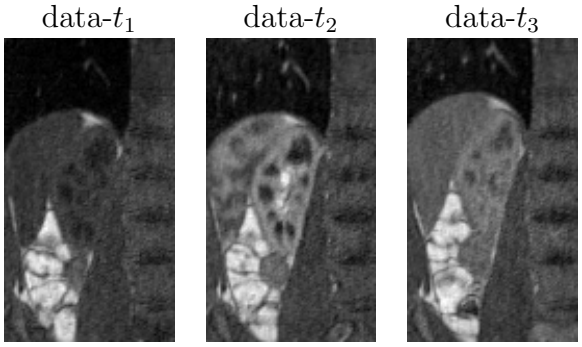


Figure 19: Illustrated is the left kidney from the dataset shown in Fig. 18.

The third experiment is about the registration of histological serial sectioning. Here, we use a dataset of a mouse brain, as shown with representative slices in Figure 20, with courtesy of O. Schmitt, University of Rostock, Germany; see [123] for experimental details. Furthermore, a marmoset monkey brain dataset is used. Figure 21 shows representative slices; data courtesy of Harald Möller, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; see also [92]. The fourth experiment uses datasets of a mouse kidney as well as datasets of a mouse heart for 3D reconstruction and fusion of the images from different modal-

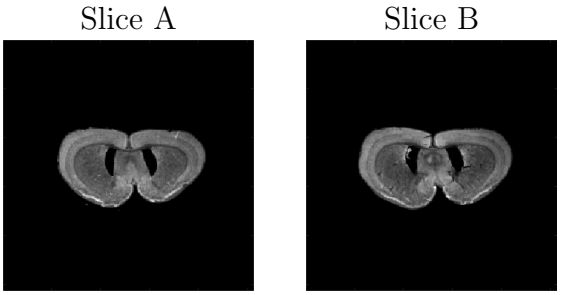


Figure 20: Two representative axial slices of a mouse brain dataset. Data courtesy of O. Schmitt, University of Rostock, Germany; see [123] for experimental details.

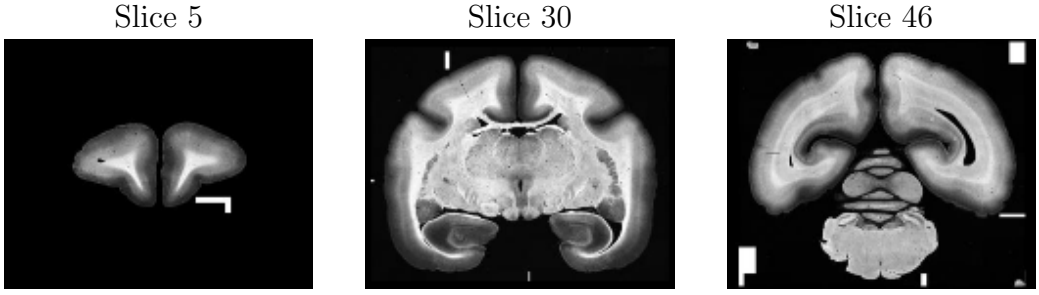


Figure 21: Three representative axial slices of a marmoset monkey brain dataset. Data courtesy of Harald Möller, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; see also [92].

ities. Figure 22 shows representative slices of the kidney, while Figure 23 shows representative slices of the heart; data courtesy of Joachim Jankowski, Institute for Molecular Cardiovascular Research, UK Aachen, Aachen, Germany.
The last experiment is about the application of SqN in a deep learning framework. For this, we utilized the DIR-LAB data, which are 4D CT datasets of human lungs and can be found on www.dir-lab.com. Data courtesy of Richard Castillo, Department of Radiation Oncology, Winship Cancer Institute, Emory University, Atlanta, USA; see [14]–[16]. Figure 24 illustrates representative, two-dimensional sections from one of the provided datasets.

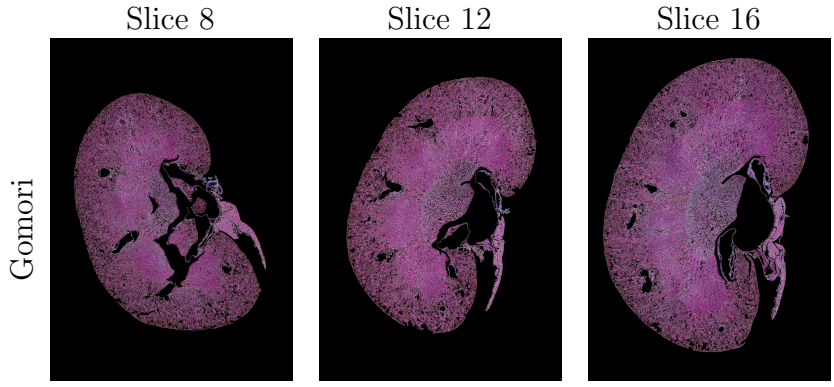


Figure 22: Illustrated are three representative histological slices of a mouse kidney. Data courtesy of Joachim Jankowski, Institute for Molecular Cardiovascular Research, UK Aachen, Aachen, Germany.

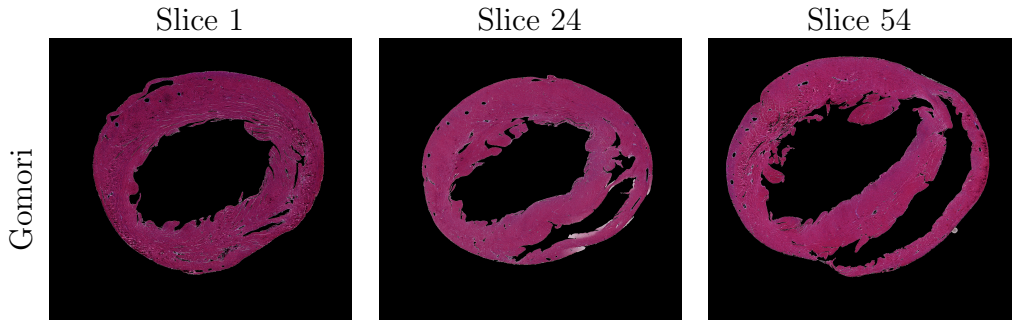


Figure 23: Illustrated are three representative histological slices of a mouse heart. Data courtesy of Joachim Jankowski, Institute for Molecular Cardiovascular Research, UK Aachen, Aachen, Germany.

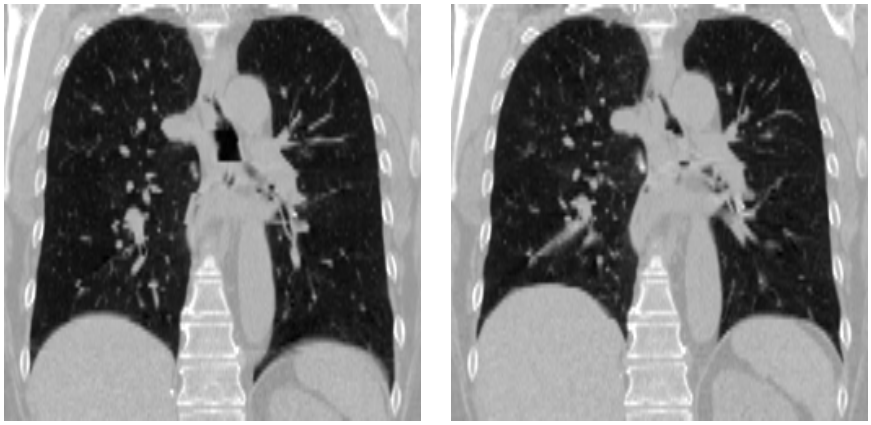


Figure 24: Visualization of CT data from DIR-Lab [14] (see www.dir-lab.com). Two representative axial slices at a representative time t of a human lung are shown. Data courtesy [14]–[16].

4.2 Academic Examples

The academic example we show involves the registration of manually rotated x-ray images of a human hand. With this example, we show that SqN is capable of achieving registration results of comparable quality to standard methods using NGF or SSD. The dataset we use is from the FAIR-Toolbox [100]. Amit originally published the images in 1994; see [2]. Here, we use the reference image of the hands, which is manually annotated with landmarks. For registration, we use five versions of this hand. In a pre-processing step, we rotate four of the images using the angles $\pm 10°$ as well as $\pm 9°$. One of the images remains un-transformed in the pre-processing. Our primary focus for the experiments is on the distance measure SqN. For comparison, we use NGF and SSD. For SqN we use local as well as global normalization of gradient-features with different values for q. Here, we choose $q \in \{0.5, 1, 4, \infty\}$. The regularization parameter α is ranging from $1 \cdot 10^{-5}$ to $1 \cdot 10^{2}$ in steps of 10^{1}. The edge-descriminating parameter η is fixed to $\eta = \sqrt{35}$, chosen manually by optical inspection. As an interpolation model, we choose a standard cubic-spline model to provide an accurate interpolation of the medical data. Our maximum iterations are set to 222 for each level, such that the buffer is large enough that the optimization stops due to the other stopping criteria. The gradient-tolerance is set to $1 \cdot 10^{-8}$. The results for NGF and SSD are computed sequentially using only one sweep. Our results are computed from level 3 with a resolution of 8×8 pixels to level 7 with a resolution of 128×128 pixels, which is the full resolution of the images. Figure 25 illustrates the x-ray image overlayed with the respective landmarks. The image domain is $\Omega = (0, 20) \times (0, 25)$. Figure 2 and Figure 3 show the results of our computations. The tables contain several abbreviations: J is the energy of the objective function. D is the energy of the distance measure. S is the regularization-energy. The time is measured as a sum of all iterations and all levels, just as the iterations are accumulated for all levels. E1 is the mean landmark-error before registration, and E2 is the mean landmark error after alignment. Max E1 is the maximal landmark-error and Max E2 the corresponding error after alignment. The results are computed using our Python-Toolbox GReAT that has been developed by the author. Table 2 and Table 3 show the best results for each distance measure. The computations for SqN with $q = 0.5$ and $q = 1$ turned out to be worse than for $q = 4$ and $q = \infty$, which is why we don't show them in the tables. We computed all results for all mentioned α. Illustrated are only the results for the lowest mean landmark error. Therefore, the tables show the respective regularization parameter α for the lowest mean landmark error. We achieved the overall lowest error using SqN with $q = \infty$ and $\alpha = 10$. For a fair comparison, we computed the results for NGF and SSD in a sequential process, registering pairs of images. Other than with the standard two-image registration, both images were transformed, meaning that there was

no fixed reference image. This is reflected in the results for NGF, for which we typically expect comparable results to SqN with $q = 4$, as a close relation was pointed out in Sec. 3. The approach without a fixed reference seems to cause trouble, such that the landmark error slightly increases after registration. SSD has no problems without a fixed reference image and is able to lower the mean landmark error, but not in a comparable quality as SqN does. Most likely, due to the absence of a fixed reference image as well as due to the pairwise process, the number of iterations for NGF and SSD was much higher than for SqN. This also causes a notable difference in computation time. A graphical illustration of the results is presented in Figure 26.

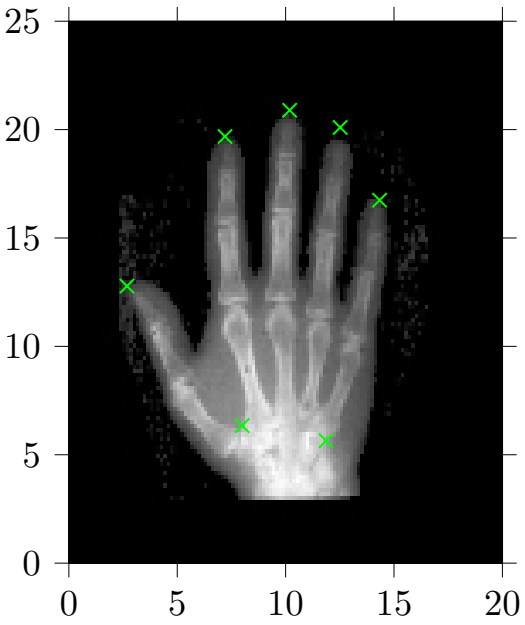


Figure 25: X-Ray image of a human hand manually annotated with landmarks. The image domain is $\Omega = (0, 20) \times (0, 25)$. The image data and the landmarks are taken from the FAIR-Toolbox [100].

J	Objective energy
D	Distance energy
S	Regularization energy
Iter	Total number of iterations (accumulation)
E1	Mean landmark error before registration
E2	Mean landmark error after registration
Reduc.	Reduction of the error (in percent): $(1 - \frac{\text{E2}}{\text{E1}}) \cdot 100$
Max {E1, E2}	Maximal error

Table 1: Abbreviations used in Tables 2 and 3.

Distance	Normal.	q	Alpha	Time (s)	J	D	S
SqN	global	4	0.01	241.5394	-2.1240	-2.1501	0.0260
SqN	local	4	0.1	300.7015	-213.8486	-219.1441	5.2955
SqN	local	∞	10.0	304.4160	-752.9994	-807.4317	54.4322
iSqN	local	4	0.001	217.8871	-2.2341	-2.2352	0.0011
NGF	local	-	100.0	509.1613	345.1176	333.7306	11.3870
SSD	local	-	100.0	334.0477	362.7093	278.5980	84.1113

Table 2: Illustrated are the results of our computations for registration of five rotated X-Ray images of a human hand. J is the objective-energy. D is the distance-energy and S is the regularization-energy. The time is measured in seconds. Further numbers can be found in Figure 3.

Distance	Normal.	q	Iter.	E1	E2	Reduc.	Max E1	Max E2
SqN	global	4	754	0.9795	0.3591	63.3384	1.1269	0.8265
SqN	local	4	828	0.9795	0.5855	40.2246	1.1269	1.7230
SqN	local	∞	960	0.9795	0.3281	66.5033	1.1269	0.6072
iSqN	local	4	975	0.9795	0.1447	85.2272	1.1269	0.3185
NGF	local	-	3823	0.9795	1.0143	-3.5528	1.1269	1.1842
SSD	local	-	2140	0.9795	0.7210	26.391	1.1269	1.1058

Table 3: Illustrated are the results of our computations for registration of five rotated X-Ray images of a human hand. Here E1 is the mean landmark error before registration. E2 ist the mean landmark error after registration. Max E1 is the maximal landmark error before registration and Max E2 is the maximal landmark error after registration. The iterations are counted over all five levels. Reduc. is the reduction of the error in percentage.

