



UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR MEDIZINISCHE INFORMATIK

Aus dem Institut für Medizinische Informatik
der Universität zu Lübeck
Direktor: Prof. Dr. rer. nat. habil. Heinz Handels

Deep Learning für die Verlaufsvorhersage von Schlaganfallläsionen mittels multivariater CT-Bilder

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck

Aus der Sektion Informatik/Technik

vorgelegt von
Christian Lucas
aus Rostock

Lübeck, 2020

1. Berichtstatter: Prof. Dr. Mattias Heinrich

2. Berichtstatter: Prof. Dr. rer. nat. Thomas Martinetz

Tag der mündlichen Prüfung: 9. Oktober 2020

Zum Druck genehmigt. Lübeck, den 12. Oktober 2020

Zusammenfassung

Der Schlaganfall ist eine der häufigsten vaskulären Erkrankungen in Deutschland und die dritthäufigste Todesursache. Den größten Anteil daran haben ischämische Schlaganfälle, bei denen Teile des Gehirns aufgrund eines Arterienverschlusses mit zu wenig sauerstoffreichem Blut versorgt werden. Als Folge sterben die Hirnzellen ab und es ergeben sich funktionelle Beeinträchtigungen, die je nach Ausmaß der Infarktläsion unterschiedlich ausfallen können. Dem Grundsatz „Zeit ist Hirn“ folgend kann man das Ausmaß der Läsion bei rechtzeitigem Eingriff begrenzen und Teile des Gehirns wieder reperfundieren. Dafür stehen zwei wesentliche Behandlungsmethoden zur Verfügung, die entsprechend bestimmter Auswahl- und Ausschlusskriterien eingesetzt werden. Die spezifische Risiko-Nutzen-Abschätzung basiert dabei unter anderem auf der Einschätzung des Verhältnisses vom nekrotischen Kern zum umgebenden Gewebe, welches potenziell gerettet werden kann. Eine verlässliche Quantifizierung dieser Bereiche über multivariate Perfusionsbildgebung ist notwendig, um die Selektionskriterien präzise anwenden zu können. Algorithmen können diese Aufgaben automatisiert ausführen und entsprechend komplexere Methoden zur Vorhersage von Behandlungsergebnissen implementieren. Dabei stellt die bildbasierte Prädiktion des räumlich aufgelösten Gewebestatus' nach erfolgreicher Behandlung einen Zwischenschritt zur darauf basierenden Einschätzung des klinischen Ergebnisses durch den Neuroradiologen dar. Dies wird in den letzten Jahren insbesondere durch maschinelles Lernen ermöglicht, welches auf einer Datenbasis von Expertenannotationen vielfältiges Wissen über den Verlauf von Schlaganfallläsionen lernen und dann effizient anwenden kann.

Die vorliegende Arbeit konzentriert sich in diesem Kontext auf die Entwicklung von neuronalen Netzwerken des Deep Learnings, wobei zunächst die robuste U-Net-Architektur erweitert und evaluiert wird. Zum einen kann diese zur Automatisierung und Unterstützung des klinischen Selektionsverfahrens eingesetzt werden. Zum anderen stellt es den bisherigen Stand der Technik von direkten Prädiktionsverfahren für das Gewebeergebnis mit finaler Läsion auf MR-Perfusionsdaten dar. Daraufhin wird ein neues Verfahren zur Modellierung des Läsionswachstums vorgestellt. Dieser Ansatz orientiert sich stärker am klinischen Vorgehen und begreift das räumliche Läsionsergebnis als Zustand während des Läsionswachstums, welches durch einen erfolgreichen Eingriff gestoppt wird. Die Methode übertrifft den Stand der Technik auf Basis der U-Nets, indem a-priori Expertenwissen über die Form von Kern und Risikogewebe in der Akutbildgebung trainiert und in einen generativen Ansatz integriert wird. Über Zeitparameter kann die Entwicklung der Läsion schließlich visuell simuliert und

die Entstehung des vorhergesagten Läsionsergebnisses durch die Ärztin nachvollzogen werden.

Die Vorstellung der Verfahren gliedert sich in zwei Teile zu je zwei Kapiteln. Im ersten methodischen Kapitel werden U-Nets vorgestellt, welche schon heute zur Unterstützung eingesetzt werden können. Damit ist es möglich, sichtbare Verschlüsse bereits zu Beginn in der nativen Kopf-CT automatisch zu detektieren und dies für die Triage zu berücksichtigen. Potenzielle Verschluss-Kandidaten werden segmentiert und die Ärztin kann so schnell den tatsächlichen Pfropf unter diesen Kandidaten feststellen. Daneben eignen sich U-Nets auch zur initialen Segmentierung der Kern- und Risikogewebebereiche, um so das Verhältnis dieser als klinisches Auswahlkriterium für die Behandlungsmethoden zu quantifizieren. In einem weiteren Kapitel wird eine Erweiterung des U-Nets mit zusätzlichen Abkürzungen innerhalb des Netzwerkes vorgeschlagen, welche im Segmentierungswettbewerb „ISLES 2017“ den zweiten Platz erreicht hat und den Stand der Technik zur Prädiktion der finalen Läsion mittels diskriminativer U-Nets darstellt.

Da in der Evaluation der U-Nets mit ihrer vor dem Nutzer verborgenen Inferenz der räumlichen Gewebeanfarzierung die Grenzen dieser Herangehensweise deutlich werden, konzentrieren sich die folgenden zwei Kapitel auf die Entwicklung eines generativen Ansatzes. Dieser macht die Modellierung des Wachstums, aus dem die finale Läsionsvorhersage resultiert, für den Nutzer transparent. Dabei wird zunächst ein Formraum aus Schlaganfallläsionen trainiert, in welchem die Interpolation der finalen Läsionsform abhängig von Zeitparametern visuell simuliert werden kann. Zusammen mit automatischen Schätzungen von Kernläsion und Perfusionsläsion des Risikogewebes durch ein U-Net ergibt sich eine signifikant verbesserte, robuste, automatische Vorhersage der erwarteten Segmentierung der Folgeläsion bei erfolgreicher Intervention. Durch diese kann das potenziell schädliche Wachstum abgeschätzt werden. Im Gegensatz zu den U-Net-Schätzungen korrelieren die abgeleiteten Volumen mit den tatsächlichen Volumen der Grundwahrheit. Das letzte Kapitel stellt schließlich eine über die Zeit durchgehend definierte Modellierung des Wachstums über Registrierungsfelder in einem PCA-Modell vor, welches noch plausiblere Deformationen der Läsionsform über die Zeit generieren kann und diese nachvollziehbarer macht. Es erreicht jedoch nicht die Genauigkeit der robust einsetzbaren, zuvor beschriebenen Interpolationsmethode.

Die vorliegende Arbeit leistet einen wichtigen Beitrag zur zukünftigen Verbesserung der Prädiktionsmethoden basierend auf Deep-Learning-Modellen. Dafür wird explizit Vorwissen von Experten berücksichtigt und das Zustandekommen der Vorhersagen kann vom Anwender transparent nachvollzogen werden, sowohl durch Wissen über den Aufbau des Modellraums, als auch durch eine mögliche Visualisierung der Wachstumsimulation. Dank der generierten Gewebeanfarzierungskarten können weitere Kennzahlen, wie das Volumen, abgeleitet werden oder funktionelle Beeinträchtigungen, die räumlich lokalisierbar sind, durch die Ärztin abgeschätzt werden.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation und Kontext	1
1.2	Ziele	5
1.3	Annahmen	7
1.4	Aufbau und Beiträge dieser Arbeit	8
2	Medizinischer Hintergrund	11
2.1	Neuroanatomie und Neuropathologie des Gehirns	11
2.1.1	Blutversorgung	11
2.1.2	Ischämischer Schlaganfall	13
2.2	Stand der klinischen Forschung	14
2.2.1	Bildgebende Diagnostik	15
2.2.1.1	Perfusionsbildgebung	17
2.2.1.2	Analyse der Perfusion zur Mismatch-Bestimmung	21
2.2.2	Akutbehandlung	23
2.2.2.1	Mechanische Thrombektomie nach Mismatch-Kriterien	24
2.2.2.2	Zeitrelevanz	26
2.2.3	Gewebeergebnis als Prädiktor	27
2.3	Klinische Bewertungssysteme (Scores)	28
2.4	Schlaganfalldatensätze zum Training und Test von Algorithmen	30
3	Methodischer Hintergrund	33
3.1	Grundlagen aus der Bildverarbeitung	33
3.1.1	Faltungsoperation (Convolution)	34
3.1.2	Diffeomorphismen	35
3.1.3	Metriken	36
3.2	Maschinelles Lernen	37
3.2.1	Hauptkomponentenanalyse (PCA)	39
3.2.2	Support Vector Machines (SVM)	40
3.2.3	Random Forests (RDF)	41
3.3	Deep Learning mit künstlichen neuronalen Netzen	42
3.3.1	Faltungsnetzwerke (CNN)	46
3.3.2	Multi-Skalen-Architekturen	50
3.3.3	Autoencoder (AE)	52

3.4	Segmentierung und Prädiktion für den Schlaganfall	53
3.4.1	Segmentierung bei multivariater Bildgebung	54
3.4.2	Prädiktion des Geweberesultats mittels neuronaler Netze	56
4	U-Nets zur Segmentierung akuter Infarkte in multivariater CT	57
4.1	Thrombusdetektion auf nativen CT	58
4.1.1	Literaturüberblick: Detektion kleiner Zielstrukturen	58
4.1.2	Vorgeschlagene Methodik	59
4.1.3	Experimente	60
4.1.3.1	Ergebnisse	61
4.1.3.2	Diskussion	62
4.2	Segmentierung der Akutläsion auf Perfusionsbildern	63
4.2.1	Literaturüberblick: Differenzierung von Risikogewebe und Kern	63
4.2.2	Vorgeschlagene Methodik	64
4.2.3	Experimente	65
4.2.3.1	Ergebnisse	65
4.2.3.2	Diskussion	66
5	U-Net zur Prädiktion der finalen Infarktläsion	67
5.1	Literaturüberblick	67
5.1.1	Kurze Skip-Verbindungen	68
5.1.2	Zusätzliche Überwachung	70
5.2	Vorgeschlagene Methodik	71
5.3	Experimente	74
5.3.1	Ergebnisse	75
5.3.2	Diskussion	76
6	Formrauminterpolation zur Vorhersage von Infarktwachstum	79
6.1	Literaturüberblick	81
6.1.1	Mathematische und biomedizinische Modellierung	81
6.1.2	Maschinelles Lernen des klinischen Schlaganfallverlaufs	82
6.1.3	Deep Learning von Läsionsformen und deren Evolution	83
6.2	Vorgeschlagene Methodik	86
6.2.1	Zweiteilige Architektur aus U-Net und CAE	86
6.2.1.1	Aufgabe von U-Net und CAE	87
6.2.1.2	Aufbau des CAE	88
6.2.2	Zweistufiges Training von Formraum und Prädiktion	89
6.2.2.1	Lernen eines Läsionenformraums	89
6.2.2.2	Interpolation geschätzter Läsionssegmentierungen	91
6.3	Experimente	92
6.3.1	Ergebnisse	92

6.3.2	Diskussion	97
7	Formdeformationsräume zur Modellierung von Infarktwachstum	99
7.1	Literaturüberblick	100
7.1.1	Integration von PCA und CNN	101
7.1.2	Integration von räumlichen Transformationen im CNN	101
7.1.3	Bildregistrierung mittels CNN	103
7.2	Vorgeschlagene Methodik	104
7.2.1	Aufbau des Deformationsraums	105
7.2.2	Schätzung der Modellparameter	105
7.3	Experimente	108
7.3.1	Datensatzerweiterung	108
7.3.2	Ergebnisse	109
7.3.3	Diskussion	111
8	Zusammenfassung und Ausblick	113
8.1	Vorteile und Grenzen	117
8.2	Ausblick	118
A	Diagnostik	123
B	Ergebnisse	125
C	Eigene Publikationen	135
	Literaturverzeichnis	137

1 Einleitung

Der Schlaganfall ist eine typische Alterserkrankung und tritt mit einer Inzidenz von fast 300 Fällen pro 100.000 Personen in der deutschen Gesamtbevölkerung auf [Stahmeyer u. a., 2019]. Als dritthäufigste Todesursache [Heuschmann u. a., 2010] stellt er eine höchst relevante Erkrankung im Rahmen der Gesundheitsversorgung in Deutschland dar. Am häufigsten wird der Schlaganfall durch einen embolischen Thrombus ausgelöst, der als Blutgerinnsel eine der Hauptarterien im Gehirn verstopft. Damit kommt es zum akuten Hirninfarkt, dessen direkte Folge eine Minderdurchblutung des nachgeordneten Versorgungsgebietes ist. Sofern dem Zellsterben nicht schnellstmöglich entgegengewirkt wird, folgt in der Regel der baldige Zelltod des infarzierten Gewebes.

1.1 Motivation und Kontext

Nach Symptombeginn bilden sich schnell unterschiedliche Bereiche des Gewebes heraus, die vom Schlaganfall betroffen sind [Thomalla u. a., 2009]. Zum einen gibt es einen Infarktkern, in dem die Hirnzellen durch anoxische Depolarisation innerhalb kürzester Zeit absterben (Nekrose). Dieser Bereich ist umgeben von der minderperfundierten Penumbra, wo die Zellen zwar unterversorgt, aber noch strukturell erhalten sind. Die Folge beschädigter Hirnbereiche sind starke funktionelle Beeinträchtigungen, oft z. B. der Sprache, der visuellen Sensorik, Motorik, oder Kognition [Kasten, 2007]. Entscheidend für die Rettung des Penumbrawebes und dementsprechend verbesserte funktionelle Resultate nach dem Schlaganfall ist die zeitige Behandlung („Zeit ist Hirn“) bis zu 24 Stunden nach Symptombeginn innerhalb der Akutphase mit möglichst vollständiger Reperfusion. In der Literatur wird diese nochmals grob in ein frühes Zeitfenster bis 4,5 bzw. 6 Stunden und ein erweitertes Zeitfenster bis wenigstens 16 Stunden unterteilt, wobei die Zeit nach etwa 12 Stunden oft als „spät“ angesehen wird (vgl. Kapitel 2.2.2.2). Die folgenden Zeitpunkte sind dabei relevant (Abb. 1.1):

t_0 Zeitpunkt des Schlaganfalls, d. h. Symptombeginn (oft geschätzt, ggf. unbekannt)

t_A Zeitpunkt der Aufnahme in der (Neuro-)Radiologie mit Akutbildgebung

t_I Zeitpunkt der Intervention bzw. Rekanalisation oder Reperfusion

t_V Zeitpunkt der Verlaufskontrolle (Bildgebung der Folgeläsion, *engl.* Follow-Up)

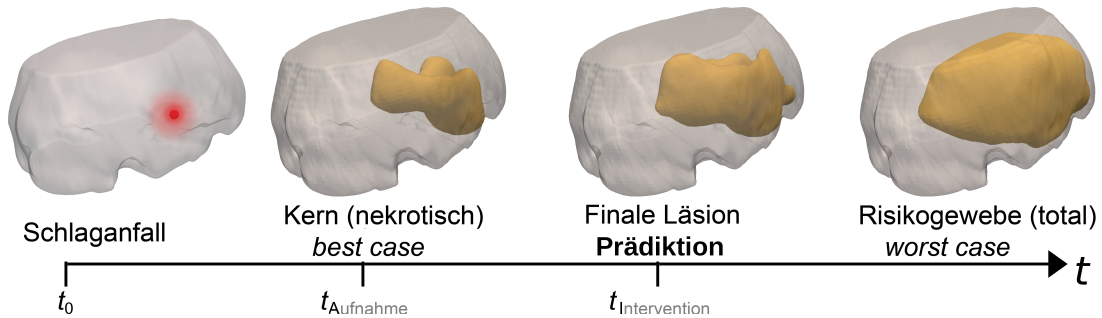


Abb. 1.1: Problembeschreibung: Wichtige Zeitpunkte sowie angenommene untere (Kernläsion) und obere (Perfusionsläsion) Schranken in der Vorhersage (Prädiktion) des Gewebestatus bei einem Verschluss der Mediaarterie in der rechten Hemisphäre. Es wird zudem angenommen, dass bei erfolgreicher Intervention die finale Läsion bei t_I ungefähr der Folgeläsion in der späteren Verlaufskontrolle bei t_V entspricht.

Entsprechend relevant ist damit vor allem die Zeitdauer $t_{0 \rightarrow I} = t_I - t_0$, welche die Zeitdauer von Beginn der Minderperfusion bis zur Reperfusion repräsentiert. Diese soll durch eine schnelle Rettungskette bis zum Krankenhaus sowie eine kurze Zeit von Aufnahme bis Intervention innerhalb sogenannter Stroke Units spezialisierter Krankenhäuser möglichst gering ausfallen.

Computerassistierte Schätzung der Läsionsbereiche eines Schlaganfalls Vor dem Eingriff ist für eine genauere Lokalisation und Diagnostik jedoch oft eine multivariate Computertomographie (CT) notwendig, welche meistens eine native Kopf-CT beinhaltet und darüber hinaus auch CT-Angiographie (CTA) und/oder CT-Perfusion (CTP). Letztere basiert auf der Messung eines Kontrastmittelbolus' im Gehirn und erlaubt die Berechnung verschiedener Parameter der Perfusion (Kapitel 2.2.1.1). Auf diesen kann man etwa das irreversibel geschädigte Gewebe (Kern, Notation c für *engl.* core) oder das gesamte Risikogewebe (Notation cp für *engl.* core+penumbra) auf zeitbasierten Parametern abschätzen. Es verbleibt mit der Penumbra jener Teil des Risikogewebes, der potenziell gerettet werden kann.

Abhängig von der bereits vergangenen Zeit seit Symptombeginn bis zur Bildgebung entscheidet das Verhältnis von Kern zu Penumbra darüber, ob die zwei Hauptbehandlungsmethoden Thrombolyse und/oder Thrombektomie zum Vorteil des Patienten angewandt werden können. Dabei ist das spezifische Risiko der jeweiligen Behandlungsmethode (z. B. Blutungsrisiko) abzuwägen. Im Notfall müssen dabei unter hohem Zeitdruck die Bild- und Patientendaten ausgewertet und mit quantifizierbaren Ergebnissen zur optimalen Risikoabwägung aufbereitet werden. Die Probleme bei rein manueller Interpretation sind dabei vielfältig: Unterschiede zwischen verschiedenen sowie ein- und denselben Experten, der zeit- und kostenintensive Aufwand durch das nicht überall verfügbare Fachpersonal (z. B. auf dem Land), welches dennoch anfällig

für subjektive Fehler ist (z. B. Übersehen kleiner Verschlüsse), oder ursächlich unterschiedliche Hirnläsionen, welche ähnliche Marker in der Bildgebung aufweisen [Crimi, 2016], die darüber hinaus noch durch die Parameterwahl bei der Bildgebung (MRI) unterschiedlich aussehen können.

Daher besteht das Ziel, die darauf basierenden Behandlungskriterien oder Vorhersagen der Behandlungsergebnisse (*engl.* Outcome) computerunterstützt und objektiv reproduzierbar zu schätzen. Dafür stehen im Rahmen der Schlaganfallbehandlung einige Endpunkte zur Auswahl, etwa die Verlaufseinschätzung in Form einer klinischen Skala (*engl.* Score), wie der mRS, das finale Infarktvolume in *ml*, oder der Gewebestatus als Infarktkarte des Hirns.

Eine hinreichend gute Schätzung des Kern- und Penumbrabereichs vorausgesetzt, besteht eine grundlegende Annahme darin, dass das Kernareal den *best case* der Behandlung darstellt, da dieses Gewebe in jedem Fall als nekrotisch angesehen wird, und das totale Risikogewebe den *worst case*, falls der Schlaganfall unbehandelt bliebe. In einer idealisierten Vorstellung breitet sich innerhalb der Akutphase von etwa 24 Stunden die Infarktläsion nichtlinear von dem einen Extremum zum anderen aus (Abb. 1.1), wenn dieses „Wachstum“ nicht durch einen Eingriff dazwischen gestoppt wird.

Für das oben genannte starre Verhältnis von Kern zu Penumbra bzw. der Abweichung des gesamten Risikogewebes zum Kerninfarkt (*engl.* Mismatch) benötigt man die Segmentierung des Kerns und Risikogewebes. Automatisiert bzw. computerassistent existieren vor allem in der Breite für die native CT seit längerer Zeit (semi-)automatische Methoden [Maldjian u. a., 2001; Olive-Gadea u. a., 2019; Rekek u. a., 2012; Ušinskas u. a., 2004], die aber bei weitem nicht mehr ausreichen für die oben genannte Patientenselektion nach neuester Leitlinie [Hennerici u. a., 2017] zur Diagnostik akuter zerebrovaskulärer Erkrankungen (vgl. Kapitel 2.2.1.2). Stattdessen benötigt man Methoden, welche auf der multivariaten Bildgebung, und dabei vor allem den Perfusionparametern, eine entsprechende Schätzung der Bereiche ermöglichen (Kapitel 4.2).

Dafür existieren zwar Schwellenwert-Methoden [Straka u. a., 2010], die teilweise zur Bestimmung des gesamten Risikogewebes in großen Studien angewandt wurden [Lansberg u. a., 2012], aber auch diese benötigen die Definition einer Zielregion (*engl.* Region-Of-Interest, ROI) sowie einige manuelle Eingriffe, wodurch die Reproduzierbarkeit stark eingeschränkt wird. Für Segmentierungen wurden daher auch schon maschinelle Lernalgorithmen auf annotierten Stichprobendaten trainiert. Bereits im Jahre 2015 wurden im Wettbewerb der „ISLES“ nur noch 3 von 16 Ansätzen komplett ohne maschinelles Lernen eingereicht [Maier u. a., 2017].

Neben der Bestimmung des Mismatch ist es vorstellbar, geeignetere Selektionsmethoden zu entwickeln, welche stattdessen zunächst die Evolution des Infarktstatus' im Hirngewebe vorhersagen und in dieser Arbeit behandelt werden (Kapitel 5–7). Da sich unter anderem klinisch gezeigt hat, dass der Einsatz der mechanischen Rekanalisationsmethode zwar bis 24 Stunden sinnvoll sein kann (Kapitel 2.2.2.2), aber der Nutzen

bzw. die Reperfusion mit zunehmender Dauer $t_{0 \rightarrow I}$ abnimmt, wäre ein patientenindividuelles und zeitabhängiges Wachstumsmodell gegenüber den festen Mismatchkriterien und Schwellenwerten zu bevorzugen.

Deep Learning für die Verlaufsvorhersage von Schlaganfällen Parallel zur schrittweisen Etablierung der mechanischen Thrombektomie als neuer Behandlungsmethode mit erforderlicher Patientenselektion anhand der Perfusionsbildgebung, begann auf dem Feld der Künstlichen Intelligenz (*engl.* Artificial Intelligence, AI) die „Renaissance der CNNs“ durch das sogenannte *Deep Learning* von besonders tiefen künstlichen neuronalen Netzwerken [Krizhevsky u. a., 2012], dessen Potenzial zu Beginn dieser Arbeit im Jahr 2016 bereits deutlich wurde. Die CNNs besitzen gegenüber anderen maschinellen Lernverfahren einige Vorteile, wie z. B. die Optimierung der Filterkerne zur Merkmalsextraktion durch das neuronale Netzwerk [Schmidhuber, 2015]. Sie können oft sogar direkt auf den Rohdaten angewandt werden, ohne das bekannt wäre, welche visuellen Merkmale einen geeigneten Prädiktor darstellen, und sind unter anderem deshalb oft traditionellen Verfahren überlegen (Kapitel 3.3).

Insbesondere Netzwerke, welche inhärent mit verschiedenen Auflösungsstufen und Skalen arbeiten, können sowohl die groben Muster als auch im Kleinen die individuell genaue Ausprägung einer Pathologie bevorzugt erfassen. So etwa die U-förmigen Architekturen von Ronneberger u. a. [2015], welche sich als äußerst robuste Technik für viele biomedizinische Probleme erwiesen haben und teilweise den Stand der Technik bzw. manuelle Expertensegmentierungen in ihrer Genauigkeit übertreffen konnten (Kapitel 3.3.2). Diese „U-Nets“ stellen – z. B. gegenüber o.g. Schwellenwertverfahren – einen fortgeschrittenen diskriminativen Ansatz zur voxelweisen Segmentierung dar, welcher eine Entscheidungsgrenze zwischen Zielstruktur und Hintergrund modelliert.

Ihr Einsatz für Prädiktionsaufgaben scheint allerdings begrenzt, wie sich in den letzten Jahren unter anderem auf den wissenschaftlichen Wettbewerben der ISLES im Rahmen der MICCAI-Konferenz angedeutet hat (Kapitel 5). Generative Methoden stellen einen anderen Ansatz dar, der auch die Generierung oder Simulation von nicht vorhandenen Läsionen auf Basis von Akutbilddaten ermöglichen könnte. Damit sind reproduzierbar generierbare Wachstumsprognosen etwa einer Infarktläsion denkbar, auf der dann weitere Endpunkte abgeleitet werden können, die für den Kliniker von Interesse sind. Dies ist damit eine Strategie, um das klinische Ergebnis vorherzusagen, statt der direkten Prädiktion von bspw. zusammenfassenden mRS-Outcomes (Kapitel 6.1.2).

Die vorliegende Arbeit entstand im Rahmen des „Trave-Stroke“-Projekts, bei dem genau dieses Vorgehen untersucht werden soll und zur Entwicklung entsprechender maschineller Lernmethoden auch ein Teil des dabei gesammelten Datensatzes retrospektiv mit umfangreicher Annotation versehen wurde (Kapitel 2.4). Es geht in dem Projekt u. a. darum, den Gewebestatus vorherzusagen, d. h. konkret um die Läsionsprädiktion

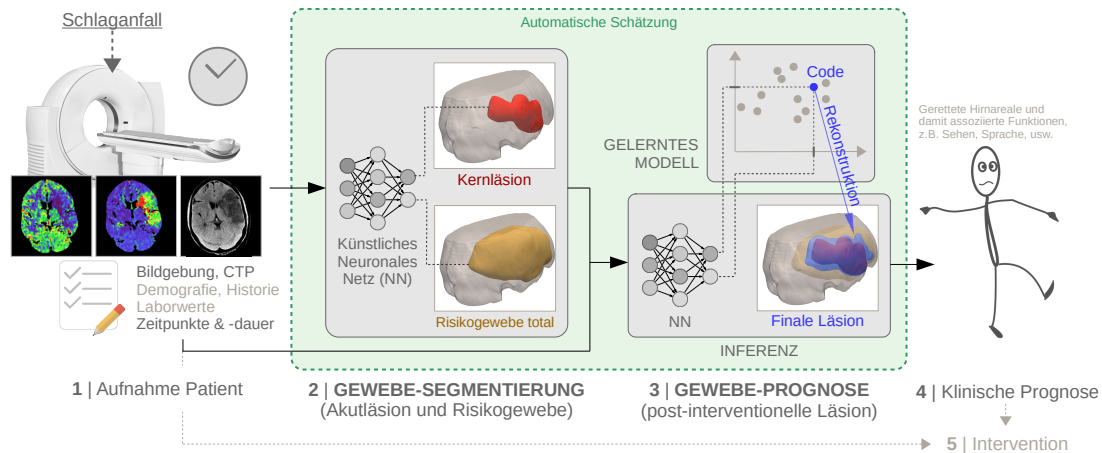


Abb. 1.2: Vorliegende Arbeit im Kontext der Schlaganfallbehandlung: Zur Risiko-Nutzen-Abschätzung von Therapieoptionen besteht die Notwendigkeit der Prädiktion des Geweberesultats als Grundlage einer klinischen Prognose. In dieser Arbeit wird die (grün unterlegte) automatische Lösung unter Verwendung künstlicher neuronaler Netze mit maschinell gelernten Prädiktionsmodellen untersucht. Der bisherige Stand der klinischen Praxis sieht u. a. vor: (1) Aufnahme der Patientin, (2) manuelle oder schwellenwertbasierte Kern- und Penumbra-Schätzung, (3) Selektion der Behandlung anhand eines Mismatch-Kriteriums nach Leitlinie, sodass sich der größte (4) therapeutische Nutzen der entsprechenden (5) Intervention ergibt.

in Form einer binären Hirnkarte mit Status „Infarkt“ gegenüber „kein Infarkt“, sodass im zweiten Schritt eine funktionale Symptom-Vorhersage basierend auf der Ausbreitung der finalen Läsion ermöglicht werden kann.

Unter der Voraussetzung, dass das klinische Ergebnis direkt von der Rettung des Gewebes abhängt, ließe sich so der therapeutische Nutzen zuverlässig quantifizieren. Damit wäre die Gewebivorhersage eine Voraussetzung für die klinisch relevanten Endpunkte (Abb. 1.2). Aufgrund unterschiedlicher funktionaler Bereiche des Hirns vermutet man, dass das klinische Ergebnis nicht nur vom Volumen, sondern zum Teil auch von der konkreten Läsionslokalisierung und -form abhängt. Dies wird mittels „Lesion-Symptom-Mapping“ (Kapitel 2.2.3) bei der Überlappung einer Läsion mit bestimmten standardisierten Hirnbereichen untersucht.

1.2 Ziele

Die übergeordnete Zielsetzung dieser Arbeit ist die Schätzung des therapeutischen Nutzens der Thrombektomie-Behandlungsmethode gegenüber den spezifischen Risiken. Die Vorstellung ist ein Framework zur objektiven Verlaufsprognose einer individuellen Schlaganfallläsion, welches möglichst schnell reproduzierbare Ergebnisse zur Quantifizierung des potenziellen Behandlungsnutzens in Form von gerettetem Gewe-

be liefern kann. Das schrittweise Vorgehen basiert dabei auf einzelnen Zielstellungen, die sich ausgehend vom klinischen und methodischen Stand der Forschung wie folgt formulieren:

Objektive, reproduzierbare Quantifizierungen für die klinische Selektion auf multivariater CT Im Rahmen der Leitlinie müssen zur Selektion vor der Behandlung auf den Akutdaten entsprechende Frühzeichen oder Perfusionsbereiche detektiert und segmentiert werden (Kapitel 2.2.2.1). Die Segmentierung des Verschlusses im CT erlaubt etwa die Vermessung der Länge oder eine Klassifikation als großer Verschluss (*engl.* Large Vessel Occlusion, LVO) mit potenziell schlechtem klinischen Ergebnis [Zhu u. a., 2016]. Risikobereich und Kerninfarkt sollen anhand verschiedener Perfusionsparameter segmentiert werden können, um den Mismatch zwischen beiden bestimmen zu können. In dieser Arbeit werden dafür in Kapitel 4 zwei Methoden mittels U-Net vorgestellt, die ebenfalls als Voraussetzung für den darauffolgenden Prädiktionsansatz in Kapitel 6 dienen.

Entwicklung eines Prädiktionsverfahrens zur Vorhersage des räumlichen Gewebestatus Als Weiterentwicklung zum klinisch etablierten Vorgehen soll eine Methode entwickelt werden, welche den Infarzierungsstatus der finalen Läsion als räumliche Wahrscheinlichkeitskarte bzw. die binarisierte Segmentierungsmaske der Folgeläsion auf einem Verlaufsbild vorhersagt. Dafür existieren bereits Vorarbeiten, etwa auf Basis eines linearen Regressionsansatzes, manuell erstellten Merkmalen und Atlas-Vorwissen [Kemmling u. a., 2015]. In dieser Arbeit sollen mittels maschinellem Lernen geeignete Merkmale aus den Perfusionsparametern automatisch extrahiert werden, während darauf aufbauend eine Verlaufsprädiktion auf Basis des retrospektiven Datensatzes gelernt wird. Die Gewebestatuskarte als Ergebnis erlaubt daraufhin die Ableitung anderer Endpunkte, wie Volumen [Boers u. a., 2018; Boers u. a., 2019; Yoo u. a., 2012; Zaidi u. a., 2012] oder das Lesion-Symptom-Mapping [Rangaraju u. a., 2015], und damit weiterer potenzieller Prädiktoren für das funktionelle oder klinische Ergebnis. In Kapitel 5 wird eine Erweiterung der U-Nets vorgeschlagen, um entsprechende Gewebestatuskarten zu prädizieren.

Einarbeitung von Expertenwissen zur Verbesserung der Vorhersage Der Ärztin steht natürlicherweise bei der Begutachtung der multivariaten CT auch ein reicher Erfahrungsschatz an persönlichem Expertenwissen zur Einschätzung einer Schlaganfallläsion und deren Prognose zur Verfügung. Es soll dargestellt werden, wie das valide Vorwissen über Läsionssegmentierungen, bestehend aus den a-priori gegebenen Expertensegmentierungen eines CTP-Datensatzes, zum Vorteil für die Prädiktion des Gewebestatus' eines ungesehenen Patienten eingebunden werden kann. Für die oben genannten Mismatch-Kriterien wurde schließlich ebenfalls das Vorwissen vieler

Studienstichproben zusammengeführt, um die einfache Selektionsregel zur erfolgreichen Behandlung mit mechanischer Thrombektomie zu ermitteln. Die Nutzung dieses Risikogewebe- und Kern-Vorwissens wird explizit im Hauptteil dieser Arbeit in Kapitel 6 und zu Teilaspekten in Kapitel 7 untersucht.

Simulation von patientenindividuellem Infarktwachstum Statt eines starren Mismatch-Kriteriums oder der Momentaufnahme von Akutsegmentierung der multivariaten CT als schlichte *Best-Case*- und *Worst-Case*-Schätzungen sowie dem Gewebestatus der Folgeläsion, soll ein Simulationsansatz die nichtlineare Evolution der Läsion abbilden. Die o.g. Momentaufnahmen sollen dabei die Evolution nach klinischer Sichtweise beschränken. Die Simulation stellt nicht nur den patientenindividuellen Verlauf anhand der multivariaten CT-Bildgebung dar, sondern auch unter Berücksichtigung (globaler) klinischer Variablen, wie der Zeit. In Kapitel 6 wird dies eingeführt und kann als Interpolation auf einer linearen Trajektorie robust visuell simuliert werden.

1.3 Annahmen

Nach Rezik u. a. [2012] treten einige besondere Herausforderungen auf Gewebeebene beim Infarktwachstum auf, das oft nur idealisiert modelliert werden kann. Das Phänomen der Schwellung in den ersten Tagen nach Symptombeginn kann sich nach Verlust der abgestorbenen Zellen später wieder zurückbilden und dazu führen, dass die Läsion in der Bildgebung über- oder unterschätzt wird. Daher hat der Aufnahmezeitpunkt t_V zur Erzeugung einer Grundwahrheit für das Prädiktionsziel einen wichtigen Einfluss auf die Plausibilität eines darauf aufbauenden Modells.

Auch ohne entsprechenden Eingriff können sich eine spontane Reperfusion oder der Kollateralfluss umgebender Blutgefäße stark auf die Ausprägung des Kerns oder der Penumbra auswirken; umgekehrt kann es passieren, dass die Wiederherstellung des Blutflusses in den Kapillaren trotz Rekanalisationseingriff ausbleibt (z. B. bei großen Verschlüssen der MCA). Diese Phänomene sind jedoch schwierig zu quantifizieren und werden damit im weiteren Verlauf der Arbeit nicht berücksichtigt. Zur Bearbeitung der oben genannten Ziele und technischen Modellierung sind aus klinischer Sicht im Vorfeld einige weitere vereinfachende Annahmen getroffen worden, die es erst ermöglichen, mit den nur begrenzt verfügbaren Daten entsprechende Vorhersagemodelle zu entwickeln:

- Durch die kollaterale Versorgung kann es auch zu dazu kommen, dass vereinzelte Läsionsherde zu erkennen sind, statt einer zusammenhängender Struktur. Die vorhandenen Daten und die vorgestellte Modellierung dieser Arbeit gehen von einer Einzelläsion als zusammenhängender Komponente aus (*engl.* connected component).

- Da die vorhandenen Daten nur Stichproben aus den ersten 10 Stunden abdecken, wird dies in der Evaluation als normierter Akutzeitraum verwendet, damit der modellierte Zeitbereich im Training auch durch Stichproben abgedeckt ist.
- Bei den geschätzten und auch von Experten angefertigten Segmentierungen von „Kern“ und „Risikogewebe“ basierend auf den Perfusionsparametern handelt es sich trotz der Benennung als solche letztlich nur um Approximationen des Kerns oder Risikogewebes, da der jeweilige tatsächliche physiologische Zustand zum Akutzeitpunkt in der Perfusionsbildgebung räumlich nicht exakt dargestellt werden kann [Turowski u. a., 2015].

1.4 Aufbau und Beiträge dieser Arbeit

Zum Direkteinstieg bei nicht-sequentieller Lektüre der Arbeit befindet sich vor jedem Kapitelanfang eine Kurzzusammenfassung des Kapitelinhalts (graue Box) als Einführung und zur Orientierung. Die Kapitel 4–7 sind in ihrer Gliederung konsistent zueinander aufgebaut, beginnend mit einem Überblick über die jeweils relevante methodische und anwendungsspezifische Literatur im ersten Abschnitt („Literaturüberblick“). Nachfolgend wird die neue Methodik vorgeschlagen und erklärt, sowie schließlich im „Experimente“-Abschnitt die Evaluation besprochen und die Ergebnisse diskutiert.

Die Arbeit gliedert sich in folgende Kapitel mit den Hauptbeiträgen dieser Arbeit:

- Kapitel 2 erläutert den medizinischen Hintergrund zur Krankheit Schlaganfall sowie den aktuellen Stand des Vorgehens in der klinischen Praxis. In Abschnitt 2.4 werden zudem die Daten aus der Klinik vorgestellt und wie diese für Training und Evaluation der Algorithmen genutzt werden.
→ Bisheriges klinisches Vorgehen für Schritte 1–5 in Abb. 1.2
- Kapitel 3 bespricht die notwendigen Grundlagen für die in dieser Arbeit vorgestellten Methoden, insbesondere die Entwicklungen rund um das Deep Learning von Multi-Skalen-Netzwerken und Encoder-Decoder-Architekturen.
→ Hintergrundwissen zur Umsetzung von Schritt 2 und 3 in Abb. 1.2
- In Kapitel 4 wird die Nutzung von U-Nets zur Segmentierung in multivariater CT beim akuten Schlaganfall vorgestellt. Dies wird zunächst mit einer zweistufigen CNN-Methodik zur Detektion von kleinen Frühzeichen der MCA bzw. ICA in nativen CT-Aufnahmen gezeigt, dessen Ergebnisse in [Lucas u. a., 2019] veröffentlicht wurden. Daneben wird auch die automatische Segmentierung der akuten Kern- und Penumbra-bereiche im Rahmen der neuroradiologischen Diagnostik sowie als Vorverarbeitungsschritt der vollautomatischen Methode aus [Lucas u. a., 2018b] mittels U-Net auf CT-Perfusionsdaten vorgestellt.
→ Umsetzung von Schritt 2 in Abb. 1.2

- In Kapitel 5 wird eine Erweiterung des U-Nets mit zusätzlichen Skip-Verbindungen für die Prädiktion der Folgeläsion als Gewebestatusresultat vorgeschlagen, wie sie in [Lucas u. a., 2018a] publiziert wurde und beim Wettbewerb „ISLES“ 2017 den zweiten Platz erreichte [Winzeck u. a., 2018]. Sie repräsentiert damit den bisherigen Stand der Technik mit U-Nets zur Schlaganfallprädiktion.
→ Direkte Umsetzung von Schritt 3 ohne Schritt 2 in Abb. 1.2
- In Kapitel 6 wird ein neuer Ansatz mittels generativer Modellierung entwickelt, welcher über die Interpolation in einem gelernten Formraum von Schlaganfallläsionen den räumlichen Gewebestatus über die Zeit simulieren kann. Es sollen damit die bisherigen Limitierungen der Modellierung von voxelweisen diskriminativen Entscheidungsgrenzen der U-Nets umgangen werden. Die Methode wurde in [Lucas u. a., 2018b] veröffentlicht und erzielt signifikant bessere Ergebnisse bei der Prädiktion als die vorherige Herangehensweise mittels U-Nets.
→ Hauptbeitrag zur kombinierten Umsetzung von Schritt 2 und 3 in Abb. 1.2
- In Kapitel 7 wird die aufwändigere Modellierung eines statistisch gelernten Deformationsraumes zur Reduzierung der spezifischen Interpolationsartefakte des zuvor vorgestellten Ansatzes entwickelt. Es kann dabei in theoretisch wohldefinierten Situationen zur qualitativen Verbesserung eingesetzt werden und wurde in [Lucas u. a., 2020a,b] vorgestellt.
→ Alternativer Modellraum für Schritt 3 in Abb. 1.2
- Kapitel 8 fasst schließlich diese Arbeit zusammen und diskutiert die wissenschaftlichen Beiträge zu den formulierten Zielen sowie noch offene Fragestellungen bezüglich der vorgestellten Methoden.

2 Medizinischer Hintergrund

Zunächst werden grundlegende Hintergrundinformationen erläutert, die zum Verständnis der Motivation, des konkreten Entwurfs und der Anwendung der in dieser Arbeit vorgestellten Methoden notwendig sind. In Kapitel 2.1 werden die anatomischen Grundlagen vor allem von jenem Teil der Blutversorgung im Gehirn erklärt, in dem ischämische Schlaganfälle am häufigsten auftreten. Kapitel 2.2 legt die momentan etablierten Techniken und Leitlinien in der klinischen Praxis bei einem solchen Schlaganfall dar, insbesondere den Einsatz multivariater Bildgebung als Grundlage von Therapieentscheidungen. Daneben werden auch die verwendeten klinischen Scores (Kapitel 2.3) und Datensätze (Kapitel 2.4) erläutert.

2.1 Neuroanatomie und Neuropathologie des Gehirns

Das Gehirn mit seinen zwei Hemisphären im menschlichen Schädel hat eine hohe funktionale Bedeutung, wie z. B. Aufgaben der kognitiven Verarbeitung, Gedächtnisbildung, oder Bewegungskoordination. Es gliedert sich nach Trepel [2017] als Teil des ZNS in graue und weiße Substanz, welche auch im Rückenmark vorkommen. Die graue Substanz enthält hauptsächlich die Nervenzellkörper (Perikarion), während im Vergleich dazu die weiße Substanz vor allem deren Fortsätze (Axone) und Gliagewebe enthält. Die graue Substanz gliedert sich im Körper so, dass sie nach außen hin von weißer Substanz umgeben ist und sogenannte Kerne bildet. Im Gehirn umschließt jedoch zusätzlich eine Schicht aus grauer Substanz (Kortex) als Groß- bzw. Kleinhirnrinde die weißen Substanzen der entsprechenden Hirnareale. Der Kortex mit seinen Windungen (Gyri) und Furchen (Sulci) ist zum überwiegenden Teil beim Blick von außen auf das Gehirn zu erkennen. Beide Substanzen haben funktional jedoch sehr unterschiedliche Aufgaben – in jedem Fall kann eine Schädigung dieser Substanzen starke funktionale Störungen hervorrufen.

2.1.1 Blutversorgung

Die extrakranielle Blutversorgung orientiert sich grob an der Symmetrie der zwei Hemisphären und speist sich nach Trepel [2017] aus der linken und rechten Arteria Carotis Interna (ICA) sowie der linken und rechten A. Vertebralis, wobei die ICA das vordere Stromgebiet (Frontal-, Parietal-, und Großteil des Temporallappens) versorgt und sich in A. Cerebri Anterior und A. Cerebri Media (MCA) aufteilt (Abb. 2.1(a)). Letztere

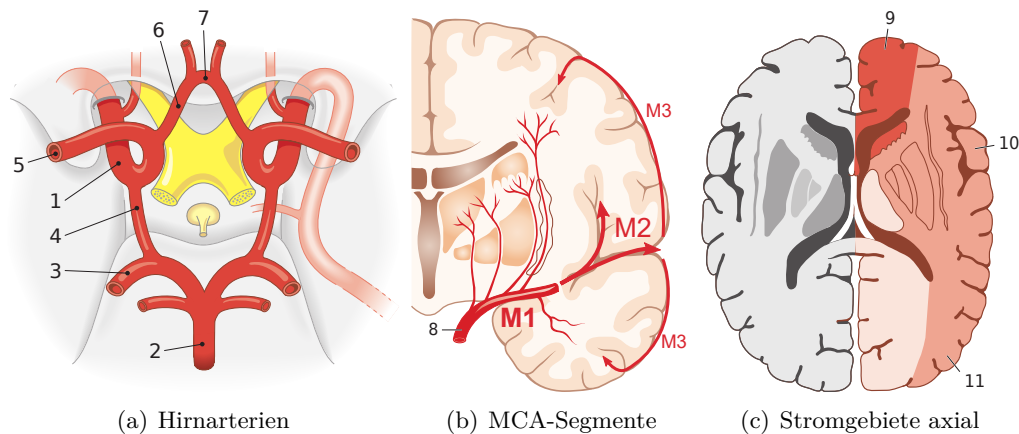


Abb. 2.1: Trepel [2017] zeigt in (a) A. Carotis Interna (1) und A. Basilaris (2), welche sich in linke und rechte A. Cerebri Posterior (3) aufgabelt. Die A. Communicans Posterior (4) verbindet diese mit der A. Carotis Interna (1), welche sich in A. Cerebri Media (5) und A. Cerebri Anterior (6) aufgabelt, wobei letztere durch die A. Communicans Anterior (7) verbunden sind. Im Frontalschnitt (b) ist die Versorgung (sub)kortikaler Bereiche durch die MCA (8) zu sehen. Die Stromgebiete (c) werden vorrangig je durch A. Cerebri Anterior (9), Media (10) und Posterior (11) versorgt. Aus: Trepel, Neuroanatomie, 7. Aufl. 2017 ©Elsevier GmbH, Urban & Fischer, München

decken den größten Teil der Blutversorgung des Gehirns ab, d. h. das mittige, laterale Stromgebiet (Abb. 2.1(b)).

Das hintere Stromgebiet wird durch die A. Basilaris aus dem Wirbelbereich versorgt, von welcher linke und rechte A. Cerebri Posterior abgehen. Die A. Cerebri Anterior, Media und Posterior stellen die drei Hauptgefäße zur Versorgung von Groß- und Zwischenhirn dar. Es gibt allerdings noch eine Verbindung in Form zweier A. Communicans Posteriores zwischen dem vorderen und hinteren Stromgebiet sowie eine Verbindung beider A. Cerebri Anterior durch eine A. Communicans Anterior, sodass sich ein sogenannter Kollateralkreislauf bildet (Circulus Arteriosus Cerebri). Dennoch werden die große Stromgebiete immer hauptsächlich durch eine der drei Hauptgefäße versorgt (Abb. 2.1(c)) und nur bei langsamen Verschlüssen eines Gefäßes ist eine Versorgung durch die Kollateralen angrenzender Gebiete möglich.

Die ICA wird in vier Abschnitte unterteilt (Pars Cervicalis, P. Petrosa, P. Cavernosa, P. Cerebralis) oder nach klinischen Gesichtspunkten auch in sieben (C1-C7: von P. Cervicalis bis P. Cerebralis). Die MCA wird ihrerseits wieder in vier Segmente gegliedert, welche von medial nach lateral mit M1 bis M4 durchnummeriert sind, wobei M4 die terminierenden Verästelungen auf der Großhirnrinde bezeichnet. Abbildung 2.1(b) illustriert grob den Verlauf von M1-M3.

2.1.2 Ischämischer Schlaganfall

Das menschliche Gehirn kann auf verschiedene Weise geschädigt werden: Ausgehend von den erblichen und damit bereits bei der Geburt vorliegenden Ursachen gestörter geistiger Entwicklung (z. B. bei Trisomie 21), über graduell manifestierende Schäden durch Entzündungen (z. B. Enzephalitis), hirneigene Krankheiten, wie Tumore (z. B. Glioblastome), Multiple Sklerose oder degenerative Erkrankungen (z. B. Alzheimer), sowie hormonell bedingten Funktionsstörungen (z. B. Hyper- oder Hypothyreose), bis hin zu traumatischen Hirnschäden (z. B. durch äußere Gewalteinwirkung), können ebenso vielfältige Funktionsstörungen die Folge sein [Kasten, 2007]. Hirnkrankheiten haben zwar grundsätzlich unterschiedliche Ursachen und Auswirkungen auf das gesunde Gehirn, jedoch besteht oft die Schwierigkeit, dass die funktionellen Beeinträchtigungen als auch Biomarker in den vielfältigen Bildgebungsmodalitäten sich stark ähneln können. Dies erschwert darauf basierende (automatische) Verlaufsvorhersagen und Krankheitsprognosen [Crimi, 2016], wie sie in dieser Arbeit besprochen werden.

Der Name des Schlaganfalls erklärt sich durch den oft schlagartigen Symptombeginn. In den allermeisten Fällen handelt es sich um eine Durchblutungsstörung aufgrund verstopfter Hirnarterien (am häufigsten betroffen ist die MCA). Oft werden diese durch einen embolischen Thrombus verstopft, der sich an anderer Stelle im Blutkreislauf an einer Gefäßablagerung bildet und als Blutgerinnsel schließlich bis ins Gehirn wandert [AllEx, 2012]. Damit kommt es zum Hirninfarkt (auch: „Apoplex“, „Insult“), womit das durch diese Arterie versorgte Gebiet im Gehirn eine Minderdurchblutung erfährt. Man spricht an dieser Stelle letztendlich von einem *ischämischen* Schlaganfall oder auch „weißen“ Schlaganfall aufgrund der Blutleere. Ischämien können jedoch ebenso allmählich etwa durch einen allgemein niedrigen Blutdruck (typischerweise im Schlaf) in Kombination mit altersbedingten Gefäßablagerungen auftreten.

Die seltenere Ausprägung (< 15 % der Fälle) des Schlaganfalls ist die *Hämorrhagie* – auch „roter“ Schlaganfall genannt. Ursache ist eine Gefäßruptur, welche etwa an einem Aneurysma auftreten kann, wenn dieses aufgrund der typischerweise nur mangelhaft ausgebildeten Muskelschicht der Arterie platzt. In der Folge fließt das Blut in das Hirnparenchym und kann nicht über die Venen wieder abfließen, was zur Erhöhung des Hirninnendrucks und schließlich zur Quetschung des Gewebes führt.

Der Schlaganfall als typische Alterserkrankung der Industrieländer hat eine Inzidenz in der deutschen Gesamtbevölkerung von jährlich 292 aus 100.000 Personen [Stahmeyer u. a., 2019], wobei die Hälfte der betroffenen Personen älter als 75 Jahre ist (Abb. 2.2). Die Mortalität ist vergleichsweise hoch, da der Schlaganfall die dritthäufigste Todesursache mit einer Gesamtzahl von circa 63.000 Toten im Jahr 2008 darstellt [Heuschmann u. a., 2010]. Weltweit erleiden nach Feigin u. a. [2015] sogar schätzungsweise 10 Millionen Menschen einen Schlaganfall pro Jahr, welcher zu 90 % auf behandelbare Risikofaktoren zurückzuführen ist [O'Donnell u. a., 2016]. Laut Diener u. a. [2018] sind

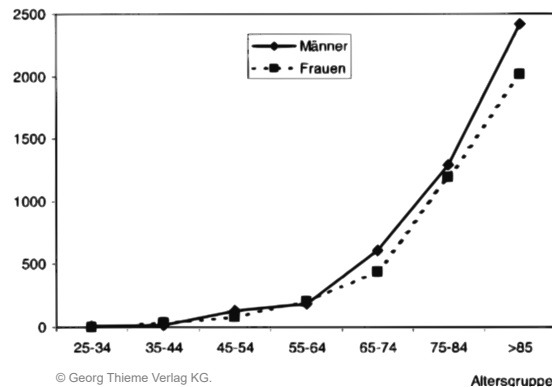


Abb. 2.2: Inzidenz pro 100.000 Einwohner nach Altersgruppen. Aus: Kolominsky-Rabas u. a., 2002

dies vor allem arterielle Hypertonie, Diabetes mellitus, Rauchen, Bewegungsmangel und Ernährung.

In der Zeit nach Symptombeginn bilden sich verschiedene Bereiche des Schlaganfalls heraus [Thomalla u. a., 2009] – siehe auch Abbildung 2.8(a) auf Seite 21. Der Infarkt-kern bezeichnet das Gebiet mit anoxischer Depolarisation der Hirnzellen, das innerhalb kürzester Zeit ohne Reperfusion nekrotisch wird. Den Kern umgibt eine sogenannte Penumbra (*lat.* Halbschatten) als Areal minderperfundierten Gewebes, welches ebenfalls einen gestörten Funktionsstoffwechsel bei erhaltenem Strukturstoffwechsel aufweist. Sobald ein kritischer Grenzwert bei der Unterversorgung mit sauerstoffführenden Blut unterschritten wird, folgt die anoxische Depolarisation – das Gebiet wird Teil des Infarkt(kern)s mit irreversiblen Schäden. Dies ist von der Schwere der Perfusionsstörung über die Zeit abhängig und für das Risikogewebe kann bei rechtzeitigem Eingriff durch eine Reperfusion verhindert werden, dass die Blutversorgung unterhalb der kritischen Schwelle fällt. Dieser Zeitbereich deckt üblicherweise die Akutphase bis zu 24 h¹ nach Symptombeginn ab, in der noch ein Eingriff zur Rettung von Gewebe möglich ist.

2.2 Stand der klinischen Forschung

Eine Hirnschädigung wie der Schlaganfall wird in der Regel durch die resultierenden neurologischen Defizite erstmals klinisch ersichtlich. Abhängig davon, welchen Bereich des Hirns eine Schlaganfallläsion mit den durch Sauerstoff unterversorgten Zellen betrifft, äußern sich unterschiedliche funktionelle Beeinträchtigungen, etwa in Sprache, der visuellen, auditiven, und taktilen Sensorik, Motorik, Konzentration, Gedächtnis, Kognition, oder auch in Verhalten und Persönlichkeit [Kasten, 2007]. Entsprechende Symptome erlauben bei deutlicher Ausprägung eine grobe Lokalisation des betroffe-

¹Stellenweise auch als Hyperakutphase bezeichnet [Allen u. a., 2014]

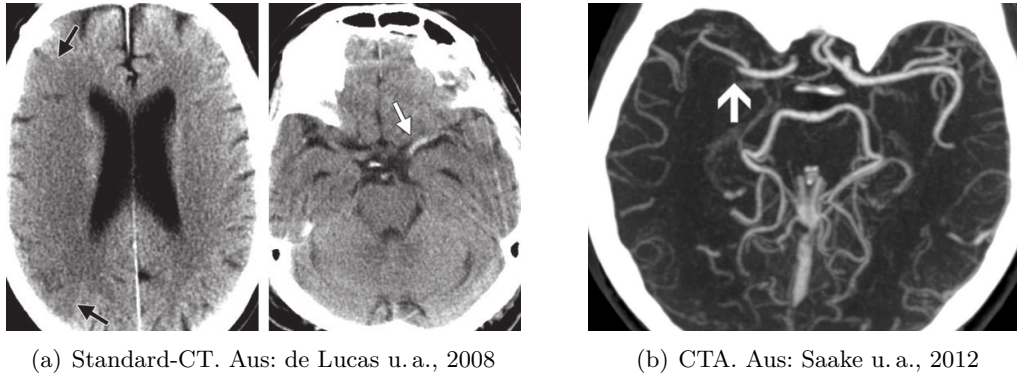


Abb. 2.3: Beispiele für Frühzeichen (Pfeile) eines Schlaganfalls im nativen CT: Ein großflächiger subtiler hypodenser Bereich im rechten Stromgebiet der MCA (a, links) und ein klar hyperdenses Frühzeichen für einen Thromboembolus (a, rechts). Verglichen damit zeigt ein CTA (b) kein Nervenzellgewebe, sondern ausschließlich Blutgefäße.

nen Hirnareals bzw. Gefäßes². Für die genauere Lokalisation und Diagnostik ist jedoch eine Bildgebung unerlässlich. Darüber hinaus kann es differentialdiagnostisch erforderlich sein, auf weitere Tests, wie EEG, Ultraschall, oder Liquor-/Lumbalpunktionen zurückzugreifen.

2.2.1 Bildgebende Diagnostik

In der klinischen Praxis wird als erster Schritt ein natives CCT (kraniale Computertomographie) durchgeführt, welches in den meisten Kliniken weit verbreitet und verfügbar ist [Hennerici u. a., 2017]. Dies dient bei Schlaganfallverdacht zu Beginn vor allem dem Ausschluss einer intrakraniellen³ Hirnblutung als Voraussetzung zur Anwendung der Thrombolyse bei ischämischem Schlaganfall. Bis in die Subakutphase (etwa 7 Tage nach Symptombeginn) stellt sich das Hämatom in der CT hyperdens dar, sodass der Nachweis von Blutungen durch den starken Kontrast gegenüber Hirngewebe bzw. Liquorräumen mit nahezu 100 % Sensitivität im CT möglich ist. Nach Thomalla u. a. [2009] lassen sich aber auch mit T2*-gewichteten Gradientenecho-Sequenzen im MRT frische intrazerebrale Blutungen „mindestens so sensitiv nachweisen wie mit der CT“ und die „Leitlinien empfehlen die Bildgebung mit kranialer CT oder MRT als gleichwertige Verfahren für den Nachweis akuter intrakranieller Blutungen“.

Darüber hinaus können auf der CT oft bereits Frühzeichen eines akuten Apoplexes sichtbar sein, etwa das hyperdense MCA-Zeichen im M1-Segment, welches im Axialschnitt des nativen CTs den Thromboembolus längs zum Verlauf der MCA darstellt und als indirekter Marker eines Infarkts im nachfolgend versorgten Gewebe gilt

²Bei Verschluss der MCA treten z. B. Gesichtsparesen, Gesichtsfeld- und Sprachstörungen auf.

³intrakraniell, d. h. intrazerebral, subarachnoidal, subdural oder epidural innerhalb des Schädels

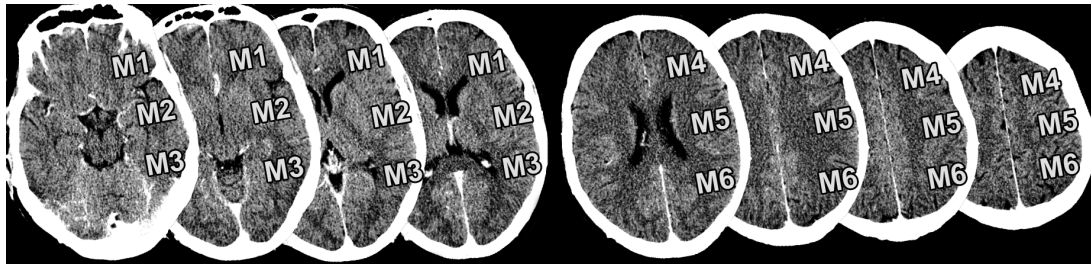


Abb. 2.4: MCA-Segmente als Teil des ASPECTS in axialen CT-Schnitten des Gehirns

(Abb. 2.3(a)). Dies erfasst zwar nur ungefähr ein Drittel aller Mediaverschlüsse, weist aber eine hohe Spezifität auf [Thomalla u. a., 2009]. Der Propf kann jedoch bei alleiniger Betrachtung der Bildaufnahme mit Gefäßverkalkungen, erhöhtem Hämatokrit, Grauwertveränderungen aufgrund eines umgebenden Tumors oder Quetschungen verwechselt werden [Jha u. a., 2009]. Manchmal ist im nachgeordneten M2-Segment im Axial-Schnitt der Thrombus als Punktzeichen (*engl.* Dot Sign) sichtbar, welches durch den Verlauf der Arterie senkrecht zum Schnitt zustande kommt. Nach Wardlaw u. a. [2005] spricht das Auftreten von Frühzeichen bereits im nativen CT für ein erhöhtes Risiko eines schwerwiegenden Krankheitsausgangs.

Im Rahmen des von Pexman u. a. [2001] vorgestellten ASPECTS (Alberta Stroke Program Early CT Score) findet eine standardisierte Beurteilung von Frühzeichen des Infarkts statt, obgleich diese immer noch subjektiven Eindrücken und entsprechender Expertise bei der Beurteilung von CT-Aufnahmen unterliegt. Eine objektivere computer-gestützte Beurteilung ist daher wünschenswert. Für die Bewertung werden zehn Hirnkompartments untersucht (vgl. Abschnitt 2.3), allerdings können nur begrenzt ischämische Veränderungen des Gewebes im hyper-akuten Stadium auf der CT-Aufnahme festgestellt werden. Daher ist nach Diener u. a. [2018] derzeit auch noch nicht ausreichend geklärt, „welche Rolle die CT-Frühzeichen, kategorisiert anhand des ASPECTS, als Selektionskriterium spielen“ können.

Üblich ist ebenso eine Gefäßdarstellung über CTA (CT-Angiographie) mittels Kontrastmittel, welche den Gefäßstatus der großen Halsschlag- und Hirnarterien abbildet (Abb. 2.3(b)). Dies hilft, den Verschluss in den einzelnen arteriellen Segmenten zu lokalisieren, den kollateralen Blutfluss einzuschätzen, sowie eventuelle Atherosklerosen festzustellen [Sporns u. a., 2018; Wardlaw u. a., 2005]. Über die „Maximumintensitätsprojektion“ (MIP) werden die Blutgefäße dank hoher Signalintensität durch das Kontrastmittel im eigentlichen Bild deutlich herausgestellt. Unterbrechungen im Verlauf der Gefäße deuten schließlich einen Verschluss an.

Eine ergänzende CT-Perfusion (CTP, Abschnitt 2.2.1.1) kann in Zweifelsfällen zur Patientenselektion der endovaskulären Therapie hilfreich sein. Bei Schlaganfällen, die vermutlich länger als 4,5 Stunden zurückliegen, oder Differenzialdiagnostik bei unklarer Ursache ist eine multivariate Bildgebung (CTA, multiparametrische und Mismatch-

Bildgebung) bereits primär angeraten. Nach Leitlinie sind „in besonderen Entscheidungssituationen die Abschätzung des irreversibel geschädigten Infarktkerns z. B. mittels ASPECTS im Nativ-CT, die Analyse der Quelldaten der CTA oder eine ergänzende CT-Perfusion zur Patientenselektion der endovaskulären Therapie hilfreich“ [Hennerici u. a., 2017].

Obgleich das MRT-Protokoll per se (zeit-)aufwändiger ist als die CT, nähern sich im Vergleich mit dem vollen multiparametrischen CT inklusive CTA und CTP die Zeitdauern an, sodass die MRT als mittlerweile gleichwertig effektiv angesehen werden kann [Thomalla u. a., 2009]. Sofern verfügbar, weist die MRT bereits wenige Minuten nach Symptombeginn bereits einen hohen Kontrast in der Diffusionsbildgebung⁴ (DWI) auf, wo hingegen die Parameter der CT mit Perfusionsbildgebung die metabolischen Folgen der Ischämie noch nicht ebenbürtig abbilden können. Auch Angiografie mittels flusssensitiver Sequenzen (*engl.* Time-Of-Flight, TOF) oder über Kontrastmittelgabe (*engl.* Contrast Enhanced, CE-MRA) ist in der MRT möglich.

Darüber hinaus ist nach Thomalla u. a. [2009] auch die neurosonologische Untersuchung in den ersten 24 Stunden häufig von praktischem Wert, da die hämodynamischen Folgen von Verschlüssen bzw. Stenosen hirnersorgender Arterien sowie die Prozesse während einer Behandlung (Rekanalisation, Reokklusion) unmittelbar nachvollzogen und direkt mit dem klinischen Status korreliert werden können. Prinzipiell ist mittels Ultraschall auch die Untersuchung der Perfusion über Signalveränderungen nach Gabe eines Echokontrastverstärkers möglich (analog zum Bolus-Tracking bei MR- oder CT-Perfusion); allerdings ist diese Technik u. a. durch die Beschränkung auf eine Ebene und fehlende Möglichkeiten der Quantifizierung der Perfusionsmessung nicht verbreitet. Weiterhin haben auch nuklearmedizinische Untersuchungsmethoden wie PET und SPECT keine Relevanz in der Routineversorgung eines Schlaganfallnotfalls, obgleich sie (bei hohem Aufwand) einige Vorteile bereithalten, wie z. B. die Möglichkeit zur quantitativen Bestimmung von Blutfluss, Sauerstoff- oder Glukosemetabolismus und anderer metabolischen Kenngrößen.

2.2.1.1 Perfusionsbildgebung

Die CT-Perfusion (CTP), bzw. die Perfusions-CT (PCT), basiert auf der Messung der Perfusion im Gewebe nach erfolgter Gabe eines Kontrastmittelbolus. Beim Kopf-CT steigt dabei mit dem Anstrom des Kontrastmittels die Röntgendichte des Hirngewebes vorübergehend an. Die Intensitätsänderung dieses Signals im zeitlichen Verlauf erlaubt die Berechnung verschiedener Parameter der Perfusion. Auch in der MRT ist eine Perfusionsbildgebung möglich, wofür ebenfalls ein Kontrastmittelbolus in Signal-Zeit-Kurven nachverfolgt wird – im Gegensatz zur CTP handelt es sich dabei um ein paramagnetisches und nicht jodhaltiges Kontrastmittel [Thomalla u. a., 2009]. Prinzipi-

⁴Durch Schrumpfung des Extrazellulärums kommt es zu einer Verminderung des Wasserdiffusionskoeffizienten (*engl.* „Apparent Diffusion Coefficient“, ADC)

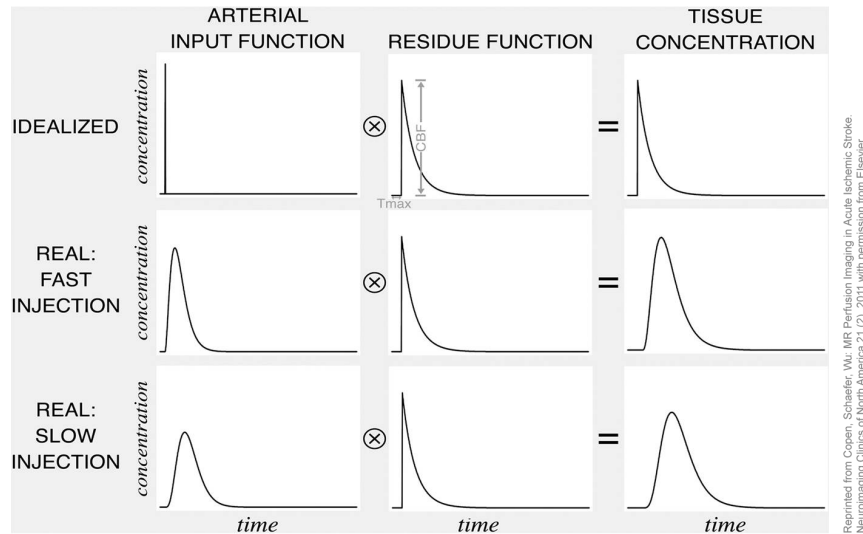


Abb. 2.5: Die „Residue“-Funktion repräsentiert den Anteil eines idealisiert eingespritzten Normbolus’ an Kontrastmittel über die Zeit (Auslaufen nach initialem Peak). Das Maximum dieser entspricht dem CBF zum Zeitpunkt T_{max} . Die AIF (*engl.* Arterial Input Function) gibt den tatsächlichen Kontrastmittelbolusverlauf an, der u. a. von der tatsächlichen Injektionsrate und Physiologie des Patienten vor Erreichen des Hirns abhängig ist. Die Gewebekonzentration (*engl.* Tissue Concentration) ist das Ergebnis aus der AIF gefaltet mit der Residue-Funktion. Aus: Copen u. a., 2011

ell werden zwar häufig einige Vorteile, z. B. eine höhere Sensitivität kleinerer Läsionen bei der MRT, genannt, welche jedoch unter Umständen durch die CT anderweitig kompensiert werden können, etwa durch höhere Auflösung bei geringerer Dosis [Otton u. a., 2013]. Vymazal u. a. [2012] reduzieren dies auf die Aussage: „Both techniques are practically equivalent in the hands of experienced personnel“.

Die resultierende Perfusionsaufnahme ist von vielen Faktoren abhängig, etwa von der Aufnahmedauer⁵, der nötigen Abdeckung des Hirns durch den Scanner, sowie der Injektionsrate des Kontrastmittels (Abb. 2.5), die Verlauf und Stärke des Bolus-Peaks beeinflusst, in Verbindung mit einer ausreichend hohen Aufnahme Frequenz zur möglichst genauen Abtastung des Peaks [Hakim u. a., 2019]. Außerdem beeinflusst die (Art der) Umwandlung der Intensitäts-Verlaufs-Messung in entsprechende Perfusionsparameterkarten deren Werte, insbesondere durch die Wahl der (nicht-)parametrischen Dekonvolutionsmethode⁶, die meist besonders sensitiv auf Rauschen reagiert, was sich zum Teil auch in den inhomogenen Parameterkarten ausdrückt (Abb. 2.6). Folgende Parameter werden üblicherweise berechnet [Hakim u. a., 2019]:

⁵Bei zu kurzer Zeitdauer $< 60s$ ist die Boluskurve unter Umständen nicht vollständig

⁶Wenn sowohl die AIF als auch der Verlauf der Gewebekonzentration aus Abb. 2.5 gegeben sind, kann man die „Residue“-Funktion z. B. mittels SVD-Dekomposition ermitteln, sodass sich dann die Parameter, wie CBF oder T_{max} , für jeden Voxel aus dieser Funktion ermitteln lassen

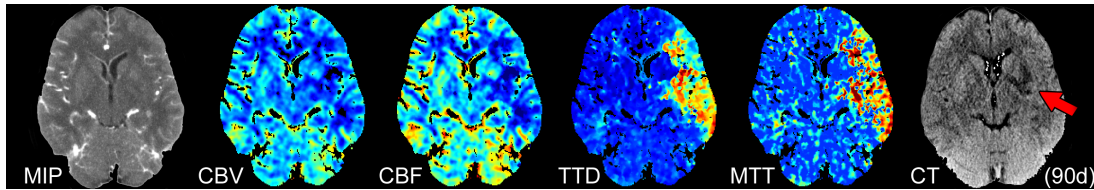


Abb. 2.6: Patient mit linkem MCA-Verschluss aus dem UKSH-Datensatz (Abschnitt 2.4). Die Perfusion deutet gering beeinträchtigte CBF bzw. CBV an, aber deutlich verlängerte Durch-/Ablaufzeiten (MTT, TTD), was einen großen Mismatch nahelegt. 90 Tage nach der erfolgreichen Behandlung ist die finale Läsion erkennbar (Pfeil).

Cerebral Blood Volume (CBV) stellt das totale Blutvolumen in einem Einheitsvolumen dar (in ml Blut pro $100 g$ Hirngewebe). Die Fläche unter der Signalverlaufskurve repräsentiert die Gesamtmenge an Kontrastmittel (Blut) in der ROI (*engl.* Region-Of-Interest). Es ist damit ein genereller Marker dafür, ob überhaupt Blut das Gewebe erreicht – unabhängig von der Zeit. Es kann vorkommen, dass das Volumen im Referenzbereich liegt, aber andere Parameter zugleich in dem betroffenen Gewebe abweichen. Außerdem kann auch Hypoperfusion zu erhöhtem Blutvolumen führen.

Cerebral Blood Flow (CBF) soll den Fluss einer Menge von Blut im Einheitsvolumen pro Zeit darstellen (in ml Blut pro $100 g$ Hirngewebe pro Sekunde). Der Fluss ist direkt proportional zum Druck (Ohm'sches Gesetz) und indirekt proportional zum Widerstand, d. h. der Fluss nimmt beispielsweise ab, sobald die Druckunterschied zwischen Arterien und Venen infolge eines arteriellen Verschlusses abnimmt. Entscheidend hat auch der Gefäßdurchmesser Einfluss auf den CBF-Parameter, sodass er bei Gefäßstenosen ebenfalls umso stärker abnimmt.

Mean Transit Time (MTT) ist die durchschnittliche Durchlaufzeit des Blutes in einer gegebenen Hirnregion und lässt sich direkt aus den ersten beiden Parametern durch $MTT = CBV / CBF$ berechnen. Es stellt besonders gut die Zirkulation der Mikrogefäße dar, weil es die Zeit misst, welche das Blut in den Kapillargefäßen verbleibt. Der in der Penumbra verminderte Blutdruck führt umgekehrt proportional zu verlängerten MTT-Zeiten und diese ist darum gut in MTT-Karten erkennbar.

Time To Peak (TTP) stellt die Zeit dar, welche das Kontrastmittel seit Injektion bis zum Erreichen seines Konzentration-Peaks benötigt. Der Parameter reagiert sehr sensitiv auf Flussänderungen, ist folglich aber auch stark abhängig von z. B. der Injektionsrate des Kontrastmittels. Eine bevorzugte alternative Definition berechnet daher stattdessen die Zeit mit dem ersten Auftreten einer Signalverstärkung durch das Kontrastmittel bis zu dessen Peak, um vorher auftretende Effekte vom Injektionsort bis zum Hirn auszuschließen.

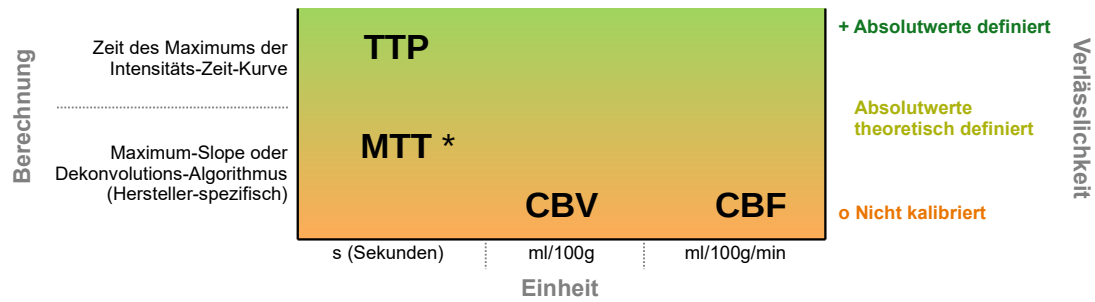


Abb. 2.7: Einordnung einiger Perfusionsparameter in eine Klassifikationsmatrix. Wichtig(*): Die Berechnung von $MTT = CBV / CBF$ basiert auf Dekonvolutions-Algorithmen.

Time To Maximum (T_{max}) deutet ebenfalls die Zeit bis zum Peak („Maximum“) an, aber auf der dekonvolutierten Kurve (*engl.* Tissue Residue Function), welche den Anteil der Kontrastmittelkonzentration zu einem Zeitpunkt gegenüber dem Gesamtbolus darstellt. Es stellt die Verzögerung des Bolus dar, wird aber durch viele Faktoren beeinflusst. Vor allem repräsentiert es den makrovaskulären Zustand, bei dem erhöhte Zeiten mit Minderperfusion und schwacher kollateraler Versorgung zusammenhängen.

Time To Drain (TTD) stellt die Zeit bis zum Auslaufen des Kontrastmittels dar und ist damit sensitiv für makro- und mikrovaskuläre Störungen, also jegliche hämodynamischen Veränderungen. Ein normalbefundetes TTD kann daher normale Perfusion mit hoher Sicherheit präzisieren.

Time To Start (TTS) resultiert aus der Subtraktion der MTT von der TTD und entspricht der Zeit von der Injektion bis zur ersten Verstärkung durch das Kontrastmittel an der beobachteten Stelle.

Turowski u. a. [2015] weisen in Bezug auf die Deutung der berechneten Werte aus der CTP darauf hin, dass – mit Ausnahme von TTP oder T_{max} – die Parameterbezeichnungen nicht die Realität des „Blutflusses“ oder des „Blutvolumens“ wiedergeben (Abb. 2.7, „Verlässlichkeit“), sondern mathematische Konstrukte aus den gemessenen Boluskurven sind, die die menschliche Vorstellung unterstützen sollen, während man solche Messungen interpretiert.

Da die Perfusionsparameter konstruierte Größen für die bessere Interpretation durch den Menschen sind, ist es in Bezug auf maschinell gelernte Algorithmen, welche multivariate Daten als Eingabe verarbeiten, zudem wichtig zu berücksichtigen, dass die Parameterwerte bei MR- und CT-Perfusion unterschiedlich zustande kommen. Fundamentaler physikalischer Unterschied ist die lineare Abschwächung der Röntgensignale durch Gewebedichte bzw. Kontrastmittelkonzentration im CT gegenüber der logarithmischen Abschwächung der Signalstärke durch den paramagnetischen Effekt des Kon-

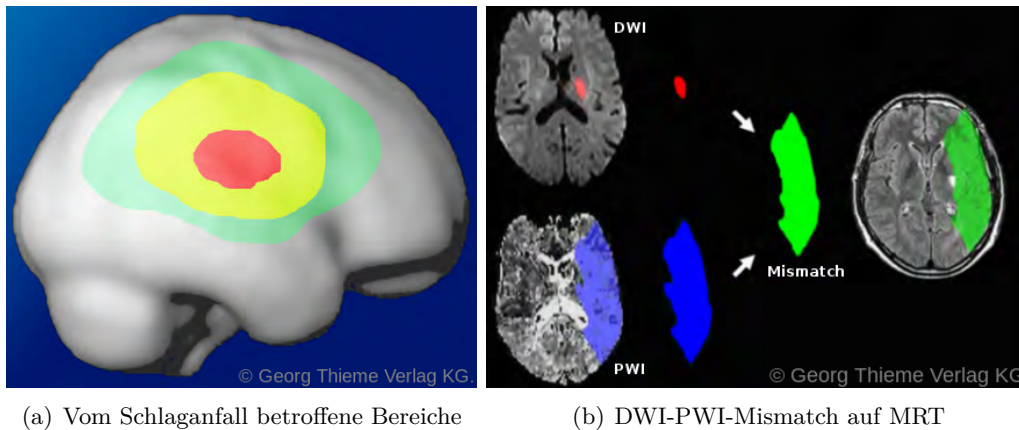


Abb. 2.8: (a) zeigt schematisch das bereits irreversibel geschädigte Gewebe im Hirn (Kern, rot), das umgebende kritisch minderperfundierte Gewebe mit erhaltenem Strukturstoffwechsel (Penumbra, gelb), sowie ein Areal oberhalb einer kritischen Grenze minderperfundierte Gewebes (benigne Oligämie, grün). In (b) definiert der Mismatch zwischen Diffusionsläsion (rot) und Perfusionsläsion (blau) potenziell rettbares Risikogewebe (grün). Im Verlaufsbild (rechts) nach erfolgter Thrombolyse ist nur eine geringe Infarzierung des Risikogewebes zu sehen. Aus: Thomalla u. a., 2009

trastmittels im MRT. Nach Turowski u. a. [2015] hängt folgerichtig auch bei der MRT die zeitliche Veränderung des Signals (Kurvenanstieg) vom Startpunkt der logarithmischen Funktion ab. Es lassen sich damit schwerlich Inter-Patienten-Schwellenwerte für die Perfusion definieren, vor allem, da MR-Intensitäten keine standardisierte, absolute Skala wie das CT besitzen (Hounsfield).

In dieser Arbeit werden vorwiegend CTP-Daten des „UKSH“-Datensatzes (siehe Kapitel 2.4) verwandt. Da die Parameter der Perfusionsbildgebung nie exakte physiologische Zustände des Gewebes darstellen können, besteht auch immer die Schwierigkeit der Lokalisation des Kerns bzw. der eigentlichen Nekrose und der Abgrenzung dieser zum minderperfundierte, aber rettbarer Gewebe in der Penumbra. Zusätzlich beeinflussen außerdem zahlreiche Effekte, wie die Schwellung des Gewebes, die Parameter. Mit Blick auf diese medizinisch nicht zu 100 % exakt fassbaren Zustände auf Basis der Bildgebung wird darum in dieser Arbeit bevorzugt die Segmentierung des „gesamten Risikogewebes“ (CTP-Perfusionsläsion inkl. Nekrose) statt der kleineren Penumbra verwendet.

2.2.1.2 Analyse der Perfusion zur Mismatch-Bestimmung

Wie zu Beginn von Abschnitt 2.2.1 bereits angesprochen, wird nach aktueller Leitlinie [Hennerici u. a., 2017] vor allem die Perfusionsbildgebung (Abschnitt 2.2.1.1) zur Selektion der Patienten für entsprechende Therapieoptionen verwendet. Die einfachs-

te Möglichkeit, die Schlaganfallbereiche abzugrenzen, stellen Schwellenwertmethoden dar. Nach Diener u. a. [2018] kann man irreversibel geschädigtes Gewebe („Kern“ in Abb. 2.8(a)) typischerweise mittels CBF (relativer CBF $< 31\%$ des mittleren kontralateralen CBF) oder des CBV ($< 2\text{ ml}$ pro 100g) identifizieren.

d’Esterre u. a. [2015] haben jedoch zeigen können, dass diese Grenzwerte für irreversible Schädigung von der Zeit bis zur Reperfusion abhängen. Die Schwellenwerte auf den Parameterkarten zur Abgrenzung der dauerhaften Infarktläsion sind umso höher, je weniger Zeit zwischen CT-Perfusionsbildgebung und Reperfusion liegt [Tsai u. a., 2018] – „Diese Ergebnisse bestätigen eindrücklich den dynamischen Charakter der Läsionsentwicklung beim akuten ischämischen Hirninfarkt.“ [Diener u. a., 2018]. Da auch schwere Perfusionsstörungen reversibel sein können, ist eine starre Anwendung ausschließlich CT-perfusionsbasierter Schwellenwerte zum Ausschluss von Patienten von z. B. einer mechanischen Thrombektomie nicht empfehlenswert.

Die Penumbra wird im MRT nach Vymazal u. a. [2012] nicht durch Subtraktion von CBF und CBV, sondern durch Subtraktion zwischen Perfusionsparametern, wie MTT, TTP oder CBF, und der DWI-Aufnahme (statt CBV⁷) ermittelt. Dabei gilt jedoch, dass dies die finale Läsion eher unterschätzt und das Volumen über die Zeit noch in die CBV-Läsion hineinwächst⁸. Die Verwendung der Zeitparameter, wie T_{\max} oder TTP, in Verbindung mit Schwellenwerten von mehr als 4 – 6s stellt jedoch recht verlässlich kritisch unterversorgtes Risikogewebe dar. Abbildung 2.9 verdeutlicht die unterschiedlichen Intensitätsveränderungen der einzelnen Parameterkarten für beide Schlaganfallbereiche. Im Anhang finden sich auf Seite 123 und 124 außerdem zwei Beispiele für unterschiedlich großen Mismatch inklusive Verlaufsbild nach Behandlung.

In nochmaligem Rückblick auf Abbildung 2.6 mit den häufig grob aufgelösten und verrauschten Parameterkarten ist es jedoch wichtig, den Kontext über die voxelweise Anwendung von Schwellenwerten hinaus zu berücksichtigen und (leichte) Abweichungen von den Schwellenwerten nicht zwingend auszuschließen. Statistische Modelle, etwa ein lineares Regressionsmodell und zusätzliche, räumlich gefilterte Eingaben (z. B. geglättete Karten) bzw. Vorwissen (Wahrscheinlichkeitsatlas) können mit diesen Abweichungen in den Daten umgehen, wie etwa in der Methode von Kemmling u. a. [2015] verwendet. Bei manueller Betrachtung bzw. Segmentierung der Parameterkarten über simple, voxelbasierte (Schwellenwert-)Verfahren sind jedoch zunächst oft nur unregelmäßige und physiologisch unplausible Läsionsbereiche zu erkennen.

Gute Protokolle und Methoden für die (manuelle) Segmentierung von Kern und Penumbra vorausgesetzt, spielt insbesondere deren (Miss-)Verhältnis zueinander – der sogenannte „Mismatch“ – eine entscheidende Rolle für die Patientenselektion, wie in vielen Studien (einige auf Seite 25 in Abb. 2.10) gezeigt werden konnte. In Abbildung 2.8(b) wird etwa die Diffusionsläsion dem Infarktkern und die Perfusionsstörung dem

⁷CBV ist insbesondere im MRT nicht sensitiv genug, um kleine Läsionen $< 10\text{ ml}$ darzustellen.

⁸Die initiale Läsion in der CBV kann tatsächlich bei zeitiger, erfolgreicher Reperfusion teilweise reversibel sein [Angermaier u. a., 2015]

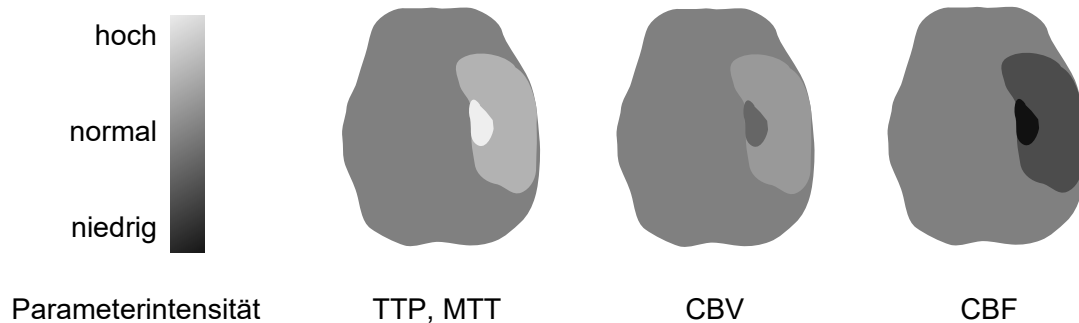


Abb. 2.9: Perfusionsparameterintensität in den beiden Schlaganfallbereichen.

kritisch minderperfundierten Areal zugeordnet und deren Missverhältnis berechnet. Alle in der Abbildung 2.10 des nächsten Abschnitts genannten Studien zeigen, dass es möglich ist, über das Mismatch-Konzept Patienten in der klinischen Praxis effektiv für das Zeitfenster über drei Stunden nach Symptombeginn zu behandeln. Zusammenfassen lassen sich die Ergebnisse grob nach folgendem zeitlichen Schema [Hennerici u. a., 2017] zur Selektion von hauptsächlich (Abschnitt 2.2.2) intravenöser Thrombolysen (IVT) bzw. mechanischer Thrombektomie (MT) bei Symptombeginn von:

- < 4,5 h Bildgebung nur zum Ausschluss intrakranieller Blutung (ICB), ansonsten Gabe der IVT; MT bei großem arteriellen Gefäßverschluss empfohlen;
- > 4,5 h Selektion der Patienten für IVT/MT durch Penumbra-Bildgebung mittels MRT oder CTP (gilt auch für unbekanntem Symptombeginn);
- < 6 – 12 h Bei Verschluss von MCA-M1/-M2 oder ICA in Bildgebung besonders MT empfohlen, unabhängig von Gabe der IVT (d. h. bei Kontraindikation IVT ist MT nicht ausgeschlossen);
- < 24 h (bzw. unbekannt) Hierbei ist besonders erst durch Selektion der Patienten mittels MRT oder CTP zur Identifizierung von Infarktkern für ein „clinical imaging mismatch“ (DAWN-Studie) bzw. Penumbra-Bildgebung (DEFUSE-3 Studie bei < 16 h) zur Ermittlung des Mismatch zwischen Perfusionsläsion und Infarktkern eine effektive Behandlung möglich:

2.2.2 Akutbehandlung

Die Akutphase des Schlaganfalls dauert bis zu einem Tag nach Symptombeginn, in der auch im späteren Zeitfenster noch häufig Reperfusion erzielt werden kann. Wie zuletzt im vorherigen Abschnitt 2.2.1.2 erläutert, ist der ermittelte Mismatch ein entscheidendes Kriterium für die Wahl der Behandlungsmethode. Dafür muss der Patient jedoch schon in einem entsprechenden Haus aufgenommen worden sein.

Die internationalen Standards kennen für die Krankenhaus-Infrastruktur verschiedene Klassifikationen (Level) von sogenannten „Stroke-Centers“ mit speziellen Behandlungsmöglichkeiten, denen Patienten aus kleineren bzw. nicht-spezialisierten Häusern zugeführt werden können [Pierot u. a., 2019]. Der Kern eines solchen Center stellt immer die „Stroke-Unit“ dar. In diesem örtlich klar definierten Bereich des Krankenhauses (oft als Teil der Neurologie) arbeitet verschiedenes, geschultes Personal zusammen, um Schlaganfälle bestmöglich zu versorgen. Erst Stufe-1-Center können endovaskuläre Therapien (wie in Abschnitt 2.2.2.1) durchführen. Niedriger klassifizierte Häuser können dann nicht die komplette Bandbreite neurovaskulärer Störungen behandeln, weshalb Patienten bevorzugt in ein Center der Stufe 1 transportiert werden sollten – es sei denn, dieses liegt zu weit entfernt (mehr als zwei Stunden). Der Zeitaspekt ist einer der wichtigsten Prädiktoren für ein gutes klinisches Ergebnis und lässt sich in dem bekannten Leitspruch „Zeit ist Hirn“ zusammenfassen: Je schneller (transportiert, diagnostiziert, therapiert), desto besser. Dies wird in Abschnitt 2.2.2.2 noch erläutert.

Campbell [2017] nennt zwei potenzielle Behandlungsstrategien beim akuten ischämischen Schlaganfall: Die Wiederherstellung des Blutflusses (Reperfusion) sowie die Neuroprotektion zur Verhinderung von (weiteren) Zellschädigungen. Zwei Reperfusionsmethoden haben sich etabliert: Die medikamentöse intravenöse Thrombolyse (IVT) zur Auflösung eines Thrombus mit sicherer Anwendung insbesondere zu einem frühen Zeitpunkt innerhalb der ersten drei bis vier Stunden nach Symptombeginn, sowie die mechanische Thrombektomie (MT) zur Entfernung des endovaskulären Propfs über einen Leistenkatheter durch die Hauptschlagader, wobei beide Methoden kombiniert werden können. Die Verwendung des synthetischen Enzyms Alteplase für die IVT ist dabei Standard; andere Medikamente (Tenecteplase, Desmoteplase) zur Verminderung spezifischer Risiken durch die Alteplase, oder die Verstärkung der Lyse (z. B. durch Ultraschall) sind derzeit Gegenstand der Forschung [Shibata u. a., 2019].

Nach Diener u. a. [2018] gelingt die Rekanalisation von Hirngefäßen bei Verschlüssen der MCA in etwa 30 % der systemisch lysierten Patienten, wobei die Wahrscheinlichkeit, ein intrakranielles Gefäß zu rekanalisieren, mit der Länge des Thrombus abnimmt. Ein ungefährender Grenzwert für die Thrombuslänge sind maximal 8 mm des Thrombus als „Dense Artery Sign“ im CCT (siehe Seite 15). Seit circa zehn Jahren wird darüber hinaus jedoch auch die mechanische Thrombektomie (MT) als Behandlungsalternative eingesetzt, die dies ermöglicht.

2.2.2.1 Mechanische Thrombektomie nach Mismatch-Kriterien

Ab dem Jahr 2015 haben mehrere Studien die signifikante Überlegenheit der endovaskulären Therapie mittels MT in Kombination mit IVT gegenüber der alleinigen Thrombolyse gezeigt [Diener u. a., 2018]. Diese sind u. a. in Abbildung 2.10 aufgeführt und bestätigen insbesondere für Patienten mit einem Symptombeginn, der länger als 4,5 h zurückliegt oder unbekannt ist, dass eine Selektion nach geeigneter Bildgebung

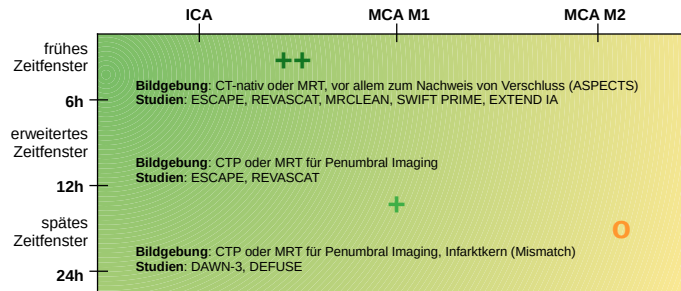


Abb. 2.10: Evidenz zum Nutzen der mechanischen Thrombektomie (MT) nach Studienlage im Jahr 2019 (siehe Barow u. a., 2019). Insbesondere im frühen Zeitfenster und bei Verschlüssen in der ACI oder dem M1-Segment der MCA ist der Nutzen unbestreitbar (++). Im erweiterten Zeitfenster bis zu 24 Stunden (oder unbekannt) gibt es deutliche Vorteile der MT gegenüber der bisherigen Standardtherapie unter bestimmten Einschlusskriterien auf Basis einer erweiterten Bildgebung (+). Für Verschlüsse im M2-Segment ist die Studienlage jedoch unsicher (o).

(MRT oder CTP) für eine intravenöse Thrombolyse und mechanische Rekanalisation essentiell ist, sodass diese Patientengruppen signifikant von einem entsprechenden positiven Outcome profitieren können. Der Effekt ist teilweise so deutlich, dass Studien bereits vorzeitig beendet wurden.

Konkrete Einschlusskriterien zum vorteilhaften Einsatz einer mechanischen Thrombektomie sehen nach Barow u. a. [2019] bei länger zurückliegendem Symptombeginn (bis 24 h) für Patienten bis 80 Jahre und einem Verschluss der ICA bzw. MCA im M1-Segment in der erweiterten Bildgebung mit MR (DWI/PWI) oder CTP wie folgt aus. Der „Mismatch“ definiert die Eignung für die MT und ist zum einen abhängig von der Quantifizierung des Infarktkerns (siehe „DAWN“-Studie):

- NIHSS⁹ > 10, Infarktvolumen < 31 ml
- NIHSS > 20, Infarktvolumen < 51 ml
- NIHSS > 10, Infarktvolumen < 21 ml

Zum anderen (siehe „DEFUSE-3“-Studie) wird darüber hinaus die Perfusionsläsion berücksichtigt und das Verhältnis zwischen Kern und umgebenden Risikogewebe als Mismatch-Kriterium bestimmt:

- Infarktkernvolumen < 70 ml
- Risikogewebevolumen > 15 ml
- Verhältnis von Kern- zu Penumbra-Volumen < 1,8

⁹für NIHSS siehe Abschnitt 2.3

In der EXTEND-Studie dagegen schlägt man ein erweitertes Mismatch-Kriterium mit einem Volumen-Verhältnis Kern (DWI bzw. CBF) zu Risikogewebe (Perfusionsläsion) größer als 1,2 vor, sowie eine absolute Mindestdifferenz von 10 ml bei Verwendung eines $T_{max} > 6$ s für die Perfusionsläsion.

Auch bei unklarem Symptombeginn ist die erweiterte Bildgebung und darauf basierende Selektion für IVT und MT empfohlen. Anzumerken bleibt ebenfalls, dass es laut Schmidt-Pogoda u. a. [2019] auch bei seltenen Schlaganfallursachen, wie Dissektionen der Halsarterien oder Tumorembolien, oft eine Thrombektomie durchführbar ist, damit diese ebenso von einem verbesserten Outcome profitieren.

Badhiwala u. a. [2015] berichten etwa in einer Zusammenfassung der Ergebnisse Rekanalisierungsraten bei der Thrombektomie von circa 75 % gegenüber 34 % bei der systemischen Thrombolyse allein, wobei eine ICB zwar bei 5,7 % in der Interventionsgruppe auftrat, aber im Vergleich zu 5,1 % in der Gruppe mit intravenöser Thrombolyse nicht statistisch signifikant war. Auch die 15,8 % Todesfälle in der Thrombektomiegruppe waren nicht signifikant erhöht gegenüber 17,8 % bei intravenöser Thrombolyse. Diener u. a. [2018] ziehen außerdem aus der HERMES-Metaanalyse die Schlussfolgerung, dass alle Subgruppen in gleichem Maße von MT profitieren und dies insbesondere Patienten über 80 Jahre, oder Patienten, die später als 5 h nach Symptombeginn randomisiert wurden, sowie Patienten mit Kontraindikationen gegen die Thrombolyse betrifft.

2.2.2.2 Zeitrelevanz

Herausstechend bei den Ergebnissen der MT-Studien ist, dass ein positiver Effekt auch im mittleren bis späten Zeitfenster gegeben ist [Albers u. a., 2018] und sich damit der Behandlungsspielraum erweitert. In der HERMES-Analyse wurde von Bourcier u. a. [2019] der Anteil erfolgreicher Reperfusion ($mTICI \geq 2b$, siehe auch Abschnitt 2.3) als Funktion von verschiedenen Intervallzeiten (Symptombeginn, Notaufnahme, Bildgebung, Behandlungsbeginn) untersucht. Dabei nahm der Anteil erfolgreicher Reperfusion mit steigender Zeitdauer seit Notaufnahme bzw. Bildgebung bis zur Behandlung deutlich ab. Dabei ist eine schnelle Reperfusion ein Faktor für ein gutes klinisches Ergebnis. Dies ist Anreiz und Grund, warum rund um die spezialisierten Stroke-Center bzw. Stroke-Units die einzelnen Intervallzeiten für Transport, Aufnahme, Bildgebung und Behandlung so gut wie möglich optimiert werden, sodass immer und überall – unabhängig von den Patientencharakteristiken (z. B. Alter) – ein Gefäßverschluss schnellstmöglich behandelt wird [Goyal u. a., 2016].

Kemmling u. a. [2015] haben den Effekt auf das Geweberesultat von erfolgreicher zu nicht erfolgreicher Rekanalisation bis zu 15 Stunden nach Symptombeginn zeigen können, wobei auch dieser nichtlinear mit der Zeit abnahm. Im Mittel betrug die Reduktion des Infarkt volumens bei einem Behandlungsintervall von fünf Stunden 53 %. Dieser Effekt nahm mit jeder Stunde um fünf Prozentpunkte bis auf unter 5 % nach

10 weiteren Stunden einer möglichen Behandlung ab, sodass eine Vorhersage über ein derartiges Modell eine Zeit-abhängige Abschätzung von geeigneten Patientenverläufen bei Einsatz einer endovaskuläre Therapie liefern könnte.

Zusammenfassend bestätigen sich die Ergebnisse aus den Studien der MT in der klinischen Routine, genauso wie der vorteilhafte Effekt einer zeitigen Reperfusion [Minnerup u. a., 2016], sodass auch eine geringere Sterblichkeit der Patienten beobachtet werden kann. In dieser Arbeit geht es folglich mit dem retrospektiven „UKSH“-Datensatz darum, dass zum Zwecke der Patientenselektion für die überlegene Thrombektomie ein objektives und schnelles Werkzeug zur quantifizierten Feststellung des potenziell rettbaren Gewebes entwickelt werden kann.

2.2.3 Gewebeergebnis als Prädiktor

Die Frage bleibt jedoch, welche Endpunkte für die behandelnde Ärztin dabei relevant sind und in welchem Zusammenhang sie stehen. Zum einen ist die Überlegenheit der MT bei geeigneten Patienten anhand klinischer Endpunkte gezeigt worden. Andererseits sind gerade die Grenzfälle und subjektive Beurteilung eines Mismatches als Selektionskriterium zur Abschätzung des Verhältnisses von Kern zu Penumbra oder des Läsionsvolumens äußerst schwierig – und objektivierte Quantifizierungen umso wünschenswerter.

Nach Nogueira u. a. [2009] werden auch „Rekanalisation“ und „klinischer Outcome“ fälschlicherweise häufig vermischt, obgleich der finale Rekanalisationsstatus (siehe Abschnitt 2.3) den besten Prädiktor für den klinischen Outcome für Patienten mit MT darstellt. Dabei kann nach Drewer-Gutland u. a. [2015] die Rekanalisation via CTP-basiertem Geweberesultat (Mismatch zwischen Penumbra und Folgeläsion) besser als der TICI (siehe Abschnitt 2.3) die Rekanalisation wiedergeben. Es kann jedoch sein, dass man den Propf mechanisch entfernt, dann aber andere Faktoren das klinische Outcome trotz guter Rekanalisation negativ beeinflussen, wie z. B. ein erhöhter Blutdruck.

Betrachtet man den Effekt der MT auf die Folgeläsion, so kann nach Compagne u. a. [2019] die Behandlung nur zu einem kleinen Teil den funktionalen Outcome erklären. Zwar waren kleinere Folgeläsionen nach der Behandlung mit besseren Outcomes assoziiert, aber sie erklärten nur 14 % des positiven Effekts der MT, sodass die Folgeläsion allein zur Abschätzung des klinischen Resultats nicht ausreicht. Rosso u. a. [2019] schreiben jedoch, dass die Lokalisation und Ausdehnung des Geweberesultats sinnvoll zum Abschätzen des klinischen Resultats sind. Nach Khan u. a. [2017] sei darüber hinaus vor allem grob die Hirnregion entscheidend für den Outcome, die Lateralität aber nachrangig.

Zu festgestellten Hirnläsionen gibt es zudem verschiedene Studien, die lokalisierte Störungen im Hirngewebe funktionalen Beeinträchtigung zuordnen. Kemmling u. a. [2013] haben z. B. mittels einer Atlas-basierten logistischen Regressionsanalyse die

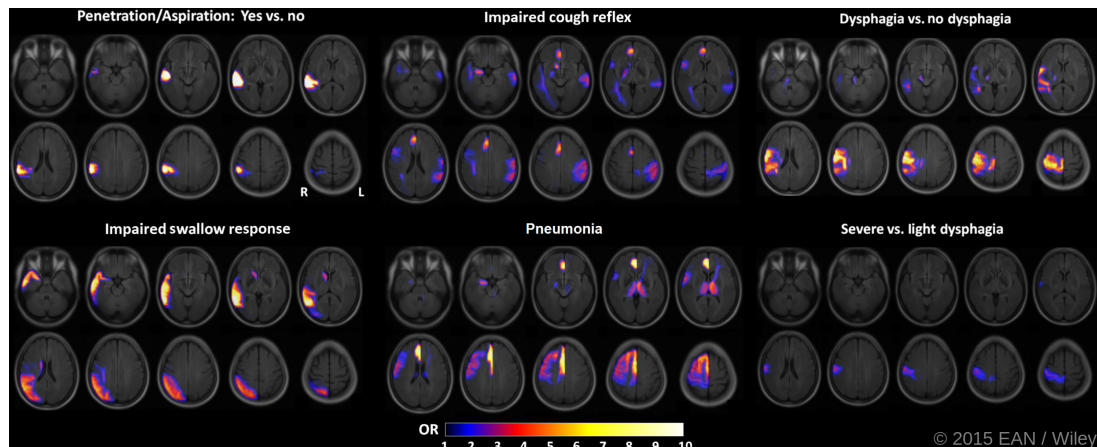


Abb. 2.11: Schlaganfallkomplikationen, die mit bestimmten Hirnarealen über farbkodierte Odds-Ratios assoziiert sind. Aus: Suntrup u. a., 2015; Suntrup-Krueger u. a., 2017

Verknüpfung von peri-insularen Infarkten in der rechten Hemisphäre und Pneumonie zeigen konnten – weitere Assoziationen sind in Abbildung 2.11 dargestellt. Machner u. a. [2009, 2012] haben darüber hinaus verschiedene visuelle Defekte und Symptome in untersuchten Schlaganfallpatienten mit deren Läsionen verknüpfen können. Es hat sich herausgestellt, dass die Läsionstopologisierung ein besserer Prädiktor für den klinischen Outcome sein kann, als beispielsweise das Volumen [Rangaraju u. a., 2015].

2.3 Klinische Bewertungssysteme (Scores)

Eine Vielzahl von Scoring-Systemen erlaubt die Klassifikation oder grobe Quantifizierung von Krankheitsbildern, sowohl vor, während, als auch nach der Therapie. Sie werden oft in Studien zur Selektion oder als Erfolgskriterien benutzt. Auf Seite 16 wurde bereits der ASPECTS vorgestellt, wovon weitere Abwandlungen existieren, etwa für die posteriore Zirkulation (pc-ASPECTS) oder für CTP-Aufnahmen. Hier wird jedoch nur der ursprüngliche ASPECTS vorgestellt.

Alberta Stroke Program Early CT Score (ASPECTS) bewertet in erster Linie anhand von CCTs ohne Kontrastmitteleinsatz, ob in zehn Regionen des Media-Strömgebiets Veränderungen sichtbar sind. Ein höherer ASPECTS ist dabei verknüpft mit niedrigerem Risiko von Blutungen und besseren funktionalen Outcomes [Barber u. a., 2000], wobei ein Wert ≥ 8 meist einen guten Nachsorge-mRS ≤ 2 prädiziert. Die Regionen sind: Nucleus caudatus, Putamen, Capsula interna, die Inselrinde, M1 (Frontal-Operculum), M2 und M3 (Anteriorer und Posteriorer Temporallappen), sowie M4-M6 im superioren MCA-Stromgebiet oberhalb von M1-M3. Bei n auffälligen Regionen ergibt sich der $ASPECTS = 10 - n$.

Beeinträchtigung		Durchblutung		Beeinträchtigung	
0	keine Symptome	0	keine	0	keine Symptome
0–4	leicht	1	minimal	1	keine relevante
5–15	mittelschwer	2	partiell:	2	leicht
16–20	mittelschwer bis schwer	2a	< 67 %	3	mittelschwer
21–42	schwer	2b	voll, aber verlangsamt	4	höhergradig
		3	vollständig	5	schwer
				6	Tod

Tabelle 2.1: NIHSS

Tabelle 2.2: TICI

Tabelle 2.3: mRS

Tabellen: Verschiedene Scores und deren Klassifikation, üblicherweise benutzt für die Aufnahme (NIHSS), das direkte Behandlungsergebnis (TICI), und die Nachsorge (mRS).

National Institutes of Health Stroke Scale (NIHSS) ist ein Defizit-Score, mit dem Patienten z. B. bei Aufnahme systematisch eingeschätzt werden. Die originale Skala unterscheidet beim Patient die Bereiche Bewusstsein (0–7 Defizitpunkte), Blickfeld (0–2), Sehfeld (0–3), Gesichtslähmung (0–3), Armbewegungen li./re. (0–8), Beinbewegungen li./re. (0–8), Gliederataxie (0–2), Sinneswahrnehmung (0–2), Sprache (0–3), Aussprache (0–2), sowie Unaufmerksamkeit/Auslöschung (0–2). Die Summe wird nach dem Schema in Tabelle 2.1 klassifiziert. Es gibt eine modifizierte Skala (mNIHSS), die die Benutzung des NIHSS vereinfachen soll und zuverlässiger eingesetzt werden kann, da redundante Items entfernt wurden, sodass nur noch maximal 31 Punkte erreicht werden [Meyer u. a., 2009].

Thrombolysis In Cerebral Infarction (TICI) beschreibt den Rekanalisationsstatus durch einen Eingriff und verursacht des Öfteren einige terminologische Verwirrung, insbesondere bei der Frage, ab wann eine „erfolgreiche Reperfusion“ gegeben ist. Dabei gilt die Definition 2b–3 mittlerweile als etabliert und strenger als der Einschluss auch von Fällen mit Rekanalisation 2a [Angermaier u. a., 2016]. Die Problematik ergibt sich auch aus dem Fakt, dass die Rekanalisation eines Gefäßes nur partiell erfüllt sein kann, aber z. B. entfernter dahinter liegende Bereiche normal reperfundiert sind, so als wäre ein TICI=3 erreicht worden. Es gibt auch einen modifizierten TICI [Higashida u. a., 2003], der vor allem die Einschätzung der partiellen Durchblutung vereinfachen soll, wobei das Kriterium nun 50 % statt 67 % für 2a bzw. 2b lautet (siehe Tab. 2.2).

modified Rankin Scale (mRS) bewertet das funktionelle Ergebnis nach Schlaganfall [Swieten u. a., 1988], um die Behinderungen des Patienten im Alltag einzuschätzen. Ein Score von 0–2 (siehe Tabelle 2.3) wird dabei üblicherweise mit einem „guten Ausgang“ gleichgesetzt.

2.4 Schlaganfalldatensätze zum Training und Test von Algorithmen

Es gibt zwei große Herausforderungen beim Deep Learning für Aufgaben in der Schlaganfallprädiktion in Bezug auf die Datengrundlage:

1. Es besteht ein allgemeiner Mangel an (öffentlich) zugänglichen, durch Experten annotierten, repräsentativen, medizinischen Daten.
2. Die Bilddaten werden in Bezug auf die zeitliche Dimension nur unregelmäßig zu den Zeitpunkten t_A bzw. t_I (schnellstmöglich) sowie t_V aufgenommen.

Es können keine regelmäßigen Observation an ein- und demselben Patienten vor einer Intervention angestellt werden, da das Vorenthalten einer Behandlung offensichtlich ethisch nicht vertretbar ist. Das macht es schwierig, die Wachstumsdynamik zu beobachten.

In den letzten Jahren wurden im Rahmen der sich weiterentwickelnden *Open Science*, beginnend mit der „Grand Challenge“-Initiative von Heimann u. a. [2009] (www.grand-challenge.org) auf der MICCAI 2007, Datensätze in Wettbewerben rund um biomedizinische Fragestellungen veröffentlicht. Im Jahr 2015 haben schließlich Maier u. a. [2017] die ISLES (*engl.* ISchemic LESion Segmentation challenge, www.isles-challenge.org) veranstaltet, die in den darauffolgenden Jahren verschiedene Modalitäten und Ziele im Rahmen der Schlaganfallsegmentierung und -prädiktion behandelt. Tabelle 2.4 listet die Ausgangscharakteristika dieser und der in dieser Arbeit benutzten Daten auf.