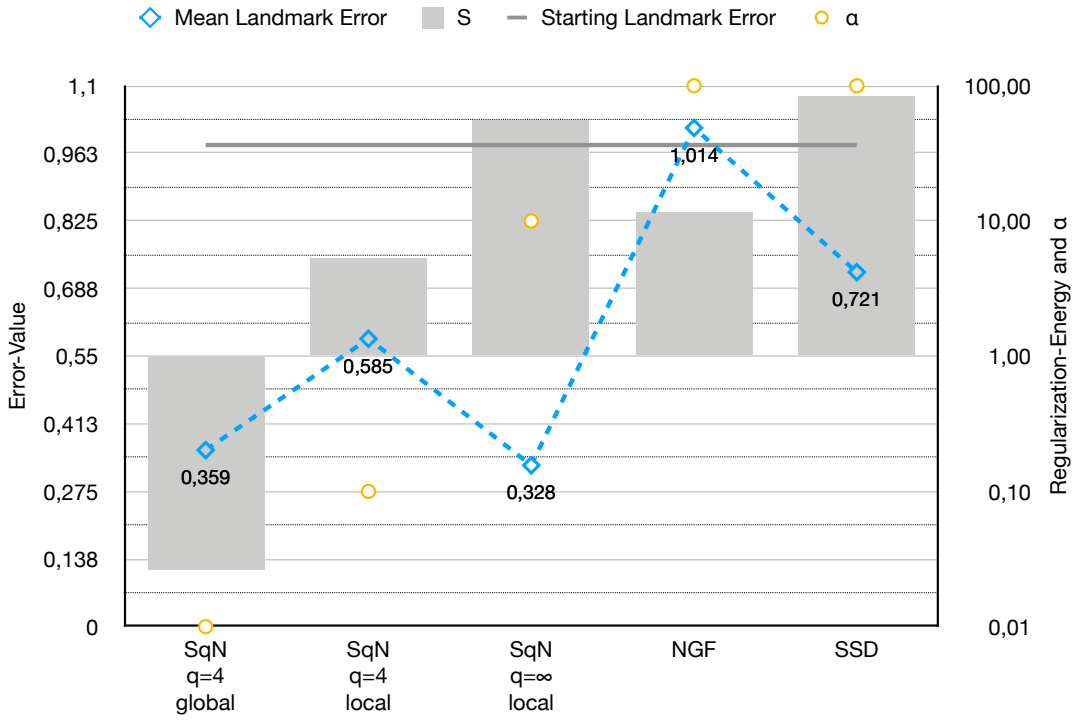


Figure 26: Illustrated are the results for registration of the X-Ray images of a human hand. The gray bars represent the regularization energy. Yellow circles illustrate the respective regularization parameter α. Both the gray bars as well as the regularization parameters correlate to the value-axis on the right-hand side. The mean landmark error for the respective distance measure is illustrated by the blue diamonds connected with a blue dashed line to illustrate the trend between the distance measures. The starting error is illustrated by the gray, horizontal bar.

4.3 Motion Correction of Dynamic MRI Sequences

Motion correction is one of the most important tasks of medical image registration and, therefore, well studied. However, motion correction for dynamic MRI sequences is a particularly challenging task as it combines spatial and temporal changes in intensities [64]. Because of this, there are still novel methods for motion correction developed and studied. In dynamic medical imaging, it is almost always inevitable to correct for the motion to examine physiological functionality within anatomical structures.

In this work, we focus on studying motion correction of dynamic MRI sequences. The studied sequences comprise a human kidney as well as a beating human heart. We begin by examining dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) of human kidneys. All kidney datasets are provided with courtesy of Jarle Rørvik, Haukeland University Hospital Bergen, Norway. The focus of DCE-MRI of human kidneys is on the estimation of clinically relevant parameters. Especially the Glomerular Filtration Rate (GFR) is an essential indicator of kidney dysfunction and diseases [62], [146], [149]. To allow for an estimation of GFR through pharmacokinetic modeling, adequate motion correction is required [62]. Correction of motion enables a voxel-wise estimation of GFR. Among the challenges we need to overcome with image registration to obtain reliable estimation results are organ motion due to breathing, physiological pulsations, and other patients movements [62], [63].

The datasets we use are 4D (3D with time-component) acquisitions. For DCE-MRI acquisitions, a contrast agent, typically gadolinium (Gd), is injected [63]. Right after injection, the first data is acquired within a short time-slot of a few seconds. This first acquisition is for the estimation of injection parameters like concentration of the agent within the blood. Before moving on with acquisition, clinicians typically wait for the contrast agent to arrive in the region of interest, after one to two minutes. This break in acquisition gives us a non-uniform time-scale for the data. Furthermore, due to the injection of a contrast agent, the data has a dynamic intensity component that is to be maintained. In conclusion, the data needs to be corrected for motion while maintaining dynamics that manifest by changing intensities over a non-uniform time scale. Additionally, the non-uniform time-scale may also produce non-continuities in motion as well as in change of intensities. Figure 27 shows three representative slices of a 4D dataset. The slices are coronal slices at pixel-position 25 of the z-direction showing different time-points. The 4D dataset comprises a total of 146-by-82-by-52 voxels in 49 time points, such that the 2D slices are 146-by-82 pixels and 49 time points. Furthermore, a maximum intensity projection (MIP) of the difference images over time is presented on the very right of the figure. The MIP illustrates the motion due to breathing and pulsation over time. This particular registration experiment is also shown in our

publication [8]. For registration we use SqN with $q = 0.5$, which in this case is motivated by the Schatten-Norm usecase in color image denoising [102]. There, an analytical solution for $q = 0.5$ is discussed. With this particular numerical example, we show that SqN is capable of outperforming standard methods. With this, we prove that SqN is capable of achieving results of comparable or even better quality compared to standard methods in a sequential manner. For ease of presentation we use our in Sec. 3.3 introduced summation of regularizers, using a standard curvature regularizer with parameter $\alpha = 0.1$. The features are globally normalized gradients. For optimization, we choose Gauß-Newton with Armijo line search within a multi-level scheme; see Sec. 2.7. A competing registration uses the NGF distance measure [47], [100] in a sequential manner with one single sweep, meaning, registering the image-pairs from image 1 to K and vice versa. We refer to Sec. 2.1 and Sec. 3 for details on the sequential registration approach. Even though the results in Figure 28 and Figure 29 look quite similar, the registration, achieved within the extended MATLAB-Toolbox FAIR [100], was about six times faster for SqN than for NGF. An optical inspection of Figure 28 and Figure 29 reveals well-aligned structures. Not only the MIPs in Figure 28 are very smooth and do not show any doubled structures anymore. Also, the sagittal and axial cuts of the stack of time-point images show straight lines. For time-dependent data, the straight lines are the structures that are expectable for motion-correction. Each row of the images in Figure 29 represents the same structure changing over time. In a best-case scenario, the structures are completely the same after correction, such that straight lines appear. We can see that this is not the case for the uncorrected data. The corrected data shows many straight columns, which is an indicator of a good motion-correction in this case. Note that the illustrated tissue-structures do not necessarily correspond due to shiftings within the hidden axis during registration. A second experiment, which we have also elaborated in [7], shows comparable results for a different dataset. Here, we use 178-by-95 pixel large coronal slices of a 178-by-95- by-30 voxel comprising volume with 45 time points for z-slice 18; see Figure 30 for representative slices. We compute the registration results for this particular dataset using SqN with $q = 4$ and $q = \infty$. Again, we compare the results to NGF because of our choice of features. We use locally normalized gradient features. For the computations, we use our own developed Python-Toolbox GReAT. We use L-BFGS with Armijo line search for optimization within a multi-level scheme over three different levels, up to half the original resolution. Again, the registration with the NGF distance measure is sequential with only one single sweep as it was already significantly slower than SqN. This time the results for SqN were achieved in around 8 minutes, compared to twice the time for the single sweep of NGF. Figure 31 illustrates the mentioned straight lines after correction. We can further see that NGF smoothed out the movement

over time. Since the structures are not fully straight, NGF does not seem to be converged after just one single sweep, but a similar result to the one of SqN is to be expected when giving more time on the computation. Note that the dataset has an irregular time-axis, so a few large jumps are visible on the time-axis in the unregistered plots in Figure 31.

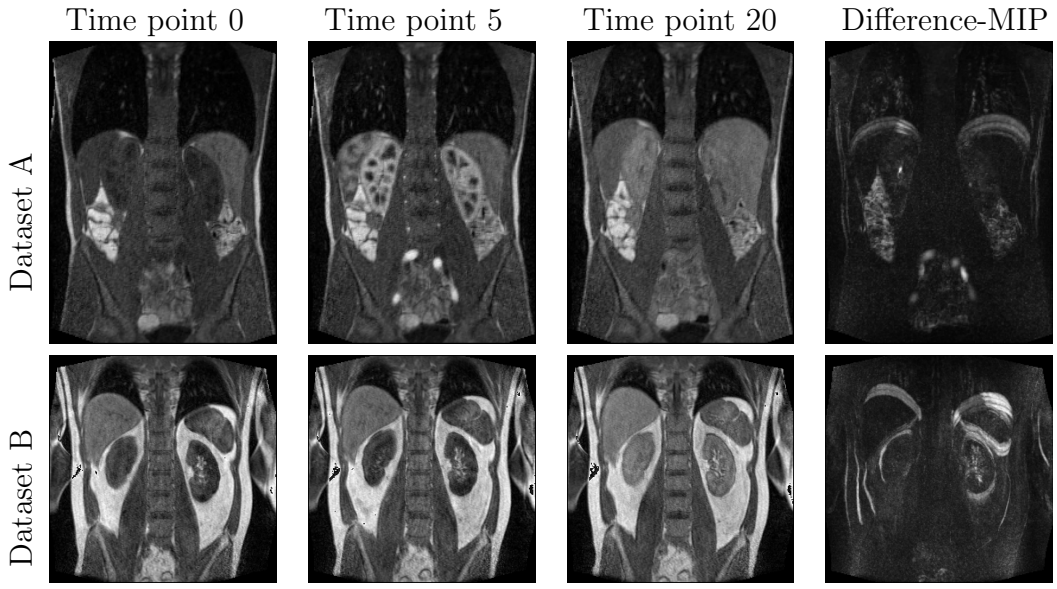


Figure 27: Illustrated are three representative 2D coronal slices of two different 4D DCE-MRI datasets of human kidneys; data courtesy of Jarle Rørvik, Haukeland University Hospital, Bergen, Norway. The slices are shown at three different time points. The datasets are 192-by-256-by-52-by-49 (Dataset A) and 256-by-256-by-30-by-45 (Dataset B) volumes, respectively. The shown slices are the corresponding number of pixels of the first two spatial dimensions. On the very right of both rows, a maximum intensity projection (MIP) of the difference images over time is shown. The MIP showcases the variations over time.

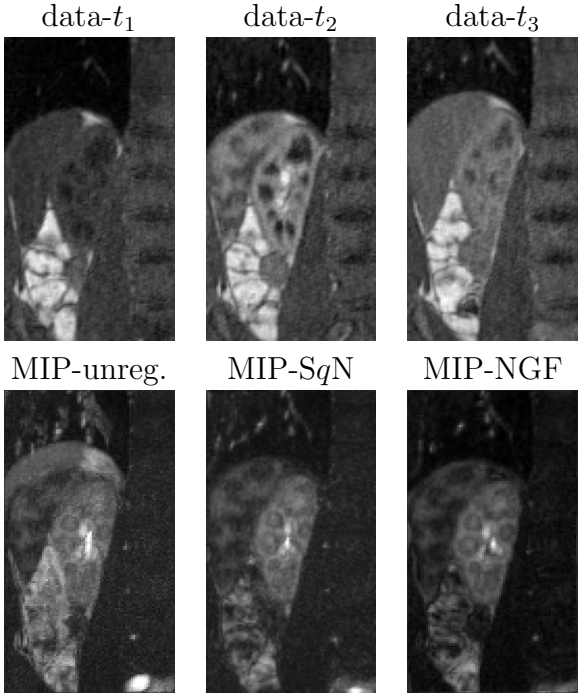


Figure 28: Illustrated is the left kidney from dataset A of Fig. 27. Top row: Displayed are 2D slices of the left kidney at three arbitrary, representative time points during contrast agent uptake. Bottom row: Coronal view of maximum intensity projections $\sum_{j>i} |I_j - I_i|$ for unregistered, SqN-registered, and NGF-registered data. Note the blurred and doubled structures in the non-registered data.

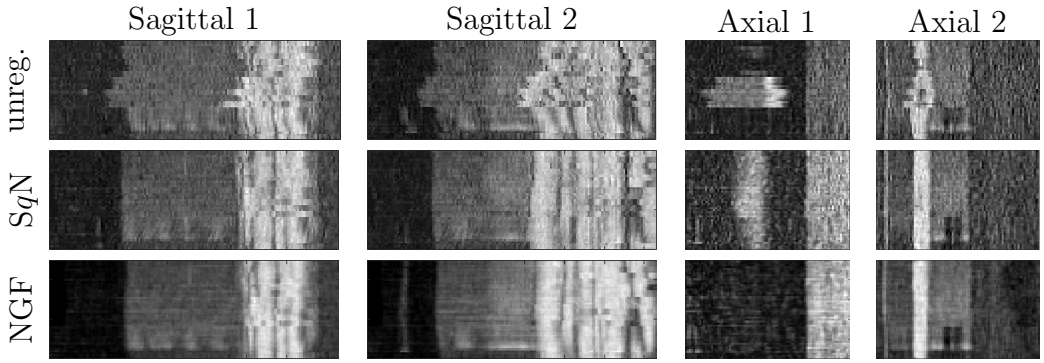


Figure 29: Two exemplary sagittal and axial slices at different positions of the data, each; see also Fig. 28. Note that for ease of presentation, the sagittal slices are rotated by 90 degrees. The top row shows the original unregistered data. The middle row shows SqN-registered data ($q = 0.5$), and the bottom row shows NGF-registered data using $\eta = 25$. Note that the laminar structure of the tissue is only visible after registration.

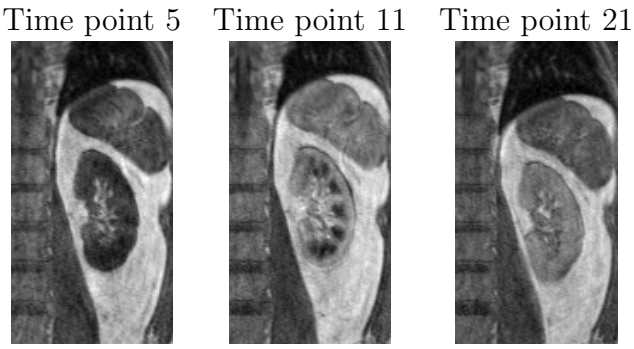


Figure 30: Shown are three representative 2D coronal slices of the 4D DCE-MRI dataset of a human kidney, data courtesy of Jarle Rørvik, Haukeland University Hospital, Bergen, Norway. The slices are shown at three different time points. The dataset is a 178-by-95-by-30-by-45 volume; the shown slices are 178-by-95.

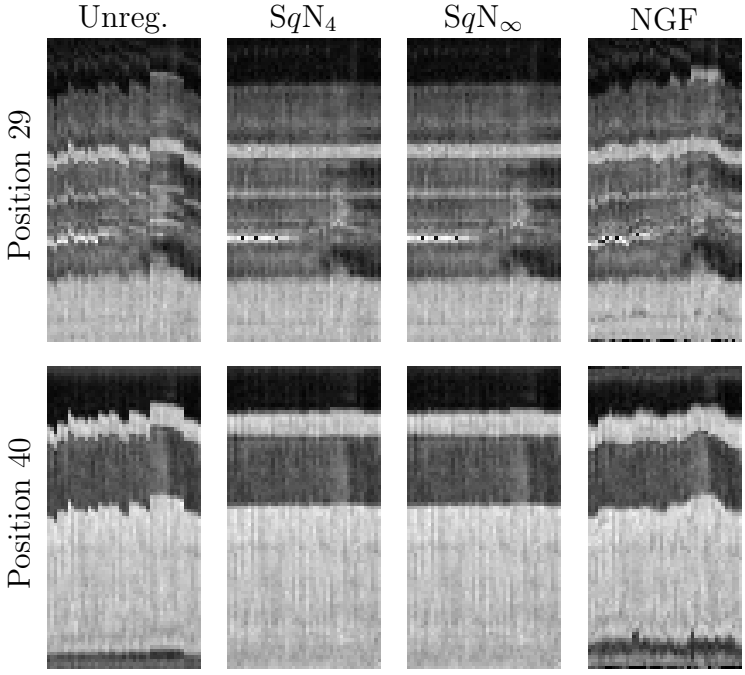


Figure 31: Illustrated are sagittal cuts through the stack of 2D slices from a 4D DCE-MRI dataset of a human kidney at positions 29 and 40. The first column shows the unregistered stack. Right next to this, the results of the different registration approaches are illustrated.

4.4 Reconstruction and Fusion of Histological Sections

In this chapter, we study the registration of histological serial sectioning. Since traditional histology does not examine the sectioned tissue across multiple sections [85], [86], [101], [124], [142], image registration provides a way of extension of the classical histology. For this purpose, the tissue sections are registered, such that a 3D volume is reconstructed. With the help of appropriate registrations methods, this reconstructed volume shows details across the slides. This would not have been possible from examining the sections only in their given two dimensions. Different from what we see in our first numerical experiment, histological sections typically are differently stained [85], [86]. These stainings reveal several details on a macroscopic and microscopic scale. For example, this involves the examination of cell nuclei and other cell components or the differentiation of healthy and corrupted tissue in biopsy samples, to name a few.

Tissue-Volume Reconstruction

We elaborate on the tissue reconstruction of histological serial sections using SqN in the following experiments. Here, we show that SqN is capable of registering a stack of consecutive slices independent of its order in a quality comparable to standard methods like NGF or SSD. Furthermore, common error behaviors of sequential registration processes, like a z-drift, can be improved by the global approach of SqN. The pure reconstruction of tissue illustrated in Figure 33 and Figure 35, we have also elaborated in [7], [8]. The sectioned datasets are animal brains. Figure 32 illustrates a sectioned mouse brain, data courtesy of O. Schmitt, University of Rostock, Germany. This dataset comprises 189 coronal slices of 512-by-512 pixels. Figure 35 illustrates a marmoset-monkey brain dataset; data courtesy of Harald Möller, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany [92]. This dataset comprises 69 sections with various high-resolution slices of different sizes. The sizes vary from 2252×3957 pixels up to 7655×9965 pixels. For proof of concept, we reduced the number of pixels per slice to reduce computation time to a reasonable level [7]. For registration we opt for SqN with $q = 0.5$ as presented in [8] for the mouse brain dataset with sequential SSD for comparison. For the monkey brain dataset we opt for SqN with $q = 4$ and $q = \infty$ and NGF for comparison, as elaborated in [7]. Again, the sequential approaches involve an alternating optimization framework which is based on a fixed initial and a fixed final slice to avoid the banana-effect [34], [123], [130], [131], [138]. Following [8], the sequential registration process may accumulate small errors that can cause a significant drift of the overall structure; see [138] for examples. The results illustrated in Figure 33 and Figure 35 indicate, that a global drift is not introduced using SqN. Furthermore, by optical inspection, we see that the local structure is

well aligned. Additionally, SqN needs just one pass to register all images. We compute the reults shown in Figure 33 using FAIR [100] within MATLAB. For optimization, we use a standard Gauß-Newton with Armijo line search within a multi-level scheme and a curvature regularizer, following the summation extension from Sec. 3.3. The sequential competitor SSD is computed with one single sweep. Both approaches are able to align the images, as expected. Although, the result of SqN looks smoother since the edges of the tissue are more straight. On the one hand, it seems that the sequential competitor has not converged yet. It is to be expected that the structures in the bottom of the axial and sagittal cuts are straightened as well. On the other hand, the result using SqN may be too straight such that it would not match the real, unknown shape of the tissue. The result of SSD may be closer to reality, as it isn't straightening out everything. The result illustrated in Figure 35 is computed using the python-toolbox GReAT. For optimization, we opt for L-BFGS with Armijo line search within a multi-level scheme. Again, we use a curvature regularizer for the regularity of the grids, as this provides smooth results that seem to be sufficiently realistic. Furthermore, the curvature regularizer was successfully used in different previous publications about similar application scenarios; see e.g., [85], [86], [101]. Figure 35 shows that the local structure is well aligned in just one pass with SqN, but the process for NGF has not converged yet. It still is closer to the unregistered stack of images, and the local structures are not as smooth as with SqN. The single sweep of NGF took about two times longer than the registration using SqN.
To show the invariance to permutations, we present in Figure 36 a registration result using SqN$_4$ applied to a randomly permuted stack of histological slices. After registration, the stack was permuted back. The permutation didn't influence the registration results, as can be seen by comparison of Figure 36 to Figure 35.

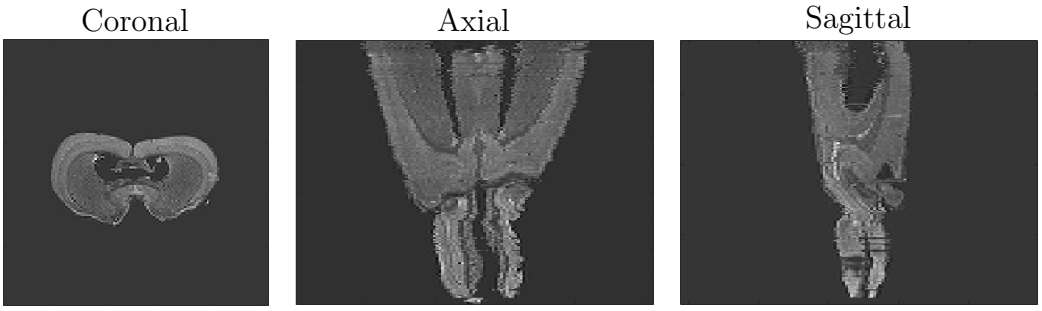


Figure 32: For an overview of the dataset, different anatomical planes (coronal, axial, sagittal) of a stack of histological mouse-brain slices are shown. On the very left, a single slide is illustrated. In the middle and right positions, the stack of histological sections is cut at an arbitrary spatial position to show the alignment of the single slices. Here, the slices are linearly pre-registered, but a variation can still be seen, which is to be corrected. The dataset comprises 189 coronal sections of 512-by-512 pixels; data courtesy of O. Schmitt, University of Rostock, Germany; see [123] for experimental details.

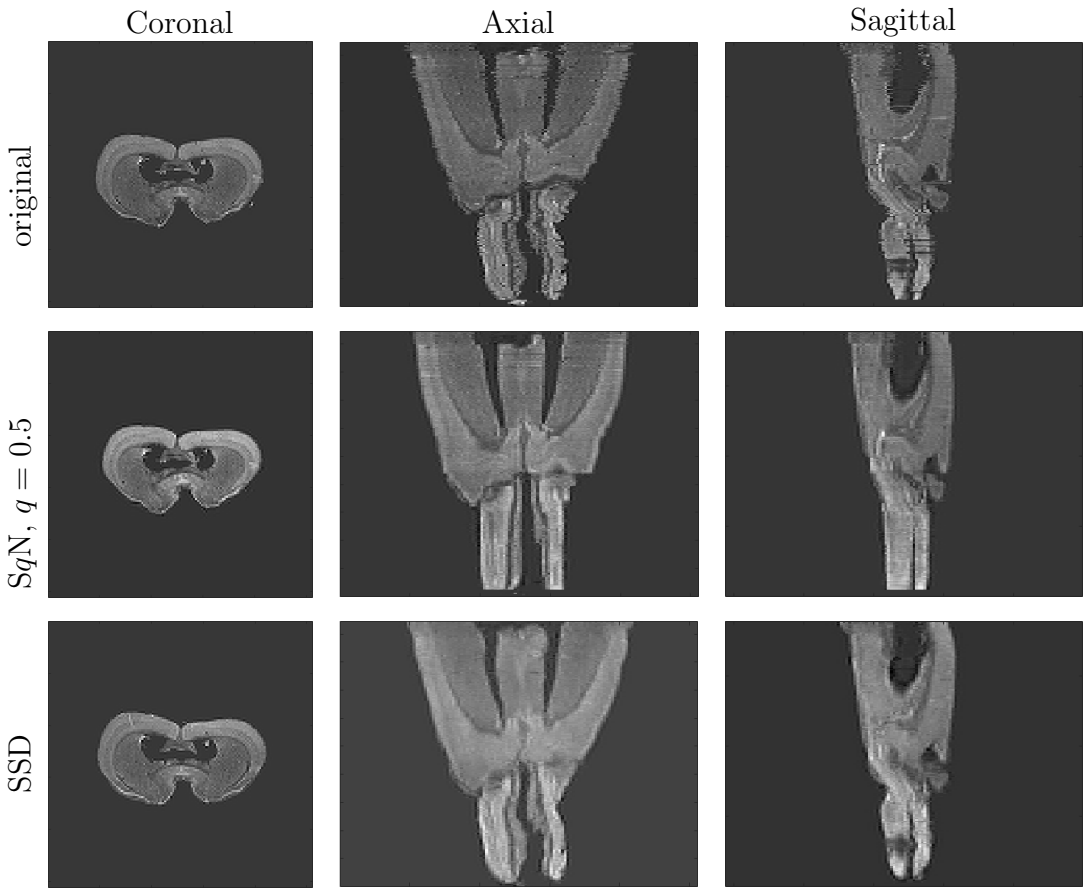


Figure 33: Displayed from left to right are exemplarily an axial, coronal, and sagittal slice of the 3D data of size 512-by-512-by-189. Displayed are non-registered data (top row), SqN-registered data (middle row) and SSD-registered data (bottom row). Note that the different cut-positions do not necessarily correspond.

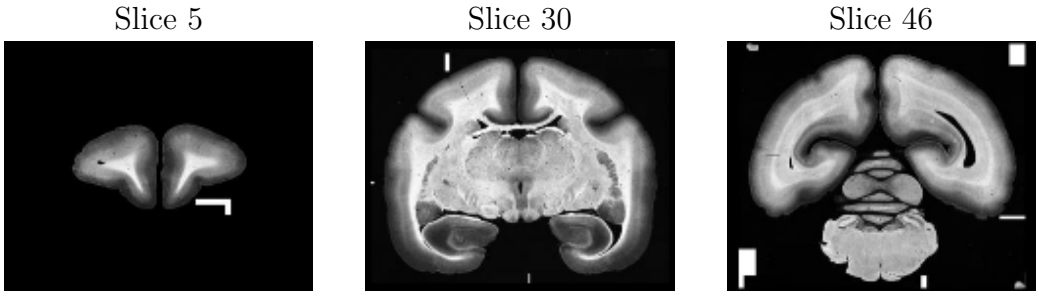


Figure 34: Three representative axial slices of a marmoset monkey brain dataset. Data courtesy of Harald Möller, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; see also [92].

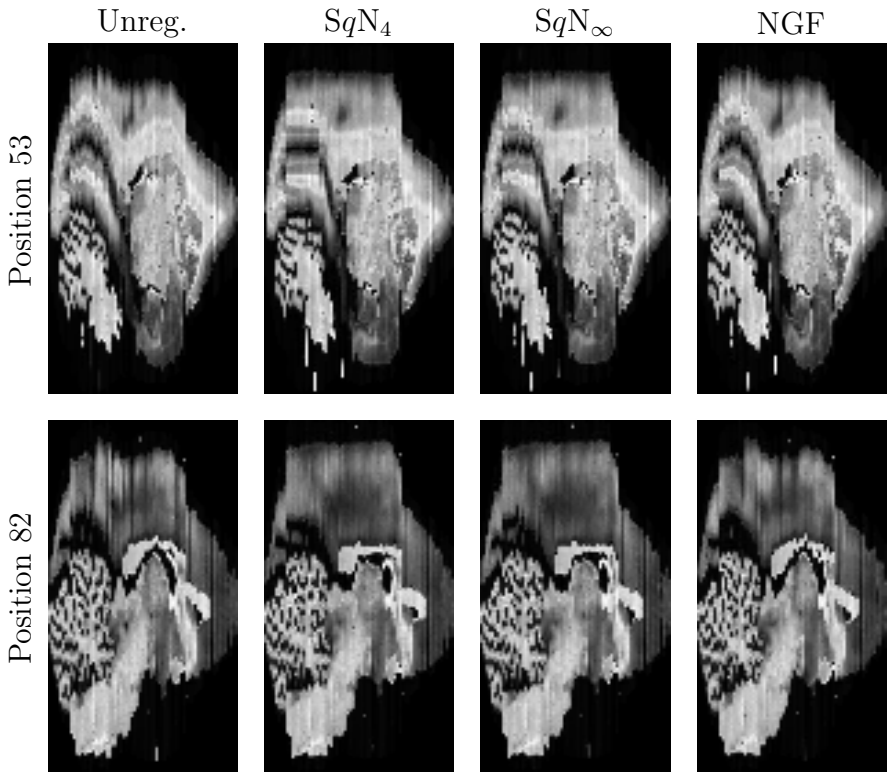


Figure 35: Shown are the registration results for 3D reconstruction of the monkey brain datasets. For illustration, we show only 2D slices that are sagittal cuts at two positions, i.e., 53 and 82. Data courtesy of Harald Möller, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; see also [92].