MR-Perfusion In der ISLES 2015 gab es zwei Aufgaben: Die Segmentierung subakuter Läsionen (nicht nur des MCA-Stromgebiets) in MR-Sequenzen, genannt „SISS“, sowie die Schätzung der akuten Infarkt- und Penumbra-Läsionen auf MR-Diffusion und -Perfusion, die die Grundlage zur Bestimmung des Diffusions-Perfusions-Mismatch bilden („SPES“). Bei der ISLES 2016 und 2017 wurden ebenfalls Diffusions- und Perfusionen der MRT bereitgestellt, bei letzterer mit Segmentierungen der Läsion bei der Verlaufskontrolle ohne Akutmasken als Grundwahrheit. Es handelt sich dabei um andere Daten als bei der ISLES 2015, sodass diese sich nicht gegenseitig ergänzen.

CT-Perfusion Für die ISLES 2018 schließlich wurde die MRT nur noch als Grundlage für die Erstellung der Grundwahrheit benutzt, denn ein Experte hat den Infarktkern direkt nach Intervention auf der DWI-Aufnahme segmentiert. Dagegen stammen die Akutbilder vom Perfusions-CT. So stellt auch dieser Datensatz wieder nur die Daten der Läsion zu genau einem Zeitpunkt bereit.

Am Universitätsklinikum Schleswig-Holstein (UKSH) wurde vorher schon im Rahmen des „TRAVE-Stroke“-Projekts ein retrospektiver Datensatz zur Prädiktion von

2.4 Schlaganfalldatensätze zum Training und Test von Algorithmen

Median (IQA)	SISS15	SPES15	ISLES17	ISLES18	UKSH
Modalität	MR-D	MR-D/P	MR-D/P	CT-P	CT-P
Probanden n	28 + 36	30 + 20	43 + 32	63 + 40	29
Männlich, n (%)	n/a	n/a	n/a	n/a	16 (55%)
Alter, Jahre	n/a	n/a	n/a	n/a	70 (63–77)
Aufnahme NIHSS	n/a	n/a	n/a	n/a	15 (12–16)
Intervention TICI	n/a	n/a	2b (2a–3)	n/a	$\geq 2b$
Verlauf (90d) mRS	n/a	n/a	1 (1–2)	n/a	n/a
Kern Volumen, ml	17 (2–87)*	10 (5–19)*	n/a	17 (7–46)*	27 (3–86)
Risikogew. Vol., ml	n/a	152 (80–211)*	n/a	n/a	164 (149–199)
Folgeläsion Vol., ml	n/a	n/a	15 (5–59)*	n/a	31 (18–93)
$t_{0 \rightarrow A}$, Std.	n/a	n/a	2,2 (1,8–3,0) [†]	≤ 8	1,7 (1,4–3,4)
$t_{A \rightarrow I}$, Std.	n/a	n/a	1,8 (1,5–2,3) [†]	≤ 3	1,7 (1,4–2,1)
$t_{0 \rightarrow I}$, Std.	n/a	n/a	4,2 (3,6–5,5) [†]	≤ 11	3,9 (3,2–4,9)
Annotierte Zielstrukturen (Binärmasken)	Subakute Läsion	Kern, Penumbra	Folgeläsion	Kern (DWI)	CBV-Läsion, TTD-Läsion, Folgeläsion
Eingabeparameter	FLAIR, T1w TFE/TSE, T2w TSE, DWI	T1c, T2, CBV, TTP, DWI	ADC, DWI, CBF, CBV, TTP, Tmax, MTT	CBF, CBV, MTT, Tmax, Rohdaten	CBV, TTD CBF, MTT

Tabelle 2.4: Ausgangsmerkmale der in dieser Arbeit angesprochenen Datensätze: In ISLES17 werden die Daten aus ISLES16 erweitert, weshalb dieser hier nicht aufgeführt ist. *Die aufgeführten Zahlen betreffen die Trainingsdaten. [†]Zudem sind die Meta-informationen der ISLES17-Trainingsdaten nur für circa 2/3 der Probanden verfügbar. Die Datensätze enthalten MR-Diffusion (MR-D), MR-Perfusion (MR-P), oder CT-Perfusion (CT-P). Die Zeiten zwischen Symptombeginn (t_0), Aufnahme in der Radiologie (t_A), sowie Intervention (t_I) sind aufgeführt. Die Probandenzahl ist für *Training* + *Test* aufgeschlüsselt, falls im Voraus getrennt.

Schlaganfallläsionen angelegt. Dieser wurde für 29 Patienten mit MCA-Infarkten um Expertensegmentierungen auf CBV und TTD in der Akutbildgebung erweitert, neben der bestehenden Segmentierung der finalen Läsion bei der Verlaufskontrolle im nativen CT und wird in dieser Arbeit als „UKSH-Datensatz“ referenziert. Der Datensatz wurde initial mit FSL-FLIRT¹⁰ von Jenkinson u. a. [2012] für eine affine Registrierung vorverarbeitet, um geneigte Köpfe zu korrigieren und in einen gemeinsamen Raum zu transformieren. Ein Downsampling wurde angewendet, sodass die Eingabegröße der CBV- und TTD-Karten mit $128 \times 128 \times 28$ Voxeln eine gute Abwägung zwischen reduziertem Rechenaufwand und ausreichender Auflösung der Daten darstellt.

Alle Patienten wurden mit der mechanischen Thrombektomie behandelt und in Bezug auf einen TICI $\geq 2b$ erfolgreich rekanalisiert. Damit sind Start- und Endzustand der Läsion, zwischen dem das Läsionswachstum stattfindet, gegeben: Zum einen bevor der Eingriff passiert (Kern und Risikogewebe) und zum anderen mit Eingriff, der das Wachstum der Läsion stoppt. Unter idealisierten Annahmen (z. B. keine zwischenzeitli-

¹⁰Nutzung von „Mutual Information“ als Ähnlichkeitsmaß für einen bereits existierenden CT-Atlas

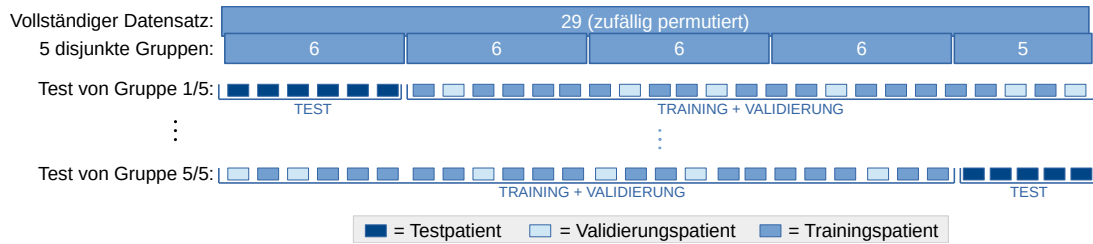


Abb. 2.12: 5-fache Kreuzvalidierung von Algorithmen anhand der 29 Patienten des UKSH-Datensatzes: Eine zufällige aber feste Permutation der Probanden wird in 5 Gruppen unterteilt, sodass jede 6 Probanden (bzw. 5 in der letzten) enthält. Jede Gruppe wird einmal als Testgruppe verwendet, während die jeweils restlichen 4 Gruppen Trainings- und Validierungsdaten im Verhältnis 3:1 bereitstellen.

chen Rückbildungseffekte) entspricht diese Läsion schließlich dem finalen, dauerhaften Gewebestatus, der im Verlaufsbild zu einem späteren Zeitpunkt beobachtet werden kann. Die Behandlung wurde dabei stets innerhalb von 10 Stunden nach Symptombeginn abgeschlossen, d. h. der Datensatz enthält nur frühe und mittlere Zeitfenster, aber keine späteren bis zu 24 Stunden (siehe Tab. 2.4). Da im Training keine Stichproben für den späteren Zeitraum der Akutphase vorliegen, wird eine maximale Zeit von 10 Stunden zur Normierung der Methode in Kapitel 6 verwendet.

Zur Evaluation auf dem UKSH-Datensatz wird in verschiedenen Kapiteln eine 5-fache Kreuzvalidierung (*engl.* 5-Fold Cross Validation) wie in Abbildung 2.12 durchgeführt. Dies stellt nach James u. a. [2013] erfahrungsgemäß eine sinnvolle Abwägung zwischen Bias und Korrelation der Testsätze dar (vgl. auch Isensee u. a. [2019]). Dabei wird jeweils ein Viertel jeder Trainingsstichprobe zur Validierung verwendet, sodass sich 17 bzw. 18 Trainingsfälle und 6 Validierungsfälle ergeben.

Eine Überanpassung von Modellen im Training wird durch frühzeitigen Abbruch verhindert, sobald der Fehler für die Schätzung oder Prädiktion auf den Validierungsfällen nicht mehr kleiner wird. Das Modell mit dem geringsten Validierungsfehler wird anschließend zum Testen verwendet. Da das Modell an nochmals komplett anderen Probanden getestet wird (Gruppe „TEST“ in Abb. 2.12), die nicht für das Training oder die Validierung verwendet wurden, vermeidet man eine Optimierung der Ergebnisse auf den Probanden der Validierungsgruppe.

Zur weiteren Steigerung der Robustheit des Trainings wurden die Modelle in dieser Arbeit mittels „Adam“-Gradientenabstieg von Kingma u. a. [2015] optimiert, wobei typischerweise eine Lernrate von 0,0001 sowie $\beta_1 = 0,99$ und $\beta_2 = 0,999$ für die Momente verwendet wurde. Aus medizinischer Sicht generell plausibel ist auch die in jedem Durchlauf durchgeführte zufällige Spiegelung an der Hirnmittellinie, um Präferenzen der Lernalgorithmen für eine der beiden Hemisphären entgegenzuwirken. Die Größe der Trainingsmenge wurde immer so gewählt, dass die 11 GB Speicher zur Berechnung auf der Grafikkarte ausgefüllt wurden und betrug mindestens vier Stichproben.

3 Methodischer Hintergrund

Das maschinelle Sehen (*engl.* Computer Vision) in der Medizin basiert auf Techniken der medizinischen Bildverarbeitung, woraus ausgewählte Grundlagen für die in dieser Arbeit genutzten Methoden in Abschnitt 3.1 kurz erläutert werden. Die Entwicklung des maschinellen Lernens in Abschnitt 3.2 eröffnet neue Möglichkeiten für die in Kapitel 2 erläuterte klinische Anwendung, bei einem Schlaganfall multivariate Bildgebung als Grundlage für die Therapieentscheidungen einzusetzen. Jene neuen Ansätze auf Basis von Deep Learning (Abschnitt 3.3) sind Gegenstand dieser Arbeit und werden in den Kapiteln 4–7 besprochen.

3.1 Grundlagen aus der Bildverarbeitung

In der (medizinischen) Bildverarbeitung gibt es eine typische Abfolge an Schritten, an deren Stelle auch maschinelle Lernverfahren verwendet werden können. Bereits während der Bildgebung können neuronale Netze zur Bildrekonstruktion eingesetzt werden. Elementare Vorverarbeitungsschritte beinhalten etwa das Ausblenden des Schädels bei Hirnaufnahmen (*engl.* Skull Stripping) oder Normalisierung der Intensitätswerte bei MRT. Gegebenenfalls müssen die Bilder auch registriert werden, bevor typischerweise Klassifikations-, Detektions- oder Segmentierungsaufgaben ausgeführt werden.

Die Repräsentation der Eingabe ist nicht immer geeignet, um die komplexen Muster zu erkennen, welche zum Beispiel einen Tumor kennzeichnen und z. B. segmentiert oder detektiert werden können. Deshalb müssen oft zuerst Merkmale (*engl.* Features), wie Kanten im Bild, etwa durch Filterung extrahiert oder durch Überführung der Eingabe in höhere Dimensionalität mittels „Kernel-Trick“ bzw. Dimensionsreduktion, wie in der PCA (Abschnitt 3.2.1), neue Repräsentationen der Eingabe erzeugt werden. In Bezug auf bestimmte Homogenitätskriterien der Merkmale lassen sich dann beispielsweise Regionen definieren, welche schließlich klassifiziert werden können und so eine Segmentierung ergeben. Beim Deep Learning spielt das tiefe Hintereinschalten von aufeinander aufbauenden Repräsentationen (*engl.* Feature Hierarchy) noch eine entscheidende Rolle, da erst mit steigender Länge (daher „Deep“) die hohe Genauigkeit der neuronalen Netzwerke erreicht wird.

Die Filterung ist eine entscheidende Möglichkeit, um bessere Daten- und Bildrepräsentationen zu erhalten. Das Wort „Filter“ bezieht sich ursprünglich auf den Frequenzbereich, in dem niedrige (*engl.* Highpass) oder hohe Frequenzen (*engl.* Lowpass) herausgefiltert werden können. Analog zu den Filterungen im Frequenzbereich (z. B. Lowpass

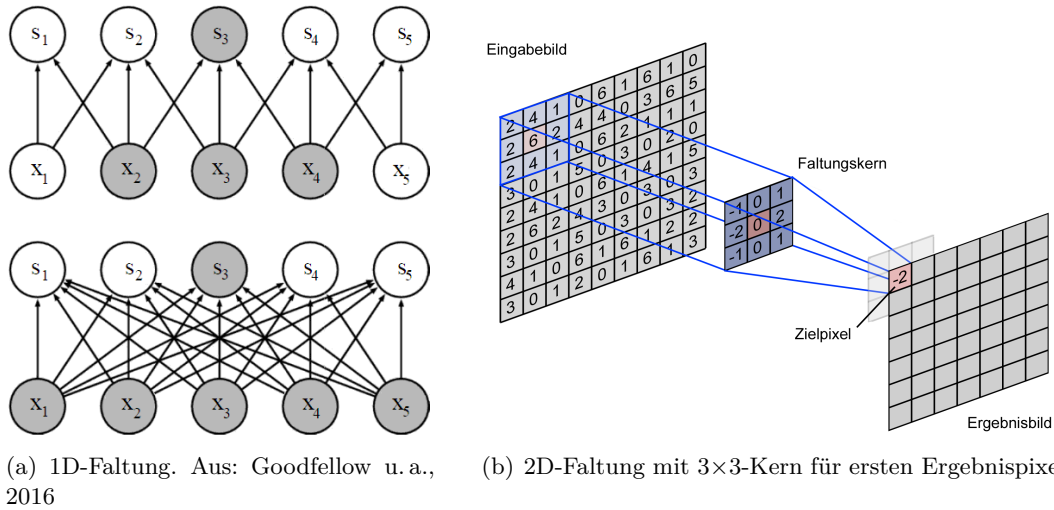


Abb. 3.1: Faltung auf einer ausgewählten Ausgabeeinheit, wie sie für alle weiteren Ausgabeeinheiten via „Sliding Window“ analog durchgeführt wird. (a) zeigt ein Beispiel in 1D mit 3er-Kern: Im Vergleich zur vollverbundenen Eingabe pro Bildpunkt (unten) gibt es bei der Faltung (oben) nur eine spärliche Verbindung der Ausgabeeinheit s_3 zu x_2 , x_3 und x_4 der Eingabe. (b) zeigt ein Rechenbeispiel in 2D mit 3×3 großem Kern für den ersten Pixel ohne zusätzliche Randbehandlung.

zum Weichzeichnen) kann man auch räumliche Filter im Bildbereich einsetzen (z. B. Gauß’scher Weichzeichner), welche über räumliche Filterkerne umgesetzt werden. Das Arbeiten im Bildbereich ermöglicht darüber hinaus auch nichtlineare Filterung, wie Median- oder Maximumsberechnung, welche nicht direkt mittels äquivalenten Operationen im Frequenzbereich berechnet werden können, da die Fourier-Transformation nur für Linearoperationen anwendbar ist. [Gonzalez u. a., 2008]

3.1.1 Faltungsoperation (Convolution)

Räumliche Filterung errechnet für ein Pixel oder Voxel mit einer bestimmten Operation über die im Filterkern definierte Nachbarschaft einen neuen Wert des Pixels bzw. Voxels. Viele Filter werden über die sogenannte Faltungsoperation mit Operator „*“ für den zeitdiskreten Anwendungsfall zweier Funktionen I (Eingabebild) und K (Kern) definiert. Für 2D-Pixelpositionen (i, j) in einem beschränkten Definitionsbereich – außerhalb dieses werden Nullwerte angenommen – wird der Faltungskern K über die Bildpunkte von I im sogenannten „Sliding-Window“-Verfahren geschoben (Abb. 3.1(b)) und ergibt die Ausgabe O :

$$O(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \quad (3.1)$$

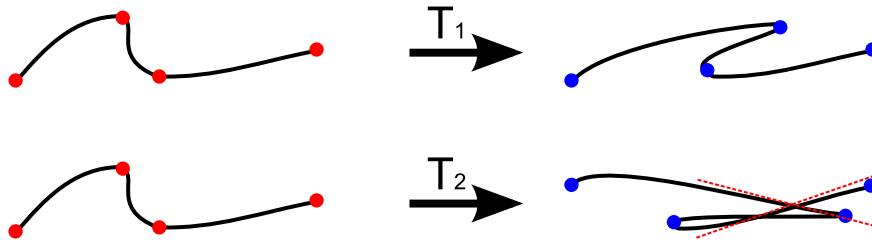


Abb. 3.2: Bijektivität beim Diffeomorphismus: Transformation T_1 ist eine Bijektion und invertierbar. Für T_2 existieren Stellen, an denen T_2^{-1} nicht differenzierbar ist (Kreuzung der roten gepunkteten Linien).

Faltungsoperationen sind linear und für sie gilt Assoziativität, Kommutativität sowie Distributivität. Zwar gibt es keine Translationsinvarianz, aber Translationsäquivalenz: Verschiebt sich die Eingabe eines Faltungsfilters, so wird sich äquivalent auch die Filterantwort verschieben. Dies ist eine in Bezug auf die Merkmalshierarchie der CNNs in Kapitel 3.3.1 wichtige Eigenschaft. Faltungen können zudem mit Translationen kommutieren, sodass die Ausgabe beider Operationen unabhängig von der Reihenfolge ihrer Anwendung gleich bleibt. [Goodfellow u. a., 2016]

Die Faltungsoperation hat neben der Translationsäquivalenz noch zwei weitere wichtige Eigenschaften: Zum einen deckt der Faltungskern nur einen kleinen Ausschnitt des Eingabebildes ab (dünnbesetzt, *engl.* sparsely connected, im Zusammenhang neuronaler Netze in Abschnitt 3.3.1), da der Kern in der Bildverarbeitung üblicherweise deutlich kleiner als das Eingabebild gewählt ist (Abb. 3.1(a)). Damit wird die Ausgabe durch die Gesamtheit aller lokalen Nachbarschaften um die einzelnen Eingabepixel bestimmt, sodass lokale Änderungen auch nur lokale Auswirkungen in der Ausgabe haben. Zum anderen teilen sich alle Positionen die Parameter, d. h. jene Gewichtung des Faltungskerns, welche an allen Bildpunkten gleich ist. Der Speicherbedarf ist damit geringer als bei vollverbundenen (*engl.* fully-connected) Ausgabeeinheiten. Diese Eigenschaften sind ebenfalls für die CNNs in Kapitel 3.3.1 von Bedeutung.

3.1.2 Diffeomorphismen

Im biologischen oder medizinischen Bildverarbeitungskontext ist es wichtig, bei jeder Art der räumlichen Verformung eines Gewebes einige grundlegende Annahmen zu berücksichtigen. Physikalisch plausibel sind dabei keine abrupten oder harten Veränderungen, sondern weiche Übergänge. Insbesondere sollen sich nicht einfach „Löcher“ innerhalb eines Gewebes auftun (sondern höchstens durch anderes Gewebe dazwischen) sowie das Gewebe in sich eine homogene Masse bleiben, die nicht einfach abrupt von benachbartem Gewebe durchstoßen wird.

Im Räumlichen haben lokale Dehnung oder Verschiebung von Gewebegrenzen unter Annahme glatter Oberflächen immer Auswirkungen auf das benachbarte (eigene oder

fremde) Gewebe. Eine elastische Verformung bei der Modellierung biologischer Strukturen sollte daher bestimmte Bedingungen erfüllen – mathematisch meistens jene eines Diffeomorphismus. Die Bijektion zwischen zwei glatten Mannigfaltigkeiten definiert diesen als eine Eins-zu-Eins-Abbildung $f : M \rightarrow N$ mit folgenden Eigenschaften:

- M, N sind differenzierbare Mannigfaltigkeiten (z. B. Domäne $\Omega \subseteq R^3$)
- f ist kontinuierlich differenzierbar
- f^{-1} existiert und ist auch kontinuierlich differenzierbar (Bijektivität, Abb. 3.2)

Wenn $f(M) = N$ gilt, sind M und N diffeomorph. In Bezug auf eine Transformation Θ der Registrierung im Vorverarbeitungsschritt von Kapitel 7 sollte $f(\Theta) = \Theta(\Omega)$ für eine diffeomorphe Deformation gelten.

3.1.3 Metriken

Metriken dienen vor allem als Maße für die Evaluation einer (automatischen) Methode, sind aber auch häufig beim Training einer Lernmethode in einer Fehlerfunktion in Gebrauch. Die hauptsächlich in dieser Arbeit angewandten Metriken werden an dieser Stelle kurz erläutert.

Die ROC-Kurve (*engl.* Receiver Operating Characteristics) kann zwei Arten von Fehlern für verschiedene Schwellenwerte darstellen [James u. a., 2013]. Für einen Binärklassifizierer ergeben zu verschiedenen Schwellenwerten, die seine Entscheidungsgrenze verschieben und damit die Richtig-Positiv- wie auch Falsch-Positiv-Rate beeinflussen, die aufgetragene Sensitivität und Spezifität eine ROC-Kurve (z. B. auf Seite 61). Die Fläche darunter (*engl.* Area Under Curve, AUC) gilt als ein Gütemaß für die Robustheit eines Klassifizierers in Bezug auf die Wahl der Schwellenwerte, wobei ein Wert von 0,5 dem Zufallsergebnis entspricht und ein Wert von 1 ideal ist.

Für Segmentierungsaufgaben wird oft eine aussagekräftigere Metrik verwendet, welche Genauigkeit und Trefferquote einer Segmentierung gegenüber einer Referenz zusammen berücksichtigt. Der Dice-Koeffizient gibt die relative Überlappung von binären Segmentierungsmasken an, wobei A und B die Menge aller Voxel innerhalb der wahren Segmentierung bzw. geschätzten Segmentierung umfassen. Er ist definiert als:

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (3.2)$$

Für weiche (*engl.* fuzzy, soft) Segmentierungen benötigt man jedoch eine andere Definition mittels *SoftDice*, die entsprechend mit den in den Segmentierungen kodierten Konfidenzen eine voxelweise gewichtete Überlappung berechnen kann. Mit einer gegebenen Grundwahrheit y und Segmentierung \hat{y} über alle Voxel $i \in \Omega$ eines Schnittbilds bzw. 3D-Volumens Ω , sowie einem kleinen $\epsilon > 0$ zur numerischen Stabilität, definiert sich die Übereinstimmung als:

$$SoftDice(y, \hat{y}) = \frac{2\sum_i y_i \hat{y}_i + \epsilon}{\sum_i y_i y_i + \sum_i \hat{y}_i \hat{y}_i + \epsilon} \quad (3.3)$$

Im Gegensatz zu relativen Überlappungsmaßen können Distanzmaße auch den Extremfall einer nicht gegebenen Überlappung von Segmentierung zu Grundwahrheit quantifizieren (Überlappungsmaße reagieren empfindlich auf die Läsionsgröße). Damit sind sie jedoch weit empfindlicher gegenüber Ausreißern. Das hier verwendete Distanzmaß $d(\cdot)$ ist der euklidische Abstand, der unter Berücksichtigung der Voxelgröße des Volumens berechnet wird. Der durchschnittliche Oberflächenabstand (*engl.* Average Symmetric Surface Distance, ASSD) ist definiert als:

$$ASSD(A_S, B_S) = \frac{ASD(A_S, B_S) + ASD(B_S, A_S)}{2} \quad (3.4)$$

Er berechnet sich durch die Mittelung aus den beiden einzelnen gerichteten Distanzen (ASD) zwischen beiden Oberflächen:

$$ASD(A_S, B_S) = \frac{\sum_{a \in A_S} \min_{b \in B_S} d(a, b)}{|A_S|} \quad (3.5)$$

Hierbei bezeichnen A_S und B_S ausschließlich die Voxel der Oberflächen (*engl.* Surface) der Segmentierung und der Grundwahrheit. Folgerichtig lässt sich darüber hinaus auch das Maximum aller Oberflächenabstände ermitteln, welches die Hausdorffdistanz (HD) definiert:

$$HD(A_S, B_S) = \max \left\{ \max_{a \in A_S} \min_{b \in B_S} d(a, b), \max_{b \in B_S} \min_{a \in A_S} d(a, b) \right\} \quad (3.6)$$

3.2 Maschinelles Lernen

Bereits Mitte des letzten Jahrhunderts wurde sich mit der Forschungsrichtung der Künstlichen Intelligenz (KI) auseinandergesetzt, für die John McCarthy, Marvin Minsky, Nathaniel Rochester und Claude Shannon auf der *Dartmouth Summer Conference on Artificial Intelligence* im Jahre 1956 den einheitlichen Begriff der „Künstlichen Intelligenz“ (*engl.* Artificial Intelligence, AI) etablierten, welcher fortan dennoch häufig äußerst vieldeutig verwendet werden sollte. Das liegt vor allem daran, dass eine einheitliche Definition von (menschlicher) Intelligenz schwierig ist. Verschiedene methodische Richtungen innerhalb der KI machen die Einordnung darüber hinaus nicht leichter – so kann sich hinter dem aktuellen Modewort eine große Vielfalt an Methoden verbergen.

In Bezug auf die Anwendung in der Medizin kann jedoch folgendes konkretisiert werden: Es handelt sich nicht um „starke“ KI, welche versucht menschliche Intelligenz generell nachzustellen, sondern um Methoden, welche eng gefasste, spezifische Aufgaben so gut wie oder gar besser als der Mensch lösen können. Entweder passiert dies bei

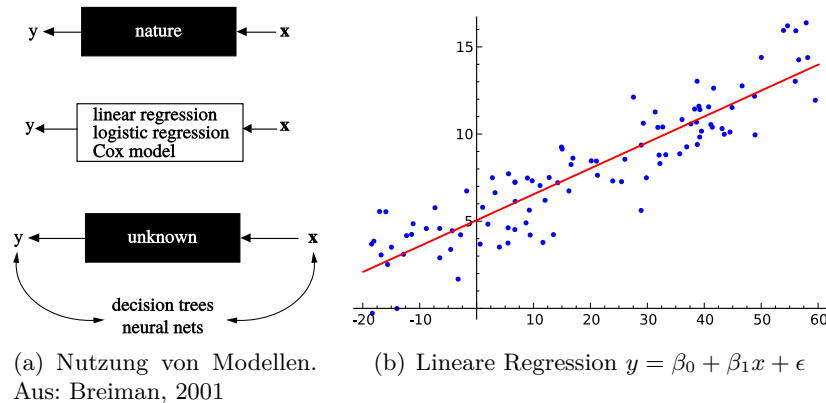


Abb. 3.3: Die realen Zusammenhänge in der Natur sind uns nicht ausreichend bekannt, um diese konkret abbilden zu können, wie bei der „Black Box“ in (a). Stattdessen nimmt man in der Statistik vereinfachte Modelle (a, Mitte), deren Parameter mittels Stichproben (b, blaue Punkte) einer wahren Verteilung geschätzt werden. Für das lineare Modell in (b) wurden $\beta_0 = 5$ und $\beta_1 = 1/6$ für dessen Regressionsfit der Gerade (rot) geschätzt. Im maschinellen Lernen werden Parameter meist komplexerer Modelle stattdessen unter definierten Nebenbedingungen algorithmisch trainiert.

äußerst komplexen Aufgabenstellungen (z. B. Diagnosesysteme) über wissensbasierte Regelsysteme, welche eine große Wissensbasis sowie ein Inferenzsystem zur Anwendung des Wissens besitzen. Für den Einsatz in Aufgabenfeldern wie Bildverarbeitung, Robotik, Sprachverarbeitung oder automatisierter Behandlungsplanung wird häufig auf den zweiten großen Bereich der KI zurückgegriffen. Das „maschinelle Lernen“ wird momentan äußerst populär beforscht und daher auch häufig synonym zum Begriff „Künstlicher Intelligenz“ benutzt. Im Gegensatz zu den wissensbasierten Systemen muss beim maschinellen Lernen nicht zwangsläufig Wissen deklariert und vorgegeben werden, sondern es kann auf Basis von Mustererkennung in Daten mittels „unüberwachtem Lernen“ zur Lösung einer Aufgabe gewonnen werden. [Wahl u. a., 2018]

Maschinelles Lernen (ML) steht zwischen den Disziplinen der Informatik und der Statistik. Dabei geht es um das Lernen anhand von Beispielen, ohne vorzugeben, wie das zugrundeliegende Modell aussehen soll. Im Rahmen der evidenzbasierten Medizin (EBM) kommt oft die klassische Statistik zur Anwendung. Dort wird davon ausgegangen, dass sich der Realität über ein stochastisches Datenmodell $y = f(x)$ mit vereinfachten Annahmen (systematische Verzerrung) angenähert werden kann. Die vorgegebene Funktion bildet die unabhängigen auf die abhängigen Variablen ab, wie in Abbildung 3.3 abstrahiert. Dafür werden die entsprechenden Parameter der Funktion f über Stichproben einer statistischen Verteilung geschätzt und diese dabei unter zufällig erscheinenden Abweichungen (Rauschen) in z. B. ein lineares Regressionsmodell, wie in Abbildung 3.3(b), eingepasst.

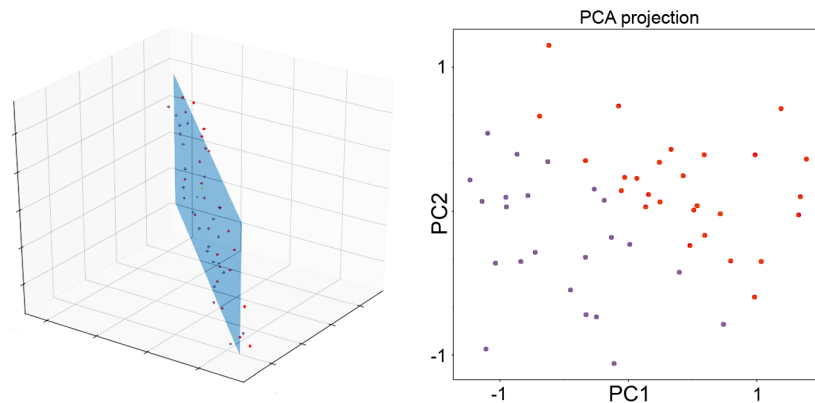


Abb. 3.4: 50 Stichproben in 3D sind links mit der 2D-Hyperebene abgebildet, welche die Summe der quadratischen Abweichung minimiert. Rechts ist die Projektion der 50 Datenpunkte auf diese Ebene mit den ersten beiden Komponenten der PCA als Achsen abgebildet (*engl.* „scores“). Die Varianz ist in der Ebene maximiert und die Datenpunkte lassen sich mit diesen zwei Dimensionen bereits gut mittels einfachster Klassifizierer (z. B. Schwellenwert) in ihre zwei Klassen linear separieren.

Breiman [2001] stellt dem gegenüber den algorithmischen Ansatz, der im Voraus von einer komplexen und unbekanntem Funktion f ausgeht. Die Algorithmen arbeiten ebenso mit denselben Daten, aber das Lernen einer Parametrisierung behält die *Black Box* bei, statt stark zu vereinfachen. Im Gegensatz zur Regularisierung über den „Bias“ des statistischen Modells wird hierbei der Optimierungsprozess zum Finden („Lernen“) der richtigen Parameter regularisiert und damit die Mächtigkeit des eigentlichen Modells nicht per se eingeschränkt. Dies ist insbesondere für hochkomplexe Probleme, wie Sprachverarbeitung oder Bildinterpretation, notwendig. Der Grundkonflikt statistischen Lernens besteht darin, dass einfachere Modelle besser zu interpretieren und robuster anzuwenden sind (und daher traditionell nach „Occam’s Razor“ zu bevorzugen), aber durch den damit einhergehenden strukturellen Fehler nicht so genau prädizieren. Dieses Problem wird nun zum Teil vom Modell weg und hin zu den Eigenschaften des Trainings durch Formulierung des Lernalgorithmus’ verschoben.

3.2.1 Hauptkomponentenanalyse (PCA)

Die Hauptkomponentenanalyse (*engl.* Principal Component Analysis, PCA) ist eine unüberwachte, generische¹¹ Methode zur Dimensionsreduktion eines Datensatzes. In Bezug auf die Tatsache, dass m statt p Parameter mit $m < p$ leichter zu schätzen sind, z. B. die Koeffizienten eines linearen Modells wie in Abb. 3.3(b), erklärt sich auch die Frage nach dem Nutzen dimensionsreduzierter Repräsentationen [James u. a., 2013]. Es handelt sich dabei um ein statistisches Lernverfahren für Repräsentationen, wobei der

¹¹Generisch: Unabhängig von der eigentliche Aufgabe, wie z. B. Klassifikation der Daten.

Algorithmus vorgegeben ist und keine zufällige Initialisierung Einfluss auf das Ergebnis hat. Die Ergebnisrepräsentation kann danach von maschinellen Lernmethoden weiter verarbeitet werden.

Statt die ursprünglichen Prädiktoren bzw. Variablen für eine Klassifikation zu nutzen, werden diese linear transformiert und in eine neue Orthogonalbasis überführt, die einen Raum von möglichst unabhängigen Merkmalsdimensionen aufspannt (Abb. 3.4). Oft sind die Eingabedimensionen abhängig voneinander und stellen demzufolge redundante Informationen dar.

Zur Berechnung der PCA muss ein Datensatz $X \in \mathbb{R}^{p \times n}$ mit p Variablen und n Stichproben mittelwertbereinigt vorliegen (im Koordinatenursprung zentriert). Daraufhin erfolgt die Berechnung der Kovarianzen in der symmetrischen Matrix $C = \langle X, X^T \rangle$. Auf dieser werden die Eigenwerte und -vektoren mit Gleichung $U^T C U = D$ berechnet. Dabei ist D die Diagonalmatrix, welche auf der Diagonalen die Eigenwerte λ von C und U in den Zeilen die Eigenvektoren von C enthält, wobei $U^T U = U U^T = 1$ gilt. Mittels der Transformationsmatrix U lässt sich der Datensatz X über $Z = U^T X$ in die neue Basis mit $Z \in \mathbb{R}^{m \times n}$ überführen.

Die berechneten Hauptkomponenten einer PCA stellen eine Linearkombination der ursprünglichen Eingabevariablen dar, wobei sie mit steigender Ordnungsnummer immer weniger Varianz aufweisen, d. h. die erste Hauptkomponente hat die größte Varianz. Die Eigenwerte λ quantifizieren diese Varianz. Die Eigenvektoren in U sind die Koeffizienten („loadings“) für die Linearkombination der ursprünglichen Variablen in X zur Berechnung von Z („scores“). Die Rekonstruktion der Daten mit einem möglichst geringen Fehler (abhängig von der Anzahl der verwendeten Hauptkomponenten) erfolgt schließlich über die inverse Transformationsmatrix $ZU = X$.

3.2.2 Support Vector Machines (SVM)

Mittels einer Hyperebene können Daten in einem Merkmalsraum in zwei Klassen unterteilt werden (z. B. rote oder grüne Gerade in Abb. 3.5(b)). Statt einer optimal trennenden, aber nur beliebig platzierten Hyperebene in einem Trainingsdatensatz, können sogenannte Stützvektoren (*engl.* Support Vectors) aus zwei Klassen in einem gleichen Abstand zur partitionierenden Hyperebene die Entscheidungsgrenze zwischen diesen Klassen definieren [James u. a., 2013]. So liegt diese Ebene in möglichst maximalem Abstand von jeder der Klassen entfernt, was die Generalisierung gegenüber einer beliebig platzierten Trennebene erhöht.

Durch Definition eines „soft margin“ können einige Ausreißer im Training trotzdem auf der „falschen“ Seite der Ebene liegen, denn nicht immer sind Daten (z. B. durch Rauschen) linear separierbar. Wie bei einer „Kernel-PCA“ kann man aber auch bei solchen „Max-Margin“-Klassifizierern die Entscheidungsgrenze im Merkmalsraum nicht-linear gestalten, falls die Daten nicht nur rauschbehaftet sind, sondern das Problem grundsätzlich nichtlinearer Natur. Der Merkmalsraum wird dafür durch Nutzung eines

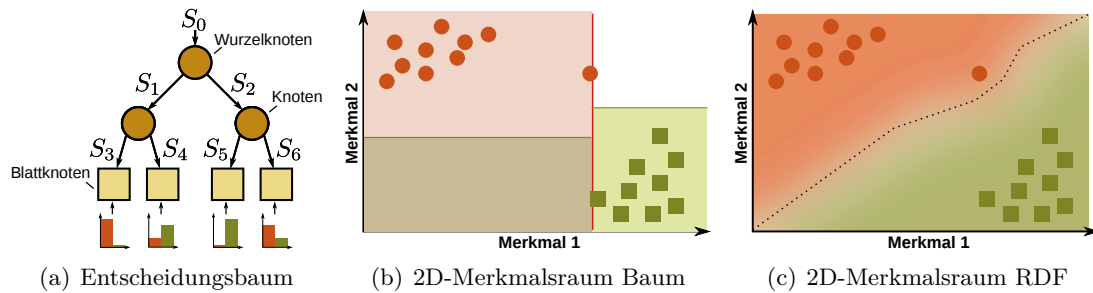


Abb. 3.5: Leicht angepasste Darstellung nach Maier [2017]: In (a) ist ein Entscheidungsbaum mit den Knoten dargestellt, an denen anhand eines Optimierungskriteriums oder via Zufallswahl („Extra Tree Forest“) ein Schwellenwert trainiert wird. Die Blätter repräsentieren eine bestimmte Klassenverteilung. Ein einzelner Baum bildet harte, lineare Entscheidungsgrenzen im Merkmalsraum (b). Durch Ensembling vieler zufälliger Bäume entsteht ein generalisierender, weicher Merkmalsraum (c) für die Klassenaufteilung mit ablesbarer nichtlinearer Entscheidungsgrenze (gestrichelt).

Kernels vergrößert (sogenannter „Kernel-Trick“), etwa mittels radialer Basisfunktion, welche die Datenpunkte vom Merkmalsraum in einen unendlichen Z-Raum überführt, in dem sich bessere Hyperebenen zur Trennung der Klassen finden lassen. Dies ist der essentielle Bestandteil einer SVM und die entscheidende Erweiterung gegenüber einem einfachen Max-Margin-Klassifizierer. Durch Nutzung mehrerer Hyperebenen ist es schließlich möglich, mehr als nur zwei Klassen zu unterscheiden.

3.2.3 Random Forests (RDF)

Bevor Deep Learning (Abschnitt 3.3) in den letzten Jahren immer populärer wurde [Krizhevsky u. a., 2012; Ravì u. a., 2017; Schmidhuber, 2015], hatte man im maschinellen Sehen für die Medizin sogenannte Entscheidungsbäume und deren randomisierten Ensembles (*engl.* Random Decision Forest, RDF) vermehrt untersucht [Maier u. a., 2015b; Mitra u. a., 2014; Wall u. a., 2012]. Diese lassen sich auf vorberechneten Merkmalen recht einfach trainieren und die Genauigkeit linearer statistischer Modelle oder auch einfacher Boosting-Ansätze haben die nichtlinearen Forests meistens übertroffen (siehe Seite 54). Der Nachteil, explizite Merkmale zu definieren oder separat Merkmalsrepräsentationen lernen zu müssen, die daraufhin jedes Mal zunächst aus den Bilddaten extrahiert werden, bleibt jedoch nach wie vor bestehen. Die neuronalen Netze in Abschnitt 3.3.1 sind dagegen in der Lage, beides gleichzeitig zu trainieren und anzuwenden – die Repräsentation und die Klassifizierung (oder Regression).

Ein Boosting-Verfahren ist eine Methode, die aus vielen einfachen und schwachen Klassifizierern einen mächtigeren formt (siehe Vergleich in Kapitel 3.4.1). In diesem Sinne stellen auch RDFs, welche mit bedingten Abfolgen einfacher Schwellenwertvergleiche an den Knoten der unterschiedlichen Entscheidungsbäume arbeiten, einen

Boosting-Ansatz dar. Durch Mehrheitsvotum der prädizierten Klassen der erreichten Blätter in den Bäumen wird die finale Klasse geschätzt. Erst durch ein Ensemble von Entscheidungsbäumen in einem „Random Decision Forest“, welche mit zufälligen Modifikationen auf den Daten gelernt werden, ergibt sich die Mächtigkeit bei gleichzeitiger Einfachheit und Interpretierbarkeit. Typischer Hyperparameter ist die maximale Tiefe der Bäume zur Steuerung von Überanpassung versus Generalisierbarkeit (Bias). Hinzu kommt eine Mindestanzahl von Bäumen im Forest, die für eine hohe Klassifikationsgenauigkeit bei möglichst vermeidbarer Redundanz der Bäume benötigt wird. Im Gegensatz zu SVMs sind RDFs recht effizient zu berechnen und aufgrund der wenigen Parameter leicht zu optimieren [Maier u. a., 2015a].

Technisch wird durch die Art des Trainings der RDFs deren Zufälligkeit („Random“) als maschinelles Lernverfahren in den Aufbau eingebracht. Anhand des Bagging-Konzepts wird eine zufällige Teilmenge der Daten für das Training jedes einzelnen Baumes neu gezogen. Daraufhin wird für jeden Knoten immer eine neue, zufällige Teilmenge an Merkmalen zu Ermittlung des Schwellenwertkriteriums berücksichtigt. Bei den „Extra Tree Forests“ wird sogar der eigentliche Schwellenwert nicht nach Informationsgehalt der entstehenden zwei Verteilungen an einem Knoten gewählt, sondern zufällig erzeugt. Durch das Bagging haben Ausreißer weniger Einfluss als etwa bei SVMs. Die zufällige Merkmalsauswahl macht den Forest robuster gegenüber schlechten Merkmalen und die Entscheidungsgrenzen werden weicher (bessere Generalisierbarkeit), was auch durch die Anzahl der Bäume im Forest beeinflusst wird.

3.3 Deep Learning mit künstlichen neuronalen Netzen

Deep Learning beschreibt die Nutzung künstlicher neuronaler Netzwerke für das maschinelle Lernen, welche besonders viele Zwischenschichten enthalten. Im Sinne einer Merkmals-hierarchie sollen sie besonders tiefgehende Konzepte lernen können, um so als Anwendung in der Medizin herausfordernde und komplexe Probleme computerunterstützt lösen zu können. Es können Netzwerke spätestens ab einer Tiefe der Größenordnung 10^1 als *deep* angesehen werden, wobei diese Attributierung aufgrund der Abhängigkeit vom zu lösenden Problem (und dessen Komplexität oder „Schwierigkeit“) nur wenig praktischen Nutzen besitzt [Schmidhuber, 2015]. Hinzu kommt, dass mit steigender Komplexität des Netzwerks – bei steigender Anzahl der Schichten – auch die benötigte Anzahl an Trainingsdaten zunimmt, diese aber ebenfalls stark vom Problem abhängig ist.

Künstliche neuronale Netze (*engl.* Artificial Neural Networks, ANN) sind ursprünglich durch ihre biologischen Pendanten, welche insbesondere im menschlichen Kortex in Schichten angeordnet sind, motiviert. Eine Reihe von Eingabeverbindungen (ähnlich den Dendriten) bildet eine gewichtete Summe, welche addiert mit einem eingabeunabhängigen Bias in eine Aktivierungsfunktion (z. B. Schwellenwert wie beim Aktivie-

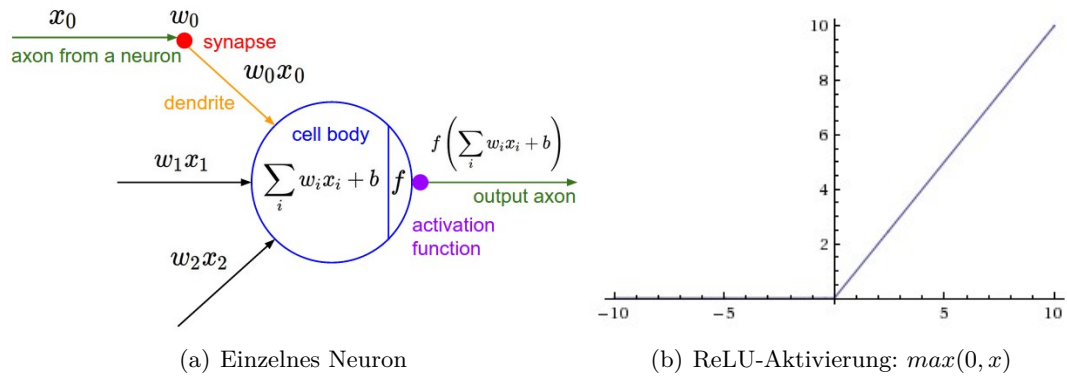


Abb. 3.6: Das Perzeptron ist ein mathematisches Modell in Anlehnung eines biologischen Neurons (a). Durch die gewichtete Summation der Eingaben wird abhängig von der Aktivierungsfunktion, z. B. der ReLU (b), eine Aktivierung in einem bestimmten Wertebereich ausgegeben. Aus: <http://CS231n.github.io/neural-networks-1> (2019)

rungspotenzial in Abb. 3.6) gegeben wird, wodurch eine entsprechende Ausgabe in die nächste Neuronenschicht weitergereicht wird (ähnlich einem Axon). Eine solche mathematische Konstruktion heißt Perzeptron (Abb. 3.6), und auch dieses kann mehrfach innerhalb einer und mehrerer Schichten strukturiert werden (*engl.* Multi Layer Perceptron, MLP). Mit der Schwellenwert-Aktivierung auf Ausgabe $[0, 1]$ ergibt sich ein einfacher binärer Klassifikator.

Die Repräsentation der Eingaben (und Ausgaben) der Schichten als Vektoren ist üblich und damit die Berechnung der Summe mit Punktprodukt möglich. Daher werden die Vorwärtsläufe von Ein- zu Ausgabe neuronaler Netze über Matrixmultiplikationen berechnet, welche sich sehr gut auf dafür optimierter Grafikkhardware berechnen lassen.

Hornik u. a. [1989] haben gezeigt, dass selbst einfache künstliche neuronale Netze sehr mächtig sein können und bereits ab einer Zwischenschicht (*engl.* Hidden Layer) theoretisch in der Lage sind, kontinuierliche Funktionen bis zu einer bestimmten Güte zu approximieren. Wenn Abbildung 3.7 von oben nach unten gelesen wird, kann intuitiv nachvollzogen werden, dass bereits wenige Schichten ausreichen, um ein generisches Modell zu formen, welches arbiträre Regionenklassifikationen auf Merkmalsvektoren als Eingabe vornehmen kann. Die Schwierigkeit liegt jedoch im Finden der nötigen korrekten Parametrisierung, ergo: dem Training eines solchen Netzwerkes. In der ersten Zeile in Abb. 3.7 modelliert ein Neuron zunächst eine einzelne Hyperebene als Entscheidungsgrenze im 2D-Merkmalsraum (zwei Eingabedimensionen), wobei eine Seite als zugehörig und die andere als nicht zugehörig zu einer Zielklasse definiert ist. Durch Hinzunahme eines weiteren Neurons (grün) ergibt sich eine weitere Hyperebene. Erst mit einem weiteren Neuron in einer weiteren Schicht (zweite Zeile) kann man jedoch die UND-Verknüpfung beider modellieren und die durch diese Entscheidungsgrenzen umrissene Region definieren. Durch Hinzufügen weiterer Neuronen in der

3 Methodischer Hintergrund


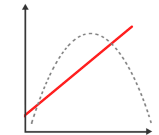
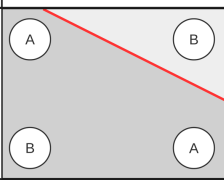
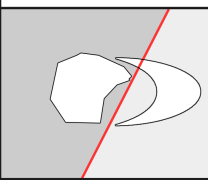
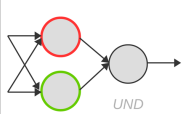
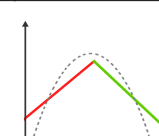
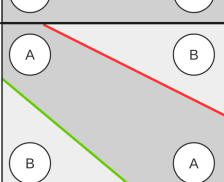
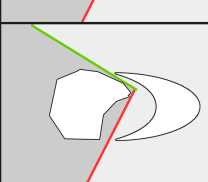
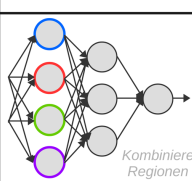
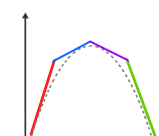
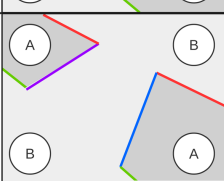
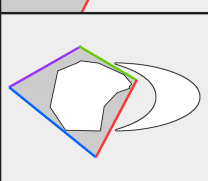
		Approximieren von 1D-Funktionen	XOR Problem auf Merkmalspunkten	Klassen von (Ober)flächen-Regionen
1-Schicht NN Einzelne Hyperebene				
2-Schicht NN (1 hidden layer) Offene/Geschlossene konvexe Regionen				
3-Schicht NN (2 hidden layers) Arbiträre Regionen				

Abb. 3.7: Es ist leicht nachzuvollziehen, wie ein neuronales Netz mit drei Schichten beliebige Regionen in 2D bis zu einer bestimmten Güte (abhängig von der Anzahl Neuronen) modellieren und klassifizieren könnte.

Ausgabeschicht könnten noch weitere Regionen klassifiziert werden. In der beispielhaften Vorstellung einer Hirnkrankheit mit auffälligen hyperdensen Regionen einer CT, könnte z. B. durch eine dritte Schicht anhand der klassifizierten Regionen wiederum die Krankheit bestimmt werden, welche durch die einzelnen Krankheitsherde definiert ist, wie etwa Multiple Sklerose oder akute und chronische Schlaganfallläsionen.

Das lineare Gleichungssystem eines jeden Neurons mit gewichteter Summe (Abb. 3.6) stellt zunächst einen einfachen linearen Klassifizierer dar. Selbst bei Anordnung dieser in Schichten, welche zahlreich hintereinander geschaltet werden, ergibt sich noch kein besonders gutes Werkzeug zur Klassifikation oder für weitergehende Aufgaben. Erst das Hinzufügen nichtlinearer Aktivierungsfunktionen verleiht neuronalen Netzen ihre außerordentliche Mächtigkeit. In Abbildung 3.8 ist dies für ein typisches Klassifikationsproblem in 2D dargestellt. Obgleich mit mehreren linearen Hyperebenen wie in Abbildung 3.7 beide Funktionen aus Abb. 3.8(a) getrennt werden können, erlaubt eine nichtlineare Transformation die Klassifikation beider Kurven in der resultierenden Repräsentation mit nur genau einer Hyperebene (ergo Neuron) in Abb. 3.8(c).

Beim Maschinellen Lernen mit Neuronalen Netzen handelt es sich um ein Optimierungsproblem: Die Parameter des neuronalen Netzwerks werden üblicherweise mit einer Variante des stochastischen Gradientenabstiegs (*engl.* Stochastic Gradient Descent, SGD, Abb. 3.9) optimiert, um eine Fehlerfunktion (*engl.* Loss, L) zu minimieren (beim überwachten Lernen die Abweichung der Vorhersage zur tatsächlichen Grund-

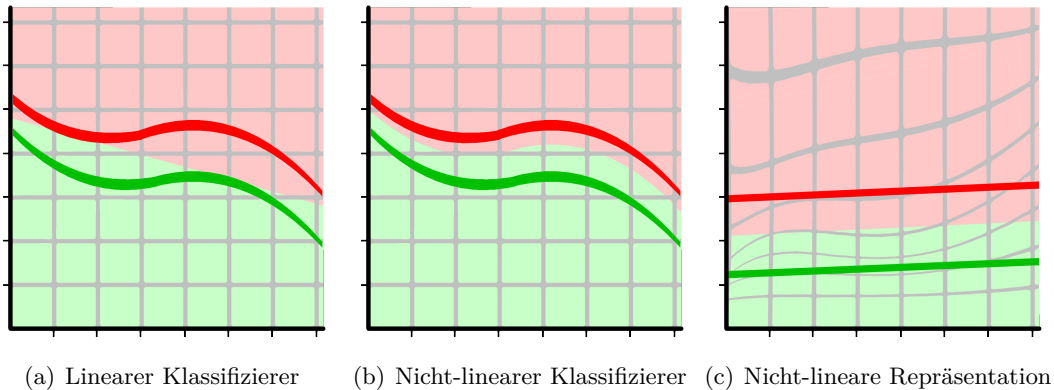


Abb. 3.8: Lineare Separierbarkeit mittels nichtlinearer höher-dimensionaler Repräsentation.

wahrheit). Das Lernen der geeigneten Netzwerkgewichte w läuft nach folgendem Rahmenalgorithmus ab, der oftmals angepasst wird:

1. Initialisierung aller Gewichte, z.B. $\sim N(0, \sigma^2)$ (randomisiert)
2. Wahl einer zufälligen Teilmenge (*engl.* Batch) aus den Eingabedaten
3. Berechnung der Netzwerkausgabe mit Batch als Eingabe
4. Berechnung der Fehlerfunktion $L(\hat{y}, y)$, z. B. mittlere quadratische Abweichung
5. Stückweise Berechnung der Gradienten als partielle Ableitungen $\frac{\partial L(\hat{y}, y)}{\partial w}$ ausgehend vom Fehler mittels Kettenregel („Backpropagation“ nach Rumelhart u. a. [1986])
6. Aktualisierung der Gewichte mit Berechnungsregel $w' = w - \eta(\alpha \frac{R(w)}{w} + \frac{\partial L(\hat{y}, y)}{\partial w})$
7. Wiederholung von (2.) mit nächstem Batch; Falls alle Batches durch, führe (8.) aus
8. Sofern kein Terminierungskriterium zutrifft, Berechnung der nächsten Epoche durch Wiederholung der Schritte ab (2.)

Die Inferenz in Schritt (3.) wird als Vorwärtslauf (*engl.* Forward Pass) bezeichnet, wohingegen Schritt (5.) als Rückwärtslauf (*engl.* Backward Pass) bezeichnet wird. In Schritt (6.) ist R der Regularisierungsterm, etwa ein $L1$ oder $L2$ „Weight Decay“ über alle Gewichte, welcher mit α gewichtet wird und damit große absolute Gewichte vermeidet. Während des Lernprozesses führt dies dann seltener zu Problemen, wie etwa rapide ansteigenden Gradienten. Zweck und Art der Regularisierung können allerdings mannigfaltig sein und sind abhängig von der zu lernenden Aufgabe des Netzwerkes.