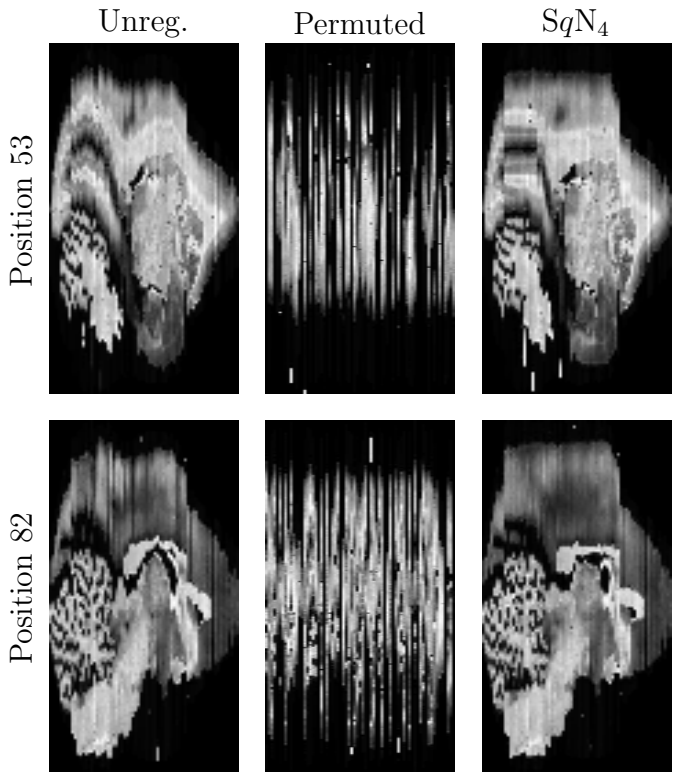


Figure 36: Registration results after random permutation of the axial slices. As expected, the results are the same as for the non-permuted image stack; also see Fig. 35 for comparison. Note that the shown positions of the permutation do not necessarily correspond to the non-permuted stack.

Fusion with Advanced Imaging for Information-Maximization

It may be interesting to examine histological sectioning in even more detail beyond the given information by the respective histological stainings. An examination of structure across the slides is possible by three-dimensional reconstruction of histological tissue from serial sectioning. Now to combine information from different stainings and to understand the correlation between macroscopic and microscopic structure, a combination of histological staining and matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI MSI) is a way to go [113], [135], [141]. While the histological staining highlights certain structures like cell nuclei or specific tissue structures, MALDI MSI provides full spectra of detailed molecular information. The reconstructed histological tissue, combined with 3D MALDI MSI, may be highly beneficial for histological and molecular examination [60]. Such a combination may be especially exciting for an examination of biopsies after tumor resection, for example. For a proof of concept and to show the potential of the fusion, we register a sectioned kidney as well as a sectioned heart of a mouse. This section is intended to show the applicability in a clinical-development scenario, as clinicians may request an examination tool.

Figure 37 shows representative histological sections of the datasets. The kidney dataset consists of 32 slices, though 128 slices are to be registered because of the four modalities. The heart dataset consists of 54 slices. For the heart, we register just three different modalities; thus, in total, 162 slices. For spatial alignment we choose SqN with $q = 4$ as a distance measure and a curvature regularizer for each images as presented in Eq. (29) and Eq. (6). We chose SqN because of successfull results in previous publications; cf. [7]–[9]. Furthermore, the fusion of the different modalities is done concomitantly with the registration of the slices. This is an application-specific assumption since the same physical slices are used for creating the images of the different modalities. Thus, in theory, there should not be any shift of tissue across the slides.
The data of the different modalities is of different pixel-resolutions, and the slices are differently rotated, which makes preprocessing inevitable. Moreover, the irregular background of the images is to be removed. Typically, there are dust and dirt, as well as stitching errors from the object-slides and the concatenation of image patches [86], [124]. The stitching errors are due to multiple photographs of the tissue-object that need to be combined to obtain the whole slice. A straightforward connected-components method [79] is used for the segmentation. This kind of segmentation is sufficient in our case since the contrast between tissue-object and background is strong enough. Scaling and rotation of the slices are linearly corrected based on the respective principal axes, as described in [99]. This prealigns the slices based on the center of mass, which is computed using the image

intensities. For our proof-of-concept study, we use two different molecular mass signals for the kidney. The molecular mass 805 focuses on localizing specific proteins around the medulla part of the kidney. Furthermore, the molecular mass 1106 localizes proteins in areas around vessel-structures. Figure 38 illustrates representative slices of the kidney, showing the mass distribution of the masses 805 and 1106. Figure 39 shows a 3D reconstruction after registration of the single slices. Besides $q = 4$, we chose a global normalization strategy with edge-parameter $\eta = 8 \cdot 10^{-5}$ and a curvature regularizer with regularization parameter $\alpha = 2 \cdot 10^{-4}$. We choose a spline interpolation model and a multi-level strategy over 3 different resolution-levels. For optimization we use L-BFGS with Wolfe-linesearch; c.f. Sec. 2.7 and [103]. The fusion of the two different molecular masses m/z 805 and m/z 1106 after 3D reconstruction are illustrated in Figure 39. For better distinction, the masses are color-coded. The red mass is m/z 805; the green mass is m/z 1106. The registration of the full stack is finally visualized using MeVisLab [52]. The examination of the tissue is possible in 3D. The advanced viewing-tools of MeVisLab based on the toolkit Open Inventor by SGI [145] allow for free rotation and free cutting of the reconstructed tissue in the 3D space. This is especially useful for the examination of tissue structures aside from the typical anatomical planes. Figure 42 shows two different, arbitrary cuts of the reconstructed Gomori volume. Apart from the 3D volumes, a split-screen-mode for examining the plain, single slices, and the slices blended with molecular mass distribution images from MALDI MSI in a side-by-side view is possible. Scrolling through the stack of sections is possible since both views are connected. Figure 41 shows a screenshot of the split-screen-view in MeVisLab for a Gomori-MALDI-Blending for molecular mass m/z 805. For better contrast, the mass is depicted in green. The same split-screen-view is shown in Figure 43 for the heart data. Again, the blended molecular mass image is depicted in green for better contrast. The reconstructed 3D volume of the Gomori-stained heart slices is illustrated in Figure 44. Note that the striped coloration of the volume is due to different lighting of the scenery while taking photos of the stained tissue-object slides in the laboratory.

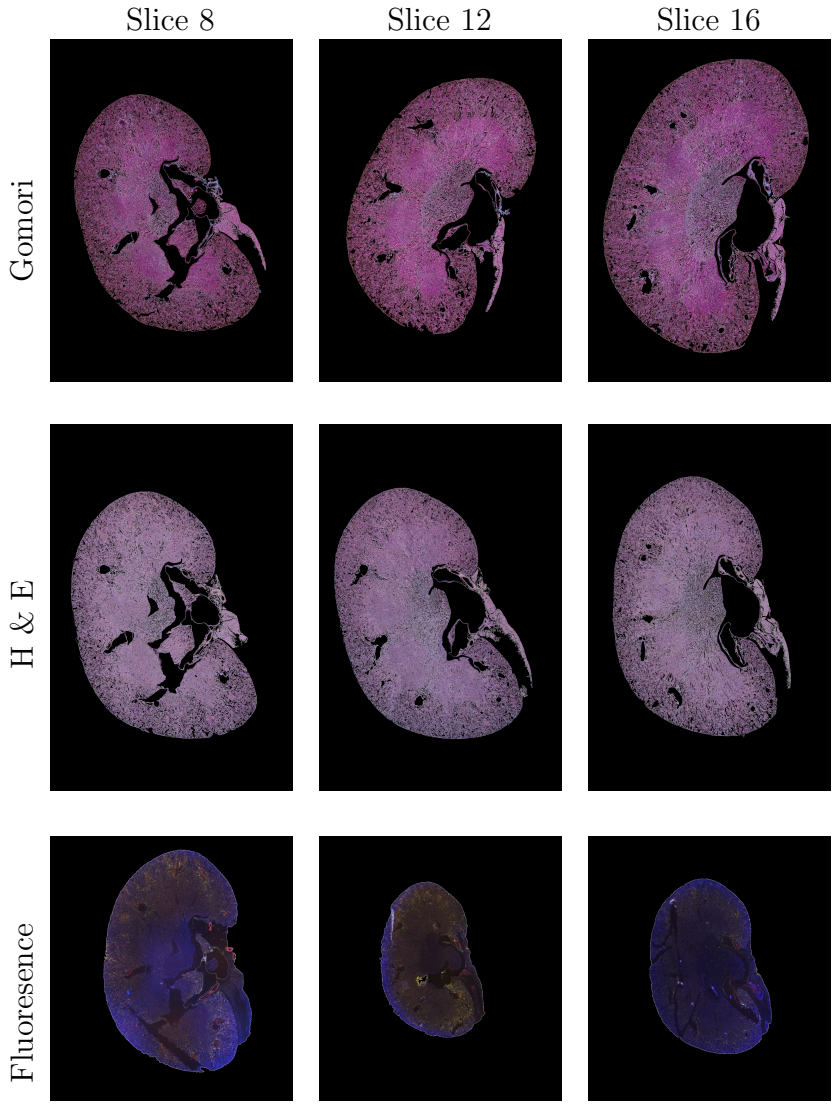


Figure 37: Illustrated are three representative histological slices of a mouse kidney. The corresponding slices are shown in three different stains. The stainings are Gomori, hematoxylin, and eosin (H & E) and a fluorescence stain. The different stainings highlight specific, for examination, relevant structures of the tissue. Data courtesy of Joachim Jankowski, Institute for Molecular Cardiovascular Research, UK Aachen, Aachen, Germany.

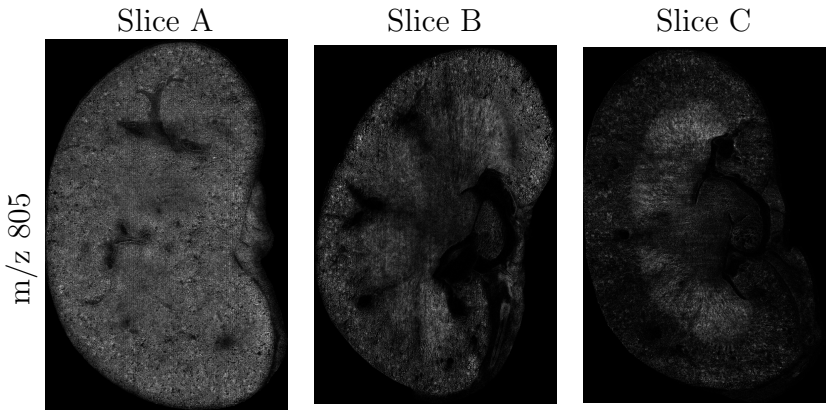


Figure 38: Illustrated are three representative MALDI MSI slices of a mouse kidney for molecular mass m/z 805. Especially slice C illustrates the focus of the intensity signal on the medulla of the kidney. From left to right, the slices are from deeper within the kidney, such that slice A shows quite a different intensity distribution. Data courtesy of Joachim Jankowski, Institute for Molecular Cardiovascular Research, UK Aachen, Aachen, Germany.

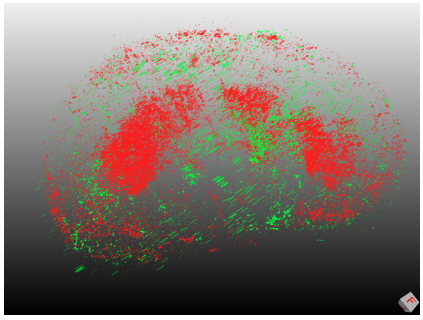


Figure 39: Illustration of the spatial mass distribution after registration of the MALDI MSI optical images. Molecular mass m/z 805 is depicted in red, m/z 1106 is depicted in green. The molecular mass m/z 1106 has a very low intensity signal and looks unspecific in the above illustration. Mass m/z 805 highlights the mass distribution within the medulla of the kidney. The depicted cube in the right bottom corner shows the actual rotation within the 3D-space.

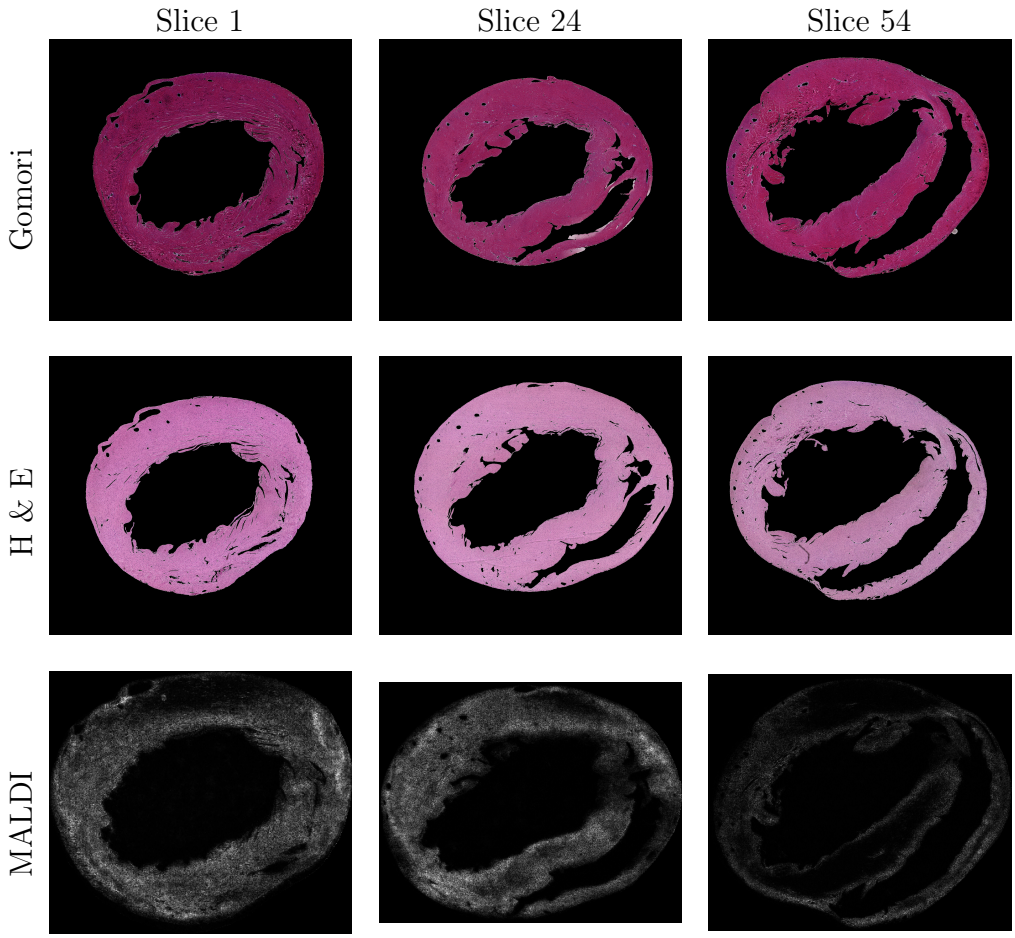


Figure 40: Illustrated are three representative histological slices of a mouse heart. The corresponding slices are shown in two different stains and with MALDI MSI. The respective stainings are Gomori and hematoxylin and eosin (H & E). Data courtesy of Joachim Jankowski, Institute for Molecular Cardiovascular Research, UK Aachen, Aachen, Germany.