Die Lernrate wird mit η bezeichnet. In Abbildung 3.9 ist ein Fehlerfunktionsraum des Optimierers abgebildet, der sich von außen dem Minimum im Zentrum annähert. Ist

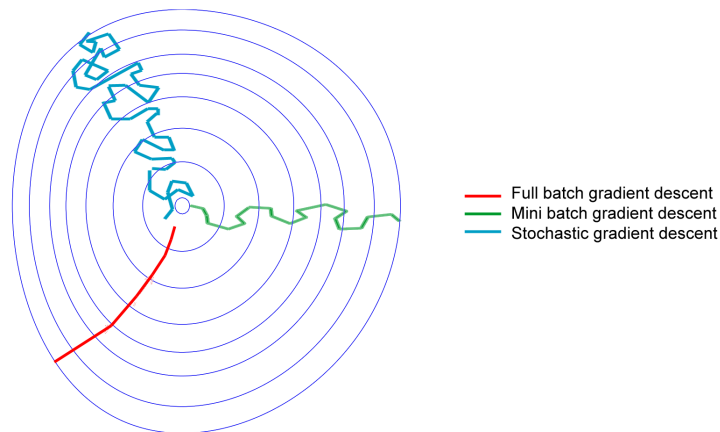


Abb. 3.9: Visualisierung der Schritte im Suchraum für den randomisierten Gradientenabstieg mit immer nur einer Stichprobe (blau), mit mehreren Teilmengen (grün), sowie der kompletten Menge an Trainingsdaten (rot) pro Schritt.

die Lernrate zu hoch, geht die Schrittweite über das Minimum hinaus und käme somit nicht nahe an dieses heran, da es ständig vor und zurück oszilliert. Ist die Lernrate dagegen zu klein, dauert es sehr lange, bis Konvergenz erreicht wird, da die Gewichte zu wenig stark angepasst werden. Es ist daher üblich, mit einer großen Lernrate zu beginnen. In der Annahme, dass der Optimierer sich bereits schnell dem Minimum genähert hat, kann die Lernrate sukzessive verringert werden, um nach und nach das Minimum der Fehlerfunktion feiner abzutasten.

3.3.1 Faltungsnetzwerke (CNN)

Convolutional Neural Networks (CNN) sind Netzwerke bestehend aus Faltungsschichten mit lernbaren Kernelgewichten. Auf einer räumlichen Eingabe (Bild Daten) werden durch die Abfolge von Filtern Merkmale extrahiert und diese in neue Repräsentationen überführt. Neben den Faltungen gehören nichtlineare Aktivierungsschichten sowie Poolingsschichten zur Unterabtastung ebenfalls zu essentiellen Bestandteilen solcher Netze, denn es gilt auch hier: Das Lernen der Gewichte der linearen Faltungsfilter (Kapitel 3.1.1) ist für sich allein noch nicht besonders mächtig, erst durch die nichtlineare Aktivierungsfunktionen können komplexere Zusammenhänge gelernt werden.

Am Ende des CNNs befindet sich typischerweise eine Klassifikations- oder Regressionsschicht, die gleichzeitig mit den Merkmalen zuvor trainiert werden kann. Dies ist ein Kernkonzept von CNNs im Vergleich zur klassischen Bildverarbeitungs Pipeline oder anderen Maschinenlernverfahren, welche auf aufwändig händisch optimierte Filter zur Merkmalsextraktion zurückgreifen (Abb. 3.10). Ein besonderer Fokus bei jeglichen Problemen besteht in der Ableitbarkeit *end-to-end* der Netzwerkstrukturen, sodass

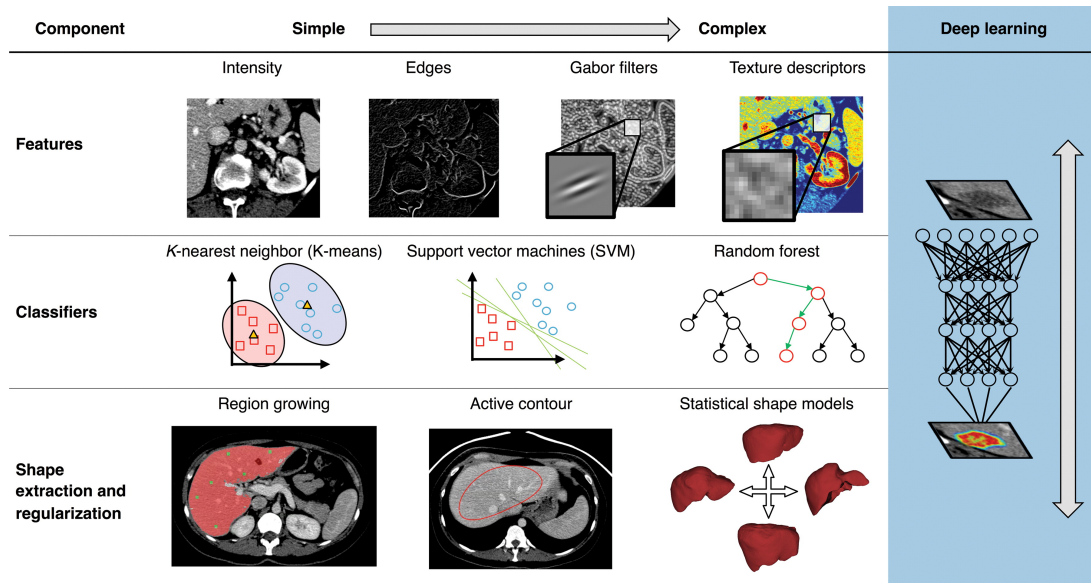


Abb. 3.10: Deep Learning bricht mit dem klassischen Bildverarbeitungsparadigma, wonach konkrete Merkmale, Repräsentationen und Regionenbeschreibungen in einer Pipeline entworfen, berechnet und klassifiziert werden. Aus: Chartrand u. a., 2017

Repräsentation und Merkmale zusammen mit der eigentliche Aufgabe (Klassifikation, Segmentierung, usw.) optimiert und gelernt werden können.

Obwohl die Faltung an sich nicht translationsinvariant ist, so werden die späteren Faltungsschichten in einem CNN durch den Einsatz von Poolingschichten doch invariant gegenüber kleinen Verschiebungen innerhalb der Poolingkerngröße. Umgekehrt spielt bei den Schichten zu Beginn eines CNNs die Translationsäquivalanz (Kapitel 3.1.1) eine größere Rolle, da dort basale Filter gelernt werden, welche lokal und unabhängig von der Position die gleiche Filterantwort basierend auf den Intensitäten berechnen sollen.

Zudem profitiert ein CNN, wie in Abschnitt 3.1.1 angesprochen, vom Teilen der Gewichte der dünnbesetzten Faltungskerne (Faltungskern kleiner als Eingabebild). Damit ergibt sich ein stark regularisierender Effekt durch kleine Faltungskerngrößen und zusammen mit der Mehrfachnutzung (*engl.* Parameter Sharing) der Kernparameter wird der Speicheraufwand gegenüber vollverbundenen neuronalen Netzen reduziert. Diese Eigenschaft der Neuronen, nur lokal auf der Eingabe zu arbeiten, führt zu den äquivalenten Repräsentationen – den Merkmalskarten (*engl.* Feature Maps).

Mit Einführung der Fully-Convolutional Networks (FCN) durch Long u. a. [2015] hat sich der effiziente und intuitive Einsatz von CNNs für semantische Segmentierung etabliert. Bisher wurde mit den Klassifikationsnetzen für jedes Pixel ein Vorwärtsweg durchgeführt, um eine dichte aber rechenaufwendige Segmentierung zu erhalten. Meistens geschieht dies bei größeren Bilddaten auf nur auf einem Ausschnitt (*engl.*

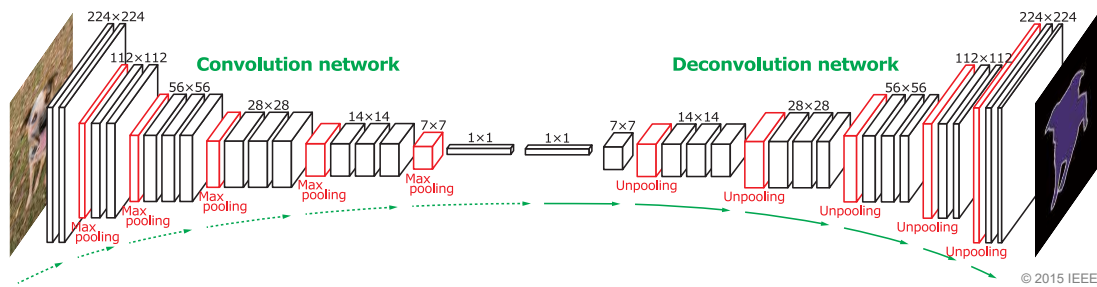


Abb. 3.11: Encoder-Decoder-Architektur zur semantischen Segmentierung in 2D. Aus: Noh u. a., 2015

Patch) statt der gesamten Eingabe. Es liegt nahe, die niedrig aufgelösten Merkmalskarten aus dem Netzwerk zum Training von Segmentierungen direkt hochzuskalieren (z. B. mittels bilinearer Interpolation). In der Vorstellung eines eingabeseitig auf räumlichen Strukturen (Bildern) arbeitenden Klassifikationsnetzes ergänzt sich intuitiv das symmetrische, ausgabeseitige Netz, welches z. B. Segmentierungsmasken oder Patches davon ausgibt. Konsequenterweise wird dafür die „Umkehroperation“ (aus Sichtweise des Lernens von Faltungsgewichten) der Faltung, die transponierte Faltung (*engl.* Deconvolution) benutzt, wobei die Gewichte nun von einem Skalar zu einem größeren Kernel gelernt werden. Analog gibt es auch für das Pooling eine Umkehroperation, das „Unpooling“, wofür die ursprünglichen Indizes aus dem Pooling-Schritt verwendet werden, um den Gradienten rückzupropagieren.

Die Encoder-Decoder-Architektur (Abb. 3.11) setzt dies generalisiert um, mit linksseitig bekanntem, kontrahierendem Pfad zur Merkmalsextraktion, welcher eine hochabstrakte Repräsentation erzeugt, sowie einem rechtsseitig davon hinterhergeschalteten extrahierenden Pfad, um wieder eine Ergebnisrepräsentation für die ursprüngliche Auflösung der Eingabe zu erzeugen. Viele Segmentierungsarchitekturen (Abschnitt 3.3.2) und die Autoencoder (Abschnitt 3.3.3) basieren auf dieser Konstruktion. Ein Problem stellt die Auflösung des Zielkonflikts zwischen der nötigen Abstraktion global auf der Eingabe und der nötigen Lösung der Merkmalskarten dar, von denen aus der Ergebnis hochgerechnet wird. Zur Verbesserung werden darum im extrahierenden Pfad immer auch Faltungsschichten eingefügt und auf zusätzliche Verbindungen vom Decoder zu basalen Merkmalen im Encoder-Pfad gesetzt (Kapitel 5.1.1).

Die Grundlagen neuronaler Netzwerke und deren effiziente Optimierung über das Backpropagation-Verfahren sind bereits seit weit mehr als 30 Jahren bekannt [Rumelhart u. a., 1986]. LeCun u. a. [1998] haben beispielsweise früh flache CNNs zur Erkennung handgeschriebener Postleitzahlen vorgestellt. Jedoch wurde nach längerer Zeit erst mit der Vorstellung des „AlexNet“ von Krizhevsky u. a. [2012] beim „ImageNet“-Wettbewerb im Zuge einer sogenannten Renaissance der künstlichen neuronalen Netze der Fokus der wissenschaftlichen Community wieder auf die CNNs gelegt. Es handelte sich dabei – zusammengefasst unter dem Begriff *Deep Learning* – um tiefere Netz-

werke, welche in akzeptabler Zeit und mithilfe handelsüblicher Grafikkhardware noch effizienter zu trainieren waren und dabei enorm an Genauigkeit im Vergleich zu bisherigen Verfahren dazugewinnen konnten. Schnell haben sich kleine Faltungskerne der Größe 3×3 für ein robustes Training wie im „VGGNet“ als vorteilhaft herausgestellt [Simonyan u. a., 2015]. Unzählige Varianten kamen seitdem auf, welche nicht nur in die Tiefe, sondern auch gezielt in die Breite gehen, wie bei Szegedy u. a. [2016] mit parallelen Pfaden innerhalb eines „Inception“-Moduls.

Viele kleinere architektonische Veränderungen halfen, dass die Netzwerke für ihre Aufgabe trotz begrenzter Datenbasis effektiver und effizienter trainiert werden konnten. Beispielhaft seien etwa Abkürzungen durch das Netzwerk genannt (*engl.* Skips), um Teilinformationen besser durch das Netzwerk zu propagieren und den Gradientenfluss verbessern zu können, wie in den „Highway Networks“ von Srivastava u. a. [2015]. Diese Abkürzungsverbindungen spielen in Abschnitt 5.1.1 noch eine Rolle. Des Weiteren seien hier noch Flaschenhalskonstruktionen (*engl.* Bottlenecks) zur Dimensionsreduktion genannt, etwa mit „Autoencoder“-Netzen (Abschnitt 3.3.3) als typischer Anwendung. Hierbei handelt sich um die oben genannte Encoder-Decoder-Struktur, welche auf Identität trainiert wird, um eine typischerweise niedrigdimensionale Repräsentation mit den wichtigsten Moden, welche die wichtigsten Variationen in den trainierten Daten beschreiben, zu erhalten.

Daneben wurden früh die sogenannten Network-In-Network-Konstrukte („NiN“) von Lin u. a. [2014] vor allem als Alternative eingeführt für die häufig zum Zwecke der Klassifikation am Ende eines Netzes eingesetzten linearen vollverbundenen Schichten (AlexNet, VGGNet). Dies spart eine Vielzahl an Parametern zur Vermeidung der Überanpassung sehr komplexer Netzwerke ein. Im Grunde stellen sie eine Generalisierung der Faltung auf Voxel-Ebene dar: Mittels Faltungen mit Kern $1 \times 1 \times 1 \times C$ (C Eingabekanäle) werden die Merkmalskanäle neu gewichtet und gegebenenfalls reduziert. Viele Architekturen haben die NiN-Faltungen adaptiert und verschiedenartig motiviert und genutzt, etwa innerhalb der Inception-Module von Szegedy u. a. [2016]. Eine neuartige Weiterentwicklung stellen die „Squeeze-and-Excitation“ Netzwerke von Hu u. a. [2019] dar, welche ebenfalls die Rekalibrierung der Merkmale ermöglichen sollen und dabei globale Informationen hervorheben sowie weniger wichtige Informationen in den Merkmalskarten unterdrücken. Hier findet wieder eine Verschiebung des Fokus' weg von der ausschließlich nur tiefen, hierarchischen Verarbeitung räumlicher Abhängigkeiten in Bildern, hin zur Breite (Merkmalskanäle) als ebenfalls wichtige zu verarbeitende Dimension von CNNs.

Neben den strukturellen Änderungen der Architektur gibt es noch grundlegende Mittel, Einfluss auf die Berechnung im Vorwärts- und Rückwärtslauf zu nehmen. Batch-Normalisierung [Ioffe u. a., 2015] und Dropout [Srivastava u. a., 2014] haben einen stark regularisierenden Effekt für ein robustes Training. Viele weitere Varianten auch der ReLUs (*engl.* Rectified Linear Unit) wurden vorgestellt, die nicht immer konsistent einen Vorteil bringen, und deren geeignete Wahl stark von der gewählten Architektur

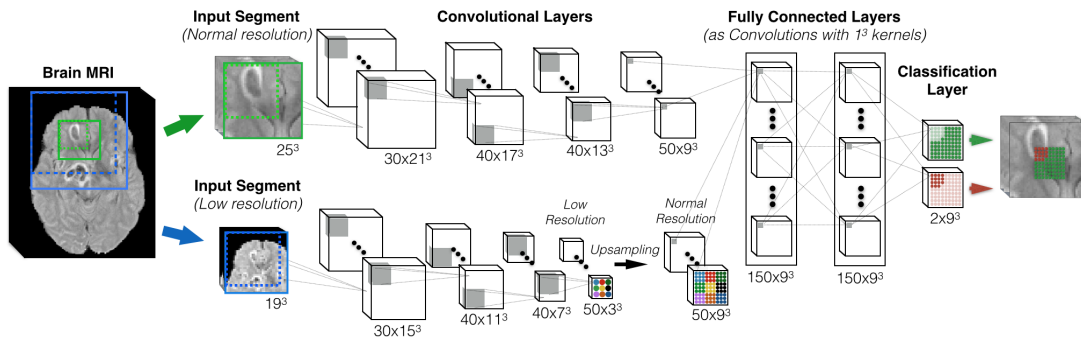


Abb. 3.12: Zwei-Pfad-Architektur von DeepMedic zur parallelen Verarbeitung von feinen Detailstrukturen und umgebendem Kontext in Hirnscans. Aus: Kamnitsas u. a., 2017

und Aufgabe abhängig sind. Es gibt noch eine Vielzahl weiterer vorgeschlagener Varianten an CNNs aus dem breiten Feld der Computer Vision und seiner Anwendungen, doch ein großer Teil der heute grundlegenden Methoden wurde gerade kurz vor Beginn der Arbeit an diesem Dissertationsprojekt (2016) veröffentlicht, sodass sich die vorliegende Schrift in erheblichem Maße an Techniken des Deep Learnings bedient.

3.3.2 Multi-Skalen-Architekturen

Mit der Wahl der Faltungsschichten ergibt sich auch immer die spezifische Größe des rezeptiven Felds, welches im Training ausreichend dimensioniert sein muss, um die relevanten Strukturen und Muster erfassen zu können. Eine besondere Herausforderung in der biomedizinischen Bildanalyse stellen unterschiedliche Größenausprägungen bestimmter Strukturen und Pathologien dar. So kann je nach Lage ein kleiner Hirninfarkt im Gewebe schon weit fortgeschritten sein, wobei ein anderer großer Infarkt noch viel Potenzial zum Wachstum und damit zur Rettung von Gewebe offenbaren kann.

Netzwerke, welche inhärent mit verschiedenen Skalen arbeiten, können grobe Muster und zugleich kleine Ausprägungen einer Pathologie genau erfassen. Der triviale Ansatz mit zwei Pfaden zur Erfassung des globalen Kontexts als auch für die feinen lokalen Informationen, wie bspw. durch Havaei u. a. [2017] vorgeschlagen, kann bereits sehr robust gute Ergebnisse erzielen. Kamnitsas u. a. [2017] haben dies mit der DeepMedic-Architektur (Abb. 3.12) bei den MICCAI-Wettbewerben etwa für Schlaganfälle auf MR-FLAIR Sequenzen (SISS-Aufgabe der ISLES 2015) oder für die Tumor-Segmentierung beim „BRATS“-Wettbewerb 2016 [Kamnitsas u. a., 2016] gezeigt.

Basierend auf den Fully-Convolutional-Netzwerken (FCN) von Long u. a. [2015] gibt es auch den generischen Ansatz, mit symmetrischen Encoder-Decoder-Netzwerken sowohl tiefe Repräsentationen zur Segmentierungen zu erhalten als auch unter Zuhilfenahme von sogenannten Skip-Verbindungen feine Detailinformationen räumlich rekonstruieren zu können. Dies sind die außerordentlich populären U-Nets (Abb. 3.13) von

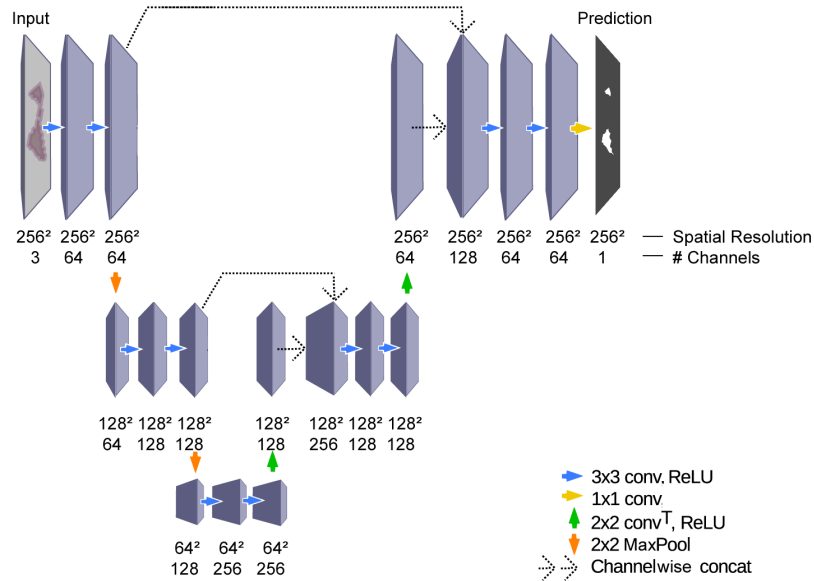


Abb. 3.13: Beispiel eines U-Nets mit linksseitigem Encoder zur Merkmals- und Kontexterfassung, rechtsseitigem Decoder zur feinen Ausgabe der vollaufgelösten Segmentierung sowie den Skip-Verbindungen dazwischen.

Ronneberger u. a. [2015], welche die CEN-Architektur von Brosch u. a. [2016] auf beliebig viele Auflösungslevel generalisieren. Der Grund für die enorme Popularität in der biomedizinischen Anwendung – obgleich bspw. DeepMedic oft ähnlich gute Ergebnisse wie das U-Net erzielt [Casamitjana u. a., 2016] – liegt vor allem an der Vielzahl bereits erzielter Ergebnisse auf dem Stand der Technik (gegenüber schwieriger zu trainierenden Architekturen), wie beispielsweise:

1. In bestimmten klinischen Domänen wurde bereits Genauigkeit auf oder über Expertenlevel nachgewiesen, etwa für Krankheiten der Retina [De Fauw u. a., 2018].
2. Isensee u. a. [2019] haben mit einer kaum modifizierten U-Net-Architektur, aber systematisch optimiertem Training¹², bei einer Vielzahl von Segmentierungswettbewerben zu medizinischen Zielstrukturen¹³ vorderste Platzierungen erreicht.
3. Als Vorstufe der krankheits- bzw. organspezifischen Segmentierung (siehe vorheriger Punkt) ist es auch für die automatisierte Triage einsetzbar, etwa um auf Kopf-CTs spezifische Auffälligkeiten zu detektieren, wie bei Chilamkurthy u. a. [2018] (teilweise mittels U-Net).

¹²Generisch verwendet wurden vor allem ein großes rezeptives Feld, ein fortgeschrittenes Optimierungsverfahren, Augmentierung, eine 5-fache Kreuzvalidierung, sowie 2D/3D-Ensembling.

¹³Hirngliome, Hippocampus, Lunge und Tumoren, Prostata, Herz, Leber(gefäße) und Tumoren, Darmkrebsprimärkulturen, Milz, Pankreas

Aufgrund der Skip-Verbindungen steigt auch der Speicherbedarf durch den wachsenden Abhängigkeitsgraphen und die dafür zwischenspeichernden Ableitungen. Das Problem verstärkt sich bei der Erweiterung auf drei Raumdimensionen. Dabei handelt es sich um eine naheliegende Adaption des U-Nets [Çiçek u. a., 2016; Milletari u. a., 2016], bei der dreidimensionale Kernel benutzt werden – die Unterschiede zu 2D sind trivialer Natur. Jede architektonische Erweiterung mit neuen Skip-Verbindungen und Schichten führt jedoch zu größerem Speicher- und Rechenaufwand. Allerdings ist der 3D-Ansatz aufgrund der Mehrinformation gerade für kleine Strukturen vorteilhaft [Roth u. a., 2018].

Eine konsequente Generalisierung gegenüber den gezielten Skip-Verbindungen stellen Entwicklungen in die Richtung dar, nicht nur durch einzelne Skalen zu verarbeiten, sondern eine Merkmals-Pyramide parallel mit Austausch zwischen den Skalen zu ermöglichen [Huang u. a., 2018; Ke u. a., 2017; Lin u. a., 2017; Zhou u. a., 2018], welche allerdings zumeist weitere Anpassungen erfordern, um auf dem stark anwachsenden Abhängigkeitsgraphen im Training mit limitierten Rechenressourcen die Optimierung durchrechnen zu können.

Ein Nachteil der U-Nets stellt damit auch die limitierte Tiefe (etwa im Vergleich mit den ResNets) dar, welche immer im Kompromiss mit der Anzahl der Auflösungsebenen steht. Durch einfache Kaskadierung von U-Nets lässt sich mit ausreichender Tiefe dennoch eine tiefere Architektur konstruieren. Tatsächlich werden etwa „Stacked Hourglass Networks“ von Newell u. a. [2016] mit entsprechender zusätzlicher Überwachung zwischen den Netzwerken dafür genutzt, menschliche Posen zu schätzen. Beim BRATS-Wettbewerb auf der MICCAI 2019 hat ebenfalls ein Ansatz mit zwei kaskadierten U-Nets den ersten Platz belegt [Jiang u. a., 2020].

3.3.3 Autoencoder (AE)

Im Gegensatz zu Abschnitt 3.2.1 kann die unterliegende Struktur eines Datensatzes auch mittels eines „echten“ unüberwachten maschinellen Lernverfahrens offengelegt werden, den sogenannten Autoencodern. Diese basieren auf neuronalen Netzstrukturen, womit sie zusammen mit diesen verwendet und trainiert werden können¹⁴. Der Name ergibt sich daher, dass das ursprüngliche Ziel des Trainings einer Repräsentation z in der versteckten (latenten) Schicht das Lernen der möglichst genauen Rekonstruktion \hat{x} von Eingabe x ist („auto“), wie in Abbildung 3.14 dargelegt. Wie in der PCA wird versucht, die Abweichung zu minimieren, hierbei nun durch Optimierung der Parameter der neuronalen Netze E und D in Bezug auf eine Fehlerfunktion, wie z. B. $L = \|x - \hat{x}\|^2$ [Berthelot u. a., 2019].

Dies gelingt durch Verringerung der Dimensionalität (Flaschenhalse) der latenten Repräsentation z gegenüber den ursprünglichen Daten – also einer Dimensionsreduk-

¹⁴In einem Faltungs-Autoencoder (*engl.* Convolutional Auto-Encoder, CAE) nach Masci u. a. [2011] besteht Encoder E in der Regel aus den typischen Faltungsschichten wie bei CNNs.

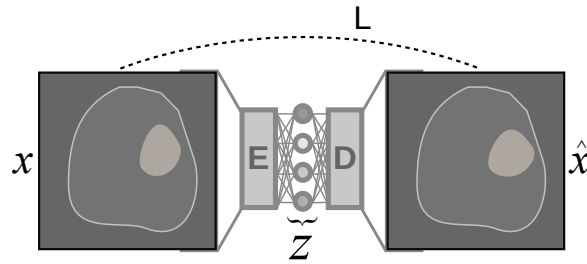


Abb. 3.14: Schematische Darstellung eines (Convolutional) Autoencoders, der Eingabe x mittels Encoder E in die latente Repräsentation z überführt, die der Decoder D unter möglichst geringem Fehler, welcher in L berechnet wird, rekonstruiert (\hat{x}).

tion wie schon bei der PCA –, sodass sich in dieser Repräsentation die wichtigsten Faktoren, welche die größte Varianz des Datensatzes verursachen, herausstellen. Damit können auch kleine Störfaktoren, die nicht die unterliegende Struktur beschreiben, herausgerechnet werden, was durch den Denoising-Autoencoder erreicht wird [Vincent u. a., 2008]. Dafür wird eine korrumpierte Version x' eingegeben (z. B. eine verauschte Eingabe) und vergleicht die Rekonstruktion nach wie vor mit dem Original in $L = \|x - \hat{x}\|^2$. So sollten die korrumpierten x' nicht auf der unterliegenden Mannigfaltigkeit von x liegen, sodass der Autoencoder lernen muss, diese möglichst nah auf die niedrigdimensionale Mannigfaltigkeit abzubilden [Berthelot u. a., 2019].

Es gibt verschiedenste Varianten und Erweiterungen, wie etwa die Variationellen Autoencoder (VAE), welche versuchen, die Verteilung der Repräsentationen aus dem z -Raum möglichst nahe einer A-Priori-Verteilung anzunähern, z. B. Normalverteilung $q \sim N(\mu, \sigma^2)$ mit Mittelwert μ und Standardabweichung σ . Dafür wird dem Fehler ein zusätzlicher Term hinzugefügt, welcher die Kullback-Leibler-Divergenz $KL(z, q)$ der Verteilungen quantifiziert. Es gibt noch einige weitere Varianten, die sich vor allem an den Adverseriellen Netzwerken (*engl.* Adverserial Generative Networks, GAN) mit ihrem zusätzlichen Diskriminator orientieren. Dieser wird trainiert, „echte“ gegenüber „falschen“ z -Kodierungen zu unterscheiden, was aber nicht zwangsläufig für jedes Problem mit Autoencodern zum Erfolg führt, wie bei Berthelot u. a. [2019] zu sehen.

3.4 Segmentierung und Prädiktion für den Schlaganfall

Die Schwierigkeit automatischer Schätzungen insbesondere von Hirntumoren und Schlaganfallläsionen erwächst nach Crimi [2016] vor allem aus der Tatsache, dass diese in ihrer Größe, Form und der Erscheinung außerordentlich variieren können. Speziell für diese Krankheiten ist häufig eine aufwändig durch Experten annotierte Bilddatenbasis zum Training von (stark) überwachten maschinellen Lernverfahren notwendig. Bevor überhaupt maschinelle Lernverfahren und neuronale Netze zur Segmentierung genutzt wurden, setzte man auf klassische Bildverarbeitungsansätze von Schlaganfäl-

Classifier	DM [0, 1]	HD (mm)	ASSD (mm)	Prec. [0, 1]	Rec. [0, 1]	Cases	Traintime
100 Nearest Neighbors	0.54**± 0.20	36.52± 22.4	7.07**± 4.25	0.82	0.45	34/37	5s
10 Nearest Neighbors	0.56**± 0.20	36.47± 25.1	6.58*± 4.01	0.82	0.46	35/37	5s
5 Nearest Neighbors	0.58**± 0.18	39.72*± 27.4	6.80*± 4.35	0.79	0.51	36/37	5s
AdaBoost	0.60*± 0.19	39.28*± 27.3	7.42*± 6.77	0.70	0.61	35/37	7m
Extra Trees	0.64**± 0.19	29.49± 18.5	5.29± 3.94	0.84	0.57	35/37	3m
Gaussian Naive Bayes	0.48**± 0.22	69.86**± 26.7	14.82**± 8.16	0.44	0.78	36/37	1s
Generalized Linear Model	0.44**± 0.25	38.77*± 21.3	8.54**± 5.76	0.87	0.34	32/37	2m
Gradient Boosting	0.63**± 0.18	32.72± 23.2	5.93± 5.28	0.72	0.62	35/37	12h
Random Decision Forest	0.67± 0.18	28.16± 20.7	4.89± 3.63	0.82	0.62	35/37	6m
Convolutional Neural Network	0.67± 0.18	29.64± 24.6	5.04± 5.28	0.77	0.64	35/37	2h

doi:10.1371/journal.pone.0145118.t002

Abb. 3.15: Vergleich maschineller Lernmethoden auf Voxel-Ebene zur Segmentierung ischämischer Schlaganfallläsionen: Bereits einfachste CNNs können die Ergebnisse vorheriger ML-Methoden erreichen oder übertreffen. Aus: Maier u. a., 2015a

len. Traditionelle Schwellenwertverfahren zur Bildanalyse haben den Nachteil, dass sie nur eine einzige univariate harte Entscheidungsgrenze zwischen betroffenem und nicht betroffenem Gewebe modellieren, und damit nicht geeignet sind die häufig verrauschten CT-Perfusionsbilder kontextabhängig zu segmentieren.

3.4.1 Segmentierung bei multivariater Bildgebung

Rekik u. a. [2012] listen 11 fortgeschrittene automatische Methoden auf, wovon 9 auf MR (vor allem T2 und DWI) und nur zwei auf CT arbeiten. Daneben wird eine einzige semi-automatische Methode für die CT-Perfusion besprochen [Contin u. a., 2010], welche allerdings das Setzen von Saatpunkten erfordert. Einige der besprochenen Verfahren führen die voxelweise Klassifikation anhand der zu jedem Voxel zur Verfügung stehenden bzw. berechneten Voxelmerkmale (Features) zur Segmentierung durch, viele der besprochenen Methoden bilden die Segmente mittels k-Means-Clustering.

Maier u. a. [2015a] haben ebenfalls systematisch verschiedene Klassifizierer auf Voxel-Ebene untersucht (Abb. 3.15), wobei vor allem statistische oder maschinelle Lernmethoden für die subakute Schlaganfallsegmentierung miteinander verglichen wurden. Bis auf ein CNN sind alle Methoden mehr oder weniger auf die Berechnung händisch definierter Merkmale¹⁵ angewiesen. Bei dem CNN handelt es sich um ein flaches Faltungsnetzwerk mit vollverbundenen linearen Schichten, welches für jeden Voxel auf einem Patch der Eingabe die Segmentklasse berechnet.

Dabei übertrifft dieses sehr einfach konstruierte CNN bereits die Ergebnisse aller anderen Vergleichsansätze im Standardszenario auf ausschließlich den FLAIR-Sequenzdaten oder erreicht mindestens das Niveau wie der RDF (Abb. 3.15). Selbst beim Direktvergleich eines flachen MLP (Abschnitt 3.3) ohne Faltungsschichten mit dem

¹⁵Intensität, gewichteter Intensitätsmittelwert über Voxelnachbarschaft, Intensitätshistogramm über Voxelnachbarschaft, Voxeldistanz zur Hirnmitte

RDF [Lucas u. a., 2017] bei gleichen Eingabemerkmale liegt auch das MLP bereits auf dem Niveau der Entscheidungsbäume. Andere fortgeschrittenere Multi-Skalen-CNNs wurden bereits in Abschnitt 3.3.2 besprochen und übertreffen die Genauigkeit jener traditionellen ML-Verfahren oft noch deutlicher, wie etwa der Gewinner der Segmentierungsaufgabe („SISS“) der ISLES 2015 [Maier u. a., 2017].

Nativ-CT und die Standard-MR-Sequenzen bilden vor allem Strukturinformationen ab (T1, T2, auch FLAIR), welche sich eher langsam herausbilden. Neben der Perfusionsbildgebung (Kapitel 2.2.1.1) als multivariater Alternative gibt es auch die vielfach in der Forschung verwandte MR-Diffusionsbildgebung (DWI), um Beeinträchtigungen in der Blutversorgung des Gehirns frühzeitig in der Akutphase des Schlaganfalls festzustellen. Chen u. a. [2017] haben beispielsweise erstmalig eine automatische Segmentierungsmethode für akute Schlaganfälle auf DWI vorgestellt. Darüber hinaus ist insbesondere bei Zentrums-übergreifender Anwendung auch die Verwendung von unüberwachten Methoden erwägenswert, welche nicht auf gerätespezifischen Daten trainiert worden sind. Doyle u. a. [2018] benutzen dafür eine MRF-basierte Segmentierung mit einem Atlas der durch die Hirnarterien versorgten Gebiete als Vorwissen. Damit erreichen sie das Niveau der besten überwachten Methoden aus dem Vergleich von Maier u. a. [2015a] in Abbildung 3.15.

In begrenztem Maße besteht die Geräteabhängigkeit auch bei CT-Perfusion, da die Hersteller unterschiedliche Dekonvolutionsalgorithmen verwenden, um aus den Rohdaten die Parameterkarten zu berechnen. Bei Verwendung der CT-Perfusion als Modalität ist es jedoch möglich, den hersteller-spezifischen Dekonvolutionsalgorithmus zu umgehen und direkt auf den Rohdaten zu arbeiten. Dabei gibt es Hinweise, dass damit teilweise bessere Ergebnisse erzielt werden können, wie kürzlich zur Segmentierung der Kernläsion sowie Schätzung des finalen Outcomes gezeigt wurde [Bertels u. a., 2019; Robben u. a., 2020].

Für die CTA als weiterer typischer Bildgebung nach Leitlinie sind bisher keine automatischen Methoden etabliert, sondern es dient eher einer qualitativen Analyse des Verschlusses. Die Ärztin kann das CTA z. B. global in 4 Stufen klassifizieren, welche in begrenztem Maße auch den Outcome prädizieren [Maas u. a., 2009]. Hilbert u. a. [2018] haben basierend auf dem DenseNet [Huang u. a., 2017] ein CNN für CTA vorgestellt, dass drei verschiedene Scores für intraarterielle Interventionen (Thrombolyse und/oder mechanische Rekanalisation) prädiziert, wobei eine AUC von 0,68 bis 0,77 im experimentellen Rahmen berichtet werden. Derartige Methoden für CTA sind bisher noch nicht in der Klinik im Einsatz und es ist darüber hinaus fraglich, inwiefern ein CTA zusätzliche Informationen zur Segmentierung von Schlaganfallläsionen bereithält. Alternativ zur CTA wurde von Livne u. a. [2019] kürzlich eine Segmentierung der Blutgefäße mittels U-Net vorgestellt, die auf Time-of-flight-Aufnahmen der MR zur Visualisierung der Blutgefäße angewandt werden kann und dabei Dice-Werte von 0,88 für die Gefäßsegmentierung erreicht.

3.4.2 Prädiktion des Geweberesultats mittels neuronaler Netze

Huang u. a. [2010] waren einer der ersten, die für die Vorhersagen des Geweberesultats bei Schlaganfällen ein einfaches MLP mit nur einer Zwischenschicht trainiert haben. Es wurde für den Verschluss der MCA von Ratten für 30 Minuten, 60 Minuten sowie den dauerhaften Verschluss die „infarction incidence map“ auf Basis von MR-Bildgebung (CBF, ADC, T2) in der Akutphase prognostiziert, wobei eine AUC von 0,89 bei 30-minütigem Verschluss und 0,95 bei dauerhaftem Verschluss berichtet wurde. Gegenüber einer SVM konnte nach Huang u. a. [2011] solch ein einfaches künstliches neuronales Netz jedoch noch keine Überlegenheit zeigen. Bagher-Ebadian u. a. [2011] haben hingegen ein mehrstufiges Netz mit ebenfalls sehr wenig Neuronen pro Schicht trainiert, um die finale Läsion bei 12 menschlichen Probanden auf Basis von MR-Bildgebung (DWI, T1, T2, PDWI) zu präzisieren. Das 3-Monats-Verlaufsbild wurde mit einer AUC von 0,89 vorhergesagt. Gemein ist beiden eben genannten ANN-Methoden das Fehlen anspruchsvoller Merkmale auf den Eingabemodalitäten, entsprechender Netzwerkkomplexität (maximal 6 Neuronen pro Schicht) und dafür notwendiger Regularisierungstechniken zum Lernen.

Stier u. a. [2015] haben erstmalig das Training eines Faltungsnetzwerks (CNN) für die Vorhersage des Geweberesultats untersucht, nachdem sie zuvor bereits gezeigt hatten, dass lokale Nachbarschaftsinformationen in Patches um einen Voxel für ein Prädiktionsverfahren basierend auf MR mit Diffusion und Perfusion wichtig sind [Scalzo u. a., 2012]. Die CNN-Architektur arbeitet auf 2D-Patches aus der T_{max} -Aufnahme einer MR-Perfusion und vergleicht die Vorhersage mit Verlaufsbild-Segmentierungen auf FLAIR von ungefähr vier Tagen danach. Die Methode konnte bei einer Kreuzvalidierung auf 19 Probanden eine klar höhere Gesamtgenauigkeit von 85 % (*engl.* Accuracy) gegenüber einem linearen Regressionsmodell mit 78 % Genauigkeit für die binäre Gewebestatuskarte erreichen. Flottmann u. a. [2017] nutzen ebenfalls ein GLM [Kemmling u. a., 2015] für die Vorhersage des Infarkt volumens auf CT-Perfusion (CBF, CBV, MTT, TTD) basierend auf der Gewebivorhersage und zeigen die Überlegenheit gegenüber ROC-optimierten Schwellenwertverfahren. Die höhere Präzision (*engl.* Precision) und der niedrigere patienten-spezifische Vorhersagefehler legen nahe, dass die bisherige auf Penumbra-Kern-Mismatch basierende Schätzung davon profitieren kann. Livne u. a. [2018] nutzen ein „Extreme Gradient Boosting“ schwacher Entscheidungsbaum-Klassifizierer und vergleichen dies wiederum mit einem GLM, wobei der Boosting-Ansatz im Gegensatz zum GLM eine AUC $> 0,9$ erzielen kann und damit den Stand der Technik für klinische einsetzbare Methoden erreicht. Es wird deutlich, dass die maschinellen Lernverfahren mindestens auf dem Niveau linearer Regressionsansätze einsetzbar sind, welche in der klinischen Forschung Anwendung finden.

4 U-Nets zur Segmentierung akuter Infarkte in multivariater CT

Für die klinische Einschätzung eines akuten Schlaganfalls muss der eigentliche Gefäßverschluss detektiert und das betroffene Hirngewebe vom gesunden abgegrenzt werden, indem man das Risikogewebe und die bisherige Nekrose segmentiert. Insbesondere zuvor eingeführte Multi-Skalen-Netzwerke (S. 50ff.), wie das U-Net, sind geeignet, um die in Form und Größe sehr variablen Pathologien in der multivariaten CT herauszustellen. Der Einsatz bisheriger maschineller Lernmethoden im Rahmen der Schlaganfallsegmentierung wurde im vorherigen Kapitel 3.4.1 bereits dargelegt. In Kapitel 4.1 wird nun der automatisierte Einsatz von U-Nets für die initiale Schlaganfall-Triage vorgestellt, welches sichtbare Media-Verschlüsse zu 100 % auf ungesesehenen Testdaten erkennen kann. Die Ergebnisse wurden in [Lucas u. a., 2019] veröffentlicht. Die Segmentierung akuter Kern- und Penumbra-Bereiche im Rahmen der neuroradiologischen Diagnostik (S. 21ff.) oder als Bestandteil der vollständigen automatischen Methode aus [Lucas u. a., 2018b] wird in Kapitel 4.2 beschrieben und liefert insbesondere für das Risikogewebe in der CTP mit einem Dice von 0,81 bereits gute Ergebnisse.

In diesem Kapitel wird die Nutzung des Fully-Convolutional-Ansatzes basierend auf der U-Net-Architektur für 2D- und 3D-Bilder aus der CT vorgestellt, welcher es ermöglicht, Multi-Skalen-Information durch das Netz zu propagieren. Damit kann – wo notwendig – auch auf Informationen aus höheren Auflösungsebenen zurückgegriffen werden. Die Wahl des U-Nets liegt einerseits in der generell vielfältigen und erfolgreichen Anwendung dieses auf biomedizinischen Problemstellungen, insbesondere bei Segmentierungsaufgaben im Gehirn, begründet. Zum anderen ermöglicht die Multi-Skalen-Konstruktion mit den Unterschieden in Größe und Form der Schlaganfallläsionen umzugehen (Abschnitt 3.3.2).

In Kapitel 3.4.1 wurde bereits der Einsatz maschineller Lernmethoden für die Schlaganfallsegmentierung besprochen. In den beiden folgenden Unterkapiteln werden im Rahmen der multivariaten CT-Bildgebung konkret zwei verschiedene Aufgaben behandelt und dafür zwei Ansätze unter Nutzung des U-Nets vorgestellt: Die Detektion

von Frühzeichen eines Verschlusses (Abschnitt 4.1) sowie die Segmentierung der Teilbereiche einer Ischämie bzw. des Infarkts für die Mismatch-Diagnostik (Abschnitt 4.2).

4.1 Thrombusdetektion auf nativen CT

Wie in Kapitel 2.2.1 dargelegt, können bestimmte Frühzeichen bereits im nativen CT einen Hinweis auf einen ischämischen Schlaganfall und das betroffene Hirnstromgebiet geben. Sieht man beispielsweise eine hyperdense Arterie in Segment M1 und M2 der MCA, so ist davon auszugehen, dass es sich um einen großen Verschluss mit entsprechend schlechter Prognose handelt. In jedem Fall ist aber bei derartigen Anzeichen eine schnelle Diagnose geboten. Die Thromben können zwar durch Hyperdensitäten im nativen CT nachgewiesen werden, aber es kann auch mehrere andere Ursachen für helle Signalintensitäten im Standard-CT geben, wie z. B. ganz normale dichte Strukturen des Gehirns und Verkalkungen außerhalb der Gehirnarterien. Deep Learning ist geeignet, diese subtilen Unterschiede der kleinen Frühzeichen im CT zu erkennen.

Damit eignet sich das native CT – eine breit verfügbare und schnell einsetzbare Modalität – als erste Stufe der Triage zur Schlaganfalldiagnostik. Die automatische Erkennung der MCA-Frühzeichen kann zur Vorselektierung von Patienten für eine schnellstmögliche spezifische Schlaganfallbildgebung genutzt werden, analog zum Vorgehen etwa von Chilamkurthy u. a. [2018] für Kopf-CTs, welche jedoch dabei noch keine Verschlusszeichen berücksichtigt haben. Darüber hinaus erfordert ein großer sichtbarer Verschluss meistens eine Thrombektomie zur mechanischen Entfernung des Pfropfes [Lisowska u. a., 2018].

4.1.1 Literaturüberblick: Detektion kleiner Zielstrukturen

Es wurden bisher wenige Arbeiten zur automatischen Erkennung von Schlaganfallfrühzeichen mittels moderner maschineller Lernverfahren publiziert, insbesondere in Bezug auf das hyperdense MCA-Zeichen von Thromben im Standard-CT. Neben dem generischen Ansatz der Bildverbesserung als Grundlage von Klassifikations- und Detektionsverfahren für kleine Zielstrukturen mit z. B. „Super-Resolution“-Verfahren [Noh u. a., 2019], oder den Deep-Learning-Ansätzen zur Multi-Instanzen-Detektion [He u. a., 2017; Redmon u. a., 2016], wird für medizinische Problemstellungen häufig eine mehrstufige Herangehensweise gewählt. Takahashi u. a. [2014] schlagen eine Methode zur automatischen Erkennung des MCA-Punktzeichens mittels SVM (Kapitel 3.2.2) vor. Zuvor (händisch) definierte Merkmale werden auf Kandidatenregionen berechnet, um falsch-positive und auch anatomisch unplausible Ausreißer zu eliminieren, bevor die SVM schließlich die restlichen Kandidaten klassifiziert.

Der kaskadierende Ansatz ist eine typische Herangehensweise für kleine Zielstrukturen. Bekannt ist dies auch aus der „InputCascadeCNN“-Architektur zur Tumorsegmentierung von Havaei u. a. [2017], welche die Ausgabe eines ersten Netzes direkt mit

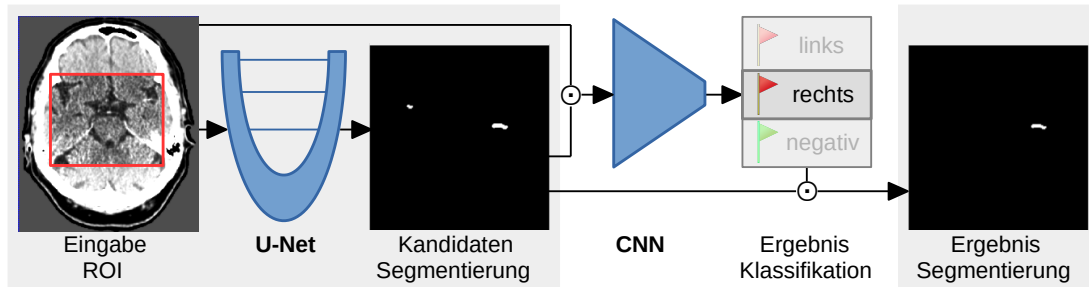


Abb. 4.1: U-Net und nachfolgendes CNN zur Thrombusdetektion und -segmentierung. Auf Basis der Segmentierung (grau hinterlegt) der Kandidaten wird ein Verschluss detektiert und die möglichen Kandidaten ausgegeben.

der ursprünglichen Eingabe verknüpft, um diese in einem weiteren Netz verarbeiten zu können. In Bezug auf die Ergebnisse erwies sich dieses Vorgehen als die beste Art der Verknüpfung gegenüber anderen Varianten, bei denen die Ausgabe des ersten Netzes erst in einer späteren Schicht des zweiten Netzes eingefügt wurde.

Das Kaskadieren hat sich noch für weitere Zielstrukturen als geeignet erwiesen, etwa zur Segmentierung von Lebertumoren [Christ u. a., 2017] oder Blutgefäßen im Abdomen [Roth u. a., 2018]. Insbesondere bei der Detektion von Knoten in den Lungen setzt man auf kaskadierende Ansätze. Setio u. a. [2016] schlagen einen Falsch-Positiv-Reduktionsansatz mit neuronalem 2,5D-Netzwerk (Schnittbilder aus verschiedenen Richtungen) vor und unterscheiden drei Typen von Knoten durch jeweils spezifisches Training. Bei der Erkennung durch Sakamoto u. a. [2017] werden drei aufeinanderfolgende 2D-Schichten ebenfalls zur Reduktion Falsch-Positiver in einem mehrstufigen neuronalen Netzwerk verwendet.

Für das in der CT stark abschwächende Thrombusfrühzeichen innerhalb der MCA sowie die frühe Erkennung einer Ischämie im CT ohne Kontrastmittelgabe haben Lisowska u. a. [2017] darüber hinaus einen Segmentierungsansatz mit CNNs vorgeschlagen. Die Methode erkennt in einer im Voraus definierten Region-Of-Interest (ROI) robust alle großen Thromben [Lisowska u. a., 2018], aber führt insgesamt auch zu einer gleich großen Zahl an Falsch-Positiven, was dem praktischen Einsatz entgegensteht. Im nächsten Abschnitt wird daher eine Methode vorgestellt, welche bei der Detektion von Verschlüssen auf erweiterten UKSH-Daten hohe Erkennungsraten bei einer geringen Falsch-Positiv-Rate ermöglicht.

4.1.2 Vorgeschlagene Methodik

Das kaskadierte neuronale 2D-Netzwerk besteht aus zwei Teilnetzen (Abb. 4.1), die jeweils auf eine bestimmte Teilaufgabe trainiert werden. Das erste Netzwerk lernt, potenzielle Gerinnsel im Bereich der MCA und ICA zu segmentieren, während das zweite Netzwerk lernt, diese Kandidaten korrekt als „Gerinnsel links“, „Gerinnsel rechts“ oder

„kein Gerinnsel“ zu klassifizieren. Der Entwurf dieses kaskadierten Netzwerks folgt dem Falsch-Positiv-Reduktionsprinzip, wie in mehreren der oben genannten Arbeiten beschrieben. Die beiden Netzwerke werden hintereinander trainiert: Zunächst wird das Segmentierungsnetzwerk trainiert und anschließend das Klassifizierungsnetzwerk, wobei jeweils zwei unterschiedliche Trainingsziele bestehen.

Das Segmentierungsnetzwerk zielt darauf ab, zunächst jedes potenzielle Gerinnsel in der MCA oder ICA zu erkennen, sodass sich ein echter Thrombus möglichst immer unter diesen Kandidaten befindet. Es wird mittels positiver Trainingsproben eine große Vielfalt an thrombotischen Gerinnseln gelernt. Für diese Aufgabe wurde eine U-Net-Architektur mit drei Auflösungsstufen (mittels 2×2 -Max-Pooling und -Unpooling) und 13 Faltungsschichten (3×3 Kernel) verwendet, wobei die Anzahl der Kanäle von 32 (oberste Ebene) bis 256 (unterste Ebene) ansteigt. Das Segmentierungsnetzwerk erwartet den gesamten ROI-Ausschnitt als Eingabe.

Das zweite Netzwerk wird mit positiven und negativen Fällen (kein Gerinnsel sichtbar) trainiert, um die Entscheidungsgrenze zwischen Gerinnsel und keinem Gerinnsel zu erlernen und zielt darauf ab, aus den potenziellen Gerinnselkandidaten eine binäre Entscheidung über das Vorhandensein eines Gerinnsels abzuleiten. Es wird ebenfalls eine CNN-Architektur genutzt, wobei diesmal das elementweise Produkt aus der nativen CT und den zuvor geschätzten Segmentierungen der Gerinnselkandidaten aus dem ersten Netzwerk als Eingabe dient. Damit dienen Position, Form, Dichte und Volumen der potenziellen Gerinnsel als Grundlage für die Entscheidung, ob ein Gerinnsel vorliegt oder nicht. Es werden nun drei 3×3 -Faltungsschichten mit zwei Max-Pooling-Schichten abgewechselt, gefolgt von einer letzten adaptiven Pooling-Schicht, die die Auflösung des Zwischenergebnisses auf $2 \times 2 \times 2$ reduziert, bevor zwei lineare Schichten die Pseudowahrscheinlichkeiten der drei Ausgabeklassen vorhersagen.

Der Entwurf des Algorithmus' unterteilt sich somit in zwei Teilaufgaben, um das Erlernen der Gerinnseldetektion zu erleichtern. Ebenso wurde die Klassifikationsaufgabe an sich vereinfacht und das Gerinnsellabel informativer annotiert, indem eine ternäre Klassifikation gelernt wird: Die Unterscheidung zwischen keinem Gerinnsel, linkem Gerinnsel und rechtem Gerinnsel ermöglicht es dem Algorithmus, während des Trainings ein Positiv-Label nur auf Kandidaten in einer der beiden Hemisphären zu beziehen. Als zusätzlicher Nutzen ergibt sich daraus, dass das Klassifikationsergebnis später zur Testzeit verwendet werden kann, um falsch-positive Segmentierungskandidaten aus der gegenüberliegenden Hemisphäre zu entfernen (setze Wahrscheinlichkeit=0 für gesamte Hemisphäre).

4.1.3 Experimente

Für eine experimentelle Evaluation des Verfahrens steht ein erweiterter UKSH-Datensatz an nativen CT-Aufnahmen mit 108 positiven Fällen (inkl. Gerinnselsegmentierung in der ICA oder MCA) und 108 negativen Fällen zur Verfügung. Während des Trainings

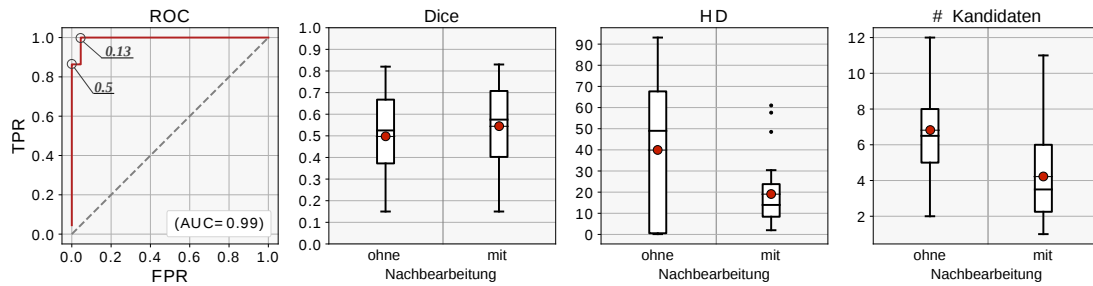


Abb. 4.2: Ergebnisse (Metriken auf S. 36) der 22 positiven und 22 negative Testdaten v.l.n.r.: ROC der Klassifikation mit angegebenen Schwellwerten, sowie drei Metriken zu den U-Net-Segmentierungen: Dice, Hausdorff-Distanz in mm (HD) und Anzahl der Kandidaten vor und nach dem Entfernen jener von der gegenüberliegenden Hemisphäre (Boxen: Interquartilsabstände inkl. Median, rot: arithmetisches Mittel).

wird jeder Fall lateral gespiegelt, um eine Präferenz für eine der beiden Hemisphären zu vermeiden, was zu insgesamt 432 Stichproben führt. Das Segmentierungsnetzwerk wird jedoch nur auf den 216 positiven Fällen trainiert. Die Auswertung basiert auf einer festen Unterteilung der Daten, mit 60 % für das Training, 30 % für die Validierung während des Trainings und 10 % für den Test.

Die Bilder liegen vorregistriert im gemeinsamen Raum wie der normale UKSH-Datensatz (Kapitel 2.4), sodass der Algorithmus innerhalb eines gültigen ROI-Bereichs arbeitet. Daher kann die ROI zur Eingabe direkt aus dem registrierten CT-Bild extrahiert werden. Diese wurde auf Basis der Trainingsdaten als minimale Hülle um alle Gerinnselsegmentierungen der MCA bzw. ICA mit einer Zugabe von 5 Voxeln in jede Richtung ermittelt und hat die Größe von $118 \times 98 \times 98$ Voxeln.

4.1.3.1 Ergebnisse

Die vollautomatische Pipeline erreicht eine Empfindlichkeit von 0,86 bei einer Spezifität von 1 (Abb. 4.2, ROC). Wenn anstelle von 0,5 ein Schwellwert von 0,13 oder niedriger gewählt wird, den die summierten Gerinnselwahrscheinlichkeiten (links+rechts) über-treffen müssen, kann das Verfahren alle positiven Fälle mit einer Falsch-Positiv-Rate von weniger als 5 % detektieren. Falsch-Positive sind in der Praxis in begrenztem Umfang tolerierbar, da der Algorithmus typischerweise als Vorauswahl für dringende Fälle dienen soll, die daraufhin von einer Ärztin gründlich diagnostiziert werden müssen. Mit einer AUC von 0,99 ist das Verfahren aber bereits sehr robust. Darüber hinaus wird die Hemisphäre des Gerinnsels für alle Richtig-Positiven korrekt erkannt.

Das erste Netzwerk erreicht einen Dice von 0,5 als Überlappung mit den Expertensegmentierungen (Abb. 4.2, Dice), wenn es einen ungesesehenen Test-Scan des Datensatzes erhält. Da das Netzwerk nur auf positiven Fällen trainiert wurde und die Zielstrukturen sehr klein sind – was etwa den Partialvolumeneffekt zu einem relevanten Faktor

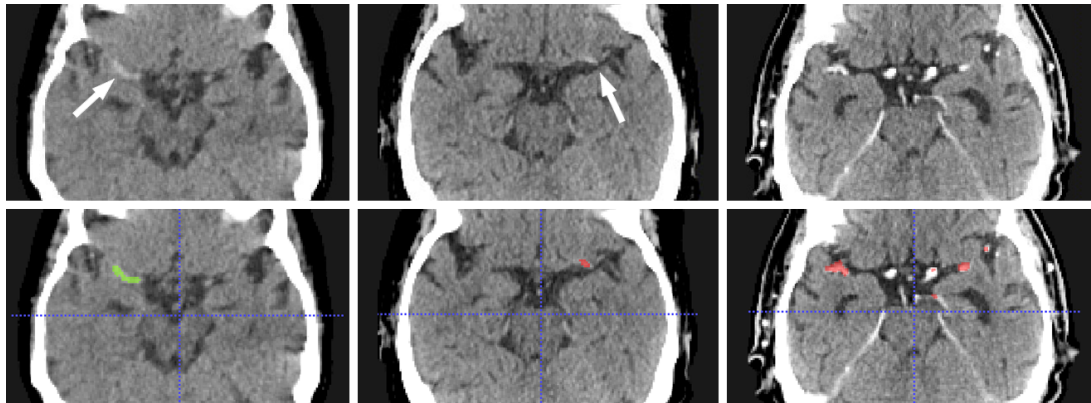


Abb. 4.3: Beispielresultate aus dem Test (oben: CT-Axialschnitt; unten: überblendete Segmentierungen). Links: Richtig-Positiv mit Segmentierungskandidat (grün) auf dem tatsächlichen Gerinnsel. Mitte: Falsch-Negativ, wobei das tatsächliche Gerinnsel (rot) zwar nicht erkannt, aber als einer von mehreren Kandidaten segmentiert wurde. Rechts: Richtig-Negativ, bei dem keiner der Kandidaten (rot) ein tatsächliches Gerinnsel darstellt.

für Fehlinterpretationen macht – ist dieses Ergebnis durchaus vielversprechend (Abb. 4.3). Die Genauigkeit kann verbessert werden, indem mithilfe des Klassifizierungsergebnisses des zweiten Netzwerks Kandidaten aus der gegenüberliegenden Hemisphäre entfernt werden, was zu einer drastischen Reduzierung der Ausreißer führt. Die durchschnittliche Anzahl der Kandidaten wird von 7 auf 4 reduziert, während der Median der Hausdorff-Distanzen von 49 mm auf 14 mm noch deutlicher abnimmt (Abb. 4.2, HD).

4.1.3.2 Diskussion

Es zeigt sich, dass mittels Deep Learning eines kaskadierten Designs inklusive eines U-Nets hochsensitive Segmentierung von Gerinnselkandidaten möglich sind, um dringend zu sichtende Schlaganfallpatienten in Bezug auf das Auftreten eines hyperdichten Frühzeichens als solche zu klassifizieren. Dies ist als Hilfsmittel für eine schnelle zielgerichtete Bildgebung eine wichtige Voraussetzung für die erfolgreiche Behandlung. Integriert in ein radiologisches Informationssystem (RIS) könnte der Algorithmus so nämlich hochwahrscheinliche akute Schlaganfallpatienten automatisch vorselektieren. Obwohl die Segmentierung in der Regel aus mehreren Objektinstanzen [He u. a., 2017], d. h. hier die potenziellen Kandidaten, besteht, können diese die Ärztin in einem CT-Scan (bei im Mittel vier Kandidaten pro Patient) direkt auf die Gerinnselkandidaten verweisen und so eine schnellere Befundung des Schlaganfalls ermöglichen. Dies stellt implizit bereits eine Segmentierung bereit, wodurch etwa die geschätzte Länge oder Dichte der Kandidaten direkt mitberechnet werden kann.

4.2 Segmentierung der Akutläsion auf Perfusionsbildern

Im vorherigen Abschnitt wurde gezeigt, dass das U-Net durch Kaskadierung bereits auf den zuerst verfügbaren nativen CT-Aufnahmen bei Schlaganfallverdacht korrekt die kleinen Frühzeichen eines Thrombus – sofern vorhanden – detektieren und lokalisieren kann. Der Einsatz von U-Nets hat sich bisher jedoch vor allem in Bezug auf größere Zielstrukturen bewährt, wie in Kapitel 3.3.2 und 3.4.1 in der Breite der erfolgreichen Anwendung zu erkennen ist. Insbesondere sind bei einer systematischen Herangehensweise und einem sinnvoll durchgeführten Training die Details der Architektur eher zweitrangig, wie Isensee u. a. [2019] schreiben. Neben einer Normalisierung von Nicht-CT-Intensitäten ist vor allem darauf zu achten, ausreichend große rezeptive Felder und initiale Merkmalskanäle für die Eingabe bereitzustellen. Empfohlen wird außerdem das Training mittels fortgeschrittenem SGD [Kingma u. a., 2015] in einer 5-fachen Kreuzvalidierung. Wo möglich, sind geometrische Transformationen als Augmentierung anzuwenden.

Es folgt als nächster Schritt die Untersuchung des Einsatzes von U-Nets für die Segmentierung der eigentlichen Ischämie und deren Kern- und Penumbra-Bereichs. Dies ist mit Blick auf die automatische Befundung erforderlich, welche sich zunächst an dem klinischen Vorbild des Mismatch-Verfahrens der Ärzte und dem ermittelten Verhältnis zwischen ausschließlich Kern und ausschließlich Penumbra orientiert.

4.2.1 Literaturüberblick: Differenzierung von Risikogewebe und Kern

Die vielfach genutzten Thresholding-Methoden zur Segmentierung können einen groben Eindruck vom Ausmaß einer Läsion auf Perfusionsbilder geben (z. B. $T_{\max} > 6s$), aber nicht als Grundlage für darauf aufbauende Verfahren genutzt werden. Selbst mit zusätzlichen Heuristiken, wie der Entfernung kleiner Läsionen von $< 3ml$ aufgrund wahrscheinlichen Bildrauschens, können Sie keine ausreichend akkurate Segmentierung bereitstellen. Das etwa von Straka u. a. [2010] genutzte Thresholding kann laut Maier u. a. [2017] bei weitem nicht mit den bei der SPES-Aufgabe (*engl.* Stroke Penumbra ESTimation) des ISLES-Wettbewerbs der MICCAI 2015 vorgestellten Methoden auf akuten MR-Perfusions-Daten konkurrieren.

Die Segmentierung des gesamten Risikogewebes stellt die einfachere der beiden Aufgaben dar, da hierfür jegliche durch den Schlaganfall minderperfundierte Areale des Hirns infrage kommen. Darüber hinaus wird für die vollautomatische Erweiterung einer Methode, wie in Kapitel 7, nur noch eine Segmentierungsschätzung benötigt und es kann somit im Test- und Anwendungsfall ausschließlich auf die zuverlässigere Segmentierung des Risikogewebes gesetzt werden.

Die Segmentierung des kompletten Risikogewebes ist demzufolge gut mit automatischen Methoden zu bewältigen. Die Abgrenzung von unterversorgtem Risikogewebe zu bereits nekrotischem Kern ist dagegen weniger trivial. In der Regel vergleicht der

Neuroradiologe längere Perfusionszeiten mit deutlich vermindertem Blutfluss und -volumen gegenüber dem nicht betroffenen Areal, was jedoch sehr schwierig sein kann zu beurteilen und stark auf der subjektiven Erfahrung des Experten basiert (siehe auch Grundlagen in Kapitel 2.2.1.1).

Dementsprechend schwierig ist die automatische Schätzung der Kernsegmentierung auf CTP. Beim ISLES-Wettbewerb 2018 erreichte einzig der Gewinner-Algorithmus von Song u. a. [2019] aus 50 teilnehmenden Methoden in der Rangliste einen Dice-Wert auf den Testdaten von über 0,5 (www.isles-challenge.org). Alle führenden sechs Methoden setzen dabei auf U-Nets (2D/3D), wobei davon zwei mehrere Netzwerke in Ensembles verwenden und eine Variante sogenanntes „Pyramid Pooling“ [Abulnaga u. a., 2019]. Mit der oben angesprochenen Argumentation von Isensee u. a. [2019] lesen sich auch die jeweils nahe beieinander liegenden Ergebnisse der vorderen Plätze in den Einzelmetriken so, dass die architektonischen Varianten nur einen geringen Einfluss haben, im Gegensatz zur übereinstimmenden Wahl des U-Nets als Rahmenarchitektur.

Hier soll darum als Grundlage für die Mismatch-Diagnostik von „Kern versus Penumbra“ (Abschnitt 2.2.1.2), aber vor allem für die vollautomatisierte Methodik im späteren Hauptkapitel 6, das unmodifizierte U-Net zur 3D-Segmentierung auf CTP untersucht werden. Dessen einfache Konfiguration stellt einen möglichst konstanten Störfaktor bei den Untersuchungen des Wachstumsmodells in Kapitel 6 dar, sodass dort vor allem der Effekt dieses Wachstumsmodells untersucht werden kann.

4.2.2 Vorgeschlagene Methodik

Das hier vorgestellte U-Net erlaubt die arbiträre Anzahl an Perfusionparametern als Eingabe zur Segmentierung von Kern und Risikogewebe. Die Verwendung eines kompakten 3D-Standard-U-Nets U , welche den Speicher- und Rechenbedarf in Grenzen hält, erlaubt die Erfassung der räumlichen Struktur dreidimensionaler Schlaganfallvolumen. Statt die vollen $128 \times 128 \times 28$ CTP-Bilder einzugeben, wird auf zufällig positionierten kubischen Ausschnitten (Patches) der Größe $64 \times 64 \times 28$ (mit zusätzlichem Rand von 20 Voxeln in jede Richtung) trainiert und muss daher vier Patches zur Segmentierung eines einzelnen Probanden verarbeiten. Das Netzwerk erhält die Patches von CBV- und TTD-Bild als konkatenierte Eingabe und berechnet zwei Schätzungen: $(\hat{S}_c, \hat{S}_{cp}) = U(I_{CBV}, I_{TTD})$.

Das U-Net besteht aus doppelten Faltungsblöcken, wie sie auch von Ronneberger u. a. [2015] bekannt sind. Jeder $3 \times 3 \times 3$ -Faltung ($Conv3d$) geht eine Batch-Normalisierung (BN) nach Ioffe u. a. [2015] vor. Jede Ebene besteht daher aus $2 \times (BN \rightarrow Conv3d \rightarrow ReLU)$ mit entweder 16, 32 oder 64 Kanälen. Im Encoder folgt darauf zweimal ein Max-Pooling (2×2), während im Decoder zwei mal ein Un-Pooling erfolgt. Dies ergibt eine Anzahl von insgesamt 355.000 lernbaren Netzwerkparametern.

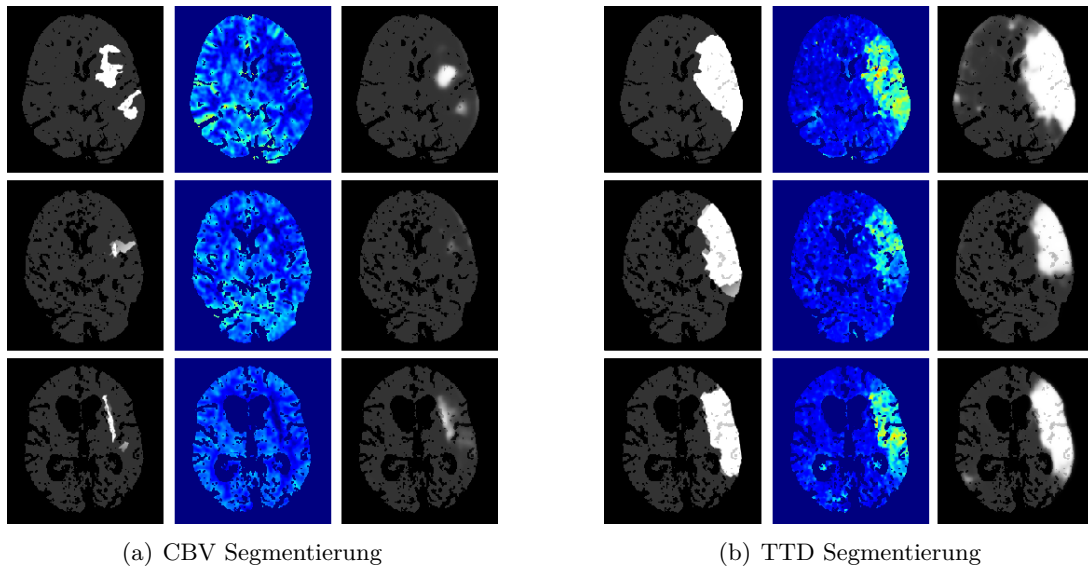


Abb. 4.4: Softmax-Karten des U-Nets für drei Patienten (Zeilen) des UKSH-Datensatzes. Gezeigt werden jeweils v.l.n.r. die Expertensegmentierung, die farbkodierte Perfusionsparameterkarte von blau (minimaler Wert) nach rot (maximaler Wert), und die Schätzung der Segmentierung für das CBV (a) und das TTD (b). Auf Seite 126ff. finden sich jeweils links unten weitere Beispielsegmentierungen.

4.2.3 Experimente

Für die Experimente wurden die Parameter CBV und TTD des Datensatzes „UKSH“ (Kapitel 2.4) verwendet und davon ausgegangen, dass ein TTD-Bild ausreichend ist zur Segmentierung des gesamten Risikogewebes, und ein CBV-Bild, um ungefähr die Segmentierung des Kernbereichs zu approximieren. Dennoch werden beide Parameterkarten gleichzeitig eingegeben und beide Segmentierungen parallel in einem Durchlauf der U-Nets berechnet, sodass jeweils die Information aus beiden Parametern für das Finden der Entscheidungsgrenze zur Verfügung steht. Eine 5-fache Kreuzvalidierung wurde auf dem UKSH-Datensatz von 29 Probanden durchgeführt, wie auf Seite 32 erläutert. Im Rahmen der umfangreichen Experimente der Gesamtmethode von Kapitel 6 inklusive dem hier vorgestellten U-Net enthält ein Batch immer genau vier Probanden.

4.2.3.1 Ergebnisse

Automatische Risikogewebesegmentierungen, die mit dem U-Net-Modell geschätzt werden, sind von hinreichend hoher Qualität bei einem Dice Score von 0,81, sodass man darauf aufbauend die Mismatch-Diagnostik automatisieren kann. Die Segmentierung der CBV-Läsion ist anspruchsvoller und ergibt Dice Scores von nur 0,45, wobei sich

grundsätzlich gleich große Abweichungen von der Grundwahrheit bei den kleinen Kernläsionen auch stärker in der Dice-Metrik auswirken als bei den großen TTD-Läsionen.

Dieses Ergebnis deckt sich – trotz hier nur zweier verfügbarer CTP-Parameter – ziemlich gut mit ähnlichen Ergebnissen der ISLES 2018 auf MR-Perfusion (vier Parameter plus Rohdaten), wie zuvor in Abschnitt 4.2.1 beschrieben. Das U-Net lokalisiert die Läsionen an der richtigen Stelle (Abbildung 4.4), aber die Genauigkeit der Segmentierung ist durch die geringe Kapazität des in diesen Experimenten verwandten Netzwerks eingeschränkt. Insbesondere ist im Gegensatz zu den TTD-Segmentierungen die Ungenauigkeit bei den CBV-Segmentierungen durch niedrigere Pseudowahrscheinlichkeiten zu erkennen, während bei TTD die Masken im Ganzen oft klarer abgegrenzt sind. Es konnten auch noch höhere Dice-Werte für die CBV-Segmentierungen mit dem Ansatz aus Kapitel 5 erzielt werden, wobei diese Segmentierungen dann weniger gut geeignet waren, um im Rahmen der Vorhersagetechnik von Kapitel 6 gute Läsionsmasken zu präzisieren (vgl. Seite B.2).

4.2.3.2 Diskussion

Die U-Net-Segmentierung kann als Grundlage für die ärztliche Mismatch-Diagnostik genutzt werden. Für die im Vergleich zur TTD schwächere Segmentierung der CBV-Läsion bietet sich die Nutzung eines „Denoising Autoencoders“ [Vincent u. a., 2008] an, der entsprechend auf verrauschten Daten trainiert wird, um kleinere Fehler in den Segmentierungsschätzungen als fehlerhaftes Rauschen zu entfernen. Dies ist auch ein Nebenergebnis aus dem späteren Kapitel 4.2.1. Im Vergleich mit den automatisierten Kern-Segmentierungen bei der ISLES 2018, wo ebenfalls nur mittlere Dice-Werte von maximal $\leq 0,51$ bei Nutzung aller verfügbarer Perfusions-Parameter-Eingaben erreicht wurden, kann offensichtlich mit den trainierten diskriminativen Segmentierungsnetzwerken kein hohes Überlappungsniveau mit den Experten erreicht werden.

Die Ärztin dagegen scheint beim Anfertigen der Expertensegmentierungen bereits anatomisches Vorwissen beim Abgrenzen der Kern- oder CBV-Läsionen mit einfließen zu lassen, welches für den Algorithmus auf den Perfusionskarten nicht ersichtlich ist. Bezogen auf die Zweiteilung des Vorhersageproblems (vgl. Einleitung auf Seite 5) ist eine Kombination von Segmentierungs- und darauffolgender Prädiktionsaufgabe denkbar, um damit die fehleranfällige (Kern-)Segmentierung im ersten der beiden Schritte zu umgehen. Dies wird im folgenden Kapitel 5 besprochen. Alternativ kann die Prädiktion auch als Wachstumsprozess unabhängig von der konkreten Schätzung von Kern und Penumbra betrachtet werden, wie in Kapitel 6 („*Experte*“) und Kapitel 7 teilweise basierend auf den Expertensegmentierungen vorgestellt.

5 U-Net zur Prädiktion der finalen Infarktläsion

Die U-Nets können nicht nur zur Schätzung der Akutläsion verwendet werden, sondern ebenso zur Prädiktion der Folgeläsion. Insbesondere die Skip-Verbindungen erlauben das Propagieren von Detailinformationen in spätere Schichten der Merkmals- und Abstraktionshierarchie des U-Nets. Gerade die in der multivariaten CT berechneten Perfusionsparameter könnten entsprechende quantitative Marker repräsentieren, welche visuell zwar nicht intuitiv interpretierbar, aber eventuell durch automatisch gelernte Merkmale eines Deep-Learning-Ansatzes unterscheidbar wären. In diesem Kapitel wird die Nutzung eines modifizierten U-Nets, vorgestellt in [Lucas u. a., 2018a], für die Vorhersage der finalen Läsion beschrieben. Beim ISLES-Wettbewerb der MICCAI 2017 erreichte diese Methode den zweiten Platz [Winzeck u. a., 2018] und zeichnet sich vor allem durch geringe Ausreißer in der Hausdorff-Metrik aus.

Unabhängig von der automatischen Schätzungen des Kern- und Penumbrabereichs zur computer-assistierten subjektiven Mismatch-Diagnostik durch die Ärztin, besteht das eigentliche Ziel darin, den geschätzten finalen Outcome automatisch und reproduzierbar durch einen Algorithmus zu quantifizieren. Da es sich bei dem Gewebere-sultat ebenfalls um eine Binärsegmentierung handelt, welche den chronisch von der Schlaganfallläsion betroffenen Bereich maskiert, kann man die Segmentierungsnetzwerke auch direkt zur Segmentierungsschätzung eines zukünftigen Verlaufsbildes auf Basis der Akutdaten nutzen.

5.1 Literaturüberblick

Es gibt einige verwandte Arbeiten, die mithilfe von Segmentierungsnetzen, wie dem U-Net mit entsprechenden Erweiterungen, die Läsionsmaske des Verlaufsbildes prä-dizieren. Beim ISLES-Wettbewerb 2017 [Winzeck u. a., 2018] zur Prädiktion des Gewe-bere-sultats wurden dafür tatsächlich viele U-Net-basierte Methoden gleichzeitig neben dem hier in Abschnitt 5.2 beschriebenen Ansatz vorgestellt. Die fünf besten Me-thoden enthalten alle eine U-Net-Variante, entweder in 2D [Mok u. a., 2017; Robben

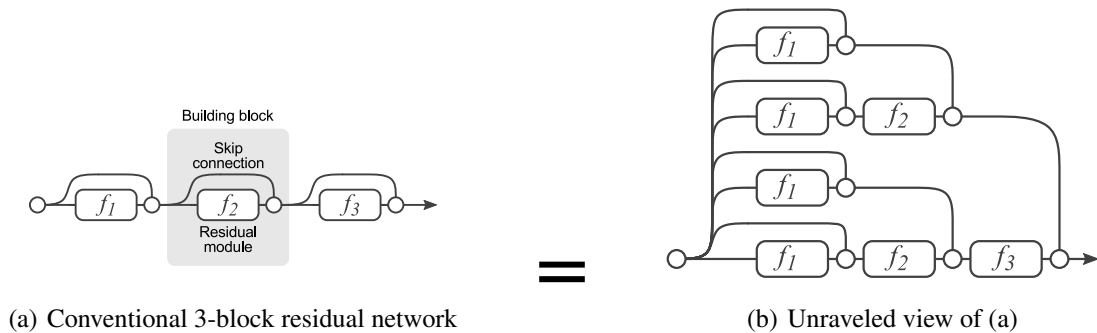
u. a., 2017] oder 3D [Monteiro u. a., 2017], und teilweise innerhalb eines Ensemble [Choi u. a., 2017; Robben u. a., 2017]. Hierbei wurden erstmalig auf einem gemeinsamen MR-Perfusionsdatensatz systematisch Prädiktionsansätze für die Verlaufsbild-Segmentierung miteinander verglichen, da bisher nur wenige Ansätze vorgestellt wurden, um das Geweberesultat zu präzisieren. Da diese Arbeit künstliche neuronale Netze vorstellt und diese heutzutage in der praktischen Forschung meist älteren Ansätzen überlegen sind, wie linearen Regressionsmodellen oder SVMs, liegt der Fokus dieses Kapitels in der Weiterentwicklung solcher Netzwerke (vgl. Kapitel 3.4.2).

Aktuelle CNN-Ansätze für die Prognose des finalen Geweberesultats setzen nach wie vor auf die U-Net-Architektur. Pinto u. a. [2018] haben versucht, dabei auch klinische Informationen mit einzubinden, einerseits auf Populationsebene in einer eigenen Fehlerfunktion während des Trainings, sowie andererseits als zusätzlicher Eingabekanal mit den patienten-spezifischen Werten. Trotz dieser Modifikationen reiht sich das Netzwerk hinter die Top5 der in [Winzeck u. a., 2018] getesteten Methoden ein und damit auch hinter dem auf den folgenden Seiten beschriebenen Ansatz (Abschnitt 5.2). Der bereits kurz genannte Ansatz von Robben u. a. [2020] direkt auf den CTP-Rohdaten konnte dagegen darlegen, dass das Auslassen klinischer Metadaten zu signifikanten Verschlechterungen führen kann. Dies betrifft allerdings nur geringe, kaum signifikante Verschlechterungen und auch nicht konsistent über allen evaluierten Metriken.

5.1.1 Kurze Skip-Verbindungen

Während die im U-Net eingeführten Skip-Verbindungen vom Encoder- in den Decoder-Pfad zum Teil große Bereiche des Netzwerk überspringen, um nach und nach die Detailinformationen von vor dem Encoding für die Rekonstruktion im Decoding wiederherzustellen, steht hinter der Einführung von Residual Networks [He u. a., 2016a] eine andere Motivation. Beim Training besonders tiefer Netze, welche potenziell mächtig genug sind, sehr komplexe Relationen zwischen Einfluss- und Zielvariablen abzubilden, kommt es schnell zum Problem verschwindender Gradienten: Je umfangreicher Backward-Durchlauf durch das Netz, umso stärker können sich die berechneten Gradienten den internen Präzisionsgrenzen der Zahlendarstellung nähern, und dabei auf einen Wert 0 gerundet werden, sodass im folgenden Trainingsdurchlauf keine Update-Berechnung mehr stattfindet. Darüber hinaus tritt nach He u. a. [2016a] das sogenannte Problem der „Degradation“ auf, bei dem eine Sättigung des Genauigkeitszuwachses im Training tiefer Netze beobachtet werden und zum Teil der Trainingsfehler sich mit zunehmender Tiefe der Netze sogar erhöhen kann.

Die strukturelle Lösung für derartige Probleme ist das Hinzufügen von Skip-Verbindungen, wobei das implizite Lernen der Residualfunktion $Y = X + F(X)$ mittels Summation des Ergebnisses eines Faltungsblöcke mit dem Pfad der Eingabeidentität zu bevorzugen ist. Es ermöglicht den unveränderten Fluss der eingegebenen Merkmalskarten durch den „Residualblock“ (*engl.* Residual Block). Typischerweise sollte man für



(a) Conventional 3-block residual network

(b) Unraveled view of (a)

Abb. 5.1: Residual Nets verhalten sich wie ein Ensemble kurzer Netze. Aus: Veit u. a., 2016

tiefe Netze auf Voraktivierung (Batch-Normalisierung, ReLU) des Faltungsblöcke setzen und jegliche Modifikation am Identitätspfad vermeiden [He u. a., 2016b], wobei etwa das Hinzufügen von 1×1 -Faltungen als Gating oder Feature-Mapping umstritten ist [Greff u. a., 2016; He u. a., 2016a,b]. Es empfiehlt sich, den Faltungsblock typischerweise aus nicht mehr als zwei Faltungen aufzubauen sowie jeweils dazugehöriger Normalisierung und Aktivierung [He u. a., 2016b; Li u. a., 2016], da die Residual-Verbindungen nur eine Abkürzung für den Gradientenfluss bei flachen Schichten darstellen, aber das Verschwinden der Gradienten in tiefen Netzwerken mit langen Verbindungen nicht verhindern können [Drozdal u. a., 2016]. Es existieren Varianten mit mehr Schichten innerhalb der Blöcke, wobei einige oftmals eine Flaschenhalsfunktion übernehmen, wie z. B. bei drei Schichten in den tieferen Architekturen des ResNet50, ResNet101, und ResNet152 von He u. a. [2016a]. Andersfalls dienen sie auch der Gruppierung innerhalb einer Schicht, wie beim „Network-In-Neuron“ der ResNeXt-Architektur von Xie u. a. [2017].

Warum Residual Networks ohne viel Konstruktionsaufwand vergleichsweise gut funktionieren, lässt sich durch die Sichtweise erklären, dass sie mittels der Identitätsabbildungen wie ein Ensemble verschiedener kürzerer Netzwerke funktionieren können. Es sind Abkürzungen über die Skip-Pfade beim Durchlauf des ResNets möglich, was man sich in einem aufgerollten Zustand, wie in Abbildung 5.1, veranschaulichen kann, bei dem aus je einem Residualblock zwei Pfade gebildet werden. Insbesondere zeigt sich experimentell, dass vor allem die kürzeren Pfade und somit relativ flache Netzwerke innerhalb eines solchen tiefen ResNets relevant sind [Veit u. a., 2016]. Damit einher geht die Interpretation, dass tiefe ResNets ihren Vorteil nicht aus besonders tiefen Pfaden ziehen, sondern aus der Vielzahl an kürzeren Pfaden, die innerhalb dieser möglich sind zu durchlaufen. Eine andere Interpretation von Greff u. a. [2016] sieht außerdem der klassischen Annahme, dass die Netzwerke zunehmend komplexere (Merkmals-) Repräsentationen modellieren, innerhalb der Blöcke widersprochen, in denen stattdessen Merkmalsidentität besteht und die aufeinanderfolgenden Schichten nur die grobe Schätzung der jeweils ersten Schicht verfeinern.

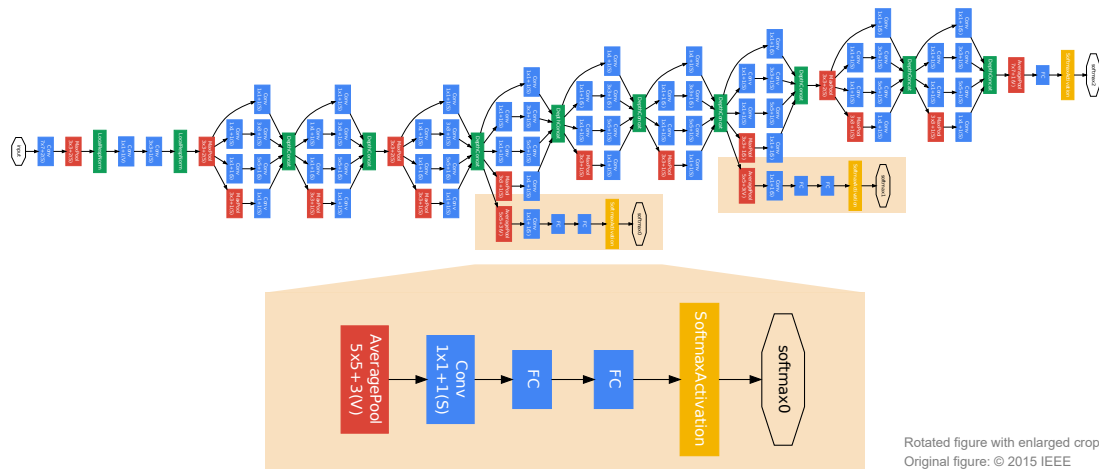


Abb. 5.2: GoogLeNet mit gelb hinterlegten Hilfsklassifizierern (vergrößerte Darstellung darunter); zu lesen v.l.n.r.: Eingabe → Netz → Klassifikation. Aus: Szegedy u. a., 2015

Wenn man neben den ResNets die Konkatenierung von Skip-Verbindungen, wie bei den U-Nets nach Ronneberger u. a. [2015] oder innerhalb der einzelnen Blöcke eines DenseNets [Huang u. a., 2017], betrachtet, so zeigt sich eine starke Evidenz für die vorteilhafte Nutzung solcher Skip-Verbindungen – sowohl kurze als auch lange – in vielen biomedizinischen Anwendungsfällen [Drozdal u. a., 2016]. Wichtig ist dabei, nicht ausschließlich sehr tiefe Netze ohne jegliche Flaschenhalse zu konstruieren, sondern sich an dem Vorgehen wie beim ResNet oder DenseNet zu halten. Bei letzterem erhält eine Schicht innerhalb eines „Dense Blocks“ zwar alle Merkmalskarten der vorherigen Schichten dieses Blocks durch Konkatenation; beim Übergang zum nächsten Block über ein „Transition Layer“ wird jedoch wieder ein Flaschenhals mit einer reduzierten Anzahl von Merkmalskarten eingeführt [Huang u. a., 2017].

Obgleich es mit „Dilated Residual Networks“ von Yu u. a. [2017] Architekturen gibt, die die räumliche Rekonstruktion mit Skip-Verbindungen umgehen, so wird doch vielfach empfohlen, Skip-Verbindungen in existierende Multi-Skalen-Architekturen (Kapitel 3.3.2) zu integrieren, wie bereits von Drozdal u. a. [2016] für das U-Net untersucht. Kamnitsas u. a. [2016] haben ebenfalls ihr „DeepMedic“-Framework um Residual-Verbindungen für die Tumorsegmentierung erweitert, und auch auf anderen Datensätzen zeigen sich bessere Ergebnisse durch die Kombination von Multi-Skalen- und Residual-Architekturen [Chen u. a., 2016; Pohlen u. a., 2017].

5.1.2 Zusätzliche Überwachung

Szegedy u. a. [2015] haben in Experimenten mit der Inception-Architektur an verschiedenen Stellen im Netzwerk Hilfsklassifizierer (Abb. 5.2) während des Trainings ausgewertet und den Fehler gewichtet zum Gesamtfehler aufaddiert. Zum einen argu-

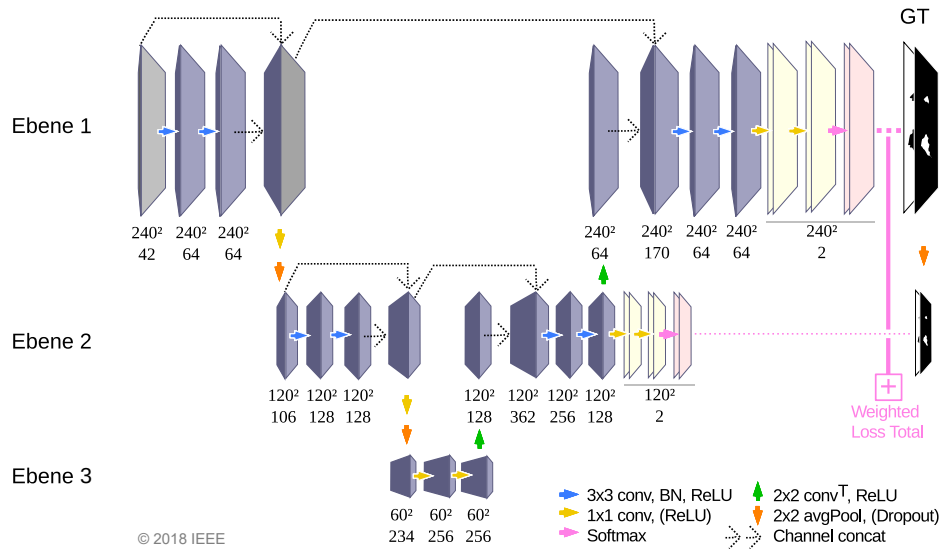


Abb. 5.3: Vorgeschlagene 2D-Netzwerkarchitektur für die Segmentierung der Folgeläsion auf Perfusionsbildern. Zusätzlich zum U-Net (Seite 51) wurden hier kurze Skip-Verbindungen innerhalb jeder Ebene des Encoders eingefügt sowie ein gewichteter Multi-Loss, der auf einzelnen Klassifikationen der Decoder-Ebenen berechnet wird.

mentieren sie, dass dies das Problem der verschwindenden Gradient gezielt angeht und trotzdem eine Fehlerrückpropagierung in frühe Schichten ermöglicht. Besonders zu Beginn sollten diese Hilfsfehler (*engl.* auxiliary loss) ausgewertet werden. Dadurch, dass direkt mit dem finalen Target verglichen wird, erzwingt man zum anderen akkurat interpretierbare Merkmalskarten bereits früher als am Ende des Netzwerks. Letztendlich ist der Genauigkeitsvorsprung beim Training mit den zusätzlichen Klassifizierern aber gering.

Wenn man sich die ursprüngliche Fully-Convolutional-Architektur von Long u. a. [2015] anschaut, wurden dort verschiedene Zwischenstufen mit der letzten Schicht der eigentlichen Basispipeline kombiniert und die ursprüngliche Auflösung aus diesen Zwischenschichten interpoliert. Eine Steuerung der Genauigkeit der Segmentierung, sowie zum Teil von grober Über- und Untersegmentierung, wäre über einen gewichteten Fehler über alle drei Ausgaben vorstellbar. Dieses soll in Abschnitt 5.3 untersucht werden.

5.2 Vorgeschlagene Methodik

Im Gegensatz zur trivialeren Aufgabe der Schätzung der Segmentierung einer akuten Perfusionsläsion (Risikogewebe), wie in Kapitel 4.2, ist bei der Prädiktion des Verlaufbild-Ergebnisses anzunehmen, dass bestimmte Informationen der Perfusionsparameter eingabe zwar keine Akut-Läsion darstellen, dafür aber bei der Dekodierung oder in tieferen Schichten des U-Nets als zukünftige Läsion erkannt werden müssen,

und umgekehrt. Das Ziel ist es daher, trotz der vergleichsweise relativ flachen (vgl. ResNets) aber robust und effizient zu trainierenden sowie erfolgreich angewandten U-Net-Architektur, möglichst alle Informationen, welche zum Modellieren der prognostizierten Segmentierung der finalen Läsion nötig sind, im gesamten Netzwerk durchpropagieren zu können.

Es wird angenommen, dass die endgültige Läsion zwischen den Grenzen des nekrotischen Kerns und den Grenzen der niedrigperfundierte Penumbra darum liegt. Das tatsächliche Resultat im Verlaufsbild hängt von der Art und Erfolg der Behandlung, der Zeit bis zur Aufnahme und Behandlung, der Qualität des Verschlusses sowie anderen, teils globalen Parametern ab. Es wird auch angenommen, dass es sich um eine zusammenhängende Läsion handelt, weshalb zusätzlich beim binarisierten Endergebnis außer der größten zusammenhängenden Komponente alle kleineren Segmente entfernt werden.

Das Hinzufügen von kurzen Skip-Verbindungen (Abschnitt 5.1.1) um jeden doppelten 3×3 -Faltungsblock im Encoder-Teil des U-Nets dient diesem Zweck (Abb. 5.3). Da U-Nets nicht so viele Schichten wie Residuals Networks verwenden, sind keine wesentlichen Vorteile durch die Nutzung eines „echten“ Residual Blocks zu erwarten, sondern konkatenieren stattdessen die Eingabe als zusätzliche Kanäle zur ausgegebenen Aktivierung des doppelten Faltungsblöcke. Dies erlaubt theoretisch den Durchlauf der ursprünglichen Eingabe bis in die tiefste Ebene des Encoder-Pfads nahezu ohne Informationsverlust, modifiziert nur durch mehrmaliges Average-Pooling sowie 1×1 -Resortierung der Kanäle.

Es werden – im Gegensatz zur allgemeinen Darstellungen in Abbildung 5.3 – vier Skalenstufen verwendet und zu Beginn in der ersten Stufe 2D-Axialschnitte in voller Auflösung eingegeben. Bei Nutzung des Datensatzes der ISLES 2017 für die Experimente müssen diese auf eine gemeinsame Auflösung von 240×240 Pixeln herunterskaliert werden. In beiden Dimensionen erfasst das rezeptive Feld 68 Voxel durch Wiederholung der Faltungen und dreier 2×2 -Average-Poolings. In der tiefsten Ebene werden ausschließlich 1×1 -Faltungen für reduzierten Speicher- und Rechenbedarf durchgeführt, da so bereits insgesamt 11 Millionen Parameter gelernt werden. Durch die zahlreichen Skip-Verbindungen steigt der Speicherbedarf gegenüber dem normalen U-Net im Training stark an, weshalb auf ein weniger speicherintensives 2D-Netzwerk zurückgegriffen wird. Konkret werden in der obersten Ebene 1 in beiden Faltungen 64 Kanäle genutzt, welche sich mit jeder darunterliegenden Ebene verdoppeln auf 128, 256 und schließlich 512 in der tiefsten Ebene 4, wo die Größe der Merkmalskarten schließlich 30×30 Pixel beträgt.

Darüber hinaus wurde in Experimenten untersucht, ob man mittels Hilfsfehler (Abschnitt 5.1.2) das Netzwerk besser auf Verlaufsbildsegmentierungen trainieren kann. Die Intention ist es, bereits früh bei der Dekodierung die grobe Größe und Form der finalen Läsion durch den Encoder-Pfad mit den zusätzlichen Skip-Verbindungen zu

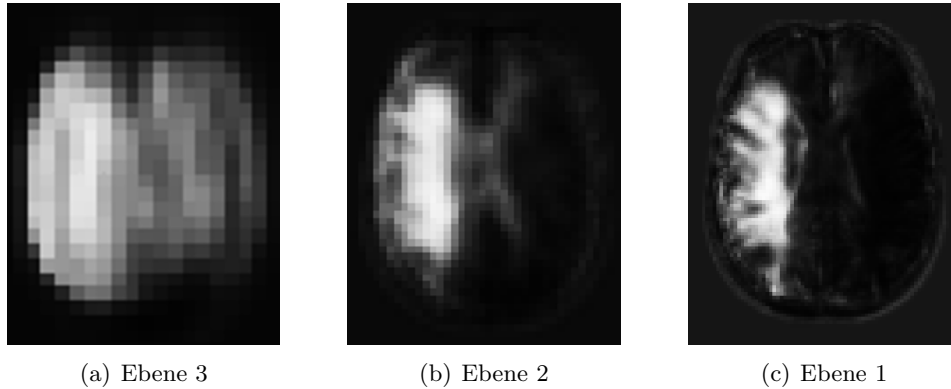


Abb. 5.4: Durch die Nutzung des Multi-Loss wird bereits auf den unteren Ebenen (a, b) eine gering aufgelöste Aktivierung erzeugt, die ungefähr das eigentliche Ziel prädiziert. Es fällt auf, dass sich die Verfeinerung in Form einer Falsch-Positiv-Reduktion vollzieht, sodass am Ende eine fein aufgelöste Schätzung ohne anfängliche Übersegmentierung ausgegeben wird (c).

erzeugen und den Dekoder-Pfad nur noch zum Verfeinern der Segmentierung ohne Übernahme des potenziellen Bildrauschens aus den Skip-Verbindungen zu trainieren.

Bis auf die tiefste der L Auflösungs Ebenen innerhalb des Decoders wird am Ende eine Klassifikationen durch zwei 1×1 -Faltungen mit ReLU und Softmax-Aktivierung vorgenommen (Abbildung 5.4). In jeder Ebene $l \in \{1 \dots L\}$ werden die Fehler dieser Klassifikationen mit einem Vektor \vec{w} multipliziert, um den Einfluss der Ergebnisse der einzelnen Auflösungsstufen auf das Training zu gewichten, sodass sich insgesamt folgender „Multi-Loss“ zwischen Grundwahrheit y und Schätzung \hat{y} berechnet:

$$MultiLoss(\hat{y}, y) = \sum_{l=1}^L -w_l \cdot SoftDice_l(\hat{y}_l, AveragePooling(y, 2^{l-1})) \quad (5.1)$$

Durch diese Gewichtung der Hilfsfehler wird der Kompromiss zwischen Abstraktion und Details (die durch die zusätzlichen Kanäle im Encoder hinzugefügt wurden) balanciert. Mit generell immer genau zwei Klassen, Vordergrund (Läsion) und Hintergrund (Gehirn), besteht ein großes Klassenungleichgewicht in den Daten – der Unterschied in der Prävalenz beträgt bei der ISLES 2017 eine Größenordnung. Mittels der SoftDice-Metrik (Seite 36) zur Fehlerberechnung lässt sich die relative Überlappung zwischen der wahren Läsionssegmentierung und deren Schätzung unabhängig der Größe quantifizieren.

Zusätzl. Skips	Ebenengewichte			Vol. > 0 ml	ASSD [mm]	HD [mm]	Präzi- sion	Treffer- quote	Dice
	w_1	w_2	w_3						
(keine)	1,00	0,00	0,00	34/43	5,29	24,59	0,42	0,32	0,31
Konkat.	1,00	0,00	0,00	43/43	4,22	21,32	0,54	0,35	0,35
Konkat.	0,67	0,22	0,11	42/43	3,45	21,48	0,52	0,35	0,35
Residual	0,67	0,22	0,11	42/43	4,05	20,35	0,47	0,42	0,36
(keine)	0,67	0,22	0,11	36/43	5,09	23,06	0,43	0,34	0,32

Tabelle 5.1: Ergebnisse untersuchter Konfigurationen von Erweiterungen eines normalen U-Nets, welches als Referenz grau hinterlegt ist: Konkatenierung oder Residual-Verbindung der Eingabe in jeder Ebene des Encoder-Pfades und Nutzung eines Multi-Loss mit Gewichtung der Ebenen-Ausgabe des Decoders wie angegeben. Resultierende Binärvolumen > 0 ml gelten als evaluierbare, erfolgreiche Fälle. Die nicht-erfolgreichen Fälle flossen *nicht* in die Berechnung der Evaluationsmetriken ein. Das beste Ergebnis pro Metrik ist jeweils **fett** markiert, die hier besprochene vollautomatische Methodik in orange („U-Net^{FL}“ in Kapitel 6).

5.3 Experimente

Beim ISLES-Wettbewerb 2017 (Daten: siehe Kapitel 2.4) wurde die vorgeschlagene Methodik für Perfusionsbilddaten zur Vorhersage des Endergebnisses einer Läsion evaluiert und belegte insgesamt den zweiten Platz mit den binarisierten Läsionsschätzungen. Die Segmentierung von akuten Schlaganfallläsionen auf Perfusionsbildern ist eine bereits recht gut zu meisternde Aufgabe (Kapitel 4.2). Der MR-Datensatz der ISLES 2017 enthält jedoch fünf Perfusions- und zwei Diffusionskarten pro Patient, sowie klinische Daten (TICI, mRS, $t_{0 \rightarrow A}$ und $t_{A \rightarrow I}$), um stattdessen mit diesen Akutdaten die finale Läsionssegmentierung des zukünftigen Verlaufsbildes vorherzusagen.

Zusätzlich wurde auf den Trainingsdaten mittels LOO-Kreuzvalidierung (*engl.* Leave One Out Cross Validation, LOOCV) vergleichend evaluiert, welchen Anteil die beiden besprochenen Erweiterungen der kurzen Skip-Verbindungen im Encoder sowie der Multi-Loss auf die Ergebnisse haben, und ob eine empirische Bestätigung für die Wahl von Konkatenierung der Eingabe bei den kurzen Skip-Verbindungen statt Residual-Verbindungen zu beobachten ist. Dies ergibt die fünf getesteten Konfigurationen in Tabelle 5.1.

Nach einer initialen Untersuchung der bereitgestellten klinischen Daten der ISLES 2017 war festzustellen, dass diese wenig zusätzlichen Wert für die automatische U-Net-Prädiktion der Folgeläsion bieten. Dafür wurden die Skalarwerte hochskaliert und je als zusätzlicher Kanal an die Eingabe konkateniert. Dies könnte auf die teilweise geringe Qualität der Daten zurückzuführen sein, z. B. fehlende Werte und eine Unterrepräsentation von Fällen mit schlechten Ergebnissen bei niedrigem TICI-Score. Die Varianz in den klinischen Werten könnte auch in Bezug auf die wenigen Proben zu hoch sein,

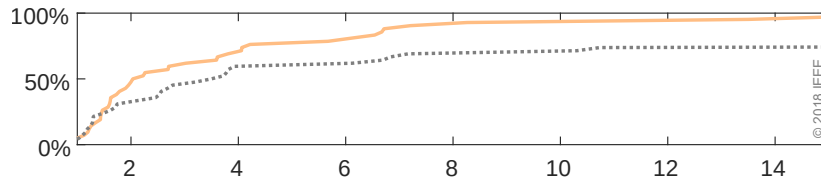


Abb. 5.5: Sortierte ASSD-Werte der LOO-Kreuzvalidierung in mm . Die Lücke von 100 % bei der vorgeschlagenen Methode (●) zu den Ergebnissen eines Basis-U-Nets (•), wie in Kapitel 4.2 ohne zusätzliche kurze Skip-Verbindungen oder Multi-Loss, erklärt sich durch 9 Fälle, für die keine Segmentierung geschätzt werden konnte.

damit der Algorithmus Unterschiede zwischen den Gruppen erkennen kann. In den Experimenten wurden daher ausschließlich die Bilddaten verwandt, sodass der Fokus auf einer automatische Extraktion geeigneter Bildmerkmale liegt (insbesondere solche, die mindestens den Infarktkern definieren). Dies bedeutet, dass das Modell zwar die potenziell unterschiedlichen Behandlungsergebnisse nicht etwa aus den TICI-Informationen antizipieren kann, aber im Durchschnitt führt dieses Vorgehen bei einem U-Net auf diesen Daten zu einem robusteren Training.

Für die Eingabe des Netzwerks ergeben sich am Ende 21 Kanäle: Es gibt sieben Eingabemodalitäten für jede der 43 verfügbaren Trainings- und 32 Testpersonen des ISLES-Datensatzes und jede 2D-Schicht wird mit ihren beiden direkt benachbarten Schichten in Axialrichtung gepaart. Dies ergibt insgesamt $7 \cdot 3 = 21$ Merkmale, was durch eine einfache Augmentierung einer zusätzlichen hemisphärischen Spiegelung aller 21 Kanäle noch einmal verdoppelt wird, um den klinisch inspirierten bilateralen Vergleich von Perfusionsunterschieden der Hemisphären direkt über die Eingabekanäle zu ermöglichen.

5.3.1 Ergebnisse

Im vorgestellten 2D-Netzwerk wurden innerhalb des Kodierungspfades weitere Skip-Verbindungen hinzugefügt, sodass nach Abbildung 5.5 50 % der Segmentierungen des vorgeschlagenen Netzwerks (blau) im Mittel nur weniger als 2 mm von der Grundwahrheit abweichen. Auch die Präzisionswerte (*engl.* precision) in Tabelle 5.1 sind bei den „Konkat.“-Skip-Varianten am höchsten, sodass diese Varianten am genauesten arbeiten. Das ist insbesondere vorteilhaft bei kleinen Läsionen, da diese präziser abgegrenzt werden können (vgl. Anhang auf S. 125) – vorausgesetzt, sie werden auch als solche erkannt, was bei der geringeren Trefferquote (auch *engl.* Recall) nicht immer der Fall ist.

Keine Vorteile lassen sich bei der Nutzung des Multi-Loss' (Abb. 5.4) ablesen. Insbesondere konnte in empirischen Tests keine Verbesserung durch eine andere Gewichtung, als die für die Experimente in Tabelle 5.1 benutzte, festgestellt. Die Gewichtung mit $w_1 = 0,67$, $w_2 = 0,22$ und $w_3 = 0,11$ summiert sich dabei zu 1 auf. Darüber hinaus

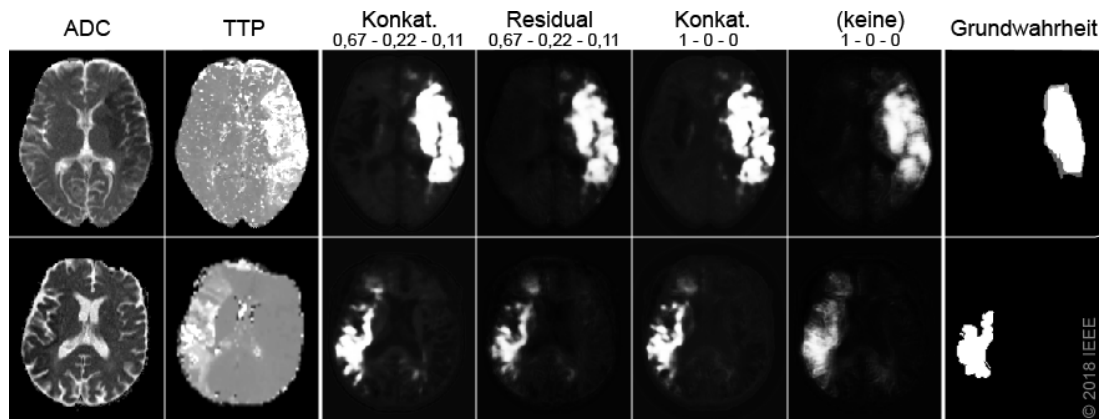


Abb. 5.6: Zwei Beispiele aus den Trainingsdaten der ISLES 2017 v.l.n.r.: ADC und TTP der MR-Bildgebung, darauf folgend die Segmentierung der Konkatenierung mit Multi-Loss, der Residualverbindung mit Multi-Loss, der Konkatenierung ohne Multi-Loss, sowie des U-Nets ohne Erweiterung. Letzte Spalte: Grundwahrheit.

ergibt auch die Wahl von Residual-Verbindungen statt der Konkatenierung der Eingabe keinen klaren Vorteil. Vor allem das Durchpropagieren der Eingabeinformation (entweder durch Konkatenation oder Residualverbindung) stellt einen direkten und auch qualitativ sichtbaren Unterschied gegenüber dem normalen U-Net dar, dessen Segmentierung weniger scharf abgegrenzt und weniger sensitiv auf kleinen Details ist (Abb. 5.6).

Es ist im Wettbewerb der ISLES aufgefallen, dass die Hausdorff-Distanz im Vergleich die niedrigste ist. Obgleich das Verfahren für alle Testfälle ein Ergebnis liefert, sind die Segmentierungsausreißer gering. Die generell niedrigen Dice-Werte des Verfahrens werden mit steigender Läsionsgröße besser und dieser lineare Zusammenhang ist im Vergleich mit den anderen Verfahren der ISLES bei der vorgestellten Methode am stärksten ausgeprägt (Korrelationskoeffizient von 0,72). Das kann durch die niedrige Trefferquote (*engl.* recall) erklärt werden: Zwar ist das Verfahren sehr empfindlich und genau, auch bei kleinen Strukturen, aber diese müssen erst einmal erkannt werden, was bei einer niedrigen Trefferquote nicht immer der Fall ist. Größere Strukturen werden bevorzugt erkannt und durch die genaue Segmentierung mit wenig Distanz zum Rand der tatsächlichen Läsion erhöht sich damit auch die relative Überlappung im Dice.