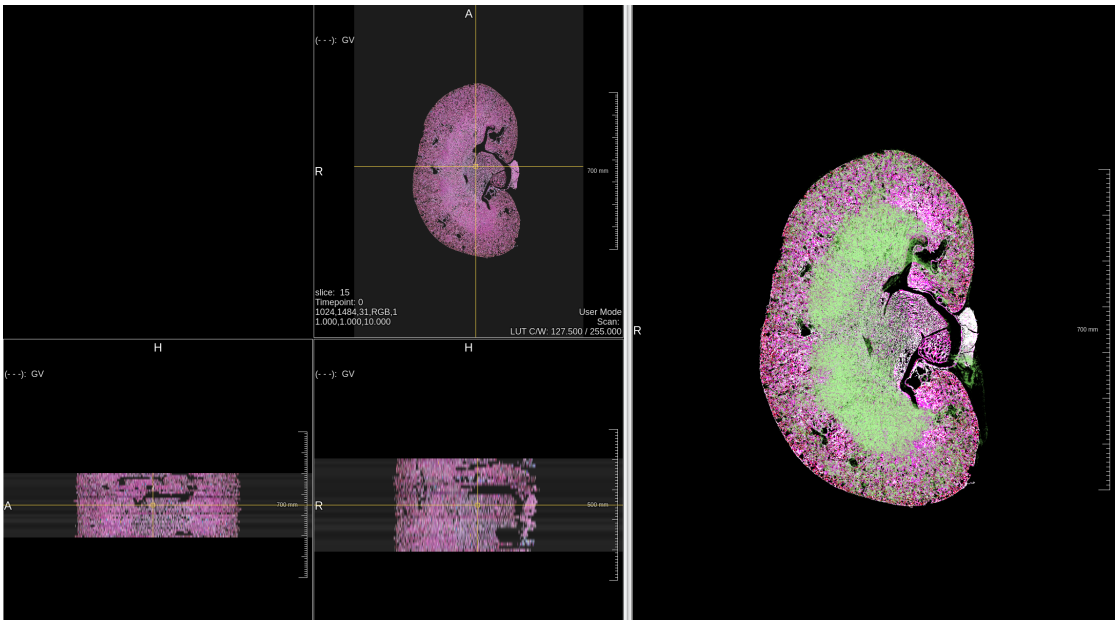


Figure 41: Screenshot of a split-screen-view within MeVisLab. The right-hand-side illustrates one of the Gomori slices blended with the corresponding optical mass image of m/z 805. On the left-hand side, the registered stack of slices is illustrated. Shown are the respective axial (top right), sagittal (bottom left) and coronal (bottom right) cuts of the stack.

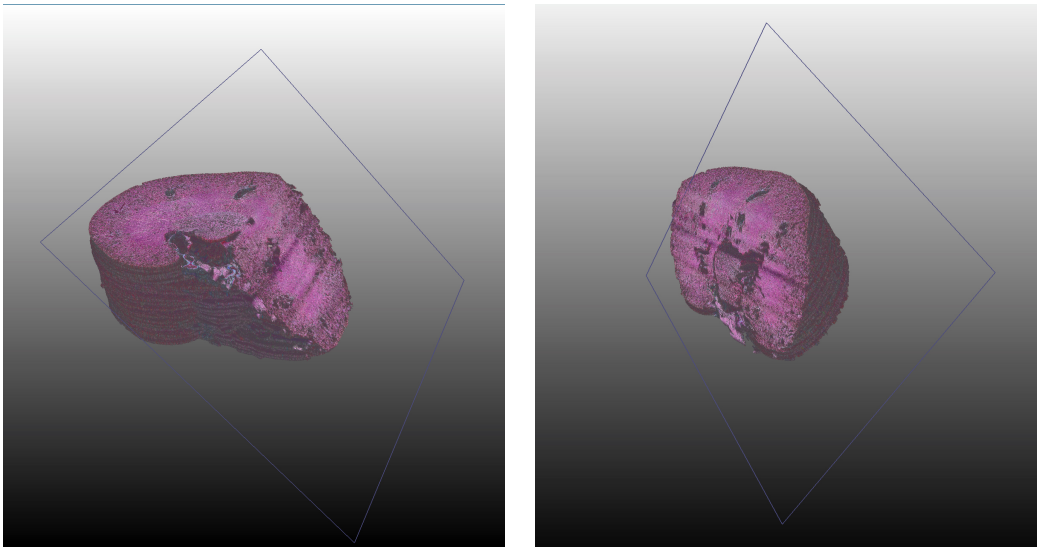


Figure 42: Illustrated is the reconstructed volume of the Gomori-stained kidney slices from different view-angles. The squares shown in the images are the cutting-planes. The planes are freely movable in the space. This allows a detailed examination of the tissue by virtual cutting.

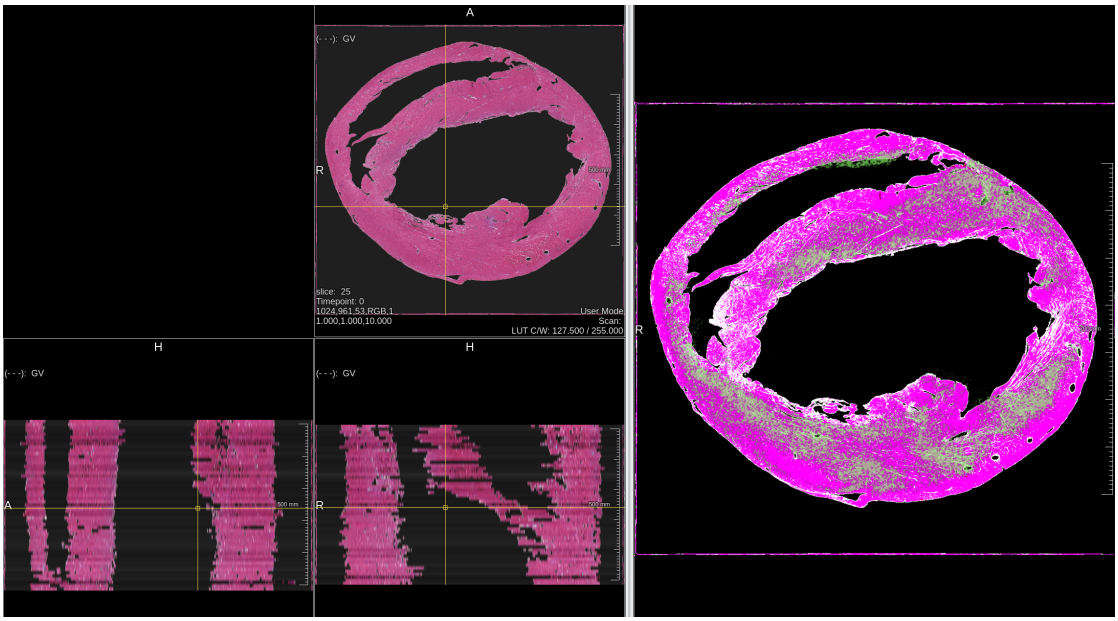


Figure 43: Screenshot of a split-screen-view within MeVisLab. The right-hand-side illustrates one of the Gomori slices of the heart blended with the corresponding optical mass image of m/z 660. On the left-hand side, the registered stack of slices is illustrated. Shown are the respective axial (top right), sagittal (bottom left) and coronal (bottom right) cuts of the stack.

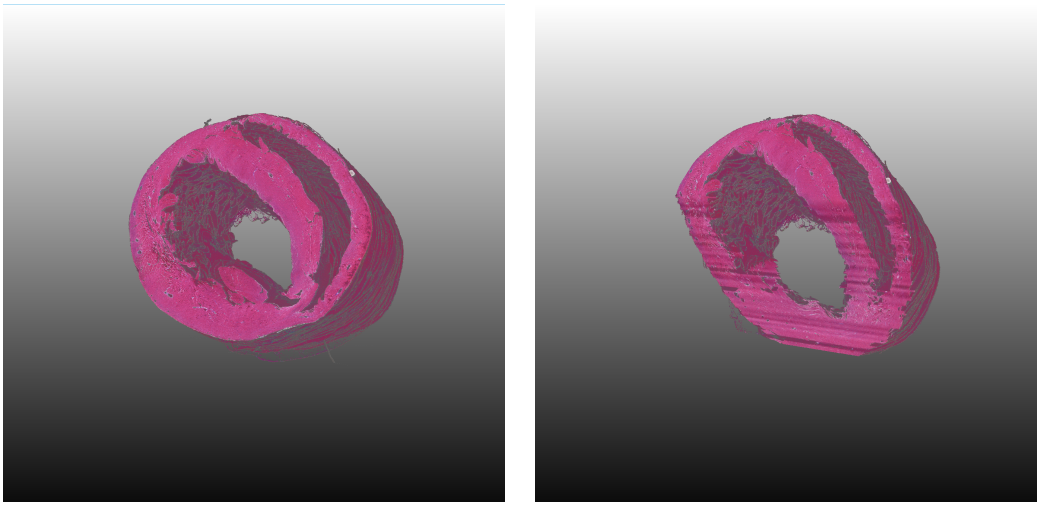


Figure 44: Illustrated is the reconstructed volume of the Gomori-stained heart slices from different view-angles. On the right, the volume is additionally cut to show the inner alignment.

4.5 Future Work: Application of SqN in Deep Learning

Considering the current state of research concerning neural networks and deep learning methods, we examine our introduced distance measurement framework within a deep learning approach. This section is intended to show the applicability of SqN within state-of-the-art deep learning approaches in a proof of concept manner. For this purpose, we take recently proposed network architectures for image registration [57]–[59], [77], [78], as a basis. Here, we give a proof of concept example for three images, which is to be studied in more detail in the future. Since neural networks and deep learning are not the focus of this work, we refer to future works that should be considered concerning groupwise image registration using deep learning methods for a more detailed analysis.

The network we use is a convolutional neural network (CNN) and follows the architecture of a U-Net [117]. More precisely, the architecture is based on [77] and is illustrated in Fig. 48; for details see [77], [78]. For general details on neural networks and deep learning methods, we refer to appropriate literature like, e.g., [36], [84], [133]. Here, we roughly outline the idea of the network and describe the outcome for our tailored approach using SqN. The network, as presented in [77], is an adaption of a U-Net. The network is a sequence of repetitive operations. These comprise several convolutions using 5×5 and 3×3 kernels, max-pooling operations, batch-normalizations, and application of rectified linear unit (ReLU) activation-functions. In summary, the images are pre-processed individually using a convolution. Then, they are concatenated such that several poolings and convolutions can be applied. This increases the number of channels of the concatenated dataset and lowers the spatial dimension, which boils down the data to relevant features. Using more convolutional layers, the final output of the network are the respective deformation grids in full resolution. Here, we refer to *spatial transformer networks*; see e.g. [71], for details.

The used network is following a so-called *unsupervised* approach. Here, this is based on the variational framework we also use for our registration, introduced in Sec. 2.1. The supervision is performed using the minimization of the variational energy composed of the distance energy and regularization energy. In our case, the distance energy is computed using SqNwith $q = 4$. For regularization, we again use curvature; see Sec. 3.3. The fact that such an approach is suitable for image registration and also might be superior to conventional registration methods has been shown in [57]–[59], [78]. As it uses SqN, our extension is not using a specific reference image. Furthermore, it handles multiple images. Our implementation uses Python and, more precisely, the package PyTorch. For our proof of concept application, we use the DIR-Lab dataset [14]–[16], which are 4D CT-scans of human lungs. Given are ten patient datasets in total, each having ten time points of three spatial dimensions. For learning, we use 3-tuples of time points from

the first five datasets. We start learning with time points $(1, 2, 3)$ from dataset 1, minimizing the energy for 450 epochs. Then, we take the next tuple, which comprises time points $(2, 3, 4)$, for the next 450 epochs, doing the same for the following tuples. This gives in total eight 3-tuples for each of the five datasets, such that the given set of tuples is $\{(k, k+1, k+2) \,|\, k \in \{1, 2, \ldots, 8\}\}$. Following this, we have 40 datasets for training. The learning rate is set to $5 \cdot 10^{-3}$, while the regularization parameter $\alpha = 5 \cdot 10^{-5}$. For SqN we use $q = 4$ and globally normalized gradient features with edge parameter $\eta = 333$. The edge parameter was chosen on a visual basis, such that the background noise does not play a major role in computing the image gradients. For optimization, we use Adam [75], which is a first-order optimization algorithm using adaptive moment-estimates, comparable to stochastic gradient descent (SGD); see [75] for details. We chose Adam because of its availability within PyTorch. For all time points, we use the resolution 128-by-128-by-64 and compute everything on a GPU. The training process takes several hours, but once learned, we are able to register three time points of the left-over datasets in about 0.006 seconds. We evaluate the trained network on dataset number six, which is unseen and therefore qualifies for a proper evaluation. In Figures 46 and 47 the results are exemplarily illustrated by absolute differences for the extreme inhale and exhale phases as well as for consecutive slices as we have used for learning using two time points each for difference computation. It can be seen that the difference is reduced after applying the network for registration. The results are better for the correction of consecutive time points, most likely because the network was trained to consecutive slices. In both Figures, we can see an improvement, although the result is far from being comparable to results shown by e.g., [76], [112]. Nevertheless, for just five used datasets within the learning process, the results seem promising. We suggest further analysis of approaches like SqN considering multiple images for image registration combined with neural networks, as these might improve results for e.g. real-time applications. A network that is capable of registering a whole series of images might be exciting in a wide variety of applications.

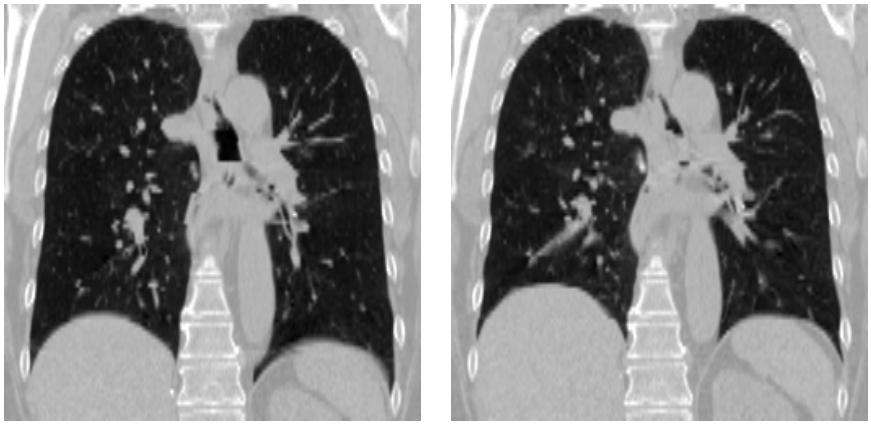


Figure 45: Visualization of 4D (x,y,z,t) CT data from DIR-Lab [14] (see www.dir-lab.com). Two representative axial slices (x,y) at a representative time t of a human lung are shown. Data courtesy [14]–[16].