5.3.2 Diskussion

Zwar zeigt sich aus dem Konfigurationsvergleich in Tabelle 5.1 keine direkte Überlegenheit der Konkatenierung zusammen mit dem Multi-Loss in Bezug auf die untersuchten Metriken, jedoch muss man dabei berücksichtigen, dass die anderen Konfigurationen ohne zusätzliche Skip-Verbindungen zum Teil tatsächlich schlechtere Ergebnisse produzieren, da weniger Fälle überhaupt segmentiert werden konnten. Die nicht-erfolgreichen

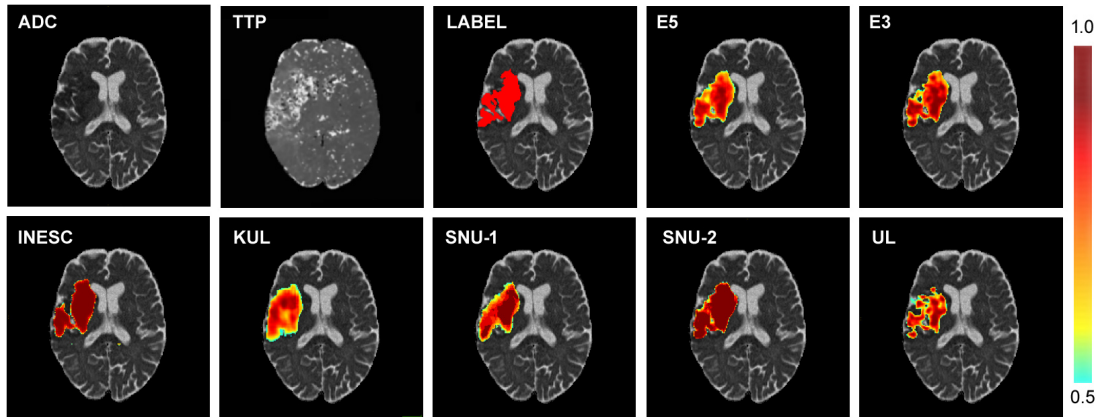


Abb. 5.7: Segmentierung der fünf erstplatzierten Netzwerke (ohne Binarisierung) bei der ISLES 2017 (untere Reihe; „UL“ hier vorgestellt), sowie ADC, TTP, Grundwahrheit („LABEL“), und Mittellung der Top3 und Top5. Aus: Winzeck u. a., 2018

Fälle wurden bei der Berechnung der Metriken jedoch gar nicht berücksichtigt. Die Nutzung der U-Net-Modifikationen kann Genauigkeit und Sensitivität erhöhen und dabei Ausreißer verringern.

Da entsprechende Untersuchungen auf dem geschlossenen Testdatensatz des ISLES fehlen, legen zumindest die obigen Ergebnisse nahe, dass mit dieser genaueren Segmentierung auch das häufige Problem der über den Dice-Fehlerterm optimierten neuronalen Netzwerke eingeführte Bias zu großer Volumina [Bertels u. a., 2020] vermieden wird. Dies müsste man allerdings explizit für alle Verfahren auf dem ISLES-Testdatensatz untersuchen und kann hier nur vermutet werden.

Winzeck u. a. [2018] zeigen, dass selbst mit verschiedenen Ensembles aus den besten drei bzw. fünf Methoden (siehe „E3“ und „E5“ in Abb. 5.7) kaum signifikant bessere Ergebnisse erreicht werden können. Dies legt nahe, dass auf dem Datensatz bereits das maximal mögliche erreicht wurde. Das wirft die Frage auf, ob diese Ansätze mit ihren niedrigen Ergebniswerten zur Prädiktion auf Basis von Perfusionsdaten geeignet sind. Da die Segmentierung von Kern und Penumbra deutlich genauer automatisch geschätzt werden kann als die Folgeläsion, folgt im nächsten Kapitel eine Wachstumsmodellierung losgelöst von den Segmentierungen auf den Perfusionsdaten. Das Hinzufügen dieser Segmentierungen als Vorwissen scheint notwendig, um die Prädiktionsgenauigkeit verbessern zu können. Nach der Evaluation zahlreicher U-Net-Varianten bei der ISLES 2017 sind in Hinblick auf weitere Modifikationen der U-Net-Architektur dagegen keine signifikanten Verbesserungen zu erwarten.

Die in diesem Kapitel vorgestellte U-Net-Architektur wird zu Vergleichszwecken mit der Interpolationsmethode auf den folgenden Seiten als „U-Net^{FL}“ („FL“ als Abkürzung für die Schätzung der „Folgeläsion“ bzw. „finalen Läsion“) referenziert.

6 Formrauminterpolation zur Vorhersage von Infarktwachstum

Die Prädiktion der finalen Läsionsgrenzen bei einem akuten ischämischen Schlaganfall ist äußerst herausfordernd und mit bisherigen Mitteln, wie den U-Nets aus Kapitel 5, noch nicht auf klinischem Niveau mit ausreichender Genauigkeit automatisiert lösbar. Im folgenden Kapitel werden verschiedene Vorgehensweisen mit Rücksicht auf die dynamische Veränderung von Gehirnläsionen diskutiert. Daraufhin wird das Problem zuerst in zwei aufeinander aufbauende Aufgaben unterteilt und die Abgrenzung des Kernbereichs zum gesamten Risikogewebe als Vorwissen eines darauf basierenden Formraums einbezogen. Eine neue Methode wird vorgestellt, welche durch Interpolation in diesem Formraum von Schlaganfallläsionen die Gewebeprognose verbessert. Das U-Net wird hierbei nur noch zur Schätzung von Kern- und Risikogewebesegmentierungen eingesetzt (Kapitel 4.2), aber nicht für die Prädiktion der Folgeläsion. Die Interpolationsmethode erreicht damit signifikant bessere Dice-Werte als die vorherige Methode sowie korrelierende Volumenvorhersagen. Teilergebnisse und Methodik aus diesem Kapitel wurden in [Lucas u. a., 2018b] veröffentlicht.

Im vorherigen Kapitel wurde ein Verfahren für die direkte Schätzung der Folgeläsion mittels U-Net vorgestellt. Dabei handelt es sich um einen aus technischer Sicht zunächst plausiblen Ansatz, da ein Segmentierungsnetzwerk zur Schätzung der Läsionssegmentierung benutzt wird. Diese Läsion stellt allerdings die finale Schädigung dar, welche erst bei der Nachsorge in der Verlaufsbildgebung und damit zu einem anderen Zeitpunkt sichtbar wird, als die vorher definierten Läsionen in der Akutbildgebung (Kern und Penumbra bzw. Risikogewebe).

Eine methodische Schwachstelle bildet hierbei die Nutzung des diskriminativen Segmentierungsmodells, welches im Gegensatz zu einem generativen Modell kein Vorwissen modelliert, sondern nur auf ein Unterscheidungskriterium trainiert werden kann, um dieses für eine Klassifikation voxelweise anzuwenden. Gerade für die Schätzung einer zukünftigen Segmentierung wird jedoch zusätzliches Wissen notwendig, welches nicht allein in den Informationen der Eingabe kodiert ist.

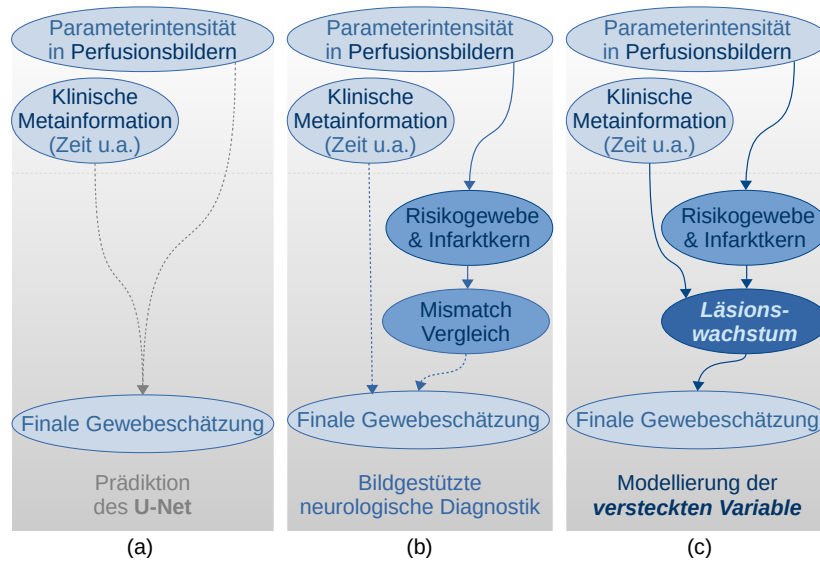


Abb. 6.1: Das Wachstum eines Infarktkerns kann im Rahmen der Outcome-Prädiktion als versteckte Variable zwischen Eingabe und dem zu schätzendem finalen Gewebesresultat erklärt werden (c), was sich an dem klinischen Vorgehen orientiert (b). Bei der Schätzung mittels U-Net (Kapitel 5) wird das Wachstum nicht explizit geschätzt, sondern versucht implizit zu lernen (a).

Die eigentliche Variable von Interesse ist das Infarktkernwachstum. Diese kann nicht direkt gemessen werden – es handelt sich um eine *versteckte* Variable (Abb. 6.1), deren zugrunde liegender Wachstumsprozess nicht bekannt ist. Das U-Net versucht implizit die Auswirkung dieses Wachstumsprozesses zu modellieren, indem es die Segmentierung der Folgeläsion im zukünftigen Verlaufsbild auf Basis der Akutbildgebung schätzt. Mit Blick auf die begrenzte Genauigkeit eignet sich dieses Vorgehen weniger zur Prädiktion des zukünftigen Ergebnisses eines Hirninfarkts. Eine Ursache lässt sich in der fehlenden Modellierung vom Wachstum der Infarktläsion erkennen, da die Prädiktion stattdessen ausschließlich auf der Unterscheidung eines wahrscheinlicheren Gewebesergebnisses von dem eines unwahrscheinlicheren in Bezug auf die Bilddaten basiert.

In diesem Kapitel sollen Vorwissen und Nebenbedingungen für den eigentlichen Wachstumsprozess definiert werden, um diesen besser lernen und für Simulationszwecke nutzen zu können („Wie entwickelt sich die Läsion über die Zeit?“). Dafür wird hier die räumlich-zeitliche Schätzung des eigentlichen Läsionswachstums basierend auf weiterhin nur schwacher Überwachung durch Grundwahrheiten in Form von Segmentierungen für Kern, Risikogewebe und finaler Läsion auf dem Verlaufsbild nach Intervention vorgestellt. Einführend wird zunächst ein Überblick über verschiedene Ansätze für die Prädiktion und Modellierung von Krankheitsverläufen gegeben.

6.1 Literaturüberblick

Die traditionelle Modellierung von Gewebewachstum basiert auf expliziten Modellformulierungen. Im Gehirn ist insbesondere die Tumorwachstumsvorhersage aufgrund der hohen Sterblichkeitsraten in der klinischen Praxis von hoher Relevanz. Gängig sind Verlaufsvorhersagen in Form sogenannter „Disease Progression Models“ (DPM) ebenso bei chronischen oder degenerativen Krankheiten, wie z. B. Herzkrankheiten, Diabetes, oder der Alzheimer-Krankheit zur Anwendung für die Medikamentenentwicklung [Mould, 2012]. Die Modellierung der Krankheitsverläufe ist mit ähnlichen praktischen Problemen verknüpft, wie die der Schlaganfallprogression. Nach Wang u. a. [2014] sind das die unvollständigen und zeitlich irregulär aufgenommenen Beobachtungen oder der Einfluss vieler versteckter Störfaktoren auf den Gesamtverlauf der Krankheit.

6.1.1 Mathematische und biomedizinische Modellierung

Neben der Modellierung von Krankheitsstadien der chronischen obstruktiven Lungenerkrankung (*engl.* COPD) durch Wang u. a. [2014] gibt es viele weitere Vorhersagemethoden, die auf der Idee von Markov-Modellen basieren, etwa für die Alzheimer-Krankheit [Sukkar u. a., 2012] oder Hirntumore [Alberts u. a., 2016; Bauer u. a., 2014]. Andere bildbasierte Ansätze versuchen sich bei der Vorhersage von Tumorwachstum an der Formulierung von Diffusions-Reaktionsgleichungen der unterschiedlichen Gewebe [Hogea u. a., 2008; Menze u. a., 2011]. Viele DPMS basieren oft aber gerade nicht auf Bilddaten, sondern auf vorher anderweitig bestimmten Biomarkern oder gar nur klinischen Ereignissen ohne das Wissen um die vollständigen Krankheitssymptome [Fonteyn u. a., 2011]. Viele dieser Modelle werden gerade deshalb für chronische oder degenerative Krankheiten formuliert, weil deren Verlauf über einen längeren Zeitraum stattfindet. Beim Schlaganfall sind dagegen üblicherweise die ersten 24 Stunden in Bezug auf den Behandlungserfolg einer therapeutischen Intervention relevant. Darüber hinaus sind die zuvor genannten Modelle immer nur spezifisch für die biophysikalischen Vorgänge einer bestimmten Krankheit entwickelt worden und nicht auf den Schlaganfall übertragbar.

Ein Schlaganfall löst eine Kaskade verschiedener Phänomene aus, von der Gewebesebene (z. B. Schwellungseffekte) bis zur Ebene der intrazellulären und molekularen Vorgänge (z. B. Zellsterben). Dabei handelt es sich um ein hochdynamisches System aus voneinander abhängigen Prozessen mit unterschiedlichen Zeitverläufen auf mehreren Ebenen, was in Abbildung 6.2 angedeutet wird. Es existieren Computermodelle, um einzelne Aspekte und Subsysteme sowie deren Zusammenwirken mathematisch zu modellieren [Seidenstein u. a., 2015]. Frühe Arbeiten von Duval u. a. [2002] oder Dronne u. a. [2004] versuchen über mathematische Formulierungen Gewebepredictionen innerhalb sehr grob aufgelöster räumlicher Matrizen zu berechnen. Letztere abstrahieren zehn Untermodule, die verschiedene Phänomene beschreiben, wobei sich bereits eine

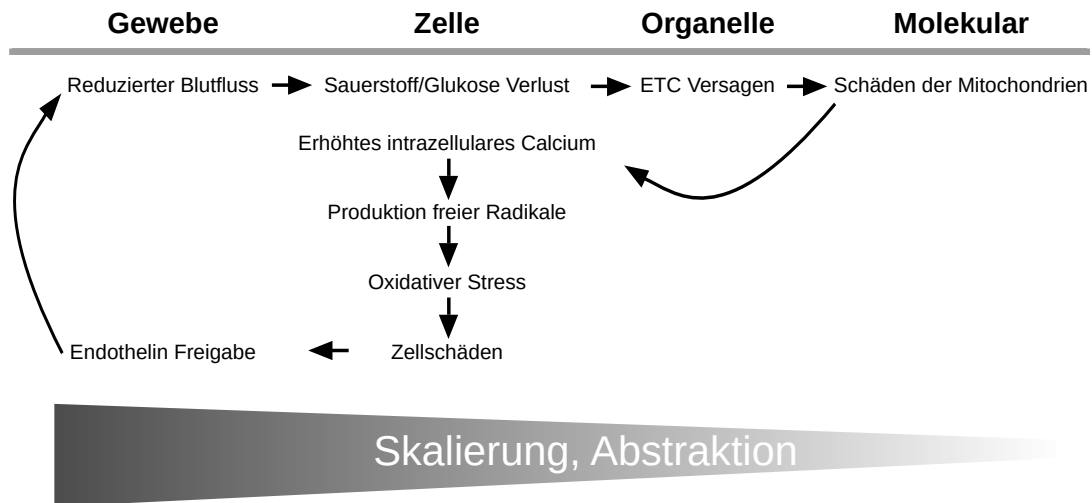


Abb. 6.2: Beispielhafte Auswahl der komplexen Prozesse und Abläufe, die an einem ischämischen Schlaganfall beteiligt sind und zur Zellschädigung führen (Ausschnitt aus größerem System in Newton u. a., 2016). Bei der Modellierung und Simulation muss die Abstraktion über die verschiedenen Skalen hinweg abgewogen werden. Daneben ist die numerische Berechnung explizit formulierter mathematischer Modelle aufwändig und resultiert oft in Ergebnissen mit begrenzter Anwendbarkeit. Man beachte in obigem Beispiel die positive Rückkopplungsschleife, welche sich durch den zirkulären Pfad ergibt.

Auswahl von drei Submodulen für eine räumliche Simulation eignete. Die Autoren haben später auch ein spezialisiertes Ionen-Bewegungsmodell entwickelt, welches die Basis für das 3D-Simulationstool von Dumont u. a. [2013] bildet, das eine Vielzahl an Diffusions-Reaktionsgleichungen für weiße und graue Substanz unterscheidet. Die Nutzung solcher Modelle zielt aber vor allem auf die pharmakologische Verwendung ab, um neue Ansätze zur Medikamentenentwicklung finden und testen zu können (pharmakologisch relevante Größen), d. h. alle genannten Modelle wurden ausschließlich *in-silico* validiert. Die Vorhersagequalität einer räumlich-zeitlichen Prädiktion bei menschlichen akuten Schlaganfällen ist dementsprechend nicht beschrieben worden. Es ist aber bemerkenswert, dass derartige biomedizinische Evolutionsmodelle die Berücksichtigung von Phänomenen auf Gewebeebene erlauben, die sonst in den bildbasierten Vorhersageansätzen selten berücksichtigt werden [Rekik u. a., 2012] – insbesondere Schwellung und Reperfusion.

6.1.2 Maschinelles Lernen des klinischen Schlaganfallverlaufs

Das funktionelle oder klinische Ergebnis, etwa in Form des mRS (Kapitel 2.3), ist zumeist der Endpunkt von Interesse im Sinne der evidenzbasierten Medizin, um den

Nutzen einer Methode einzuschätzen. In der vorliegenden Arbeit wurde – wie in der Einleitung motiviert – der Fokus auf den räumlich aufgelösten Infarktstatus des Gewebes gelegt, was insbesondere durch Radiologen, welche üblicherweise mit räumlichen Bilddaten arbeiten, bevorzugt wird. In der klinischen Forschung existieren bisher vor allem jedoch Ansätze zur Schätzung eines globalen klinischen Patientenergebnisses. Kamal u. a. [2018] listen einige Methoden für die Vorhersage bestimmter Outcomes bei Schlaganfallbehandlung auf, etwa das Risiko für eine intrazerebrale Blutung (*engl.* Intracranial Haemorrhage, ICH). Bentley u. a. [2014] haben früh einen SVM-basierten Ansatz vorgestellt, der sich prognostischen Scores („SEDAN“, „HAT“) aus der Klinik überlegen zeigt. Eine aktuelle Arbeit von Nielsen u. a. [2018] nutzt auch ein CNN für die Gewebeprognoze und Vorhersage des Behandlungsergebnisses und verzeichnet bessere Ergebnisse ($AUC = 0,88$) als ein damit vergliches GLM.

Für die mRS-Prognose haben Asadi u. a. [2014] einen Vergleich maschineller Methoden, wie ANN und SVM, durchgeführt aber noch nicht besonders hohe Ergebnisse berichtet. Bei der ISLES 2016 bestand die zweite Aufgabe ebenfalls in der Prädiktion des mRS, wobei nur drei Teams final teilgenommen haben. Neben den zwei RDF-Methoden nutzten Choi u. a. [2016] ein Ensemble aus U-Net und logistischer Regression, konnten damit aber nicht den erstplatzierten Random-Forest-Regressionsansatz von Maier u. a. [2016] schlagen. Die Autoren geben jedoch selbst an, dass die Ergebnisse des Wettbewerbs mit Vorsicht zu genießen sind, da die Datengrundlage zu schwach erscheint, um robuste Vorhersagen zu machen. Xie u. a. [2019] stellen noch einen aktuelleren Vergleich an Boosting-Ansätzen vor, die auf Basis von Bild- und klinischen Metadaten das binäre Ergebnis von $mRS > 2$ prädizieren. Mit den initial verfügbaren Bilddaten konnte eine AUC von 0,75 erreicht werden. Durch Hinzufügen zusätzlicher Informationen (z. B. NIHSS) nach 24 Stunden konnte die Vorhersage des 90-Tage-mRS mit einer AUC von 0,88 bedeutend verbessert werden. Nach wie vor werden aber bei der direkten Vorhersage eines zusammenfassenden klinischen Endergebnisses, wie dem mRS, zu viele Einfluss- oder Störfaktoren impliziert und damit nicht erklärt, sodass es nicht die erste Wahl für mögliche Handlungsempfehlungen der Ärztin darstellt.

6.1.3 Deep Learning von Läsionsformen und deren Evolution

Die Vorhersage des räumlich-zeitlichen Verlaufs von Intensitäten in einem Bildraum, die den Gewebezustand im Gehirn kodieren, ist äußerst schwierig, da sie viele versteckte Prozesse abstrahieren muss. Läsionsformen oder strukturierte Bilddaten liegen nicht im euklidischen Raum, sondern können im Allgemeinen als Riemannsche Mannigfaltigkeiten interpretiert werden, worauf einige Ansätze zur Modellierung explizit Bezug nehmen [Schiratti u. a., 2015]. Wie schon aus der Bildverarbeitung bekannt (Kapitel 3.1), bietet es sich oft an, eine andere Repräsentation zu nutzen, mit der sich ein Problem besser lösen lässt (vgl. Abbildung 6.3).

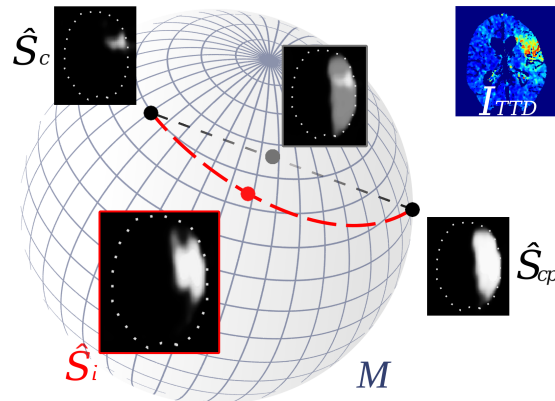


Abb. 6.3: Interpolation der Läsionsformen von Kern \hat{S}_c und Risikogewebe \hat{S}_{cp} . Die triviale lineare Interpolation (graue Linie) im Voxelraum Ω führt zu einem Ein- und Ausblenden der beiden Formen. Die nichtlineare Einbettung von Formen auf einer Mannigfaltigkeit M ermöglicht eine geodätische Interpolation (rote Linie) auf M , was zu einer nichtlinearen Interpolation der Formen im Voxelraum führt. (\hat{S}_i)

Eine datengetriebene Möglichkeit zur Beschreibung valider Läsionsformen ist das Lernen von Verteilungsparametern statistischer Formmodelle (*engl.* Statistical Shape Models, SSM). Oft wird die Form dafür zunächst in eine geeignete, reduzierte Repräsentation überführt, wie Punktwolken oder triangulierte Oberflächenmodelle. Eine wichtige Aufgabe ist dabei die Sicherstellung von Punktkorrespondenzen, um mit diesen formregularisierende Modelle trainieren zu können (z. B. mittels PCA, Abschnitt 3.2.1). Krüger u. a. [2015] haben einen probabilistischen Ansatz vorgestellt und über die Punktposition hinaus noch „Appearance“-Merkmale berücksichtigt, um damit Schlaganfallläsionen mit dem Vorwissen über die Form auf 2D-MR-Sequenzen zu segmentieren. Für die Schlaganfallvorhersage wurden derartige Modelle aber bisher nicht eingesetzt.

Vorteil der Nutzung von Deep Learning ist der aus Entwicklersicht scheinbar „modellfreie“ Ansatz, wofür keine explizite geometrische Parametrisierung gefunden werden muss, sondern die Parameter der neuronalen Netze können komplett datengetrieben gelernt werden – im starken Kontrast etwa zu der händischen Modellformulierung aus Abschnitt 6.1.1. In Bezug auf die zuvor besprochenen Repräsentationen statistischer Formmodelle mittels PCA kann man auch analog durch einen CAE (Kapitel 3.3.3) kompakte und darüber hinaus nichtlineare Räume mit Vorwissen aus Formannotationen maschinell lernen.

Ravishankar u. a. [2017] schlagen eine kaskadierte Architektur aus U-Net und CAE vor, wobei die Segmentierung des U-Nets noch einmal durch einen CAE rekonstruiert werden muss. Ein kombinierter Fehlerterm sorgt dafür, dass die Projektion der Segmentierung auf die durch den CAE erzeugte Mannigfaltigkeit möglichst ähnlich zur Segmentierung des U-Nets ist, welche zugleich einen geringen Fehler gegenüber

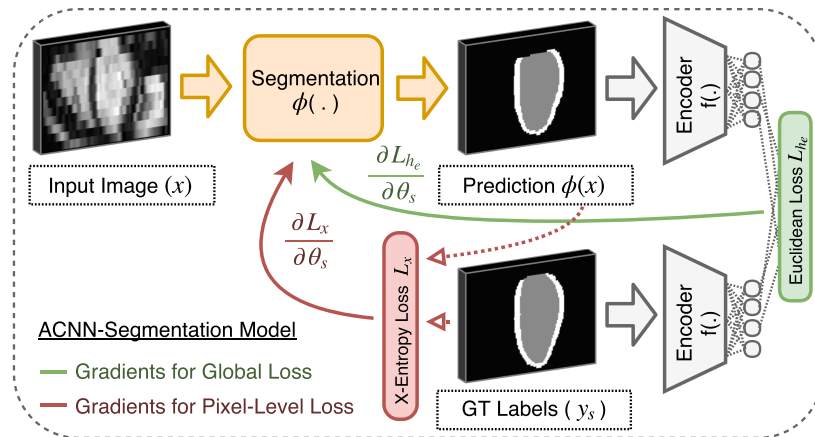


Abb. 6.4: Trainingsschema des ACNN für die Bildsegmentierung: Ein CAE als Regularisierungsmodell erzwingt, dass die Segmentierungen der Verteilung der mit dem CAE gelernten niedrigdimensionalen Repräsentationen (latente Kodierungen ganz rechts, berücksichtigt im euklidischen Fehler) folgen. Aus: Oktay u. a., 2018

der Grundwahrheit aufweisen sollte. So folgt die Segmentierung gerade der Verteilung, die durch den kompakten Formraum des CAE ermöglicht wird, um akkurate Segmentierungen zu erzeugen. Oktay u. a. [2018] stellen einen generischen Ansatz vor, um allgemein die Formbeschränkungen von anatomischen Annotationen als Vorwissen für verschiedene Aufgaben zu berücksichtigen, wie z. B. der Segmentierung oder der sogenannten Super-Resolution. Durch die gelernte A-Priori-Verteilung im CAE auf Basis medizinischer Bilddaten werden diese Aufgaben anatomisch reguliert. Dafür müssen etwa geschätzte Bildsegmentierungen nahe der erlernten latenten Darstellung der Grundwahrheit im CAE liegen (Abb. 6.4). Man kann diese formregularisierende Wirkung auch unabhängig von der sonstigen Architektur und Aufgabe als separaten Nachverarbeitungsschritt erzeugen, welcher dann generisch auf praktisch beliebige Segmentierungen anatomischer Formen angewandt werden kann [Larrazabal u. a., 2019].

Schiratti u. a. [2015] schlagen explizit das Lernen von Krankheitstrajektorien als Geodäten auf riemannschen Mannigfaltigkeiten vor, auf denen die beobachteten longitudinalen Daten liegen (etwa MR-Bilder). Mittels gemischter Modelle aus der Statistik lassen sich feste Effekte (durchschnittliche Trajektorie der Gruppe von Patienten einer Krankheit) und zufällige Effekte (dazu parallele patientenindividuelle Trajektorie) unterscheiden, wodurch letztere zu verschiedenen zeitlichen oder räumlichen Verschiebungen sowie unterschiedlichen relativen Geschwindigkeiten führen. Die Schwierigkeit besteht allerdings in der Konstruktion entsprechender Mannigfaltigkeiten für eine konkrete Krankheit, wofür es kein pauschales Vorgehen gibt.

Kürzlich haben Petersen u. a. [2019] auf Basis des probabilistischen U-Nets zur Segmentierung mehrdeutiger Bilddaten einen variationellen Ansatz genutzt, um Verteilungen möglichen Wachstums eines Hirntumors (Gliom) zu modellieren, welches umfang-

reiche Trainingsdatensätze erfordert. Die Methode entspricht ohne Skip-Verbindungen grob der Verwendung eines VAE – der variationellen Variante des in Kapitel 6.2 genutzten CAE. Allerdings ist beim Schlaganfall diese „Vieldeutigkeit“ durch das Wissen um das gesamte Risikogewebe als *worst case* sowie die Abhängigkeit von der zugrundeliegenden Gefäßanatomie nicht in dieser Form gegeben. Ebenso fehlen beim Schlaganfall longitudinale Beobachtungen, wodurch aus technischer Sicht der Einsatz regulärer Sequenzmodelle, wie Markov-Ketten, rekurrenten neuronalen Netzen oder dem Long-Short-Term-Memory von Hochreiter u. a. [1997] erschwert wird. Die Convolutional-LSTM von Zhang u. a. [2019] zur Prädiktion von länger andauerndem Tumorwachstum stellt eine Ausnahme der seltenen Verwendung dieser Ansätze dar.

6.2 Vorgeschlagene Methodik

Es wird in diesem Kapitel eine Methode vorgeschlagen, bei der zunächst die Segmentierungen von Kern und Risikogewebe vorliegen müssen. Wie schon in Kapitel 4.1 setzt sich der vollautomatische Ansatz aus zwei Netzwerken zusammen, um unterschiedliche Teilaufgaben zu lösen: Zum einen ein U-Net zur Segmentierungsschätzung von Risikogewebe und Kern auf den Perfusionskarten, so wie in Kapitel 4.2 vorgestellt; zum anderen ein CAE zur Wachstumsmodellierung der Läsionsmaske. Der CAE trainiert auf augmentierten Läsionsformen (zufällig elastisch deformierte Grundwahrheiten) einen Formraum, welcher die Interpolation der finalen Läsionsform ermöglicht und der die Hauptvariation der Kern- und Risikogewebesegmentierungen ohne Rauschen modelliert.

6.2.1 Zweiteilige Architektur aus U-Net und CAE

Das vollautomatisierte Verfahren besteht aus einem zweiteiligen neuronalen Netzwerk, das drei Hauptbestandteile zur automatischen formbeschränkten Vorhersage von Folgeschäden kombiniert (Abbildung 6.5):

1. U-Net-Schätzung von Kern (\hat{S}_c) und Risikogewebe (\hat{S}_{cp}) auf CBV und TTD.
2. CAE-Transformation der Segmentierungen in einen Formraum und zurück.
3. Lineare Interpolation im Formraum zur Vorhersage der Folgeläsion (\hat{S}_i).

Zunächst werden die Perfusionsbilder I_{CBV} und I_{TTD} von einem U-Net U verarbeitet, um die Segmentierungsschätzungen \hat{S}_c und \hat{S}_{cp} zu berechnen. Zweitens wandelt der Encoder E des CAE jede Segmentierung in eine niedrigdimensionale Form und bettet damit deren Kodierungen \hat{z}_c und \hat{z}_{cp} in einen Formraum ein, der vorher gelernt werden muss. Die lineare Interpolation (\hat{z}_i , Gleichung 6.1) zwischen den latenten Kern- und Risikogeweberepräsentationen \hat{z}_c und \hat{z}_{cp} erfolgt nach der erwarteten $t_{A \rightarrow I}$, die um

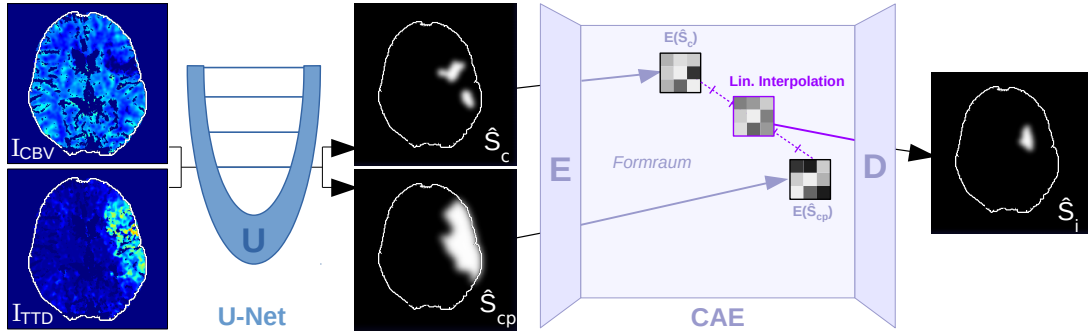


Abb. 6.5: Übersicht über das vorgeschlagene Verfahren, welches das U-Net zur Segmentierung von Kern (\hat{S}_c) und Risikogewebe (\hat{S}_{cp}) zeigt, die an den Encoder E des CAE weitergeleitet werden, um sie in niedrigdimensionale Formraumdarstellungen umzuwandeln, so dass sie linear interpoliert und durch D dekodiert werden können.

die verbleibende Zeit normalisiert werden muss, um 10 h nach Beginn zu erreichen (entspricht hier dem Auswachsen der Läsion auf das gesamte Risikogewebe).

$$\hat{z}_i = \hat{z}_c + \eta(\hat{z}_{cp} - \hat{z}_c), \quad \eta = \frac{t_{A \rightarrow I}}{10 - t_{0 \rightarrow A}} \quad (6.1)$$

Diese lineare Interpolation im Formraum ist physiologisch begründet, da sie einer nichtlinearen Interpolation der rekonstruierten Formen auf der Mannigfaltigkeit entspricht (Abbildung 6.3). Der Decoder D des CAE ist erforderlich, um diese Rekonstruktion der interpolierten Kodierung \hat{z}_i im Bildraum mit einer Segmentierung \hat{S}_i für die endgültige Läsion als Ergebnis zu berechnen (Seite 130 veranschaulicht eine binarisierte 3D-Segmentierung).

6.2.1.1 Aufgabe von U-Net und CAE

Der Aufbau eines zweiteiligen Netzwerks zielt auf verschiedene Teilaufgaben ab, die das Erlernen der hochdimensionalen und komplexen Gesamtaufgabe der Vorhersage von Folgeschäden beinhaltet. Wenn man nur die endgültige binäre Läsionsklasse von Hintergrund zu Vordergrund durch hochdimensionale multivariate Eingaben unterscheiden wollte, wird es für einen maschinellen Lernalgorithmus schwierig, die Beziehung zwischen Eingabe und Ergebnis richtig zu verallgemeinern. Statt eine Segmentierung der Folgeläsion aus den Eingangsdaten direkt zu schätzen, wird daher zuerst ein U-Net zur korrekten Segmentierung von Kern und Risikogewebe genutzt, was – wie bereits erläutert – für die Vorhersage des endgültigen Läsionsergebnisses von großer Bedeutung ist. Sobald diese Daten zur Verfügung stehen, kann als zweites der CAE die wichtigsten Formmerkmale in einer latenten, vereinfachten Darstellung in Bezug auf die Aufgabe (Formwahrscheinlichkeitskarten in Abhängigkeit von verschiedenen physiologischen

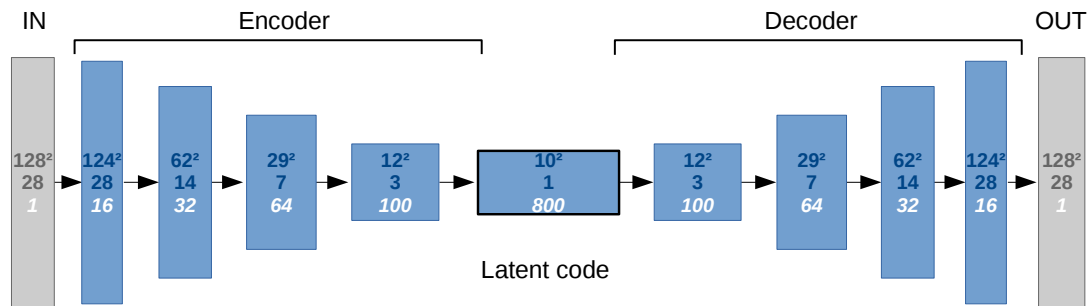


Abb. 6.6: CAE-Architektur mit Größe der ersten beiden Raumdimensionen, der dritten Raumdimension, sowie der Merkmalsanzahl (jeweils von oben nach unten gelesen).

CTP-Parametern) zusammen mit klinischen Daten zur Abschätzung der Folgeläsion lernen.

6.2.1.2 Aufbau des CAE

Mit Blick auf den anschließenden CAE sollte eine ausreichende Zahl von Schichten (Abbildung 6.6) sichergestellt werden, um die markantesten, beschreibenden Merkmale der einzelnen Läsionsformen erkennen zu können. Gleichzeitig sorgt eine entsprechende Regularisierung dafür, dass eine im Mittel gut generalisierte Schätzung der Form rekonstruiert wird. Dies erfordert einen Flaschenhals zwischen E und D , um die niedrigdimensionalen, latenten Kodierungen zu erzeugen, die eine begrenzte, aber ausreichende Dimensionalität aufweisen müssen. Folglich sollten in einer Kodierung nur die Hauptmodi der wesentlichen Formvariationen von Schlaganfallläsionen dargestellt werden, die verschiedene Läsionsformen ohne Rauschen oder Überanpassung von Trainingsproben beschreiben können. Was die Überpassung betrifft, so ist der vorgestellte lineare Ansatz für die Interpolation auf den niedrigdimensionalen Kodierungen grundsätzlich auch robust gegenüber Rauschen in den gegebenen klinischen Zeitpunkten.

Die Eingabe des CAE wird ähnlich wie beim U-Net durch doppelte Faltungsschichten pro Block einschließlich Batch-Normalisierungsschichten verarbeitet. Statt eines Max-Poolings folgt auf diesen jeweils eine $2 \times 2 \times 2$ -Faltung mit Laufweite 2 (*engl.* Stride), während es sich beim letzten Block nur noch um eine einzelne $3 \times 3 \times 3$ -Faltung handelt, die die Merkmalskarte auf die Größe $10 \times 10 \times 1$ des Flaschenhalses abbildet. Wie beim U-Net werden innerhalb der Blöcke des CAE ausschließlich $3 \times 3 \times 3$ -Filterkerne verwendet, um die Komplexität im Vergleich zu Netzwerken mit größeren Kernen zu verringern, während die gleichen rezeptiven Feldgrößen durch einfaches Stapeln mehrerer Schichten mit kleinerem Kernel erreicht werden können – dies führt zu mehr Nichtlinearitäten und generalisiert in der Regel besser. Im Decoder als Spiegelbild zum Encoder werden schließlich die $2 \times 2 \times 2$ -Faltungen durch Dekonvolutionsschichten ersetzt. Mit dieser Architektur kann ein einzelnes Segmentierungsbild mit $128 \times 128 \times 28$

großen Formen in eine niedrigdimensionale Darstellung kodiert und dann wieder zu einem Segmentierungsbild dekodiert werden.

Die latente Repräsentation ist idealerweise ein Vektor, welcher durch eine vollverbundene oder Faltungsschicht mit entsprechender Kernelgröße davor erzeugt wird. Dies erleichtert die Quantifizierung der Dimensionalität des Formraums. Allerdings gestaltete sich in den Experimenten die Regularisierung eines solchen äußerst schwierig – auch bei Verwendung von Dropout [Srivastava u. a., 2014] – und die Rekonstruktion war weniger genau als bei Verwendung latenter räumlicher Merkmalskarten der Größe $10 \times 10 \times 1$. Der CAE besteht folglich ausschließlich aus Faltungsschichten zusammen mit der Kombination η der beiden Zeitprädiktoren als klinischen Variablen für die Interpolation gemäß der Definition in Gleichung 6.1. Zudem war eine Mindestanzahl von etwa 200 latenten Kanälen für eine einigermaßen genaue Rekonstruktion notwendig. Das Erhöhen der Kanalanzahl auf 800, wie in Abbildung 6.6, bringt dann nur noch geringfügige Verbesserungen unter gleichbleibend robustem Training.

6.2.2 Zweistufiges Training von Formraum und Prädiktion

Das Training wird in drei aufeinander folgenden Phasen durchgeführt, die sich durch unterschiedliche Ziele auszeichnen, die in ihren entsprechenden „ L “-Fehlerfunktionen formuliert sind:

0. Training des U-Net U mittels $L_{SoftDice}$ (Vorverarbeitung, Kapitel 4.2)
1. Training von E_1 und D des CAE zum Erlernen eines Formraums aus manuellen Segmentierungen (L_{Shape})
2. Training von E_2 bei festem D zur Einbettung automatischer Segmentierungsschätzungen von U in den gelernten Formraum ($L_{Prediction}$)

Das U-Net wird zunächst wie in Kapitel 4.2 trainiert und stellt einen Vorverarbeitungsschritt dar („Schritt 0“). Die Segmentierungsgrundwahrheiten S_c und S_{cp} von Kern und Risikogewebe werden verwendet, um ihre Vorhersagen mit einem größeren Fehler bei weniger Überlappung im $SoftDice$ -Maß zu bestrafen, wie in Kapitel 3.1.3 definiert. Darauf erfolgt das eigentliche Lernen der Forminterpolation in zwei Schritten: Das Trainieren eines Formraums und die Anpassung für die durch das U-Net geschätzten Formen zur Forminterpolation, was in Abbildung 6.7 zusammengefasst wird.

6.2.2.1 Lernen eines Läsionenformraums

Zuerst wird ein niedrigdimensionaler Formraum erlernt, der die Formgrundwahrheiten von Kern- und Risikogewebe-Segmentierungen, S_c und S_{cp} , einbettet. Dies wird durch

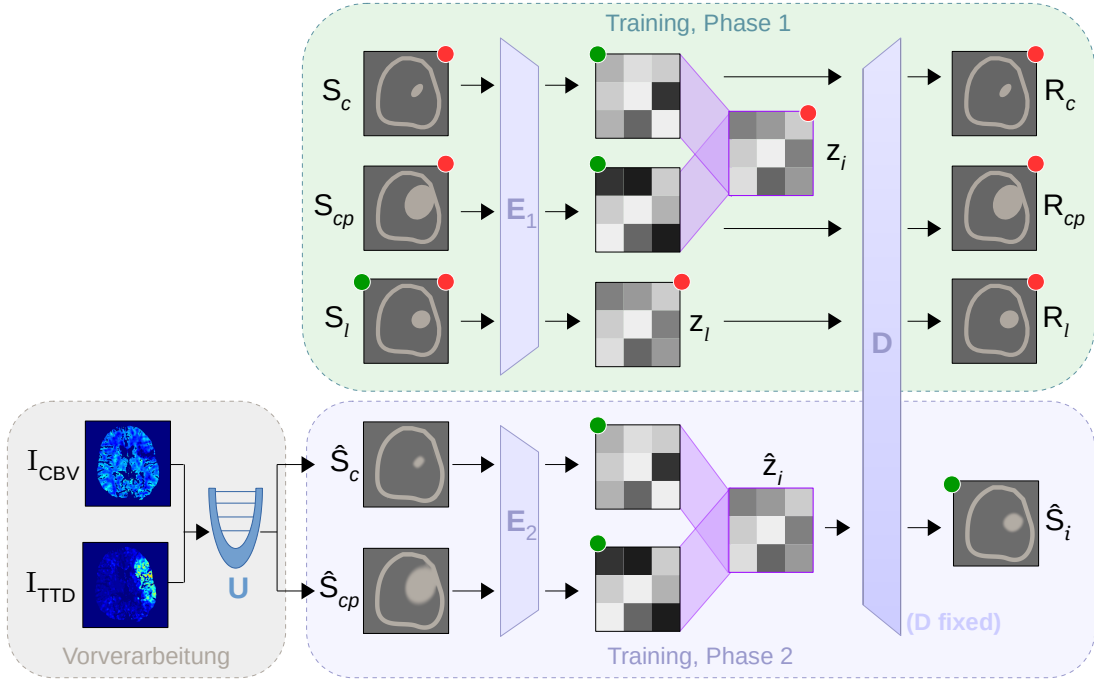


Abb. 6.7: Training des Forminterpolationsmodells: Zunächst wird der Formraum (Phase 1) über Rekonstruktionen R von Segmentierungen S gelernt, wobei die Interpolation z_i der Kodierungen von Kern- und Risikogewebe (Kern+Penumbra) nahe der Kodierung der tatsächlichen finalen Läsion z_l liegen muss. In Phase 2 werden dann die zuvor von U geschätzten Segmentierungen \hat{S} in den Formraum kodiert, sodass diese nahe den wahren Kodierungen aus Phase 1 liegen. Die farblichen Marker geben an, ob der entsprechende Daten-Tensor bei der L_{Shape} -Berechnung (•) oder bei $L_{Prediction}$ -Berechnung (•) berücksichtigt wird. Die Interpolation ($>$) in beiden Phasen wird dabei nach Formel 6.1 berechnet.

den Fehler L_{Shape} in Gleichung 6.2 erzwungen, der aus drei Teilen besteht: (1.) Die Rekonstruktion R_c , R_{cp} und R_l von Kern, Risikogewebe und endgültiger Läsionsform, (2.) die Eigenschaft des rekonstruierten Kern- bzw. Läsionvolumens noch eine Teilmenge des gesamten Risikogewebe-Volumens zu sein, und (3.) ein L_1 -Fehler zum Positionieren der latenten Kodierungen z_l der Läsionsform nahe der linearen Forminterpolation z_i aus Kern- und Risikogewebe-Kodierung:

$$L_{Shape} = \sum_{s \in \{c, l, cp\}} L_{SoftDice}(R_s, S_s) + \sum_{s \in \{c, i\}} L_{mono}(R_s, R_{cp}) + \alpha L_1(z_l, z_i) \quad (6.2)$$

Für die ersten 25 Epochen gilt $\alpha = 0$, ansonsten $\alpha = 1$. Es konnte beobachtet werden, dass dies der nicht-konvexen Optimierungsfunktion hilft, zuerst zu lernen, die

Form in die richtige Hirnhälfte zu rekonstruieren (Abb. B.11 auf Seite 131). Das α hilft robust zu verhindern, dass das Netzwerk zu Beginn des Trainings in unplausiblen Minima gefangen wird. Für das Minimierungsproblem wird der *SoftDice* (Kapitel 3.1.3) als Fehler $L_{SoftDice}(S, \hat{S}) = -SoftDice(S, \hat{S})$ definiert. Es ist wünschenswert, dass die Interpolation aus dem Formraum monoton entlang der Zeitachse vom Kern zum Risikogebiet wächst. Daher definiert man für zwei Segmentierungsbilder A und B eine Randbedingung L_{mono} für alle Voxelpositionen i , so dass die Rekonstruktionen der Form von Kern-Segmentierung und Zwischeninterpolation der Läsion hin zur gesamten Risikogewebe-Segmentierung monoton zunehmen:

$$L_{mono}(A, B) = \sum_i \max(A_i - B_i, 0) \quad (6.3)$$

6.2.2.2 Interpolation geschätzter Läsionssegmentierungen

In der zweiten Trainingsphase werden die Gewichte von Encoder und Decoder aus der vorhergehenden Formraum-Lernphase fixiert. Ein zweiter Encoder wird dann für die U-Net-Vorhersagen \hat{S}_c und \hat{S}_{cp} von Kern und Risikogewebe gelernt, um latente Darstellungen \hat{z}_c und \hat{z}_{cp} zu berechnen, die sich in der Nähe der Formrepräsentationen z_c und z_{cp} der dazugehörigen Segmentierungsgrundwahrheiten befinden (in Bezug auf die L_1 -Norm). Der $L_{Prediction}$ -Fehler in Gleichung 6.4 erzwingt weiterhin die monotonen Eigenschaften für die rekonstruierten Segmentierungen und stellt das Hauptziel einer hohen Überlappung für die Vorhersage \hat{S}_i aus der interpolierten Kodierung \hat{z}_i gegenüber der tatsächlichen Grundwahrheit S_l sicher:

$$L_{Prediction} = L_{SoftDice}(\hat{S}_i, S_l) + \sum_{s \in \{c, i\}} L_{mono}(\hat{S}_s, \hat{S}_{cp}) + \sum_{s \in \{c, cp\}} L_1(\hat{z}_s, z_s) \quad (6.4)$$

Damit transformiert der Decoder D der ersten Phase eine Darstellung aus dem Formraum derart, dass die Rekonstruktion der Kern-Schätzung $(D \circ E_2)(\hat{S}_c) \approx (D \circ E_1)(S_c)$ und der Risikogewebe-Schätzung $(D \circ E_2)(\hat{S}_{cp}) \approx (D \circ E_1)(S_{cp})$ nahe an der Rekonstruktion der Grundwahrheiten von Kern und Risikogewebe liegen sollte. Darüber hinaus bleibt das Hauptziel, eine Interpolation so nah wie möglich an der echten finalen Läsionssegmentierung zu erreichen:

$$\hat{S}_i = D \left(\eta \cdot E_2(\hat{S}_{cp}) + (1 - \eta) \cdot E_2(\hat{S}_c) \right) \approx (D \circ E_1)(S_l) \approx S_l \quad (6.5)$$

Die einzelnen Bestandteile des Trainingsfehlers wurden nicht durch weitere manuelle Parameter gewichtet, da kein Nutzen aus anderen als den einheitlich gewichteten Fehlerbestandteilen für L_{Shape} und $L_{Prediction}$ erkenntlich wurde. Darüber hinaus wurde auch versucht, beide Trainingsfehler alternierend oder *end-to-end* zu optimieren, aber damit konnten während der Optimierung keine sinnvollen Minima gefunden wer-

den. Insbesondere das ausschließliche Auftreten eines Schlaganfalls auf einer der beiden Hemisphären konnte dann nicht frühzeitig gelernt werden, was das Netzwerk jedoch grundsätzlich bereits in den ersten Epochen des Trainings erkennen sollte.

6.3 Experimente

Es wird die 5-fache Kreuzvalidierung mit dem UKSH-Datensatz von 29 Probanden durchgeführt (Seite 32). Von jedem Probanden wurden die beiden Zeiträume $t_{0 \rightarrow A}$ und $t_{A \rightarrow I}$ verwendet, die entsprechend der Gleichung 6.1 normiert worden sind. Alle Formsegmentierungen wurden in jeder Epoche elastisch deformiert, um die begrenzten verfügbaren Trainingsdaten zum Erlernen allgemeiner Merkmale des CAE zu ergänzen, so dass die Darstellungen im Formraum robust rekonstruiert werden können.

Es wird der hier vorgestellte Ansatz mit dem in Kapitel 5 vorgestellten U-Net verglichen bzw. einer einfacheren Variante, die dem U-Net aus Kapitel 4.2 entspricht. Bei beiden wurden keine klinischen Zeitpunkte als Eingabe berücksichtigt, da sich dies initial bereits als nachteilig herausgestellt hat (der Vollständigkeit halber ist ein Vergleich mit zusätzlichen Zeit-Eingabekanälen in der Tabelle auf Seite 132 als „*Metad.*“ aufgeführt).

Unsere Teilaufgabe der linearen Interpolation entlang einer Trajektorie zwischen Kern und Risikogewebe erfordert Darstellungen dieser in einem geeigneten nichtlinearen Formraum, der vorher gelernt werden muss. Um den Vorteil der Durchführung in einem solchen Formraum auf einer Mannigfaltigkeit zu zeigen, wird das Verfahren mit der naiven Durchführung einer linearen Interpolation der Formsegmentierungen im beobachteten Bildraum Ω verglichen. Diese kann einfach mit dem gleichen η , wie zuvor in Gleichung 6.1 definiert, berechnet werden:

$$\hat{S}_i = \hat{S}_c + \eta(\hat{S}_{cp} - \hat{S}_c) \quad (6.6)$$

6.3.1 Ergebnisse

Die Anwendung des U-Nets aus Kapitel 5 auf dem UKSH-Datensatz führt zu einem mittleren Dice von 0,35 für die Prädiktion der Folgeläsion (Tabelle 6.1). Dies stellt ein ähnliches Ergebnisniveau dar, wie bei den ISLES-Wettbewerben auf den MR-Perfusionsdaten und kann als Referenz des Stands der Technik herangezogen werden.

Die Rekonstruktionsergebnisse unseres neu erlernten Modells zeigen zunächst die Fähigkeit, zeitabhängige Vorhersagen zu treffen und die Vorteile unseres CAE-Ansatzes unter Verwendung von Kern- und Risikogewebe-Segmentierungen darzustellen. Damit kann selbst mit einem kleinen Trainingsdatensatz eine gute Verallgemeinerung erzielt und eine nichtlineare Interpolation der Folgeläsion geschätzt werden. Im Vergleich zu einer manuellen Segmentierung wurde eine Dice-Überlappung von 0,46 erreicht. Betrachtet man zunächst nur die Rekonstruktion der Grundwahrheiten mit dem CAE

Segm.	Methode Prädiktion	Anzahl Parameter	Dice		
			Kern	Risikogew.	Prädiktion
<i>Experte</i>	CAE1-Orakel	$4,7 \cdot 10^6$	0,68 (R_c)	0,90 (R_{cp})	0,53 (R_i)
<i>Experte</i>	CAE1-Interp.	$4,7 \cdot 10^6$	0,68 (R_c)	0,90 (R_{cp})	0,46 (R_i)
U-Net	CAE1-Interp.	$3,6 \cdot 10^5 + 4,7 \cdot 10^6$	0,43	0,80	0,40
U-Net	CAE2-Interp.	$3,6 \cdot 10^5 + 4,7 \cdot 10^6$	0,55	0,81	0,43 (\hat{S}_i)
U-Net	Segm.-Interp.	$3,6 \cdot 10^5$	0,45 (\hat{S}_c)	0,81 (\hat{S}_{cp})	0,36
–	U-Net ^{FL}	$1,1 \cdot 10^7$	–	–	0,35

Tabelle 6.1: Dice-Resultate: Das „U-Net“ (Kapitel 4.2) berücksichtigt hier keine klinischen Metadaten. Zusätzlich wird die Prädiktion des U-Net^{FL} aus Kapitel 5 verglichen. *Experte* bezeichnet die Grundwahrheit der Segmentierungsmasken. Die Interpolation („Interp.“) ist auf einen Zeitbereich von 0–10 Stunden normiert. Für den CAE wurde jeweils angegeben, ob nur die erste (CAE1) oder beide Phasen (CAE2) trainiert wurden. Das Orakel optimiert den Interpolationszeitpunkt anhand der bestmöglichen Rekonstruktion der Grundwahrheit. Das beste Ergebnis für Segmentierung/Prädiktion ist jeweils **fett** markiert, die vorgeschlagene vollautomatische Methodik in **blau** (vollständige Tabelle auf Seite 132).

von 0,68 bzw. 0,90 für Kern und Risikogewebe, so stellt dies ein gutes Ergebnis dar. In einem Anwendungsfall, in dem ein klinischer Experte Nekrose und Risikogewebe manuell segmentiert, kann dies bereits eine hilfreiche Schätzung für die Beurteilung des erwarteten Behandlungsergebnisses nach einer Thrombektomie sein.

Allerdings garantiert auch eine sehr gute Rekonstruktion von Kern und Risikogewebe nicht automatisch eine sehr gute endgültige Läsionsschätzung (Abb. 6.8). Die Qualität und Schwere des Schlaganfalls in klinischen Routinedaten ist nicht immer vollständig in der Abgrenzung von Kern- und Penumbra-Bereichen allein kodiert, und einige der nachfolgenden Läsionssegmentierungen sind tatsächlich kleiner als die zugehörige Kernapproximation der CBV-Segmentierungen. Dies steht der hier getroffenen Annahme entgegen, dass der Kernbereich nur nekrotisches Gewebe beinhalten sollte, das nicht wiederhergestellt werden kann. Ursächlich dafür sind die Herausforderungen der manuellen Annotationen allein auf Basis von CBV- und TTD-Parametern.

Für das trainierte Modell konnte eine obere Grenze von 0,53 für ein lineares interpolationsbasiertes Läsionsvorhersage-Orakel (Tabelle 6.1) bestimmt werden, das nicht die wahre Behandlungszeit zur Vorhersage verwendet, sondern das optimale $\eta \in [0,1]$ kennt, welches die beste Überlappung mit der Grundwahrheit ergibt¹⁶. Abgesehen von dem in der Literatur beobachteten nichtlinearen Wachstum über die Zeit [Wheeler u. a., 2015] und dem Fehlen von Informationen aus dem Perfusionssignal könnte der

¹⁶Das Wissen über die genau passende Kodierung im gelernten Formraum („Code-Orakel“) erlaubt eine noch weit höhere Überlappung: Siehe Ergebnis-Tabelle von Kapitel 7 auf Seite 110.

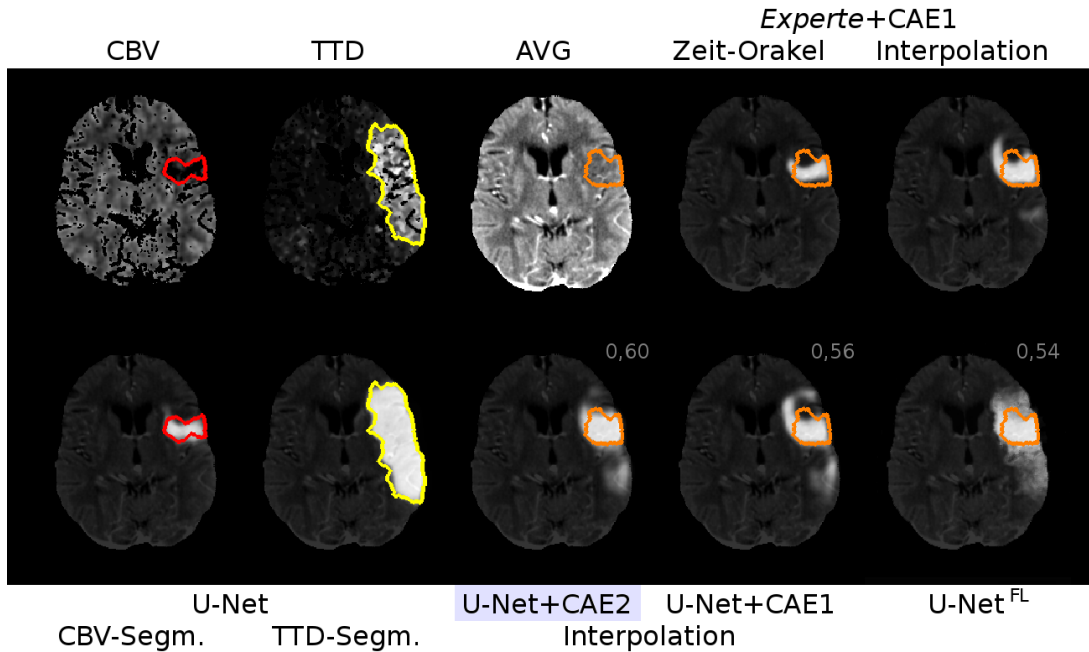


Abb. 6.8: Fall 17 ($t_{0 \rightarrow A} = 1 : 12$, $t_{A \rightarrow I} = 2 : 00$, $\eta = 0,23$): Gezeigt werden oben das CBV-, TTD- und AVG-Bild der CTP, sowie daneben die Formrauminterpolation auf den Grundwahrheiten. In der unteren Zeile werden dagegen die Schätzungen gezeigt (mittig: Ergebnis der vorgeschlagenen Methode). Zum Vergleich sind die Umrisse der Grundwahrheiten von CBV-Segmentierung (\bullet), TTD-Segmentierung (\bullet) und finaler Läsionsform (\bullet) eingeblendet (Werte in grau: Dice Übereinstimmung). Die Zeiten sind im Format *Stunden : Minuten* angegeben. Auf Seite 126-129 finden sich weitere Beispielsegmentierungen.

Unterschied zwischen 0,46 und 0,53 durch zu viel Rauschen in den Zeitdaten erklärt werden; insbesondere die Bestimmung von $t_{0 \rightarrow A}$ kann in der klinischen Praxis oft sehr ungenau sein. Während es wünschenswert wäre, die beste Interpolation aus Zeiten so nah wie möglich an der wahren Behandlungszeit zu rekonstruieren, ist der lineare Ansatz recht robust gegen ungenaue Zeiten und monotonen Wachstum. Die Ärztin ist im Wesentlichen daran interessiert zu sehen, ob es ein hohes relatives Wachstum geben wird und wie viel von dem Gewebe in den nächsten Stunden geborgen werden könnte.

Mit dem CAE lernt der zweite Encoder $E2$, die Segmentierungen aus dem U-Net in der zweiten Trainingsphase qualitativ hochwertig in den Formraum abzubilden (Tabelle 6.1). Diese Phase benötigt nur 50 Epochen für eine Konvergenz des Validierungssatz-Fehlers, verglichen mit 200 Epochen in der ersten Phase beim Training von $E1$ und D . Automatische Risikogewebe-Segmentierungen, die mit unserem U-Net-Modell berechnet werden, sind von sehr hoher Qualität (Dice von 0,81) und nahe an der optimalen Rekonstruktion angesichts der Einschränkungen des CAE (Dice von 0,90). Die Segmentierung des Kerns ist anspruchsvoller und ergibt Dice-Scores von nur 0,45, die

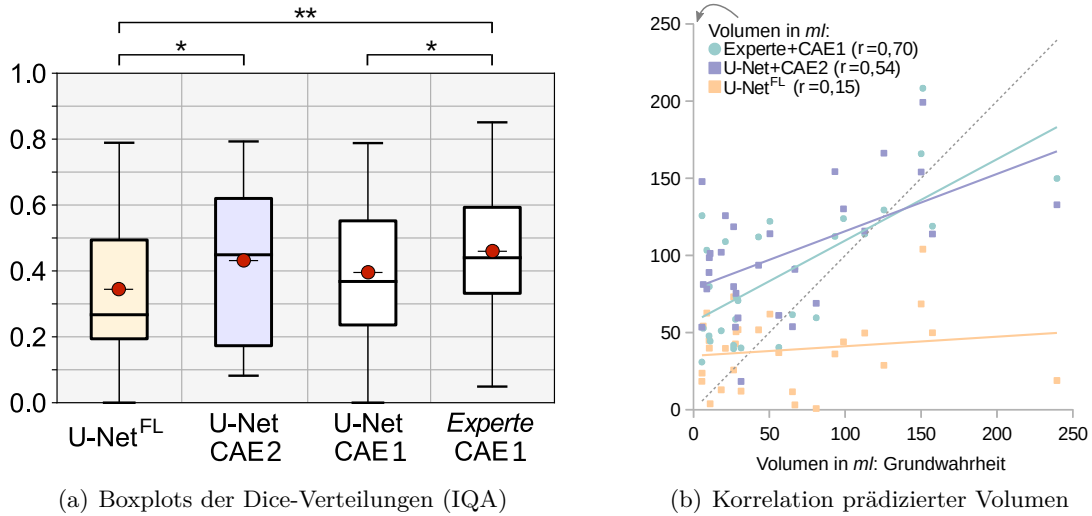


Abb. 6.9: (a) zeigt die Boxplots der Dice-Resultate (Box: Interquartilsbereich inkl. Medianlinie; roter Punkt: arithmetisches Mittel) für die Vorhersage der finalen Läsion durch die Interpolationsvarianten im Formraum („CAE-Interp.“ in Tabelle 6.1) und dem $U\text{-Net}^{\text{FL}}$ aus Kapitel 5 zum Vergleich. Die in Abschnitt 6.2 vorgeschlagene Methode $U\text{-Net}+\text{CAE}2$ erzielte dabei signifikant (\star) höhere Dice-Werte als das $U\text{-Net}^{\text{FL}}$. (b) zeigt die Korrelation ($-1 < r < +1$) der auf diesen Vorhersagen berechneten Volumene in ml (y-Achse) mit den Volumene der Grundwahrheit (x-Achse) – siehe auch Interrater-Reliabilitäten in Tab. B.1 auf Seite 131.

durch den CAE nach dem zweiten Training auf 0,55 verbessert werden. Dies bestätigt die Ergebnisse des von Ravishankar u. a. [2017] vorgeschlagenen formregularisierenden Netzwerks. Das Training des E2 führte in der Evaluation in jedem Fall dazu, dass die Dice-Ergebnisse der Läsionsvorhersage für ungesehene Testdaten nun signifikant höher sind, als mit der Vergleichsmethode aus Kapitel 5 (Abbildung 6.9(a)).

Die Interpolation zwischen den vom U-Net geschätzten latenten Formdarstellungen von Kern und Risikogewebe ist weniger genau als die Durchführung dieser Interpolation mit den latenten Darstellungen von Segmentierungsgrundwahrheiten. Dennoch wird der Vorteil der vorgeschlagenen Architektur von U-Net + CAE mit einem Dice von 0,43 für die Läsion deutlich, da das Ergebnis nahe an der Interpolation der Grundwahrheiten (Dice 0,46) liegt. Im Vergleich zur Referenzmethode aus Kapitel 5 (Dice 0,35), welche zu den besten Methoden aus dem ISLES-Wettbewerb zählt, ist eine signifikante¹⁷ Verbesserung bei Anwendung auf dem nicht-öffentlichen UKSH-Datensatz zu verzeichnen (Abbildung 6.9(a): \star bei $p < 0,05$, $\star\star$ bei $p < 0,01$).

¹⁷signifikant: $p < 0,05$ sowohl im gepaarten t-Test als auch ohne Normalverteilungsannahme im Wilcoxon-Vorzeichen-Rang-Test

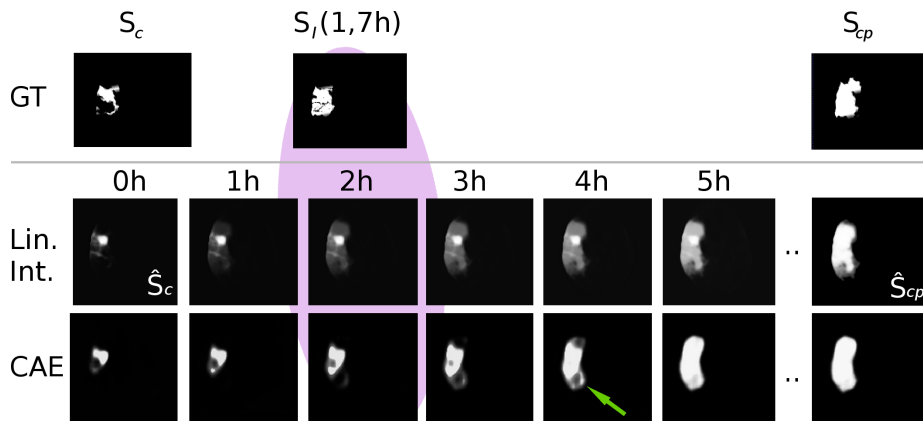


Abb. 6.10: Visueller Vergleich der linearen Interpolation in Bild- und Formraum für einen axialen Schnitt („GT“: Grundwahrheit). Man beachte die Vorhersagen bei 2 h mit der tatsächlichen Verlaufssegmentierung für die Intervention nach 1,7 h: Bei linearer Interpolation der Segmentierungen bleibt die Kernregion \hat{S}_c unverändert konstant stark, während die umliegende Penumbra region graduell eingebledet wird. Bei einer Normalisierung von 24 h würde die Einblendung über die Zeit dabei noch langsamer und damit unplausibler voranschreiten. Mit dem CAE und seiner Interpolation im Formraum wächst die Form nichtlinear, d. h. zunächst lokal und dann schnell in den umgebenden, durch das U-Net als \hat{S}_{cp} segmentierten Risikogewebereich hinein. Auch dabei lassen sich oft lokale Überblendungseffekte und topologische Löcher nicht gänzlich vermeiden (grüner Pfeil).

Darüber hinaus zeigt sich in Abbildung 6.9(b), dass die auf den endgültigen Läsionsmasken berechneten Volumen durchaus mit den tatsächlichen Volumen in *ml* korrelieren. Insbesondere bei Nutzung von Expertensegmentierungen kann man bereits davon sprechen, dass die prädizierten Volumen eine recht gute Einschätzung von der Größe des zukünftigen Infarkt volumens erlauben (Korrelationskoeffizient $r = 0,7$). Bei der vollautomatischen Variante kommt es jedoch vor allem bei kleineren Läsionen zu häufigen Überschätzungen des Volumens. Im Gegensatz zum U-Net^{FL} aus Kapitel 5 korrelieren aber auch diese Ergebnisse noch moderat mit den tatsächlichen Volumengrößen, während die Ergebnisse des U-Nets diese pauschal unterschätzen.

In der Abbildung 6.10 ist visuell zu beobachten, dass unsere lineare Interpolation im Formraum zu einem nichtlinearen Wachstum der Infarktformen führt: Zuerst im Lokalen, dann in das äußere Penumbra gebiet. Im Gegensatz dazu führt die einfache lineare Interpolation auf Segmentierungsschätzungen („Segm.-Interp.“ in Tab. 6.1) zu einem unplausiblen Verblässen der Infarkt wahrscheinlichkeiten des gesamten Risikogewebes im Voxelraum (siehe Abb. 6.10), wo die Progressionsrate zudem streng vom geeignet gewählten Normierungswert abhängt (siehe auch Tabelle Seite 132). In der Folge ergeben sich bei der Normierung mit 24 h weit schlechtere Ergebnisse als mit 10 h.

6.3.2 Diskussion

Schlussendlich ist zu sehen, dass die vollautomatische Prädiktion auf den Schätzungen von Kern und Risikogewebe mittels U-Net (Dice 0,43) nahe dem Ergebnis der Prädiktion auf Basis der Expertensegmentierungen (Dice 0,46) ist, obgleich die Rekonstruktionen der Segmentierungen von Kern und Perfusionsläsion (Risikogewebe) dabei noch besser gelingt. Demzufolge kann man zwar weiter versuchen die Abstände der \hat{z} - zu den z -Kodierungen zu verringern, aber das Potenzial für eine Verbesserung der finalen Interpolation in dem hier trainierten Formraum bleibt gering (maximal $\text{Dice} \leq 0,46$).

Dies liegt zum einen am Formraum, den man weitergehend untersuchen sollte. Es stellt sich beispielsweise die Frage, ob die Glattheit ausreichend ist, um darin Interpolationen durchzuführen. Die Nutzung eines sogenannten VAE („Variational Convolutional Auto-Encoder“) mit einer Wahrscheinlichkeitsverteilung als Nebenbedingung könnte eventuell glattere Interpolationen auch zwischen den existierenden Datenstichproben erzeugen. Insbesondere bei der Interpolation von dazwischen liegenden Punkten wäre bei Nutzung von Kern- und Risikogewebe-Schätzungen sichergestellt, dass kleine Abweichungen von der Trajektorie der Grundwahrheit keine abrupte Änderung der Interpolation verursachen.

Zum anderen liegt das begrenzte Verbesserungspotenzial an der Datengrundlage. Die Probandenzahl ist trotz großer Variation stark limitiert ($n < 30$). In Bezug auf die Zeit-Metainformationen, welche für den Interpolationszeitpunkt entscheidend sind, darf von einem nicht zu unterschätzendem Rauschen und Unsicherheit in den Daten ausgegangen werden, was insbesondere die Zeitdauer seit dem Symptombeginn betrifft (geschätzter Schlaganfallbeginn). Deutlich wird dies, wenn man den Zeitpunkt auf der genormten Zeitachse von 0–10 Stunden mittels der bekannten finalen Läsion optimiert, wo ein Dice von 0,53 statt 0,46 für die Vorhersage erreicht werden kann.

Der starke strukturelle Bias ermöglicht zwar ein gut gestelltes Lernproblem, könnte aber durch Einführung eines Offsets für die lineare Trajektorie im Z -Raum noch etwas abgeschwächt werden. Couronne u. a. [2019] haben auf Basis der Arbeit von Louis u. a. [2019] für die Anwendung multimodaler Daten gezeigt, wie man entsprechende Parameter einer gemeinsamen patientenindividuellen Trajektorie auf der Mannigfaltigkeit eines CAE lernen kann. Neben dem Offset wird dafür noch eine Progressionsrate der Trajektorie geschätzt. Eine weitere Einschränkung des hier vorgestellten CAE stellen die Eingabeinformation dar, wofür ausschließlich Binärmasken der Läsionen verwendet werden. Eine umfangreichere Untersuchung, welche Intensitätsinformationen aus den unterschiedlichen CTP-Modalitäten als zusätzliche Prädiktoren für die Vorhersage der finalen Läsionen dienen könnten, sollte vor einem klinischen Test des Verfahrens zunächst berücksichtigt werden.

In diesem Kapitel wurde eine neue vollautomatische Methode entwickelt, die die bisherigen Deep-Learning-Ansätze der traditionellen Segmentierungsnetze signifikant

übertrifft. Das zweistufige Verfahren hat sich als guter Ansatzpunkt für die Modellierung der Läsionsevolution durch Nutzung von Formvorwissen herausgestellt. Die Vergleichbarkeit vom „U-Net+CAE2“-Ansatz für andere Daten als den hier verwendeten proprietären UKSH-Datensatz wurde durch eine indirekte vergleichende Evaluation mit dem bei der ISLES 2017 zweitplatzierten U-Net gezeigt, welches hier nur einen geringfügig höheren Dice-Durchschnitt (0,35) als auf den ISLES-Daten erzielt, aber signifikant unter der CAE-Methodik liegt. Es ist anzunehmen, dass „U-Net+CAE2“ mit entsprechend verfügbaren Annotationen für die ISLES-Daten (Kern- und Risikogewebe-Segmentierungen) ebenfalls bessere Ergebnisse als die reinen U-Net-Ansätze bei einem Vergleich mit der ISLES 2017 erreichen würde.

Das nächste Kapitel basiert ebenfalls auf dem Vorwissen um Kern und Risikogewebe, setzt aber den Schwerpunkt nicht weiter auf eine Vollautomatisierung, sondern befasst sich mit der Frage, wie ein noch besser geeigneter, abstrakter Raum für die Wachstumsmodellierung aussehen könnte.

7 Formdeformationsräume zur Modellierung von Infarktwachstum

Auf Grundlage einer kleinen Datenbasis ist es schwierig, die in Kapitel 6 vorgestellten Formräume zur Prädiktion von finalen Schlaganfallläsionen derartig zu trainieren, dass ein plausibles Auswachsen der Läsionsform stattfindet. In diesem Kapitel wird ein alternativer Modellraum für das Schlaganfallwachstum entwickelt und untersucht. Im Gegensatz zur Interpolationsmethode aus dem vorherigen Kapitel basieren die Schätzungen der finalen Läsion durch Verformung der Kern- oder Risikogewebesegmentierung entlang zuvor trainierter Hauptverschiebungsrichtungen auf topologieerhaltenden Deformationen. Es zeigt sich, dass dies auf idealen Daten, die den klinischen Annahmen entsprechen, eine geeignetere Prognose der finalen Läsionsform ermöglichen kann. Teilergebnisse und Methodik aus diesem Kapitel wurden in [Lucas u. a., 2020a,b] veröffentlicht.

Die Evolution der Läsion nach einem Schlaganfall ist ein dynamischer Prozess über einen bestimmten Zeitraum, der auf den initialen, plötzlichen Auslöser und Symptombeginn folgt. Die Behandlungsplanung nach der neuroradiologischen Aufnahme basiert jedoch auf einem Mismatch-Verfahren statischer Bilder, vor allem dem Vergleich von nekrotischem zu potenziell heilbarem Gewebe, welches die Dynamik dieser Evolution nicht realistisch abbilden kann. Ein Ziel dieses Kapitels ist es, eine stärkere Überwachung (*engl.* Supervision) des Prozesses für ein maschinelles Lernverfahren zu erzeugen. Mehr Beobachtungen als zur initialen Aufnahme oder von Nachsorge-Terminen können jedoch nicht erhoben werden. Darum herrscht ein Mangel an entsprechend verfügbaren Daten für die kontinuierliche Modellierung des Gewebezustandes, wie in Kapitel 6 deutlich geworden ist: Räumlich-zeitliches Läsionswachstum wird dort auf Basis von sehr begrenzt vorhandenem Vorwissen um die Form von Kern, Risikogewebe sowie der Läsion bei der Nachsorge nach erfolgter Intervention versucht zu schätzen, was unter anderem zu unplausiblen Zwischenzuständen der Evolution führen kann (siehe Abb. 6.10 auf S. 96).

Daher müssen andere Möglichkeiten zur Modellierung des Läsionswachstums gefunden werden. Die räumlich-zeitliche Evolution kann durch Vektoren mit unterschiedlichen Richtungen und Magnituden als eindeutige Abbildung von Ursprungsform

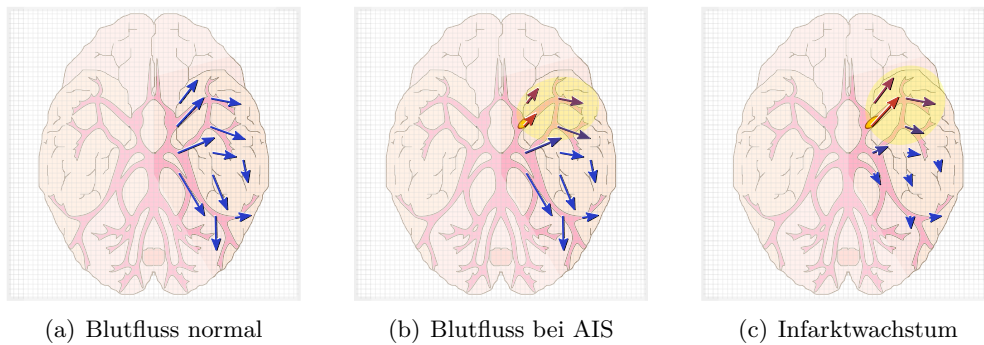


Abb. 7.1: Motivierung des Methodenwechsels: Der sauerstoffreiche Blutzufluss soll durch Vektoren an einigen Abtastpositionen in der linken Hemisphäre repräsentiert werden. Interpretiert man die Vektorrichtung und -magnitude als Blutfluss, folgen die Vektoren den Hirnarterien (a) und kodieren so indirekt die patientenindividuelle Hirngefäßanatomie. Bei einem akuten ischämischen Schlaganfall (AIS) reduziert sich durch den Verschluss (gelb) der Fluss und die Sauerstoffzufuhr (b). Je weniger Sauerstoff, desto schneller das Läsionswachstum (c).

zu Endform der Läsion dargestellt werden. In diesem Kapitel soll das Schlaganfallwachstum durch Vektorfelder regularisiert werden. Das oftmals fehlende Wissen um die vorliegende patientenspezifische Hirnanatomie kann dabei implizit abgebildet werden. Aus einer Menge von diffeomorphen Registrierungsfeldern werden die Hauptrichtungen der Deformation von der Kernläsion bis zum Erreichen der gesamten Risikogewebeausbreitung bestimmt. Es wird idealisiert angenommen, dass die finale Läsion genau einen Zwischenschritt dieser Deformation darstellt und nicht einzelne nekrotische Herde entstehen, sondern die Läsion entlang der unterversorgten Arterien auswächst (Abbildung 7.1).

7.1 Literaturüberblick

Der typische Einsatz von Vorhersagetechniken in der Bildverarbeitung betrifft das Präzifizieren der Segmentierungen einzelner Frames in einem Video. Auch dabei hat man das Problem lediglich schwach überwachter Trainingsdaten, da es enorm aufwändig ist, die Vielzahl an Frames händisch zu annotieren. Es sollen dabei (semi-)automatisch Segmentierungen möglichst genau und zueinander konsistent über die gesamte Videosequenz geschätzt werden. Das geschieht etwa durch Berücksichtigung initial segmentierter Frames für die darauffolgenden Frames [Caelles u. a., 2017; Perazzi u. a., 2017] oder über die Berücksichtigung des optischen Flusses [Cheng u. a., 2017]. Ein Problem in Bezug auf die Schlaganfallsegmentierung stellt die sehr geringe zeitliche Dichte der Informationen aus den eigentlichen Bildframes zu den verschiedenen Zeitpunkten in der

regulären Abfolge dar, da hierbei nur einzelne Segmentierungen bzw. Bilder vorliegen (zum einmaligen Aufnahmezeitpunkt bzw. vom Zeitpunkt des Verlaufsbildes).

7.1.1 Integration von PCA und CNN

Die Hauptkomponentenanalyse (PCA, Abschnitt 3.2.1) eignet sich gut, um repräsentative, niedrigdimensionale Kodierungen zu erzeugen, die die Hauptinformationen der Daten enthalten. Als robuster, linearer Ansatz eignet sie sich in einem ersten Schritt zur Erstellung darauf aufbauender, trainierbarer neuronaler Netzwerke für komplexe Probleme oder Aufgaben, bei denen Vieldeutigkeiten in den Eingabedaten die Schätzung erschweren, sodass damit zusätzliches Vorwissen in das PCA-Modell integriert wird. Foote u. a. [2019] erstellen etwa vor der Intervention aus 4D-CT-Atemsequenzen ein patientenindividuelles Bewegungsmodell, für welches während der Intervention über ein neuronales Netzwerk die korrekten Parameter aus einer Radiographie-Projektion geschätzt werden, um die Bewegungskomponenten des Modells für die Rekonstruktion der Atemsequenz-Bewegungen in Echtzeit zu gewichten.

Die Integration von Vorwissen über die Herzform haben Milletari u. a. [2017] für die herausfordernde Detektion von Landmarken oder Segmentierungen des Herzens auf Ultraschallsequenzen vorgeschlagen, wobei sie das Formmodell mittels einer PCA erzeugen. Mithilfe einer passenden PCA-Schicht direkt in einem CNN kann durch implizites End-To-End-Lernen der passenden Koeffizienten die Eigenbasis im Training eines doppelarmigen Netzwerks für die grobe und feine Bestimmung von Key-Points berücksichtigt werden, mit denen Landmarken detektiert und die Zielstruktur segmentiert werden kann. Bemerkenswert ist noch das zusätzliche Schätzen einer Verschiebung in der PCA-Schicht, die auf die rekonstruierten Key-Point-Positionen addiert wird. Die Evaluation ist jedoch schwierig zu interpretieren, da die Durchschnittswerte ungefähr gleich sind, sich aber die Standardabweichung sowie Minimum und Maximum im Vergleich zum normalen U-Net unterscheiden, was jedoch nicht konkret besprochen wird.

7.1.2 Integration von räumlichen Transformationen im CNN

In der obigen Arbeit wurde für den Verfeinerungs-Arm ein Bereich rund um einen Key-Point zugeschnitten. Diese Transformation ist beim Training end-to-end im Netzwerk ableitbar und orientiert sich am STN (*engl.* Spatial Transformer Network) von Jaderberg u. a. [2015], welches es ermöglicht, Transformationen implizit mitzulernen, um für bestimmte Aufgaben vor allem geeignetere „kanonische“ Darstellungen der Eingabe zu erhalten. Dabei besteht ein STN aus drei Teilen (Abb. 7.2): Ein Lokalisierungsnetzwerk, welches auf der Eingabe die Parameter der Transformation schätzt; ein Abtastgitter-Generator, der ein Gitter an Koordinaten auf dem Eingabebild definiert, welches zu den Pixeln des Ergebnisbildes korrespondiert; sowie ein Interpolator, welcher die Eingabe auf dem Gitter neu abtastet und damit die Ausgabe interpoliert. Der

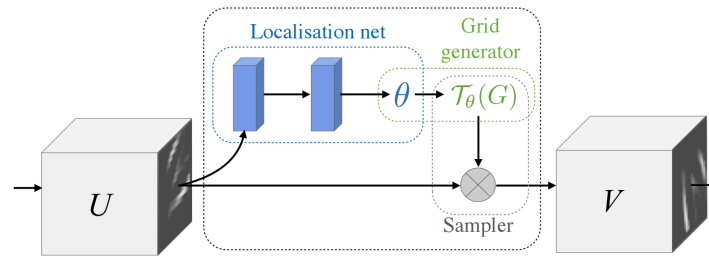


Abb. 7.2: Aufbau eines Spatial Transformer Netzwerks: U ist eine Merkmalskarte als Eingabe, welche in das Lokalisierungsnetz gereicht wird, um die Parameter θ zu regrestriieren. Das reguläre Gitter G der Ausgabe V wird zum Abtastgitter $\mathcal{T}_\theta(G)$ auf U transformiert, um die deformierte Ausgabe V zu interpolieren. Aus: Jaderberg u. a., 2015

entscheidende Vorteil der STN ist dabei das implizite Lernen geeigneter geometrischer Transformationen in Bezug auf die jeweilige Aufgabe des Netzwerkes. Die Transformationen sind dabei nicht auf 2D oder affine Verformungen beschränkt – z. B. lassen sich mit dem „GridSampler“ der PyTorch Bibliothek für ein dreidimensionales Abtastgitter beliebige Verschiebungen im kontinuierlichen Raum interpolieren [Paszke u. a., 2017].

Yu u. a. [2016] nutzen STN ebenfalls als Basis zur Regression von Landmarkenpositionen in einem „Point Transformer Network“ (PTN), welches der zweite Teil eines kaskadierten end-to-end trainierbaren Netzes ist. Mit diesem kann schrittweise zuerst eine typische Form rekonstruiert werden, welche dann feiner deformiert wird. Den ersten Teil bildet dafür ein sogenanntes „Shape Basis Network“, welches die Koeffizienten eines gegebenen PCA-Modells schätzt und als Initialisierung für das PTN dient. Mit diesem wird dann eine TPS-Transformation geschätzt (diffeomorphe „Thin Plate Spline“), welche die Landmarkenpositionen in Bezug auf die Eingabe korrigiert.

Ein STN-artige Lösung benutzen auch Qin u. a. [2018] zum gemeinsamen Lernen der Herzbewegung und Segmentierung in MR-Sequenzen. Sie verzichten dabei auf Vorwissen etwa in Form von PCA-Modellen. Sie versuchen durch das Aufteilen der Parametergewichte im Segmentierungs- wie Bewegungsschätzungsarm gemeinsame Merkmale zu lernen. Es ermöglicht durch die Bewegungsschätzung eines rekurrenten Netzwerkes für die Herzsequenz auch das Training mit schwacher Überwachung, bei dem nicht für alle Frames Annotationen vorhanden sind. Rekurrente Netzwerke eignen sich generell sehr gut für reguläre Sequenzprobleme, sodass sie vor allem für zeitlich-reguläre Sequenzdaten genutzt werden können. Beispielhafte Anwendungen finden sich in der Herzsegmentierung [Poudel u. a., 2017], wobei nicht zwangsläufig 4D-Daten zu nutzen sind, sondern auch für iterative Ansätze zur einzelnen 3D-Segmentierung verwendet werden können [Pace u. a., 2018; Wang u. a., 2016].

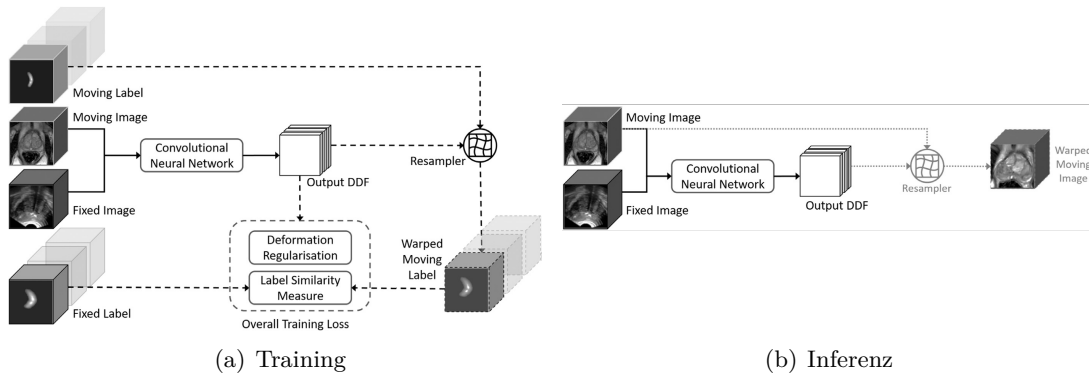


Abb. 7.3: Das Registrierungslernen basiert zumeist auf dem Lernen von Merkmalen auf Intensitätsbildern für die Schätzung von Transformationsparametern (a), wobei im Trainingsfehler auch die Label berücksichtigt werden. Im Anwendungsfall werden für die Inferenz dann (nur) die Intensitätsbilder benötigt (b). Aus: Hu u. a., 2018

7.1.3 Bildregistrierung mittels CNN

Mittlerweile werden auch end-to-end trainierbare Netzwerke für die reine Bildregistrierung genutzt, die aufgrund der Verlagerung des iterativen Optimierungsaufwandes vom Anwendungsfall hin zum Training ausgesprochen schnelle Ergebnisse liefern.

Idealerweise sollte als Ziel ein Deformationsfeld berechnet werden und demzufolge dieses in einem klassischen stark-überwachten Training auch als Grundwahrheit zur Verfügung stehen. Dabei besteht immer die Frage, wie man eine tatsächliche Grundwahrheit erhält, ohne diese durch ein anderes Verfahren (etwa ein klassischer Registrierungsalgorithmus) nur anzunähern und dessen Fehler im Training zu übernehmen. Rohé u. a. [2017] stellen das „SVF-Net“ vor, welches stationäre Geschwindigkeitsfelder (*engl.* Stationary Velocity Fields, SVF) anhand von Trainingsreferenzen schätzt. Diese werden über ein morphometrisches Optimalitätsverfahren zum Matching der Segmentierungslabels erzeugt. Statt des SVFs können auch beliebige andere Vektorfelder als Trainingsziel genutzt werden.

Hu u. a. [2018] präsentieren eine Methode, die auf schwach-überwachten Daten in Form von Segmentierungslabels trainiert werden kann, weil diese die Berechnung des Trainingsfehlers auf korrespondierenden Strukturen ermöglichen. Für die Inferenz im Anwendungsfall werden diese Annotationen dann nicht benötigt, um eine Registrierung durchzuführen. Die Merkmale zur Schätzung der Deformation werden dabei direkt auf den Bilddaten extrahiert, d. h. auch hier bilden Intensitätsbilder die Arbeitsgrundlage (Abb. 7.3).

Schließlich kann man auch gänzlich unüberwacht den traditionellen Optimierungsprozess im Training eines CNNs nachahmen und direkt eine Ähnlichkeitsmetrik als Trainingsfehler auf den Bildeingaben nutzen. Dalca u. a. [2019] haben dies für ihren

variationellen Ansatz vorschlagen, welcher mittels sogenannter Integrationsschichten (differenzierbare Umsetzung von „Squaring and Scaling“ der SVF) in einer Architektur während des Trainings end-to-end sicherstellt, dass die resultierenden Deformationsfelder diffeomorph sind. Die Anzahl der dafür benötigten Integrationsschichten muss zunächst grob problembezogen empirisch bestimmt werden.

Aus der Literatur sind verschiedene Herangehensweisen zur Überwachung des Trainingsprozesses ersichtlich, um Eingabedaten und Labels zu deformieren. Allerdings basieren die Registrierungsansätze zum überwiegenden Teil auf zu jedem Label korrespondierenden Intensitätsbilddaten, auf denen fortgeschrittene Merkmale zur Regression der Deformationsparameter extrahiert und gelernt werden können. Das Lernen geeigneter Merkmale allein anhand von Binärmaskensegmentierungen homogener Bereiche gestaltet sich aufgrund mangelnder diskriminativer Merkmale weitaus schwieriger. Dies ist bei der Läsionsvorhersage des Schlaganfalls für ein zukünftiges Verlaufsbild der Fall.

Zusammenfassend kann aus den vorgestellten Ansätzen abgeleitet werden, dass das direkte Schätzen eines dichten Verschiebungsfeldes über Lernverfahren generell schwierig ist, da die Voraussetzungen bei dem hier besprochenen Problem der Läsionsvorhersage des Schlaganfalls nicht gegeben sind. Es fehlen zum einen regulär aufgenommene Intensitäts- oder Labeldaten. Zum anderen besteht ein anderer zeitlicher Zusammenhang (Risikogewebe stellt z. B. den approximierten *worst case* zu einem späteren als den Aufnahmezeitpunkt dar, zu dem gar keine Bilddaten real existieren). Aus diesem Grund wird die robuste Hauptkomponentenberechnung der PCA zur Vereinfachung des Problems verwendet, um die Komponenten-Gewichtung in einem möglichst niedrigkomplexen Modell regressieren zu können. Zudem werden die Trainingsdaten durch Nutzung eines diffeomorphen Registrierungsmodells topologieerhaltend generiert.

7.2 Vorgeschlagene Methodik

Die hier verfolgte Idee für eine robuste und plausible Prognose des Verlaufs der Schlaganfallläsion ist es, die Formänderungen der Segmentierungsmaske (räumliches Wachstum) durch ein Registrierungsproblem ihrer Segmentierungen zu verschiedenen Zeiten abzubilden (Kern bei $t = 0$, Verlaufsbild bei $0 < t < 1$, Risikogewebe bei $t = 1$). Damit wird die Formveränderung, die in Kapitel 6 durch eine räumlich nur dünn regularisierte Interpolation von Formabstraktionen im Formraum durchgeführt wurde, nun stärker im Training durch die Registrierungsinformation überwacht. Damit kann die CNN-basierte Vorhersage ein plausibleres Wachstum schätzen bzw. modellieren. Der Erfolg dieses Ansatzes hängt damit entscheidend vom Registrierungsverfahren ab, welches plausible Deformationen der Ausgangsläsion hin zur finalen Läsion initial bereitstellen muss. Deshalb werden in der Vorverarbeitung ausschließlich diffeomorphe Verschiebungsfelder genutzt.

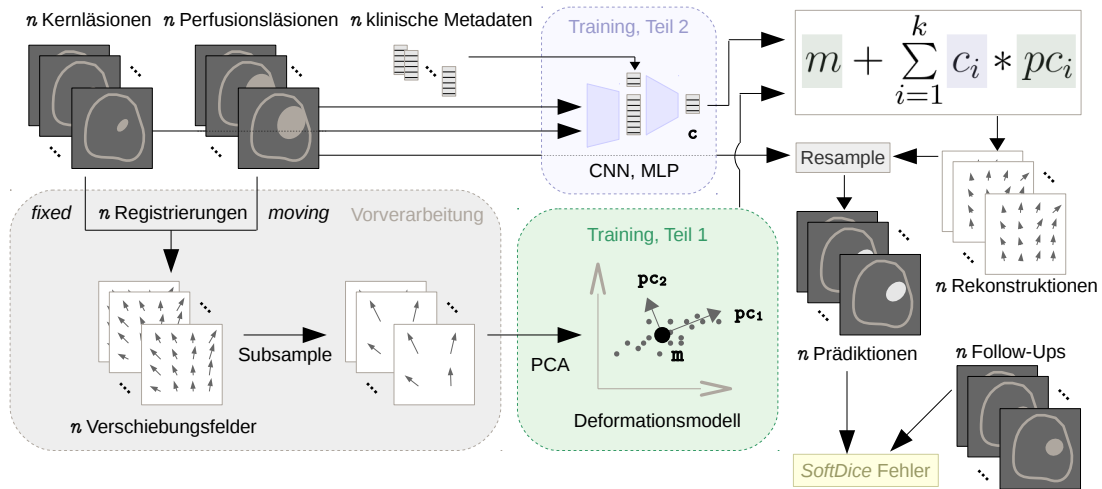


Abb. 7.4: Trainingsübersicht: n Paare von Kern- und Risikogewebe-Läsionsmasken werden jeweils in einem Vorverarbeitungsschritt registriert. Die resultierenden Verschiebungsfelder werden herunterskaliert bevor über alle die Hauptmoden pc_i ($1 \leq i \leq k$) berechnet werden und die durchschnittliche Verschiebung m . Mit einem CNN werden die Koeffizienten c_i geschätzt, um die Verschiebungen zu rekonstruieren, welche die Verlaufsläsion prädizieren. Teil 1 wird einmalig berechnet, während Teil 2 über mehrere Epochen auf den n Paaren trainiert wird.

7.2.1 Aufbau des Deformationsraums

Das Training (Abb. 7.4) wird in zwei separaten Schritten hintereinander ausgeführt, um die Hauptrichtungen des Schlaganfallwachstums zu schätzen:

- Ein Surrogat der wichtigsten Verschiebungsmodi zwischen Kern und Risikogewebe aus einem unabhängigen Registrierungsverfahren wird verwendet. Diese Modi werden durch statistisches Lernen mittels einer PCA (Teil 1 in Abb. 7.4, Abschnitt 3.2.1) der Verschiebungsvektorfelder modelliert, welche das Risikogewebe auf die Kernsegmentierung transformieren.
- Anstelle einer hohen Anzahl von Deformationsparametern eines zu schätzenden Vektorfeldes wird die Aufgabe durch das robuste PCA-Modell entlang einer kleinen Anzahl von $k \leq n$ Hauptverformungsrichtungen bei insgesamt n Trainingsfällen vereinfacht.

7.2.2 Schätzung der Modellparameter

Das CNN (Teil 2) erlernt die Koeffizienten c_i des PCA-Modells für eine geeignete Gewichtung jeder Komponente $i \in \{2 \dots k\}$, um die Verschiebungen für die finale Läsion auf dem Verlaufsbild vorherzusagen. Insbesondere werden hier *nicht* die Koeffizienten gesucht, um Kern und Risikogewebe zu registrieren (deren Verschiebungsfelder

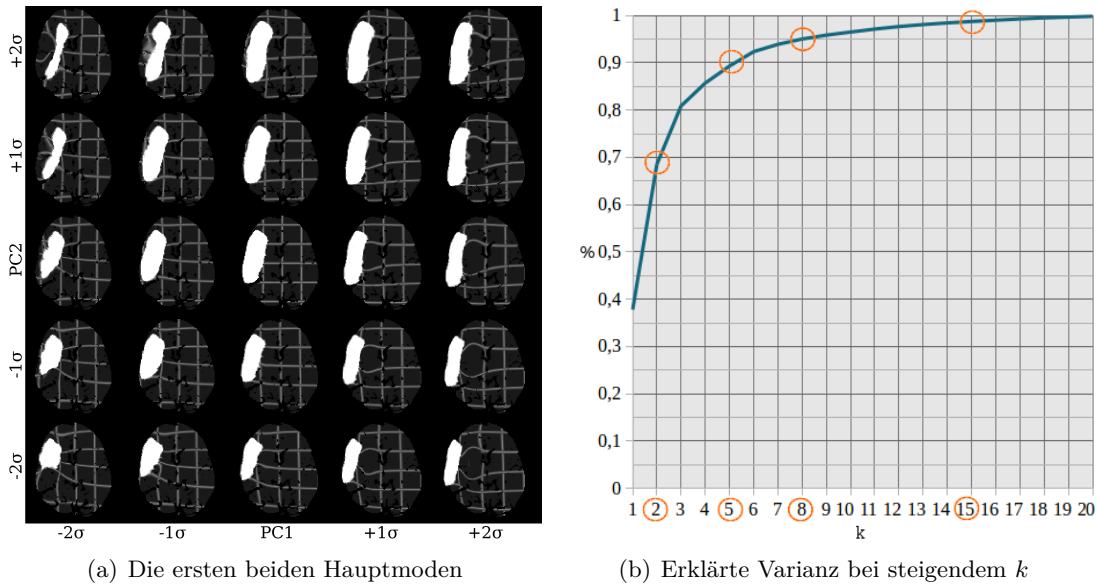


Abb. 7.5: Deformationsmodell aus allen 29 Probanden des UKSH-Datensatzes. (a) zeigt die ersten beiden Hauptmoden der Deformation einer mittleren Läsionsform für $k=8$, wobei (b) den akkumulierten Anteil an Varianz der Daten zeigt, die durch das Modell mit seinen k ersten Moden abgedeckt wird: $\sim 90\%$ bei $k=5$, $\sim 95\%$ bei $k=8$, und $\sim 99\%$ bei $k=15$.

jedoch im Training genutzt werden), sondern um Kern bzw. Risikogewebe auf den „Zwischenschritt“ einer finalen Läsion zu verformen. Diese Deformation ermöglicht eine dicht abgetastete Interpolation der Läsion durch das „GridSampler“-Modul, das im PyTorch-Paket [Paszke u. a., 2017] implementiert ist, und bei topologisch korrekter, zusammenhängender Ausgangssegmentierungen bei einem diffeomorphen Verschiebungsfeld (Kapitel 3.1.2) ebenfalls eine zusammenhängende finale Läsion erzeugt.

Die Linearkombination aus den Koeffizienten c_i und den Hauptrichtungen pc_i der PCA zusammen mit dem Mittelwert m in Gleichung 7.1 ergibt ein rekonstruiertes Verschiebungsfeld d , das das Risikogewebe auf die Folgeläsion verformt. Unter Verwendung invertierbarer, diffeomorpher Verschiebungs-Grundwahrheit als Überwachung kann das Wachstum vom Kern zur Folgeläsion (vorwärts in der Zeit, d^{-1}) auch in umgekehrter Vektorrichtung vom Risikogewebe zur Folgeläsion (rückwärts in der Zeit, d) gedacht werden. Dies ermöglicht es uns, die Risikogewebe-Segmentierung als zu verformendes Bild (*engl.* Moving Image) mit mehr Forminformationen zu nutzen. Während des Trainings quantifiziert der SoftDice-Fehler (Kapitel 3.1.3) den Unterschied zur tatsächlichen finalen Läsionsmaske aus der Grundwahrheit. Um die Parametergewichte des CNN zu aktualisieren, kann der am Ende berechnete SoftDice-Fehler durch die Ableitung von Gleichung 7.1 durch das CNN zurückpropagiert werden.

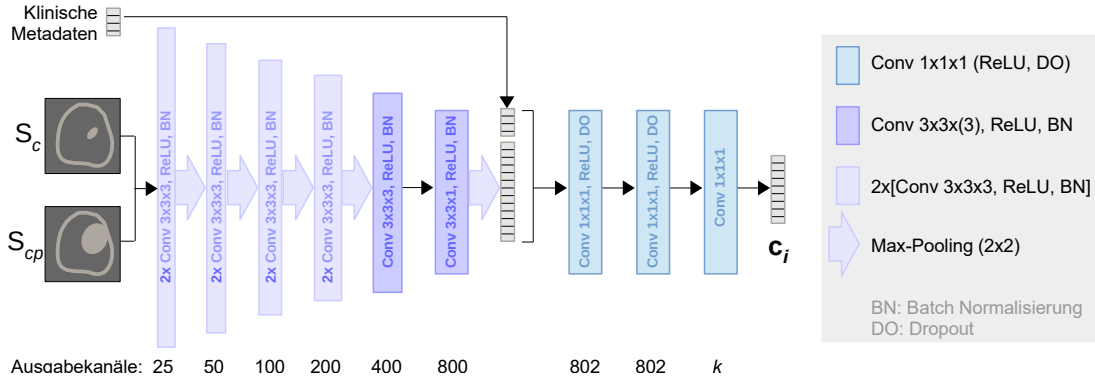


Abb. 7.6: Die Architektur des CNNs aus Abb. 7.4, Teil 2, besteht aus zwei Seiten: Vor dem Zusammenführen wird die Eingabe soweit gefaltet, herunterskaliert und die Anzahl der Kanäle erhöht, bis diese als Vektor mit den klinischen Metadaten konkateniert werden kann. Der entstehende Vektor wird neu gewichtet und auf die Dimensionalität k des Vektor-Deformationsraums heruntergerechnet.

$$d = m + \sum_{i=1}^k c_i * pc_i \quad (7.1)$$

Das Vorgehen in zwei Schritten ist begrenzt mit dem in Kapitel 6 vergleichbar: Zunächst wird der Modellraum trainiert, wie hier der Deformationsraum (PCA) oder zuvor ein Formraum (CAE). Während bei der Interpolationsmethode schon inhärent vorgegeben ist, wie die Schätzung der finalen Folgeläsion berechnet wird (Interpolation anhand der Zeit im Formraum zwischen zwei Kodierungen als definierte Endpunkte), muss im Deformationsmodell direkt eine einzelne Kodierung für das Verschiebungsfeld der finalen Läsion geschätzt werden. Ein weiterer Architekturbestandteil, wie das CNN aus Teil 2 der Abb. 7.4, ist daher zwingend erforderlich.

Das CNN bekommt Kern- und Risikogewebe-Masken als Eingabe, erhöht durch stückweise Verdopplung die Anzahl der Kanäle von 25 auf 800 und reduziert die räumliche Auflösung auf $1 \times 1 \times 1$ (Abb. 7.6, vor der Konkatenierung). Die Zeiträume $t_{0 \rightarrow A}$ und $t_{A \rightarrow I}$, das heißt beide Zeiträume von Beginn des Schlaganfalls bis zur Bildgebung und von der Bildgebung bis zur Intervention, werden mit diesem Vektor verknüpft und durch zwei 802-kanalige $1 \times 1 \times 1$ Faltungsschichten verarbeitet, gefolgt von Dropout [Srivastava u. a., 2014] zur robusteren Schätzung, um den Ausgangsvektor c_i , der die PCA-Modellkoeffizienten enthält, neu zu gewichten (Abb. 7.6, nach der Konkatenierung).

7.3 Experimente

Die Vorverarbeitung in Abb. 7.4 verwendet ein diffeomorphes B-Spline-Transformationsmodell, das durch stochastischen Gradientenabstieg [Kingma u. a., 2015] optimiert wird, um die Grundwahrheiten der Verschiebung zwischen Risikogewebe und Kern zu erzeugen. Die Wahl der Registrierungsmethode ist jedoch modular austauschbar und lässt sich unabhängig von der vorgeschlagenen Methode anpassen. Die Verschiebungsvektorfelder werden trilinear von $256 \times 256 \times 28 \times 3$ auf $32 \times 32 \times 28 \times 3$ Voxel interpoliert, um den Rechenaufwand für die darauffolgende Modellberechnung zu reduzieren.

Das Ziel ist die Hauptdeformation entlang des zerebralen Gefäßsystems zu lernen, weshalb sich keine kanonische Basis mit zentrierten Segmentierungsmasken eignet, sondern die Position in einem gemeinsamen Standardraum des Gehirns für die Bilder beibehalten wird. Darüber hinaus wird das Modell nur für eine Hemisphäre erstellt. Die Blutversorgung ist grob symmetrisch (Kapitel 2.1.1) und so muss der PCA-Raum nicht unterschiedliche Deformationen für die Hemisphären modellieren können, sondern eine Läsion kann im Standardraum bei Bedarf gespiegelt werden. In Abb. 7.5(a) ist die Läsion in der rechten Hemisphäre zu erkennen.

Analog zu Kapitel 6 wurde zur Vergleichbarkeit wieder die 5-fache Kreuzvalidierung (S. 32) mit identischen Gruppen durchgeführt. Die ersten beiden Komponenten eines erlernten Verschiebungsmodells zeigen grob die Hauptvariation in Höhe und Breite, was etwa $2/3$ der Varianz in den Daten erklärt (Abb. 7.5(b)). Zur Verbesserung der Robustheit der PCA kann auf die Verschiebungsfelder eine affine Augmentierung angewandt werden (in den Experimenten 9 zufällige Transformationen pro Proband), für die wiederum k erhöht werden muss, um eine ähnliche Varianz abzudecken wie ohne die Augmentierung.

7.3.1 Datensatzerweiterung

Für die Experimente wird wieder der UKSH-Datensatz mit 29 Patienten verwendet, wobei deren TTD-Segmentierung dabei das gesamte Risikogewebe von Kern+Penumbra abgrenzt und die CBV-Segmentierungsmaske als Näherung für die Kernsegmentierung dient (vgl. Herausforderungen auf S. 20f). Dies entspricht jedoch nicht immer der klinischen Hypothese einer monoton wachsenden Segmentierungsform, da die Folgeläsion in der Verlaufs-CT auch kleiner sein kann als die akute CBV-Läsion in der CTP, und wurde als relevante Ursache für problematische Ergebnisse in Kapitel 6 identifiziert. Darum ist der UKSH-Datensatz hier nochmals angepasst worden, sodass die Trainings- und Testdaten auch zur klinischen Annahme einer monoton wachsenden Schlaganfall-läsion passen und die Modelle damit ebenfalls getestet werden können.

Das Trainingsziel (Folgeläsion) wurde für den ersten Datensatz UKSH-A durch die Vereinigung der ursprünglichen finalen Läsionssegmentierung (FL) und der Kernsegmentierung (CBV Segm.) ersetzt – wie in Abb. 7.7, links, dargestellt. Im Datensatz

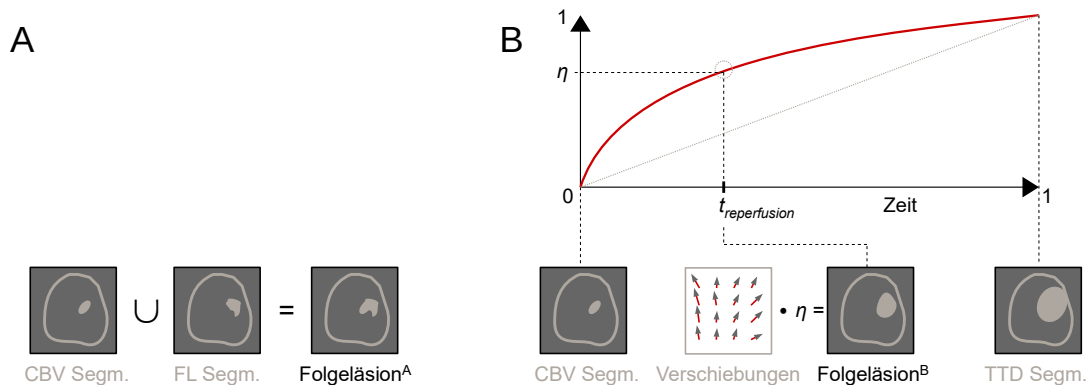


Abb. 7.7: Zusätzliche Modifikationen des UKSH-Datensatzes für die Experimente. UKSH-A (links) vereint die CBV- mit der Verlaufsbild-Segmentierung (FL) zu einer neuen Folgeläsion. UKSH-B (rechts) generiert stattdessen einen Zwischenzustand von der Deformation von Kern zu Risikogewebe entsprechend dem nichtlinearen Faktor („Wachstumsrate“) des Verschiebungsfelds, welchem die Dauer bis zur Behandlung zugrunde liegt.