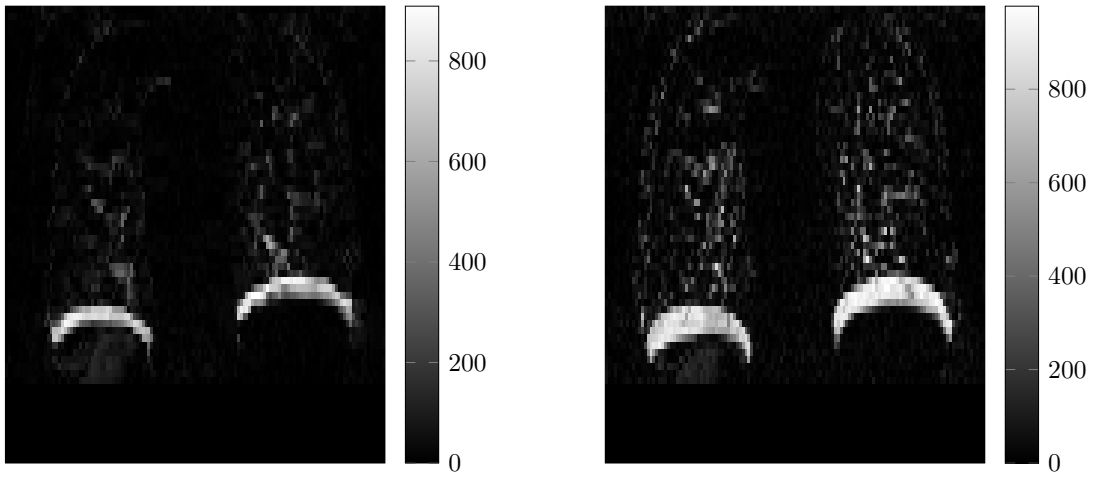


Figure 46: Shown are the absolute differences of the maximum inhale and maximum exhale phase as 2D slice (x,y) of the human lung from the 4D-CT (x,y,z,t) DIR-Lab dataset [14] after (left) registration using the neural net with SqN and before (right) registration.

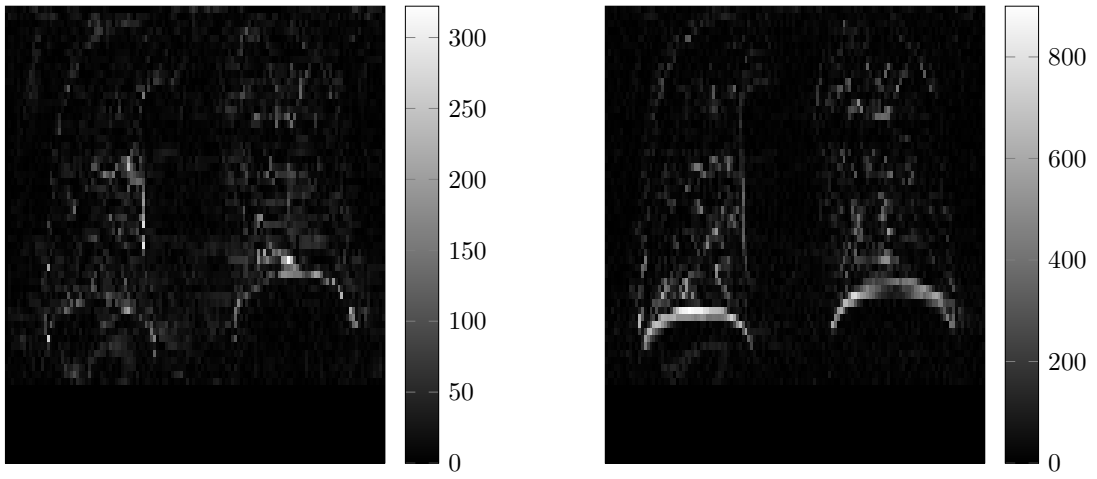


Figure 47: Shown are the absolute differences of two consecutive time points as 2D slice (x,y) of the human lung from the 4D-CT (x,y,z,t) DIR-Lab dataset [14] after (left) registration using the neural net with SqN and before (right) registration.

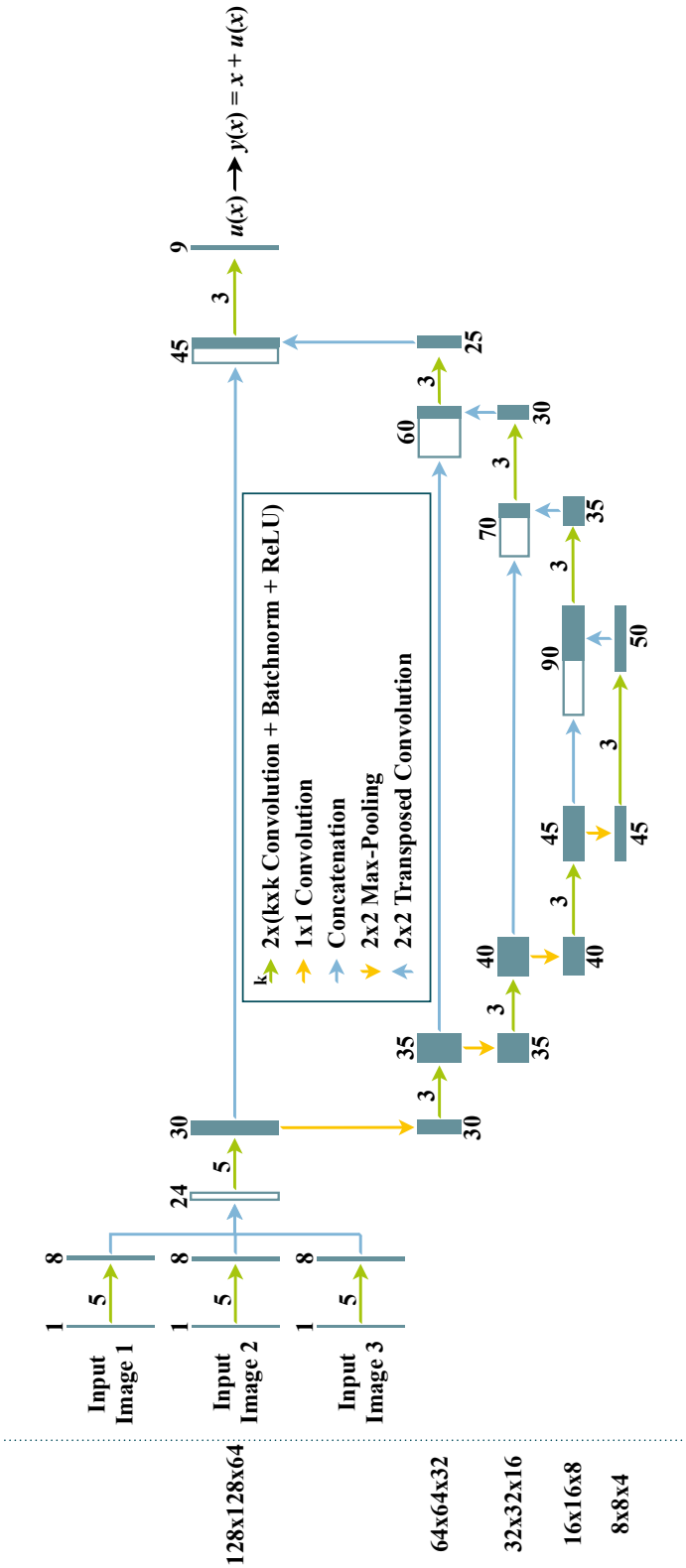


Figure 48: Illustration of the U-Net-shaped convolutional neural network for three images. The illustration is adapted from [77].

5 Discussion, Conclusion and Future Research

In this thesis, we presented a distance measurement framework for image registration of multiple images. Based on the variational image registration framework for two images presented in Section 2, we derived a novel, versatile distance measurement framework in Section 3. We started by defining the notation of the variational framework for multiple images and multiple transformations as well. Thereupon, we examined a straightforward extension to use two-image based registration methods for multiple image registration in a sequential manner. This extension is based on the summation of pairs of images. Despite the definition using a sum of two-image based distances, this approach may be used without a fixed reference image. Typically, one of the transformation grids is fixed [99], [100], such that the standard methods, reviewed in Sec. 2 can directly be used. A drawback is that only the similarity information of couples is considered for the computation of the energy.
In order to derive our novel multiple image distance framework in the most versatile way possible, we used a feature matrix, built from the vectorized feature images. This feature array allowed us to examine the properties of our distance measure without early specification of what features are used for measurement of image similarity. Furthermore, different normalization techniques for the features were discussed to overcome differences in magnitude concerning different choices of features. The particular choice of gradient features is a prominent example. For these, the magnitude is different for different modalities. Therefore, an application of gradient features in multi-modal image registration makes an appropriate normalization indispensable. Before we headed for the central part of this thesis, we briefly described how we extend the regularization to multiple images. The regularization for registration of multiple images does not lie within the scope of this work but is an important topic that needs to be examined. Our approach is to regularize each transformation individually. The resulting energy is a summation of all individual energies. This is the same regularization as for the sequential image registration. The major drawback of this approach is that there is no coupling across the transformation girds. Time-dependent data, like dynamic imaging from DCE-MRI, may suffer from a simple, individual regularization of the transformation grids since the time dependency is not considered. This may lead to smooth individual frames but jumps and discontinuities over time, e.g., across the frames. First, investigations concerning image registration of multiple images are made in, e.g., [1], where a time-dependent smoothing extends a spatially coupling. After this, we considered three different geometrically inspired approaches for multiple image registration in order to derive a general distance measurement framework, which is the central result of this work. We started with an idea from color-image

regularization [102].
The idea is to align the directions of the intensity gradients. In color-image regularization, this minimizes color noise since edges across the different color channels should match, and noise only occurs if they do not. This idea is transferred to image registration by using it for distance measurement with normalized gradient features. The Schatten-Norms, as a result of this, act as a smooth approximation to the rank-function. So the first result is the rank-minimization of the feature matrix consisting of the feature-images as columns. The geometrical background is an alignment of the gradients by making them linearly dependent. The major drawback of this approach is the fact that the energy is minimal for the trivial solution. Without proper regularization and an additional constraint concerning the translation of the transformation grids, this approach tends to shift all images out of the domain, such that all columns become zero. The output is an optimum for the feature matrix of rank zero, meaning the feature matrix is the trivial solution. The idea of coupling the color channels may also be interesting for non-gray-valued medical image data. Besides multiple images, the feature matrix may also incorporate multiple channels of the image data, such that each channel can be transformed on its own. For this, the later introduced weighting concept might be handy to control the influence of coupling channels and images.
A different, geometrically inspired point of view is to consider the vectorized feature-images to span a parallelotope of which the volume is to minimize. This follows the same idea of minimizing the rank of the feature matrix but is modeled slightly differently. The volume is given by the determinant of the correlation matrix built by the quadratic form of the feature matrix multiplied with itself. In this case, the determinant is rephrased using the product of singular values to which additionally the logarithm is applied to convert the product to a sum. Finally, the rank minimization from the first approach, using the Schatten-Norms as a rank-function approximation, is rephrased via modeling the volume of the spanned feature-parallelotope without using the Schatten-Norms. In general, the volume minimization could also be expressed using Schatten-Norms. The derivation without them via a geometrical idea shows a justification to use the Schatten-Norms for the first approach. Comparing both approaches, both aim for a minimization of the rank of the feature matrix, which implies that small singular values vanish. Following [136], the best rank-one approximation is a maximum singular value, meaning all smaller singular values vanish, and the full weight is put on the largest singular value. Here, we have the major drawback of the volume minimization approach. Since we are using the logarithm function, the singular values can not be zero; otherwise, the energy is undefined and the approach becomes numerically unstable for singular values tending to zero. Because of the weaknesses of these two approaches, we discussed the alignment of images using

the correlation of features. For this purpose, we took the correlation matrix that we already built for computing the volume of the feature parallelotope as a basis. Due to the assumption that our approach is based on normalized features, the correlation matrix is a concatenation of the inner products expressing the angles between the respective feature vectors. Here, we presented the main idea of the correlation approach. Since the features are normalized, the main diagonal of the correlation matrix always consists of ones. The best case, in which all feature maps are aligned is that all inner products are either 1 or -1. The worst case is the identity matrix, such that we wanted to be as far as possible from this solution. From this maximization problem, we built the minimization problem using the negative sign. By a straightforward computation, we saw that we do not need to subtract the identity for minimization.
To go even further, we added a weighting matrix, such that the considered neighborhood of images, meaning the number of images for comparison, can be controlled. Different weightings may be interesting for specific applications to incorporate prior knowledge of the connection between images. A gaussian weighting or even a weighting, considering only three to four images, may be worthwhile testing. With Theorem 2, we furthermore saw that the correlation approach could be converted to the rank minimization approach by adapting q, meaning that all approaches are closely related. The novel distance measure framework can be derived from various geometrical ideas, it is straightforward to interpret and extremely versatile. Especially our given overview of invariances and the energy-behavior of the novel distance measurement framework revealed different interesting results. In order to use the approach, we do not need to choose a specific reference image. Furthermore, the ordering of images is completely irrelevant due to the use of singular values. The energy is also quite smooth compared to others, and for $q > 2$, our numerical experiments show that there is always a global minimum. Concludingly, the examination of the first derivative and the error estimates show that the approach is straightforward to use also with higher-order optimization schemes. Moreover, the error behavior is as expected since it tends to zero when increasing the number of grid points, which is suitable for use with multi-level schemes. All statements of Sec. 3 have been proven by numerical application. This was done in Sec. 4 using medical applications.

Future Research

This thesis introduces a novel distance measurement framework. Here, we will give an overview of what may and should be done in future works to gain even more insight into SqN and its properties for image registration. We begin with a few obvious things. The choice of the parameter q for the Schatten-Norms plays a significant role. A more detailed analysis – building on the given results – of

the behavior of the Schatten-Norms with respect to q should be done. Especially the behavior of the Schatten-Norms with respect to q concerning the changing transformation fields during the optimization process may be interesting to understand. Furthermore, the regularization was just extended in a pertinent way. Here, the focus should lie on investigating regularization methods for multiple images, as started in [1]. Concerning the applications part, different features should be tested in medical applications. Here, we focused on intensity- and gradient-features. More advanced features may be interesting for applications combined with deep-learning methods. Given the present situation concerning neural networks, artificial intelligence and deep learning methods, it may be interesting to see how the distance measurement framework performs. A specifically designed network can probably learn relevant features to give the framework another degree of freedom. Overall, the SqN distance measure is very promising and should be investigated further.

BIBLIOGRAPHY

[1] H. O. Aggrawal and J. Modersitzki, "Accelerating the Registration of Image Sequences by Spatio-temporal Multilevel Strategies", *arXiv preprint*, 2020.

[2] Y. Amit, "A nonlinear variational problem for image matching", *SIAM Journal on Scientific Computing*, vol. 15, no. 1, pp. 207–224, 1994.

[3] D. P. Bertsekas, "Nonlinear programming", *Journal of the Operational Research Society*, vol. 48, no. 3, pp. 334–334, 1997.

[4] K. K. Bhatia, J. Hajnal, A. Hammers, and D. Rueckert, "Similarity metrics for groupwise non-rigid registration", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2007, pp. 544–552.

[5] K. K. Bhatia, J. V. Hajnal, B. K. Puri, A. D. Edwards, and D. Rueckert, "Consistent groupwise non-rigid registration for atlas construction", in *2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821)*, IEEE, 2004, pp. 908–911.

[6] R. Bhatia, *Matrix analysis*. Springer Science & Business Media, 2013, vol. 169.

[7] K. Brehmer, H. O. Aggrawal, S. Heldmann, and J. Modersitzki, "Variational Registration of Multiple Images with the SVD Based SqN Distance Measure", in *International Conference on Scale Space and Variational Methods in Computer Vision*, Springer, 2019, pp. 251–262.

[8] K. Brehmer, B. Wacker, and J. Modersitzki, "A novel similarity measure for image sequences", in *International Workshop on Biomedical Image Registration*, Springer, 2018, pp. 47–56.

[9] ——, "Simultaneous Registration of Image Sequences–a novel singular value based images similarity measure", *PAMM*, vol. 18, no. 1, 2018.

[10] C. Broit, "Optimal registration of deformed images", PhD thesis, Computer and Information Science, University of Pensylvania, 1981.

[11] L. G. Brown, "A survey of image registration techniques", *ACM Computing Surveys (CSUR)*, vol. 24, no. 4, pp. 325–376, 1992.

[12] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping", in *European conference on computer vision*, Springer, 2004, pp. 25–36.

[13] M. Burger, J. Modersitzki, and L. Ruthotto, "A hyperelastic regularization energy for image registration", *SIAM Journal on Scientific Computing*, vol. 35, no. 1, pp. 132–148, 2013.

[14] E. Castillo, R. Castillo, J. Martinez, M. Shenoy, and T. Guerrero, "Four-dimensional deformable image registration using trajectory modeling", *Physics in Medicine & Biology*, vol. 55, no. 1, p. 305, 2009.

[15] R. Castillo, E. Castillo, D. Fuentes, M. Ahmad, A. M. Wood, M. S. Ludwig, and T. Guerrero, "A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive", *Physics in Medicine & Biology*, vol. 58, no. 9, p. 2861, 2013.

[16] R. Castillo, E. Castillo, R. Guerra, V. E. Johnson, T. McPhail, A. K. Garg, and T. Guerrero, "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets", *Physics in Medicine & Biology*, vol. 54, no. 7, p. 1849, 2009.

[17] B. Chalmond and S. C. Girard, "Nonlinear modeling of scattered multivariate data and its application to shape change", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 422–432, 1999.

[18] G. Charpiat, O. Faugeras, and R. Keriven, "Approximations of shape metrics and application to shape warping and empirical shape statistics", *Foundations of Computational Mathematics*, vol. 5, no. 1, pp. 1–58, 2005.

[19] T. Che, Y. Zheng, J. Cong, Y. Jiang, Y. Niu, W. Jiao, B. Zhao, and Y. Ding, "Deep group-wise registration for multi-spectral images from fundus images", *IEEE Access*, vol. 7, pp. 27 650–27 661, 2019.

[20] U Clarenz, M Droske, and M Rumpf, "Towards fast non-rigid registration", *CONTEMPORARY MATHEMATICS*, vol. 313, pp. 67–84, 2002.

[21] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multi-modality image registration based on information theory", in *Information Processing in Medical Imaging*, vol. 3, 1995, pp. 263–274.

[22] S. S. Collis and M. Heinkenschloss, "Analysis of the streamline upwind/Petrov Galerkin method applied to the solution of optimal control problems", *CAAM TR02-01*, vol. 108, 2002.

[23] T. F. Cootes, C. J. Twining, V. S. Petrovic, K. O. Babalola, and C. J. Taylor, "Computing accurate correspondences across groups of images", *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 11, pp. 1994–2005, 2009.

[24] D. Cremers, T. Kohlberger, and C. Schnörr, "Shape statistics in kernel space for variational image segmentation", *Pattern Recognition*, vol. 36, no. 9, pp. 1929–1943, 2003.

[25] C. De Boor, *A practical guide to splines*. Springer, 1978, vol. 27.

[26] M. De Craene, G. Piella, O. Camara, N. Duchateau, E. Silva, A. Doltra, J. D'hooge, J. Brugada, M. Sitges, and A. F. Frangi, "Temporal diffeomorphic free-form deformation: Application to motion and strain estimation from 3D echocardiography", *Medical image analysis*, vol. 16, no. 2, pp. 427–450, 2012.

[27] J. E. Dennis Jr and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*. Siam, 1996, vol. 16.

[28] O. Faugeras, G. Adde, G. Charpiat, C. Chefd'Hotel, M. Clerc, T. Deneux, R. Deriche, G. Hermosillo, R. Keriven, P. Kornprobst, *et al.*, "Variational, geometric, and statistical methods for modeling brain anatomy and function", *NeuroImage*, vol. 23, S46–S55, 2004.

[29] B. Fischer and J. Modersitzki, "Curvature based image registration", *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 81–85, 2003.

[30] ——, "Fast image registration: a variational approach", in *Proceedings of the International Conference on Numerical Analysis & Computational Mathematics, G. Psihoyios (ed.), Wiley*, 2003, pp. 69–74.

[31] ——, "FLIRT: A flexible image registration toolbox", in *International Workshop on Biomedical Image Registration*, Springer, 2003, pp. 261–270.

[32] ——, "Ill-posed medicine—an introduction to image registration", *Inverse Problems*, vol. 24, no. 3, 2008.

[33] M. A. Fischler and R. A. Elschlager, "The representation and matching of pictorial structures", *IEEE Transactions on computers*, vol. 100, no. 1, pp. 67–92, 1973.

[34] S. Gaffling, V. Daum, and J. Hornegger, "Landmark-constrained 3-D Histological Imaging: A Morphology-preserving Approach.", in *VMV*, Citeseer, 2011, pp. 309–316.

[35] P. E. Gill, W. Murray, and M. H. Wright, *Practical optimization*. SIAM, 2019.

[36] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT Press, 2016.

[37] A. A. Goshtasby, *Image registration: Principles, tools and methods*. Springer, 2012.

[38] A. A. Goshtasby, *2-D and 3-D image registration: for medical, remote sensing, and industrial applications*. Wiley, 2005.

[39] A. Guimond, J. Meunier, and J.-P. Thirion, "Average Brain Models: A Convergence Study", *Comput. Vis. Image Underst.*, vol. 77, no. 9, 192–210, 2000.

[40] M. E. Gurtin, *An introduction to continuum mechanics*. 1982.

[41] J.-M. Guyader, W. Huizinga, V. Fortunati, D. H. Poot, M. van Kranenburg, J. F. Veenland, M. M. Paulides, W. J. Niessen, and S. Klein, "Total correlation-based groupwise image registration for quantitative MRI", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 186–193.

[42] J.-M. Guyader, W. Huizinga, D. H. Poot, M. van Kranenburg, A. Uitterdijk, W. J. Niessen, and S. Klein, "Groupwise image registration based on a total correlation dissimilarity measure for quantitative MRI and dynamic imaging data", *Scientific reports*, vol. 8, no. 1, pp. 1–14, 2018.

[43] R. Haase, S. Heldmann, and J. Lellmann, "Deformable Groupwise Image Registration using Low-Rank and Sparse Decomposition", *arXiv preprint*, 2020.

[44] E. Haber, S. Heldmann, and J. Modersitzki, "Adaptive mesh refinement for nonparametric image registration", *SIAM journal on scientific computing*, vol. 30, no. 6, pp. 3012–3027, 2008.

[45] E. Haber and J. Modersitzki, "Volume preserving image registration", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2004, pp. 591–598.

[46] ——, "A multilevel method for image registration", *SIAM Journal on Scientific Computing*, vol. 27, no. 5, pp. 1594–1607, 2006.

[47] ——, "Intensity gradient based registration and fusion of multi-modal images", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2006, pp. 726–733.

[48] J. Hadamard, "Sur les problèmes aux dérivées partielles et leur signification physique", *Princeton University Bulletin*, pp. 49–52, 1902.

[49] B. Hafner, S. Zachariah, and J. Sanders, "Characterisation of three-dimensional anatomic shapes using principal components: application to the proximal tibia", *Medical and Biological Engineering and Computing*, vol. 38, no. 1, pp. 9–16, 2000.

[50] V. Hamy, N. Dikaios, S. Punwani, A. Melbourne, A. Latifoltojar, J. Makanyanga, M. Chouhan, E. Helbren, A. Menys, S. Taylor, *et al.*, "Respiratory motion correction in dynamic MRI using robust data decomposition registration–Application to DCE-MRI", *Medical image analysis*, vol. 18, no. 2, pp. 301–313, 2014.

[51] M. Hanke-Bourgeois, *Grundlagen der numerischen Mathematik und des wissenschaftlichen Rechnens*. Springer, 2009.

[52] F. Heckel, M. Schwier, and H.-O. Peitgen, "Object-oriented application development with MeVisLab and Python", *Informatik 2009–Im Focus das Leben*, 2009.

[53] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, M. Brady, and J. A. Schnabel, "MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration", *Medical Image Analysis*, vol. 16, no. 7, pp. 1423–1435, 2012.

[54] M. P. Heinrich, M. Jenkinson, B. W. Papież, M. Brady, and J. A. Schnabel, "Towards realtime multimodal fusion for image-guided interventions using self-similarities", in *International conference on medical image computing and computer-assisted intervention*, Springer, 2013, pp. 187–194.

[55] S. Heldmann, *Non-linear Registration Based on Mutual Information: Theory, Numerics, and Application*. Logos-Verlag, 2006.

[56] S. Henn and K. Witsch, "Iterative multigrid regularization techniques for image matching", *SIAM journal on scientific computing*, vol. 23, no. 4, pp. 1077–1093, 2001.

[57] A. Hering, B. van Ginneken, and S. Heldmann, "mlVIRNET: Multilevel Variational Image Registration Network", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 257–265.

[58] A. Hering and S. Heldmann, "Unsupervised learning for large motion thoracic CT follow-up registration", in *Medical Imaging 2019: Image Processing*, International Society for Optics and Photonics, vol. 10949, 2019.

[59] A. Hering, S. Kuckertz, S. Heldmann, and M. P. Heinrich, "Enhancing label-driven deep deformable image registration with local distance metrics for state-of-the-art cardiac motion tracking", in *Bildverarbeitung für die Medizin 2019*, Springer, 2019, pp. 309–314.

[60] J. Hermann, K. Brehmer, V. Jankowski, M. Lellig, M. Hohl, F. Mahfoud, T. Speer, S. J. Schunk, T. Tscherning, H. Thiele, and J. Jankowski, "Registration of image modalities for analyses of tissue samples using 3D image modelling", *submitted to PROTEOMICS–Clinical Applications*, 2020.

[61] D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration", *Physics in medicine & biology*, vol. 46, no. 3, 2001.

[62] E. Hodneland, E. A. Hanson, A. Lundervold, J. Modersitzki, E. Eikefjord, and A. Z. Munthe-Kaas, "Segmentation-driven image registration-application to 4D DCE-MRI recordings of the moving kidneys", *IEEE transactions on image processing*, vol. 23, no. 5, pp. 2392–2404, 2014.

[63] E. Hodneland, Å. Kjørstad, E. Andersen, J. A. Monssen, A. Lundervold, J. Rørvik, and A. Munthe-Kaas, "In vivo estimation of glomerular filtration in the kidney using DCE-MRI", in *2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA)*, IEEE, 2011, pp. 755–761.

[64] E. Hodneland, A. Lundervold, J. Rørvik, and A. Z. Munthe-Kaas, "Normalized gradient fields for nonlinear motion correction of DCE-MRI time series", *Computerized Medical Imaging and Graphics*, vol. 38, no. 3, pp. 202–210, 2014.

[65] B. K. Horn and B. G. Schunck, "Determining optical flow", in *Techniques and Applications of Image Understanding*, International Society for Optics and Photonics, vol. 281, 1981, pp. 319–331.

[66] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[67] Y. Hu, M. Modat, E. Gibson, W. Li, N. Ghavami, E. Bonmati, G. Wang, S. Bandula, C. M. Moore, M. Emberton, *et al.*, "Weakly-supervised convolutional neural networks for multimodal image registration", *Medical Image Analysis*, vol. 49, pp. 1–13, 2018.

[68] W. Huizinga, "Groupwise Image Registration in Diffusion Weighted MRI using a PCA based Dissimilarity Metric", PhD thesis, Delft University of Technology, 2013.

[69] W. Huizinga, C. T. Metz, D. H. Poot, M. de Groot, W. J. Niessen, A. Leemans, and S. Klein, "Groupwise registration for correcting subject motion and eddy current distortions in diffusion MRI using a PCA based dissimilarity metric", in *Computational Diffusion MRI and Brain Connectivity*, Springer, 2014, pp. 163–174.

[70] W. Huizinga, D. H. Poot, J.-M. Guyader, R. Klaassen, B. F. Coolen, M. van Kranenburg, R. Van Geuns, A. Uitterdijk, M. Polfliet, J. Vandemeulebroucke, *et al.*, "PCA-based groupwise image registration for quantitative MRI", *Medical image analysis*, vol. 29, pp. 65–78, 2016.

[71] M. Jaderberg, K. Simonyan, A. Zisserman, *et al.*, "Spatial transformer networks", in *Advances in neural information processing systems*, 2015, pp. 2017–2025.

[72] D. G. Kendall, "Shape manifolds, procrustean metrics, and complex projective spaces", *Bulletin of the London mathematical society*, vol. 16, no. 2, pp. 81–121, 1984.

[73] H. Y. Kim and J. Orchard, "Registering a non-rigid multi-sensor ensemble of images", in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, IEEE, 2010, pp. 5935–5938.

[74] T. Kim and Y.-J. Im, "Automatic satellite image registration by combination of matching and random sample consensus", *IEEE transactions on geoscience and remote sensing*, vol. 41, no. 5, pp. 1111–1117, 2003.

[75] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", *arXiv preprint*, 2014.

[76] L. König and J. Rühaak, "A fast and accurate parallel algorithm for nonlinear image registration using normalized gradient fields", in *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*, IEEE, 2014, pp. 580–583.

[77] S. Kuckertz, M. Heinrich, and A. Hering, "Schwach überwachtes Lernen nichtlinearer medizinischer Bildregistrierung mit neuronalen Faltungsnetzwerken",

[78] S. Kuckertz, N. Papenberg, J. Honegger, T. Morgas, B. Haas, and S. Heldmann, "Deep learning based CT-CBCT image registration for adaptive radio therapy", in *Medical Imaging 2020: Image Processing*, International Society for Optics and Photonics, vol. 11313, 2020.

[79] N. Kung, L. E. Albertelli, and D. L. Ii, *Grayscale Image Connected Components Segmentation*, US Patent 6,898,314, 2005.

[80] T. Lange, N. Papenberg, S. Heldmann, J. Modersitzki, B. Fischer, H. Lamecker, and P. M. Schlag, "3D ultrasound-CT registration of the liver using combined landmark-intensity information", *International journal of computer assisted radiology and surgery*, vol. 4, no. 1, pp. 79–88, 2009.

[81] J. Le Moigne, N. S. Netanyahu, and R. D. Eastman, *Image registration for remote sensing*. Cambridge University Press, 2011.

[82] V. S. Lee, H. Rusinek, M. E. Noz, P. Lee, M. Raghavan, and E. L. Kramer, "Dynamic three-dimensional MR renography for the measurement of single kidney function: initial experience", *Radiology*, vol. 227, no. 1, pp. 289–294, 2003.

[83] P. Leškovskỳ, A. Alekseychuk, A. Stanski, O. Hellwich, K. Schlüns, N. Zerbe, and P. Hufnagl, "Point based registration of high-resolution histological slices for navigation purposes in virtual microscopy", *Annals of the BMVA 2012, 10: 1*, vol. 18,

[84] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis", *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[85] J. Lotz, J. Berger, B. Müller, K. Breuhahn, N. Grabe, S. Heldmann, A. Homeyer, B. Lahrmann, H. Laue, J. Olesch, *et al.*, "Zooming in: high resolution 3D reconstruction of differently stained histological whole slide images", in *Medical Imaging 2014: Digital Pathology*, International Society for Optics and Photonics, vol. 9041, 2014.

[86] J. Lotz, J. Olesch, B. Müller, T. Polzin, P Galuschka, J. Lotz, S. Heldmann, H. Laue, M. González-Vallinas, A. Warth, *et al.*, "Patch-based nonlinear image registration for gigapixel whole slide images", *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 9, pp. 1812–1819, 2015.

[87] D. G. Lowe, "Object recognition from local scale-invariant features", in *Proceedings of the seventh IEEE international conference on computer vision*, IEEE, vol. 2, 1999, pp. 1150–1157.

[88] ——, "Distinctive image features from scale-invariant keypoints", *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[89] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision", IJCAI'81, 674–679, 1981.

[90] J. A. Maintz and M. A. Viergever, "A survey of medical image registration", *Medical image analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[91] T. Makela, P. Clarysse, O. Sipila, N. Pauna, Q. C. Pham, T. Katila, and I. E. Magnin, "A review of cardiac image registration methods", *IEEE Transactions on medical imaging*, vol. 21, no. 9, pp. 1011–1021, 2002.

[92] H. Marschner, A. Pampel, R. Müller, N. A. Bock, M. Weiss, S. Geyer, and H. E. Möller, "High-resolution quantitative magnetization transfer imaging of post-mortem marmoset brain", in *22nd Annual Meeting of the International Society for Magnetic Resonance in Medicine*, 2014.

[93] S. Marsland, C. J. Twining, and C. J. Taylor, "A minimum description length objective function for groupwise non-rigid image registration", *Image and Vision Computing*, vol. 26, no. 3, pp. 333–346, 2008.

[94] J. Martin and J. L. Crowley, "Comparison of correlation techniques", in *Intelligent Autonomous Systems*, 1995, pp. 86–93.

[95] C. R. Maurer and J. M. Fitzpatrick, "A review of medical image registration", *Interactive image-guided neurosurgery*, vol. 17, 1993.

[96] F. Mémoli and G. Sapiro, "A theoretical and computational framework for isometry invariant recognition of point cloud data", *Foundations of Computational Mathematics*, vol. 5, no. 3, pp. 313–347, 2005.

[97] I. Mendichovszky, M. Pedersen, J. Frøkiær, T. Dissing, N. Grenier, P. Anderson, K. McHugh, Q. Yang, and I. Gordon, "How accurate is dynamic contrast-enhanced MRI in the assessment of renal glomerular filtration rate? A critical appraisal", *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 27, no. 4, pp. 925–931, 2008.

[98] C. T. Metz, S. Klein, M. Schaap, T. van Walsum, and W. J. Niessen, "Nonrigid registration of dynamic medical imaging data using nD+ t B-splines and a groupwise optimization approach", *Medical image analysis*, vol. 15, no. 2, pp. 238–249, 2011.

[99] J. Modersitzki, *Numerical methods for image registration*. Oxford University Press, 2004.

[100] ——, *FAIR: flexible algorithms for image registration*. SIAM, 2009, vol. 6.

[101] J. Modersitzki, O. Schmitt, and S. Wirtz, "Registration of histological serial sectionings", in *Mathematical Models for Registration and Applications to Medical Imaging*, Springer, 2006, pp. 63–80.

[102] T. Möllenhoff, E. Strekalovskiy, M. Möller, and D. Cremers, "Low rank priors for color image regularization", in *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, Springer, 2015, pp. 126–140.

[103] J. Nocedal and S. Wright, *Numerical optimization*. Springer, 2006.

[104] F. P. Oliveira and J. M. R. Tavares, "Medical image registration: a review", *Computer methods in biomechanics and biomedical engineering*, vol. 17, no. 2, pp. 73–93, 2014.

[105] G. Opfer, *Numerische Mathematik für Anfänger*. Springer, 1993, vol. 4.

[106] J. Orchard and L. Jonchery, "Ensemble registration: aligning many multisensor images simultaneously", in *Image Processing: Algorithms and Systems VII*, International Society for Optics and Photonics, vol. 7245, 2009, 72450O.

[107] T. Papadopoulo and M. Lourakis, "Estimating the Jacobian of the Singular Value Decomposition: Theory and Applications", in *European Conference on Computer Vision*, Springer, 2000, pp. 554–570.

[108] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert, "Highly accurate optic flow computation with theoretically justified warping", *International Journal of Computer Vision*, vol. 67, no. 2, pp. 141–158, 2006.

[109] X. Pennec, R. Stefanescu, V. Arsigny, P. Fillard, and N. Ayache, "Riemannian elasticity: A statistical regularization framework for non-linear registration", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2005, pp. 943–950.

[110] D. Perperidis, R. Mohiaddin, and D. Rueckert, "Construction of a 4D statistical atlas of the cardiac anatomy and its use in classification", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2005, pp. 402–410.

[111] M. Polfliet, S. Klein, W. J. Niessen, and J. Vandemeulebroucke, "Laplacian eigenmaps for multimodal groupwise image registration", in *Medical Imaging 2017: Image Processing*, International Society for Optics and Photonics, vol. 10133, 2017.

[112] T. Polzin, M. Niethammer, M. P. Heinrich, H. Handels, and J. Modersitzki, "Memory efficient LDDMM for lung CT", in *International conference on medical image computing and computer-assisted intervention*, Springer, 2016, pp. 28–36.

[113] W. Pusch, M. T. Flocco, S.-M. Leung, H. Thiele, and M. Kostrzewa, "Mass spectrometry-based clinical proteomics", *Pharmacogenomics*, vol. 4, no. 4, pp. 463–476, 2003.

[114] J. Quanico, L. Hauberg-Lotte, S. Devaux, Z. Laouby, C. Meriaux, A. Raffo-Romero, M. Rose, L. Westerheide, J. Vehmeyer, F. Rodet, *et al.*, "3D MALDI mass spectrometry imaging reveals specific localization of long-chain acylcarnitines within a 10-day time window of spinal cord injury", *Scientific reports*, vol. 8, no. 1, pp. 1–11, 2018.

[115] A. Roche, G. Malandain, and N. Ayache, "Unifying maximum likelihood approaches in medical image registration", *International Journal of Imaging Systems and Technology*, vol. 11, no. 1, pp. 71–80, 2000.

[116] R. T. Rockafellar, *Convex analysis*. Princeton University Press, 1970, vol. 28.

[117] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.

[118] D. Rueckert and J. A. Schnabel, "Medical image registration", in *Biomedical image processing*, Springer, 2010, pp. 131–154.

[119] M. Rumpf and B. Wirth, "A nonlinear elastic shape averaging approach", *SIAM Journal on Imaging Sciences*, vol. 2, no. 3, pp. 800–833, 2009.

[120] L. Ruthotto, "Hyperelastic Image Registration - Theory, Numerical Methods, and Applications", PhD thesis, Fachbereich Mathematik und Informatik, Westfälische Wilhelms-Universität Münster, 2012.

[121] D. Salomon, *Computer graphics and geometric modeling*. Springer, 2012.

[122] J. Salvi, C. Matabosch, D. Fofi, and J. Forest, "A review of recent range image registration methods with accuracy evaluation", *Image and Vision computing*, vol. 25, no. 5, pp. 578–596, 2007.

[123] O. Schmitt, "Die multimodale Architektonik des menschlichen Gehirns", PhD thesis, 2001.

[124] O. Schmitt, J. Modersitzki, S. Heldmann, S. Wirtz, and B. Fischer, "Image registration of sectioned brains", *International Journal of Computer Vision*, vol. 73, no. 1, pp. 5–39, 2007.

[125] R. Schönmeyer, N. Brieu, N. Schaadt, F. Feuerhake, G. Schmidt, and G. Binnig, "Automated whole slide analysis of differently stained and co-registered tissue sections", in *Bildverarbeitung für die Medizin 2015*, Springer, 2015, pp. 407–412.

[126] H. Siebert and M. P. Heinrich, "Deep Groupwise Registration of MRI Using Deforming Autoencoders", in *Bildverarbeitung für die Medizin 2020*, Springer, 2020, pp. 236–241.

[127] M Söhn, M Birkner, D Yan, and M Alber, "Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation", *Physics in Medicine & Biology*, vol. 50, no. 24, p. 5893, 2005.

[128] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey", *IEEE transactions on medical imaging*, vol. 32, no. 7, pp. 1153–1190, 2013.

[129] Ž. Spiclin, B. Likar, and F. Pernus, "Groupwise registration of multimodal images by an efficient joint entropy minimization scheme", *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2546–2558, 2012.

[130] M. Stille, E. J. Smith, W. R. Crum, and M. Modo, "3D reconstruction of 2D fluorescence histology images and registration with in vivo MR images: application in a rodent stroke model", *Journal of neuroscience methods*, vol. 219, no. 1, pp. 27–40, 2013.

[131] J. Streicher, W. J. Weninger, and G. B. Müller, "External marker-based automatic congruencing: A new method of 3D reconstruction from serial sections", *The Anatomical Record: An Official Publication of the American Association of Anatomists*, vol. 248, no. 4, pp. 583–602, 1997.

[132] C. Studholme, D. L. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment", *Pattern recognition*, vol. 32, no. 1, pp. 71–86, 1999.

[133] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?", *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[134] P. Thévenaz, T. Blu, and M. Unser, "Image interpolation and resampling", *Handbook of medical imaging, processing and analysis*, vol. 1, no. 1, pp. 393–420, 2000.

[135] H. Thiele, S. Heldmann, D. Trede, J. Strehlow, S. Wirtz, W. Dreher, J. Berger, J. Oetjen, J. H. Kobarg, B. Fischer, *et al.*, "2D and 3D MALDI-imaging: conceptual strategies for visualization and data mining", *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, vol. 1844, no. 1, pp. 117–137, 2014.

[136] C. F. Van Loan and G. H. Golub, *Matrix computations*. Johns Hopkins University Press, 1983.

[137] C. Wachinger and N. Navab, "Simultaneous registration of multiple images: Similarity metrics and efficient optimization", *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 5, pp. 1221–1233, 2012.

[138] C.-W. Wang, E. B. Gosno, and Y.-S. Li, "Fully automatic and robust 3D registration of serial-section microscopic images", *Scientific reports*, vol. 5, 2015.

[139] Q. Wang, L. Chen, P.-T. Yap, G. Wu, and D. Shen, "Groupwise registration based on hierarchical image clustering and atlas synthesis", *Human brain mapping*, vol. 31, no. 8, pp. 1128–1140, 2010.

[140] Q. Wang, G. Wu, P.-T. Yap, and D. Shen, "Attribute vector guided groupwise registration", *NeuroImage*, vol. 50, no. 4, pp. 1485–1496, 2010.

[141] U. Weber and H. Thiele, *NMR-spectroscopy: modern spectral analysis*. Wiley, 2008.

[142] N. Weiss, J. Lotz, and J. Modersitzki, "Multimodal image registration in digital pathology using cell nuclei densities", in *Bildverarbeitung für die Medizin 2015*, Springer, 2015, pp. 245–250.

[143] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information", *Medical Image Analysis*, vol. 1, no. 1, pp. 35–51, 1996.

[144] ——, "Multi-modal volume registration by maximization of mutual information", *Medical image analysis*, vol. 1, no. 1, pp. 35–51, 1996.

[145] J. Wernecke, *The Inventor mentor: programming object-oriented 3D graphics with Open Inventor, release 2*. Citeseer, 1994, vol. 1.

[146] C. A. White and A. Akbari, "The estimation, measurement, and relevance of the glomerular filtration rate in stage 5 chronic kidney disease", in *Seminars in dialysis*, Wiley, vol. 24, 2011, pp. 540–549.

[147] G. Wu, H. Jia, Q. Wang, and D. Shen, "SharpMean: groupwise registration guided by sharp mean image and tree-based registration", *NeuroImage*, vol. 56, no. 4, pp. 1968–1981, 2011.

[148] G. Wu, Q. Wang, H. Jia, and D. Shen, "Feature-based groupwise registration by hierarchical anatomical correspondence detection", *Human brain mapping*, vol. 33, no. 2, pp. 253–271, 2012.

[149] J. L. Zhang, H. Rusinek, H. Chandarana, and V. S. Lee, "Functional MRI of the kidneys", *Journal of Magnetic Resonance Imaging*, vol. 37, no. 2, pp. 282–293, 2013.

[150] B. Zitova and J. Flusser, "Image registration methods: a survey", *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[151] L. Zollei, "A unified information theoretic framework for pair-and group-wise registration of medical images", 2006.