UKSH-B wurden zunächst Kern und Risikogewebe registriert und die resultierenden Verschiebungen mit einem Gewichtungsfaktor η skaliert, um die Vektormagnituden entsprechend der Zeit t bis zur Reperfusion anzupassen (Abb. 7.7, rechts). Dies simuliert eine nichtlinear wachsende künstliche Läsion mit einer Wachstumsrate von $\eta = t^{0,5}$ innerhalb der Zeitspanne $t \in 0 \dots 1$ zwischen der anfänglichen Nekrose ($t = 0$) und dem *worst case* ($t = 1$). Mit dem Datensatz UKSH-B kann die allgemeine Anwendbarkeit der Methode für ein gut spezifiziertes Wachstumsproblem, welches sich an der üblichen klinischen Beobachtung eines höheren Wachstums zu Beginn der Läsionsevolution orientiert [Brooks u. a., 2019], nachgewiesen werden, da eine nichtlineare Wachstumsrate dabei berücksichtigt wurde.

7.3.2 Ergebnisse

Die Dice-Ergebnisse der binarisierten Segmentierungen finden sich in Tab. 7.1: Das „Zeit-Orakel“ ignoriert die tatsächlichen (u.U. verrauschten) Zeit-Metadaten und optimiert gerade den Zeitpunkt für eine CNN-Prädiktion möglichst nahe der Grundwahrheit. Der Wert des „Code-Oracle“ quantifiziert die generelle Fähigkeit der PCA zur möglichst perfekten Registrierung des Risikogewebes mit der finalen Folgeläsion, indem die Kodierung bzw. Parametrisierung c des PCA-Modells passend zur Grundwahrheit optimiert wird. Bei ausschließlicher Nutzung der Eingabedaten von Kern, Risikogewebe sowie Metadaten ohne Berücksichtigung der Grundwahrheit handelt es sich um eine echte „Prädiktion“. Diese quantifiziert die Fähigkeit des CNNs, die richtigen Koeffizienten für das gegebene PCA-Modell für eine ungesehene finale Läsionssegmentierung des Verlaufsbildes zu finden.

Modell	k	Datensatz	Prädiktion	Zeit-Orakel	Code-Orakel
PCA augm.	60	UKSH-B	0,74	0,77	0,92
PCA	20	UKSH-B	0,71	0,75	0,88
CAE1	(800)	UKSH-B	0,62	0,68	0,88
PCA augm.	60	UKSH-A	0,52	0,53	0,84
PCA	20	UKSH-A	0,48	0,55	0,76
CAE1	(800)	UKSH-A	0,60	0,70	0,87
PCA augm.	60	UKSH	0,37	0,41	0,84
PCA	20	UKSH	0,38	0,38	0,74
CAE1	(800)	UKSH	0,46	0,53	0,89

Tabelle 7.1: Dice-Resultate der zwei Wachstumsmodelle: CAE mit Formmodell (Kapitel 6) und PCA mit Deformationsmodell (Kapitel 7), wobei „PCA augm.“ für eine affine Augmentierung der Verschiebungsfelder im Training steht. Die höchsten Werte pro Datensatz sind **fett** markiert. Im Vergleich zu vorher publizierten Ergebnissen wurden hier zwei Orakel zur besseren Vergleichbarkeit nachgerechnet und dabei die Berechnungsmethode vereinheitlicht (vorher: uniformes Sampling des Zeitstrahls; jetzt: Optimierung auf bekannter Grundwahrheit).

Zunächst ergibt sich bei der diffeomorphen Registrierung der Risikogewebe-Segmentierung, die vom Vorverarbeitungsschritt auf die Kernläsion-Segmentierung deformiert wurde, ein Durchschnitts-Dice von 0,84. Die dabei generierten Verschiebungsfelder werden dann zum Training benutzt. Aus diesem Vorgehen ergibt sich grob eine obere Schranke für die maximal erreichbare Überlappung der Prädiktion. Die Registrierung auf dem Kern ist dabei nicht perfekt, aber diffeomorph, so dass die nachfolgende PCA faltungsfreie, d. h. anatomisch plausible, Verschiebungsfelder als Eingabe für das Training des Deformationsraums erhält.

Methodisch sind die beiden Verfahren CAE und PCA schwierig direkt miteinander zu vergleichen – insbesondere in Bezug auf ihre Mächtigkeit bzw. Komplexität. Die k Dimensionen im PCA-Raum und die Flaschenhals-Größe bzw. Dimensionalität der latenten niedrigdimensionalen Repräsentation im CAE-Raum stellen keine äquivalente Bausteine ein- und desselben Verfahrens dar. Bei der PCA-Methodik findet darüber hinaus z. B. keine Interpolation statt. Das reine Austauschen des CAE durch eine PCA in Kapitel 6 würde einen technischen Rückschritt bedeuten und ist nicht von Interesse, da ein CAE in der Praxis einen sehr viel mächtigeren Lernalgorithmus darstellt (u. a. durch komplexe nichtlineare Deformationen in Encoder und Decoder).

Unter Beachtung dieser methodischen Vorbehalte lässt sich bei einem idealen Datensatz, der der klinischen Hypothese einer monoton wachsenden einzelnen Läsion entspricht, feststellen, dass der evaluierte PCA-Raum besser generalisiert und bei ungesehenen Daten bessere Überlappungen erzielen kann (Dice 0,74 > 0,62 bzw. 0,92 > 0,88

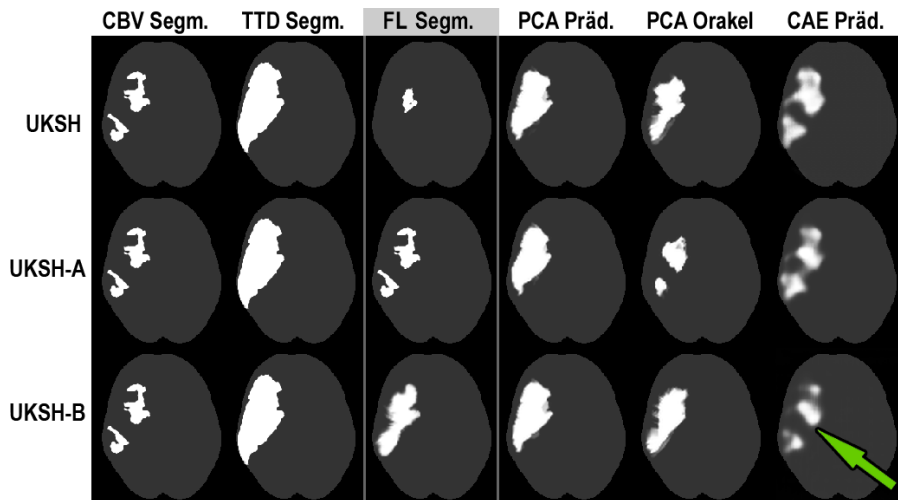


Abb. 7.8: Ergebnisse für einen Beispielfall in allen drei Datensätzen: Die ersten drei Spalten zeigen die Segmentierungsgrundwahrheiten, während die letzten drei Spalten die Vorhersagen zur tatsächlichen finalen Läsion („FL Segm.“) zeigen: Prädiktion von PCA und CAE, sowie Code-Orakel der PCA. Unplausible Interpolationen (grüner Pfeil) können insbesondere bei der CAE-Interpolation aus Kapitel 6 auftreten.

in Tab. 7.1). Jedoch ist bei den originalen UKSH- sowie den UKSH-A-Daten die Methode aus Kapitel 6 stets der in ihrer Vorverarbeitung viel komplexer modellierten Methode aus diesem Kapitel überlegen, sowohl bei der reinen Prädiktion (Dice $0,46 > 0,38$ bzw. $0,60 > 0,52$) wie auch den theoretisch motivierten Untersuchungen des Zeit- (Dice $0,53 > 0,41$ bzw. $0,70 > 0,53$) und Code-Orakels (Dice $0,89 > 0,84$ bzw. $0,87 > 0,84$).

Zum qualitativen Vergleich finden sich neben Abbildung 7.8 weitere Beispielprädiktionen auf Seite 133. Eine Vereinigung von ursprünglicher Folgeläsion und CBV-Läsion im UKSH-A-Datensatz stellt zwar die Bedingung sicher, dass die Folgeläsion mindestens gleich groß der CBV-Läsion ist; dies stellt für den Algorithmus trotz besserer Ergebnisse als im originalen UKSH-Datensatz (Dice $0,52 > 0,37$) dennoch keinen generellen Vorteil gegenüber der CAE-Methode dar, da unter Umständen zwischen der Folgeläsion und CBV-Läsion immer noch wenig Formunterschiede existieren (vgl. UKSH-A in Abb. 7.8: kein Wachstum, also keine Verformung zwischen „CBV Segm.“ und „FL Segm.“). Obgleich auch der CAE kontinuierlich höhere Dice-Ergebnisse von Datensatz UKSH über UKSH-A hin zu UKSH-B erzeugt (Dice $0,46 - 0,62$), gelingt es der PCA-basierten Methode besser, die nichtlineare Veränderung der Form abzubilden (Dice $0,37 - 0,74$).

7.3.3 Diskussion

Dieses Kapitel bietet erste Resultate eines Deformationsmodells als Basis für die Vorhersage der Schlaganfallläsion nach Intervention. Das Modell ist geeignet, die Hauptrich-

tungen vom räumlichen Wachstum des Kerns in das Risikogewebe über die Zeit zu modellieren und das basierend auf der kleinen Stichprobenanzahl des UKSH-Datensatzes. Es erlaubt die Rekonstruktion von Verschiebungsfeldern, die das Risikogewebe nahe der finalen Läsionsform ($\text{Dice} \geq 0,74$) deformieren können.

Es ist jedoch nicht primäres Ziel, die höchste Überlappung zu finden, sondern topologisch konsistentes Wachstum abzubilden, d. h. die Vermeidung von Überblendungseffekten wie bei Interpolationsansätzen (Abb. 7.8) zu erreichen. Dies wird durch das dichte Abtasten der glatten Transformation (z. B. mit dem PyTorch „GridSampler“) zur räumlichen Deformierung einer binären Läsionsform vermieden. Obwohl die Segmentierungsergebnisse auf dem originalen und dem UKSH-A-Datensatz auf einen Genauigkeitsrückgang schließen lassen, wurden keine vergleichbaren Überblendungen wie in Kapitel 6 beobachtet.

Als perspektivisch vorteilhaft könnte sich in der Praxis erweisen, dass man im Anwendungsfall nur noch eine Segmentierungsschätzung von Kern oder Risikogewebe benötigt, welche dann deformiert wird, um die finale Läsion zu prädictieren. Da das Risikogewebe besser automatisch geschätzt wird (vgl. Kapitel 4.2.3.1) und mehr Forminformationen enthält als die Kernsegmentierung, würde bei Verwendung der Segmentierung des gesamten Risikogewebes (Perfusionsläsion) eine Fehlerquelle aus dem vorherigen Schritt ausscheiden. Wenn das CNN direkt auf CTP-Parameterkarten trainiert wird, kann es zudem die benötigten Informationen direkt extrahieren, die beim hier vorgeschlagenen Training noch ausschließlich auf den Segmentierungen gewonnen wurden (Abb. 7.4 auf Seite 105).

Allerdings wird das Modell nur zur Trainingszeit mit diffeomorphen Verschiebungen angelernt. Es wäre wünschenswert zu untersuchen, wie sich das Sicherstellen des Diffeomorphismus nach der Rekonstruktion aus dem PCA-Modell – etwa nach dem Verfahren von Dalca u. a. [2019] – auf die Ergebnisse auswirkt. Sobald diffeomorphe Rekonstruktionen bereitstünden, könnte man dazu noch die Konsistenz des Wachstums in beide zeitlichen Richtungen sicherstellen, etwa durch eine „Inverse-Consistency“-Nebenbedingung im Fehlerterm, sodass sowohl von der Kernsegmentierung als auch der Risikogewebesegmentierung die finale Läsion gleichermaßen gut prädictiert werden kann.

Es sollte jedoch zunächst das Ziel sein, zuvor die Pipeline durchweg differenzierbar für das maschinelle Lernen zu gestalten, sodass die PCA-Berechnung ebenfalls in den Lernprozess miteinbezogen werden und nach einer Initialisierung von Teil 1 zusammen mit dem CNN optimiert werden kann. Dies könnte die Flexibilität in Bezug auf das Trainingsziel erhöhen, allerdings zulasten von eventuell nicht ganz optimalen Hauptrichtungen bzw. einer geringeren Varianzabdeckung.

8 Zusammenfassung und Ausblick

Der Nutzen und demzufolge die Wahl der Behandlungsmethode beim akuten ischämischen Schlaganfall ist von bestimmten Selektionskriterien abhängig. Unter Berücksichtigung therapiespezifischer Risiken gilt es, die mögliche Eignung des Patienten durch eine Quantifizierung von rettbarem Gewebe festzustellen und das Auswachsen zu einer dauerhaft geschädigten Infarktregion zu vermeiden. In der vorliegenden Arbeit wurden dafür verschiedene Methoden auf Basis des maschinellen Lernens mit künstlichen neuronalen Netzen vorgestellt. Im Vergleich zur momentan gültigen Leitlinie [Hennerici u. a., 2017] in der Praxis, handelt es sich um unterschiedlich fortgeschrittene Ansätze zur Ermittlung objektiver und reproduzierbarer, sowie durch Computerunterstützung schnell berechenbarer Ergebnisse, die das Ableiten geeigneter Kriterien des Einsatzes einer Behandlung mit dem patientenindividuell höchsten Nutzen ermöglichen sollen. Die eingangs dargelegten Zielsetzungen (Seite 5ff.) wurden dabei folgendermaßen erreicht:

Objektive, reproduzierbare Quantifizierungen für die klinische Selektion auf multivariater CT Mit Kapitel 4 wurde deutlich, dass bereits in der nativen Standard-CT als breit verfügbarer Bildgebung eine Vorselektion der Patienten möglich ist, sofern diese – neben anderen – über entsprechende Frühzeichen eines Media-Arterienverschlusses verfügen. Es wurde in Kapitel 4.1 gezeigt, dass mit Deep Learning und einem robusten kaskadierten Verfahren inklusive U-Net bereits zuverlässig Gerinnungskandidaten detektiert werden. Dies ist als Hilfsmittel für eine schnelle, zielgerichtete Bildgebung und Diagnose eine wichtige Voraussetzung für die erfolgreiche Behandlung und kann integriert in einem Radiologie-Informationssystem potenzielle Schlaganfälle oder solche mit schlechter klinischer Prognose bei großem Gefäßverschluss markieren.

Auch für die Feststellung der Penumbra und dem Risikogewebe-Kern-Mismatch (Abschnitte 2.2.1.2 und 2.2.2.1) kann auf neuronale Faltungsnetzwerke zur Segmentierung der multiparametrischen Perfusion zurückgegriffen werden. In Kapitel 4.2 konnte nachgewiesen werden, dass dies für das Risikogewebe durch die TTD-Läsion bereits hinreichend gut gelingt (Abschnitt 4.2.3.1), aber sich die Segmentierung der CBV-Läsion als Kern-Schätzung anspruchsvoller gestaltet. Es wurde gezeigt, dass durch Architektur-Erweiterungen des U-Nets (U-Net^{FL} aus Kapitel 5) oder der nachgeschalteten Regularisierung durch das Vorwissen über Schlaganfallläsionen (CAE aus Kapitel 6) die Schät-

zung der CBV-Läsion als approximierter Kern verbessert werden kann (vgl. Kern-Dice in Tab. B.2 auf S. 132).

Für die Erzeugung einer objektiven Prädiktion der finalen Folgeläsion als binäre Infarktkarte konnte ebenfalls ein Verfahren mittels U-Net vorgestellt werden (Kapitel 5), welches den Stand der Technik, der mit U-Nets und diskriminativen Segmentierungsansätzen erreicht werden kann, darstellt (vgl. ähnliche Ergebnisse der Ensembles in Abb. 5.7 auf S. 77). Weiterhin wurden in Kapitel 6 und 7 zwei weitere Ansätze vorgestellt, welche binäre Infarzierungskarten reproduzierbar präzisieren können.

Es handelt sich in allen genannten Fällen insofern um *objektive* Quantifizierungen, als dass für das Training der neuronalen Netze auf verschiedenes Expertenwissen zurückgegriffen werden kann, welches darüber hinaus im Training aus den Einzelstichproben generalisiert werden muss, damit die Netze auf neuen und ungesehenen Patientebildern Schätzungen vornehmen können, wie in den Experimenten der verschiedenen Kapitel durch Testsätze gezeigt.

Entwicklung eines Prädiktionsverfahrens zur Vorhersage des räumlichen Gewebestatus Statt des globalen binären Entscheidungskriteriums (Mismatch über Schwellwert) erlauben die in dieser Arbeit vorgestellten Verfahren (Kapitel 5–7) eine räumliche Zuordnung der erwarteten Infarzierung des Gewebes als Grundlage zur Entscheidung über eine Therapie. Anders als in den bisherigen Vorarbeiten [Kemmling u. a., 2015; Maier u. a., 2016] werden keine vordefinierten Merkmale benötigt. In Kapitel 5 wurde eine Erweiterung des U-Nets vorgeschlagen, um entsprechende Gewebestatuskarten zu präzisieren, und konnte sich gegenüber vielen anderen Architekturen behaupten [Winzeck u. a., 2018]. Es wurde deutlich, dass diese Variante vor allem genauer als normale U-Nets arbeiten kann (z. B. bei kleinen Strukturen) und dabei auch weniger Ausreißer in der Hausdorff-Metrik resultieren. In jedem Fall zeigte sich, dass Eingabeinformation, die durch zusätzliche Skip-Verbindungen im Encoder-Pfad durchgereicht werden kann, die Vorhersagen gegenüber dem Standard-U-Net verbessert (vgl. Tab. 5.1 auf S. 74).

Der Vorteil ist die direkte, vollautomatische Anwendung der Prädiktion ohne manuell Parameter einzustellen, sodass es in einem nächsten Schritt klinisch evaluiert werden könnte. Allerdings ist das Niveau der Genauigkeit im Rahmen der Folgeläsionsprädiktion dafür noch zu gering und es wurde gezeigt, dass es sich um einen Stand der Technik handelt, welcher kaum bessere Ergebnisse auf Vergleichsdaten erreichen kann [Winzeck u. a., 2018]. Darüber hinaus kann es durch einen anderen Ansatz mit Nutzung von Vorwissen (siehe nächste Zielstellung) deutlich übertroffen werden (Kapitel 6), woraufhin der herkömmliche CNN-Ansatz (Diskriminierungsansatz mittels Segmentierungsnetzwerk, wie dem U-Net) zur Prädiktion der Folgeläsion im Laufe der weiteren Arbeit verworfen und nur noch zur Schätzung der Kern- und Penumbra-bereiche eingesetzt wurde. Dort musste allerdings auch festgestellt werden, dass eine

schlichte Kombination des U-Net^{FL} aus Kapitel 5 mit dem CAE aus Kapitel 6 nicht zu noch weiter verbesserten Ergebnissen führen konnte (vgl. Tab. B.2 auf S. 132), als dies mit dem Standard-U-Net in Verbindung mit dem CAE möglich war.

Einarbeitung von Expertenwissen zur Verbesserung der Vorhersage Das Verwenden von A-priori-Vorwissen von Expertenannotationen ist eine gängige Methode zur Regularisierung von Vorhersagen oder Erstellung generativer Modelle zur Prädiktion bei medizinischen Fragestellungen. In Kapitel 6 wurde ein methodischer Rahmen vorgestellt, um eine Wachstumsprognose von Schlaganfallläsionen basierend auf den Kern- und Risikogewebesegmentierung zu generieren. Dafür wurde ein niedrigdimensionaler Formraum aus Segmentierungsmasken derart trainiert, dass bei linearer Interpolation zweier Kern- und Risikogewebe-Repräsentationen die Repräsentation der Folgeläsion auf (bzw. nahe) dieser Trajektorie liegt. Dabei konnte zuerst gezeigt werden, dass das gelernte generative Modell basierend auf Vorwissen echter Schlaganfallformen Ischämien und Läsionen der MCA zum großen Teil abbilden und rekonstruieren kann (vgl. Ergebnisse des „CAE1“ in Abschnitt 6.3.1 und Beispiele auf S. 126–129).

Die darauf aufbauende und nur von zeitlichen Parametern abhängige Wachstumsmodellierung kann daraufhin in Bezug zu einem konkreten Segmentierungsverfahren für die Perfusionsdaten durchgeführt werden, erfordert aber einen auf den geschätzten Segmentierungen trainierten „CAE2“. Damit konnte zum einen nachgewiesen werden, dass eine gelernte Interpolation in einem solchen Formmodellraum die Vorhersage der Folgeläsion auf Basis automatisch geschätzter Kern- und Risikogewebesegmentierungen ermöglicht; zum anderen wurden signifikant bessere Ergebnisse erreicht als allein mit dem U-Net^{FL} aus Kapitel 5. Das Hinzufügen dieser Segmentierungen als Vorwissen in bzw. als Prädiktor für das CAE-Modell ist mit Blick auf die erreichten Ergebnisse offensichtlich notwendig, um die Prädiktionsgenauigkeit über das Niveau bisheriger Methoden, wie in Winzeck u. a. [2018] vorgestellt, anheben zu können. Damit verbessert sich auch die Vorhersage daraus berechneter finaler Infarkt volumens in *ml* (siehe S. 95 und S. 131), wobei die prädizierten Volumens im Mittel annähernd proportional zum tatsächlichen Volumens ansteigen, aber in den Absolutwerten noch deutlich abweichen können.

Es hat sich außerdem herausgestellt, dass durch Eingabe zweier Expertensegmentierungen – sofern diese in einer Akutsituation effizient für CBV und TTD erstellt werden können – das CAE-Modell in dem nun semi-automatischen Verfahren leicht überlegene Vorhersagen erzeugen kann als die vollautomatische Methode (vgl. Ergebnisse von „*Experte*+CAE1“ mit „U-Net+CAE2“ auf S. 95). Durch den vorgestellten zweistufigen Ansatz ist es Ärzten möglich, automatisiert erstellte Segmentierungen im Zwischenschritt mit vergleichbar wenig Aufwand noch händisch zu korrigieren und dabei direkt nachzuvollziehen, auf welcher Grundlage die darauffolgende Prädiktion der Läsion basiert. Dies sorgt für Nachvollziehbarkeit im Sinne einer erklärenden KI (*engl.*

explainable AI), obgleich auch bei den manuellen CBV- und TTD-Segmentierungen der Dice-Score für die daraus vorhergesagte finale Läsion auf den UKSH-Daten im Mittel noch unter 0,5 bleibt. Dies ist teilweise auf Interpolationsartefakte im abstrakten Formraum zurückzuführen, welche mit dem alternativen Deformationsansatz aus Kapitel 7 qualitativ reduziert werden können.

Simulation von patientenindividuellem Infarktwachstum Durch die Entwicklung eines Wachstumsmodells der Infarktläsion auf Basis der Perfusions-CT, konkreter Aufnahmezeit und erwartetem Behandlungszeitpunkt wurde gezeigt, dass sich nicht nur patientenindividuelle Aussagen über die Folgeläsion und Infarktstatus des Gewebes tätigen lassen, sondern eben dieses nichtlineare Auswachsen der Läsion in einer Visualisierung simuliert werden kann. Mit dem Forminterpolationsansatz aus Kapitel 6 umgeht man dabei die Zeitabhängigkeit der Schwellwerte für Perfusionsparameter (Abschnitt 2.2.1.2) gegenüber einer reinen U-Net-Variante. Stattdessen werden die beiden Zeitparameter t_A und t_I explizit für die patientenindividuelle Progression des Wachstums berücksichtigt. Trotz unterschiedlicher Zeitpunkte jedes Einzelfalls und verschiedener Normierungen der Akutphase (10 gegenüber 24 Stunden, vgl. Tab. B.2 auf S. 132), wird ein entsprechender Formraum gelernt, der die unterschiedlichen Progressionsraten nichtlinearer Läsionsausbreitung abbilden kann.

Durch die Konstruktion der Interpolation wächst die Läsion daher deutlich vom Kerninfarkt in das Media-Stromgebiet aus, welches durch das Risikogewebe im TTD-Parameter bestimmt wird. Die Plausibilität zu jedem Zeitpunkt ist mangels Überwachung mit realen Daten damit nicht gegeben, aber ein monotoner Anwachs wird sichergestellt. Diese konsistente zeitabhängige Simulation kann dem Kliniker im Vergleich mit anderen Fällen einen ersten Eindruck vom (schädlichen) Potenzial eines ischämischen Schlaganfalls vermitteln, indem er den Akutzustand der Kernläsion mit der Vorhersage einer fortgeschrittenen Simulation bis zum angenommenen Interventionszeitpunkt begutachtet („Wie viel Gewebe wird sich verschlechtern?“) und dieses mit anderen Vorhersagemodellen z. B. anderer Therapiemethoden vergleicht (bzw. dem Auslassen der Therapie, wenn die Simulation bis zum Ende des Akutzeitraums durchläuft).

Um die Plausibilität der durch die Simulation erzeugten Läsionsvorhersagen insbesondere in Bezug auf deren Form und Topologie zu verbessern, wurde in Kapitel 7 eine umfangreichere Modellierung vorgenommen, welche Deformationsfelder aus der Registrierung von Kern zu Risikogewebe als Surrogat der Deformationsrichtungen einer Infarktläsion annimmt und damit eine räumliche dichte Überwachung des Wachstums zulässt, indem abhängig vom Zeitpunkt die entsprechende Deformation in Vektorrichtung stückweise abgetastet wird. Es wurde in theoretischen Betrachtungen gezeigt, dass dies für einen idealen, monotonen Wachstumsfall die Folgeläsion sehr gut rekonstruieren kann und die teilweise zuvor auftretenden Interpolationsartefakte vermeidet, dass

jedoch die eigentliche Prädiktion für den PCA-Raum an Deformationen schwieriger durch ein CNN zu lernen ist, als die passenden Repräsentationen im Interpolationsmodell aus Kapitel 6.

Zudem kann man sich des Problems der weniger akkuraten Schätzung der Kernläsion bei der Inferenz des Deformationsverfahrens entledigen, da dort nur eine Segmentierungsmaske des Risikogewebes zum Deformieren auf die vorherzusagende finale Läsion benötigt wird.

8.1 Vorteile und Grenzen

Es zeigt sich, dass das im Hauptteil vorgestellte Verfahren der linearen Interpolation in einem gelernten Formraum die Schlaganfallläsionsprädiktion selbst bei einem Trainingsumfang von nur 16 oder 17 Stichproben (und 6 Validierungsfällen) und tendenziell deutlich verrauschten Zeitdaten im Mittel Folgeläsionen genauer vorhersagt, als das mit den populären U-Net-Varianten der Fall ist. Insbesondere zeigen sich folgende Vorteile:

- Die Methode kann vollautomatisch und ohne Kenntnis der Wahl von Hyperparametern oder der Berechnung vorher zu definierender Merkmale angewandt werden.
- Gewebestatuskarten geben räumliche Pseudowahrscheinlichkeiten des Infarktstatus' statt einem skalaren Endpunkt (z. B. mRS) an.
- Damit können Volumenvorhersagen räumlich differenziert werden, statt wie bei einer direkten Regression eines globalen Skalarwertes des voraussichtlichen Volumens in ml, wobei auch dieses darauf berechnet werden kann.
- Die Evolution des Gewebestatus' vom Infarkt seit der Bildgebung kann für einzelne Patienten über die Zeit visuell simuliert werden.
- Es ermöglicht damit dem Anwender im Sinne erklärender KI (Explainable AI) die Entwicklung einer vorhergesagten Folgeläsion zu einem bestimmten Zeitpunkt nachzuvollziehen.
- Die Simulation lässt sich auf modernen, weit verfügbaren Grafikkarten effizient und schnell innerhalb von Sekunden berechnen.

Der klinische Anwender bekommt Anhaltspunkte zur Hand, wie plausibel ihr oder ihm die Vorhersage auf Basis der simulierten Evolution erscheint. Zudem ist die Modellierung des Wachstums im CAE im Prinzip nachvollziehbarer (vgl. Abb. B.11) als die direkte Einzelfallprädiktion mittels U-Net-Inferenz. Noch verständlicher ist die Modellierung durch Registrierungsfelder aus Kapitel 7. Durch diese Wachstumsmodelle erhöht sich auch die Akzeptanz und Nutzerfreundlichkeit der Vorhersageverfahren für eine bessere Risiko-Nutzen-Einschätzung.

Voraussetzung für diese Verlaufssimulation ist ein Training von a-priori gelerntem Formwissen aus Kern- und Risikogewebesegmentierungen von klinischen Experten. Das übliche Problem hierbei ist das Fehlen von Grundwahrheiten durch aufwändige Expertenannotationen und generell kleine Datensätze in der Größe von nur 30–60 Stichproben, die nur zu zwei Zeitpunkten Bild- oder Forminformationen bereitstellen können. Dies stellt bei Nutzung der Methoden des Deep Learnings mit seinen vorzugsweise vielen Parametern, die mit vielen Daten trainiert werden müssen, zunächst eine Schwierigkeit dar. Die U-Net-Architektur bietet dabei einen guten Kompromiss (nicht zu viel Tiefe und Parameter, dabei aber Verknüpfung von Skalenebenen), jedoch wird noch mehr Regularisierung benötigt, um den Sprung von (einfacherer) Segmentierung auf Akutdaten hin zu einer Prädiktion basierend auf Akutdaten zu vollziehen.

Die für den Interpolationsansatz notwendige automatische Kernsegmentierung ist ein noch stark fehleranfälliger Teil der Methode, da wiederum die Daten in Form der CBV-Parameterkarten den Kern nur begrenzt erlauben zu approximieren, und somit der Start der Interpolation nicht immer akkurat geschätzt wird. Eine alternative Modellierung aus Kapitel 7 kann im Anwendungsfall zwar auf Schätzungen des Kernbereichs verzichten, ist aber im Vergleich zum Formraum zu stark eingeschränkt, um eine automatische Prädiktion zu lernen. Hinzu kommt die Tatsache, dass oftmals durch verschiedene Faktoren die tatsächliche Segmentierung der Folgeläsion auf dem CT kleiner ist als die CBV-Läsion als Kern-Approximation, was bei der strikt monotonen Konstruktion der Formrauminterpolation deutlich hinderlich wirkt und den Hauptgrund für die Überschätzung insbesondere kleiner finaler Volumen darstellt. Selbst unter Voraussetzung von idealisiertem Wachstum, können Effekte wie Schwellungen während der zeitlichen Evolution auftreten, welche mit nur einem finalen Verlaufsbild als Überwachung nicht erlernbar sind.

8.2 Ausblick

Es wurde mehrere Problembereiche der hier vorgestellten Ansätze identifiziert. Daher gibt es in Bezug auf die konkrete Weiterentwicklung sowie in Vorbereitung auf erste Studien in der klinischen Anwendungsforschung diverse Fragestellungen, die als nächstes untersucht werden sollten. Zunächst betrifft dies die Datengrundlage des Trainings:

- Das Kernproblem des Datenmangels sollte idealerweise durch weitere klinische Annotationen zu Kerninfarkt, Risikogewebe und Folgeläsion für deutlich mehr der insgesamt verfügbaren Bilddaten des „Trave-Stroke“-Datensatzes (Grundlage der annotierten UKSH-Daten) erweitert werden. Die Datenqualität, insbesondere das Verhältnis zwischen Kern- und Folgeläsion könnte durch einen zweiten unabhängigen Experten quantifiziert und verbessert werden. Hinzu kommt, dass insbesondere auch Daten aus dem Zeitraum 10–24 Stunden zur Erhöhung der Validität der gelernten Modelle berücksichtigt werden sollten, denn gerade die späteren Zeiten sind mit

Blick auf die Wahl der Behandlungsmethoden und entsprechende Selektion anhand der Bildgebung relevant.

- Statt der Erweiterung des UKSH-Datensatzes durch Expertenannotationen, gibt es mittels Deep Learning auch die Möglichkeit zum Modalitätstransfer. Mit den ISLES-Wettbewerben der letzten Jahre liegen bereits verschiedene Perfusionsdatensätze (ISLES 2015 SPES, ISLES 2017, ISLES 2018) mit ausschließlich entweder Kern- und Risikogewebesegmentierungen oder der Segmentierung der finalen Folgeläsion auf einem weiteren Verlaufsbild vor. Das Zusammenbringen dieser Datensätze über eine plausible Synthese des Bildes in die jeweils andere Modalität ist bisher für CT und MR demonstriert worden [Qin u. a., 2019; Zhang u. a., 2018b], vor allem durch Berücksichtigung der Forminvarianz anatomischer Strukturen sowie nachfolgendem Rücktransfer (*engl.* Cycle Consistency) in das ursprüngliche Bild. Dies stellt bei Perfusionsparametern eine Herausforderung dar, weil diese nicht direkt anatomische Strukturen abbilden.

Ein weiteres Problem in Bezug auf die tatsächlichen Daten und den klinisch wie auch in dieser Arbeit zur Methodenentwicklung getroffenen Annahmen ist die mehrfach angesprochene Problematik, dass durch Schwellungen und Schrumpfungen des Gewebes über die Zeit die Kernläsion im gleichen Raum größer ausfallen kann als die Folgeläsion später. Dazu kommt die Unsicherheit darüber, dass die CBV-Läsion oftmals die akute Kernläsion chronisch geschädigten Gewebes überschätzt.

- Die Verwendung der richtigen Eingabedaten sollte systematisch ausgewertet werden. Es ist vorstellbar, dass sich mit weiteren als den CBV- und TTD-Parametern etwa die Kernläsion des Infarkts genauer annotieren und automatisiert schätzen lässt (z. B. zusätzlich CBF), und so bei kleinerer Kernläsion auch kleine Folgeläsionen modelliert werden können. Weiterhin ist eventuell eine Korrektur der Kernläsionsschätzung vorstellbar, welche sich auf eine Untersuchung stützen sollte, inwieweit Schwellungs- und Rückbildungsphänomene abhängig von der Zeit Einfluss auf die Segmentierung haben.
- Eine weitere Möglichkeit ist die Verwendung eines Offsets bei der Evolution, sodass auch eine initiale Zeitverzögerung gelernt werden kann. Man kann annehmen, dass der Infarkt nicht kontinuierlich sondern schlagartig wächst, sobald eine Zeitdauer der Unterversorgung einer Zelle überschritten wird und diese damit abstirbt. Dabei darf man nicht außer Acht lassen, dass die Progressionsgeschwindigkeit zu Beginn kurz nach Symptombeginn größer ist und sich am Ende der Akutphase abschwächt. Darum sollte zusätzlich das Schätzen eines Exponenten für eine nichtlineare zeitliche Wachstumsrate untersucht werden, analog zur Trajektorie-Definition von Couronne u. a. [2019]. In Anbetracht nur zweier zusätzlich zu schätzender Parameter dürfte ein derart modifiziertes Modell weiterhin stark regularisieren und gegenüber Rauschen in den Zeitdaten robust bleiben.

Es wurde in dieser Arbeit deutlich, dass das generelle Vorgehen auf einem Ansatz basieren sollte, der die Informationen über Kern und Penumbra berücksichtigt. Dennoch gibt es auch zur vorgestellten Methode in Kapitel 6, welche dieses am zweckmäßigsten umsetzt, weitere offene Fragestellungen:

- Der Encoder des CAE2, der auf den Segmentierungsschätzungen (des U-Nets) trainiert wird, kann eventuell durch zusätzliche quantitative Informationen aus den Perfusionsparametern profitieren, was in systematischen Versuchen als zusätzlicher Eingabekanal evaluiert werden müsste. Darüber hinaus ist bekannt, dass ein CT-Frühzeichen ein Prädiktor für ein schlechtes klinisches Ergebnis ist. Mit Auftreten dieses dürfte die Läsion eher deutlich (größer) ausfallen. Auch hier sollte die zusätzliche Eingabe der nativen CT-Aufnahme bzw. deren Segmentierung von Frühzeichen überprüft werden. Dazu muss jedoch im Training nicht nur die Verlaufs-CT, sondern auch die native Akut-CT neben den Perfusionsparametern verfügbar sein.
- Ebenso ist bei der ähnlichen, symmetrischen Architektur von U-Net und CAE vorstellbar, das zweistufige Verfahren zu vereinigen, indem das U-Net entfernt und der Encoder direkt auf den Intensitätsbildern der Parameterkarten trainiert wird, sodass er die zwei Repräsentationen des Formraums in einem Schritt schätzt. Da aber bereits das End-To-End-Training mit U-Net und CAE nicht erfolgreich gelang, liegt hier eine besondere Herausforderung, beide Trainingsziele (Kern und Risikogewebe als Endpunkte der Interpolation, sowie Folgeläsion aus deren Interpolation) zusammenzubringen. Es könnte daher sinnvoll sein, nicht nur einen zweiten Encoder wie bisher, sondern auch einen zweiten Decoder zu verwenden, sodass beide Trainingsziele parallel gelernt werden können; alternativ könnte das U-Net direkt parallel zum CAE trainiert werden und sich mit dem Encoder des CAE die Gewichte teilen, ähnlich zum Vorgehen von Pham u. a. [2019].

Das Netzwerk aus Kapitel 7 ist durch die Konstruktion und Einschränkung mit den Deformationsfeldern deutlich schwieriger auf passende Koeffizienten des Modells zur Läsionsvorhersage zu trainieren. Neben der naheliegenden Adaption des Trainings, um alle Parameter durchgehend in einem Schritt inklusive der PCA zu optimieren, welche dafür einmal initial berechnet werden sollte (*engl.* pre-trained), gibt es einige weitere Ansatzpunkte:

- Statt der Datensatzgenerierung auf Basis von CBV- und TTD-Läsion (UKSH-B) sollte man in Bezug auf die Unsicherheit der CBV-Läsion als Kernapproximation besser die TTD-Läsion und (sichere) Folgeläsion von der Verlaufs-CT verwenden, um stattdessen eine Kernläsion zu generieren, die plausibel zur klinischen Hypothese passt. Nach wie vor kann dies aber nur für theoretische Experimente verwandt werden, da die klinische Validität derartiger Daten fraglich bleibt.

- Die Schätzung der PCA-Koeffizienten zur Deformation des Risikogewebes auf die finale Läsion kann statt auf Basis der Segmentierungen (analog zum Ausblick für Kapitel 6) auch direkt auf Grundlage der Parameterintensitäten gelernt werden und damit mehr Informationen als die reinen Segmentierungen zur Verfügung stellen. Ein eventueller Nutzen müsste noch nachgewiesen werden.

Abgesehen von den konkreten Möglichkeiten zur direkten Verbesserung gibt es mittelfristige Ziele, die man im Rahmen der vorgestellten Methodik einarbeiten könnte:

- Das Ziel von Deep Learning sind akkurate, patientenindividuelle Vorhersagen. Um diesem Ziel nahe zu kommen, müssen möglichst viele weitere patientenindividuelle Faktoren berücksichtigt werden. Der prädizierte Gewebestatus sollte nicht nur von Zeit- und Perfusionsparametern abhängig sein, sondern auch der Qualität der Rekanalisation (etwa in Form des TICI-Score) oder eventuellen Frühzeichen im CT, sowie globalen Anhaltspunkten, wie Alter oder Geschlecht.
- Ein hoher Bedarf an Vorhersagemodellen besteht im neurologischen Rahmen vor allem auch bei Hirntumoren. Eine Adaption des vorgestellten Interpolationsverfahrens auf die Prädiktion der verschiedenen Tumorbereiche und die Frage, ob sich damit etwa die Gliomprogression geeignet simulieren ließe, sollte untersucht werden. Daneben sind auch geschlossene Trajektorien im Formraum für sich wiederholende Prozesse, wie Herzbewegungen, möglich und der Einsatz eines gelernten Formraums ist auch hier vorstellbar.
- Sofern andere Biomarker in ein Verlaufsmodell eingearbeitet werden, wäre die multimodale Verlaufsmodellierung auch mit rekurrenten Netzwerken (*engl.* Recurrent Neural Networks, RNN) denkbar, welche den Vorteil zusätzlicher Beobachtungsdaten als Überwachung eines maschinellen Lernverfahrens haben [Couronne u. a., 2019].
- Die Simulation ist ein Schritt in Richtung nachvollziehbarer Entscheidungen von Künstlicher Intelligenz zur Erhöhung der Akzeptanz durch den klinischen Anwender. Das Ableiten von Konfidenz-Werten oder Visualisierung der entscheidenden Bereiche [Zhang u. a., 2018a] in den Eingabedaten (Perfusionskarten bzw. Gewichtung skalarer Faktoren), durch die die entsprechende Simulation oder Vorhersage zustande kommt, wäre als zusätzliche Ausgabe denkbar. Die Eignung und der Nutzen für den Anwender sollte gezielt untersucht werden. Darüber hinaus ist auch die direkte, realistische Dekodierung einer möglichen Verlaufs-CT statt der binären Infarktkarte durch den CAE vorstellbar, was gegebenenfalls die Verwendung eines gegenläufig operierenden Diskriminatoren erforderlich macht (*engl.* Generative Adversarial Networks), damit die generierte Ausgabe möglichst nicht mehr von einer echten Verlaufs-CT unterschieden werden kann. Dies könnte die Akzeptanz bei den klinischen Anwendern weiter erhöhen.

All diese Erweiterungen sollten im Rahmen des ethischen Einsatzes von maschinellen Lernmethoden immer schlussendlich dazu dienen, der Ärztin als verantwortlicher Instanz eine best-informierte Entscheidung zu ermöglichen, sodass auch die spezifischen Einschränkungen der verwendeten Algorithmen durch den Anwender eingeschätzt und die Ergebnisse entsprechend eingeordnet werden können.

A Diagnostik

zu Kapitel 2.2.1.2

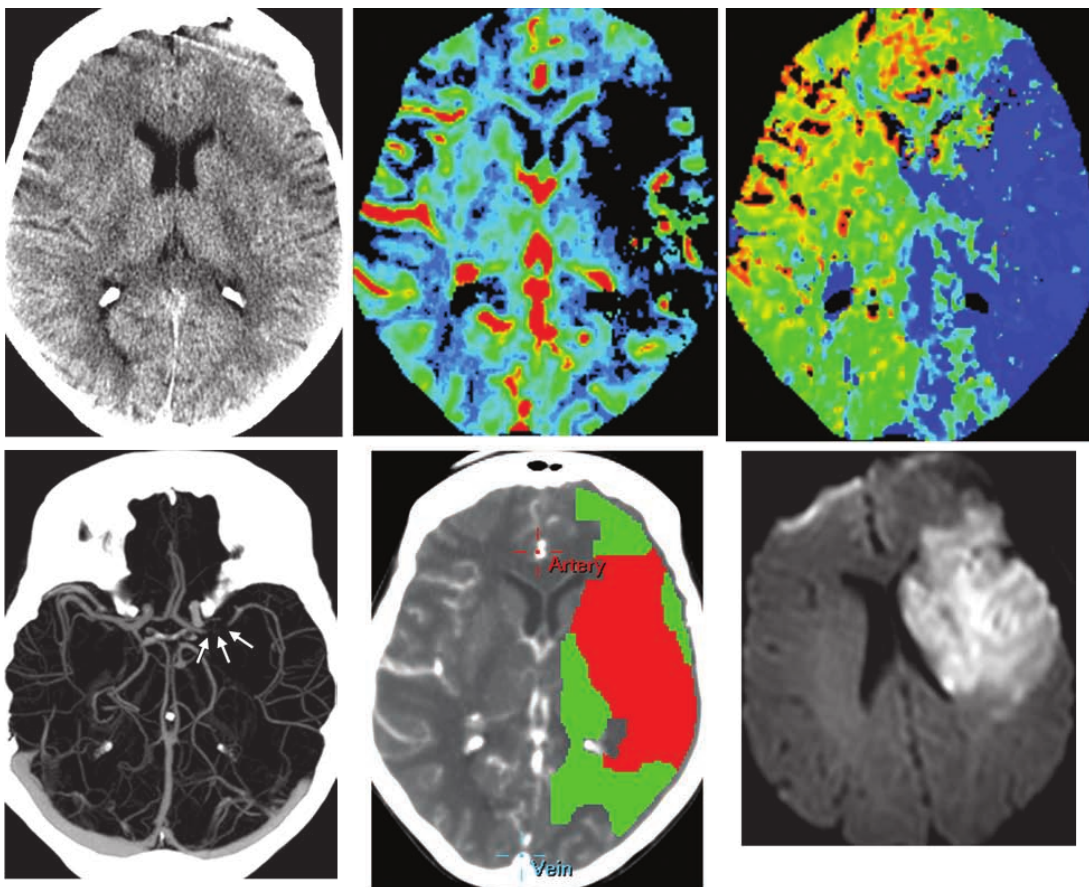


Abb. A.1: Diagnostik mittels Schlaganfallbildgebung nach de Lucas u. a., 2008, die einen geringen Mismatch aufzeigt, v.l.n.r.: Nativ-CT, CBV und MTT der CT-Perfusion (oben), sowie CTA-MIP, Mismatch-Fusionsbild und 24h-MR-Verlaufsbild (unten). Ausgewählte Modalitäten wie in: de Lucas, Sánchez, Gutiérrez et al. CT Protocol for Acute Stroke: Tips and Tricks for General Radiologists. *Radiographics* 2008;28:1673-1687.

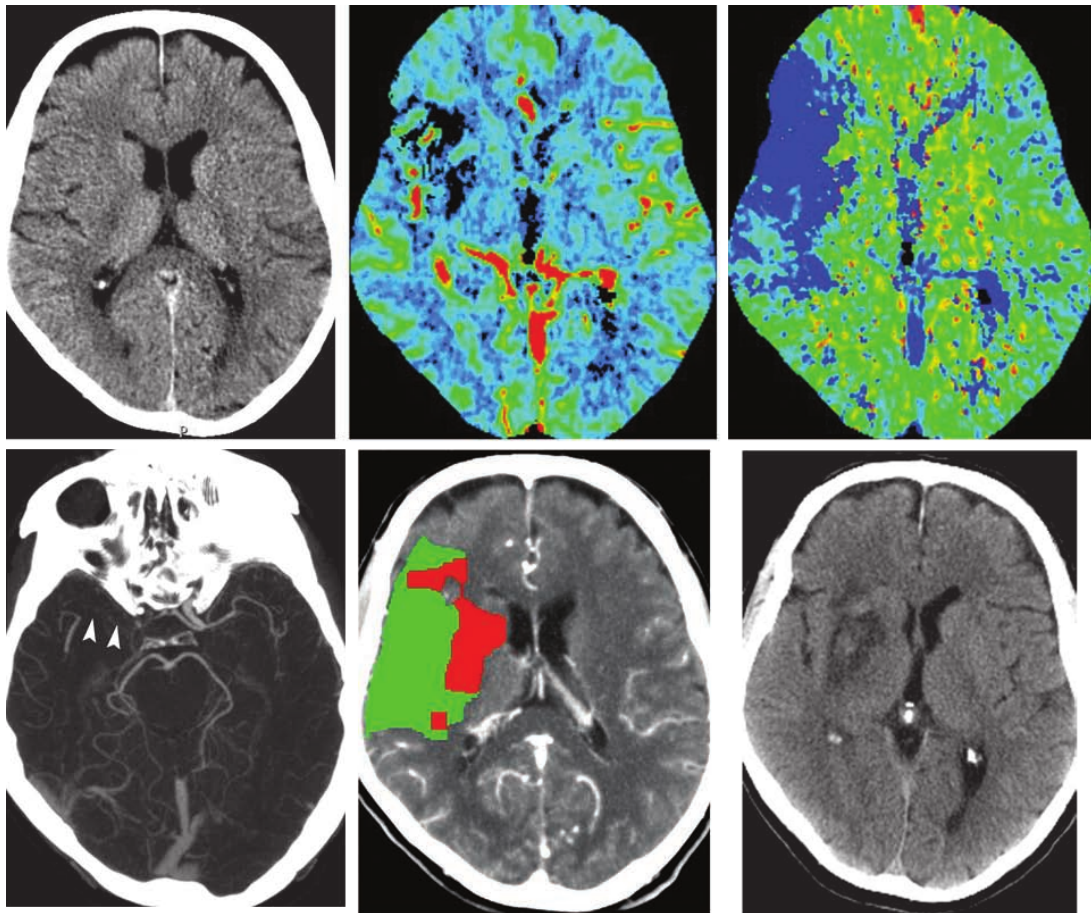


Abb. A.2: Diagnostik mittels Schlaganfallbildgebung nach de Lucas u. a., 2008, die einen großen Mismatch aufzeigt, v.l.n.r.: Nativ-CT, CBV und MTT der CT-Perfusion (oben), sowie CTA-MIP, Mismatch-Fusionsbild und 24h-CT-Verlaufsbild (unten). Ausgewählte Modalitäten wie in: de Lucas, Sánchez, Gutiérrez et al. CT Protocol for Acute Stroke: Tips and Tricks for General Radiologists. *Radiographics* 2008;28:1673-1687.

B Ergebnisse

zu Kapitel 5

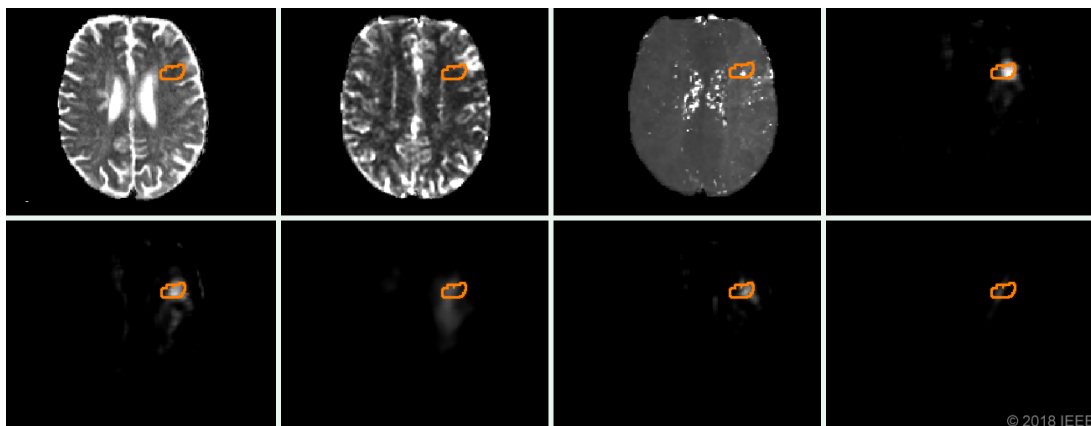


Abb. B.1: Kleine Läsionen bei Prädiktion der Verlaufsbild-Segmentierung. Gezeigt werden v.l.n.r.: ADC, CBF, TTP und die Segmentierungsschätzung mit der vorgeschlagenen Methode aus Kapitel 5. Die untere Reihe zeigt die Schätzung bei Nutzung von Residual-Skips mit Multi-Loss, bei Nutzung ausschließlich des Multi-Losses, bei Nutzung ausschließlich der Konkatination, sowie das normale U-Net ohne Erweiterung. Die Grundwahrheit ist als orangene Umrandung (●) eingeblendet. Man erkennt hier typisch das präzise Segmentieren der vorgeschlagenen Methode insbesondere kleiner Strukturen und Details, wie der hier gezeigten eng begrenzten Läsion.

zu Kapitel 6

Seite 126-129: Gezeigt werden oben das CBV-, TTD- und AVG-Bild der CTP, sowie daneben die Formrauminterpolation auf den Grundwahrheiten. In der unteren Zeile werden dagegen die Schätzungen gezeigt (mittig: Ergebnis der vorgeschlagenen Methode). Zum Vergleich sind die Umriss der Grundwahrheiten von CBV-Segmentierung (●), TTD-Segmentierung (●) und finaler Läsionsform (●) eingeblendet (Werte in grau: Dice Übereinstimmung). Die Zeiten sind im Format *Stunden : Minuten* angegeben.

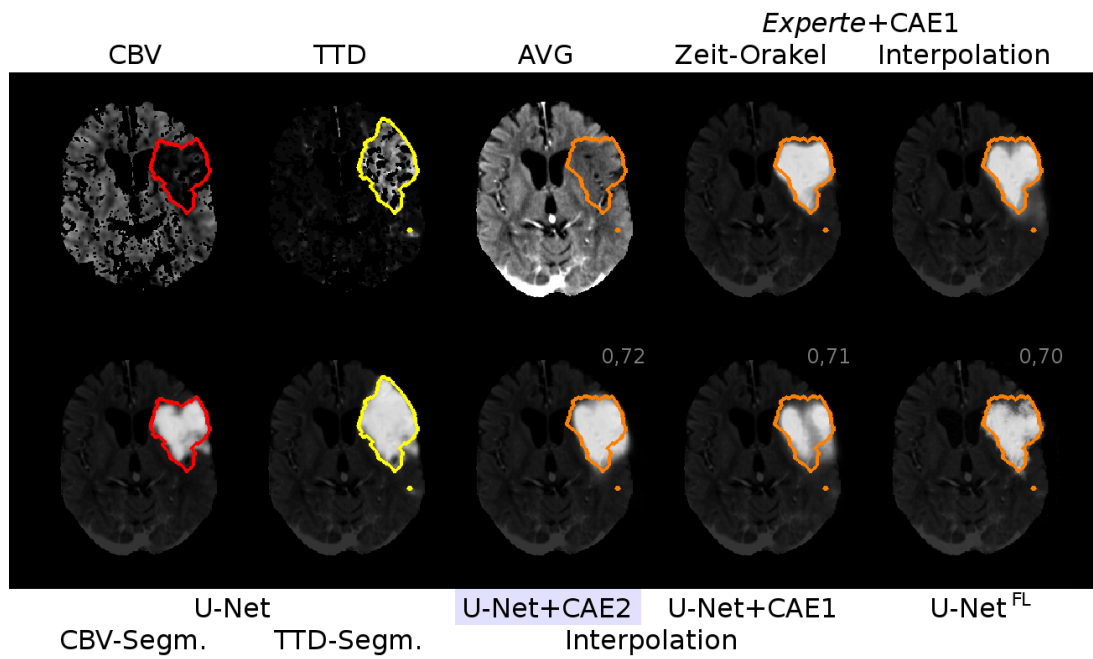


Abb. B.2: Fall 2: $t_{0 \rightarrow A} = 3 : 29$, $t_{A \rightarrow I} = 1 : 14$, $\eta = 0,19$ (Erläuterung auf Seite 125)

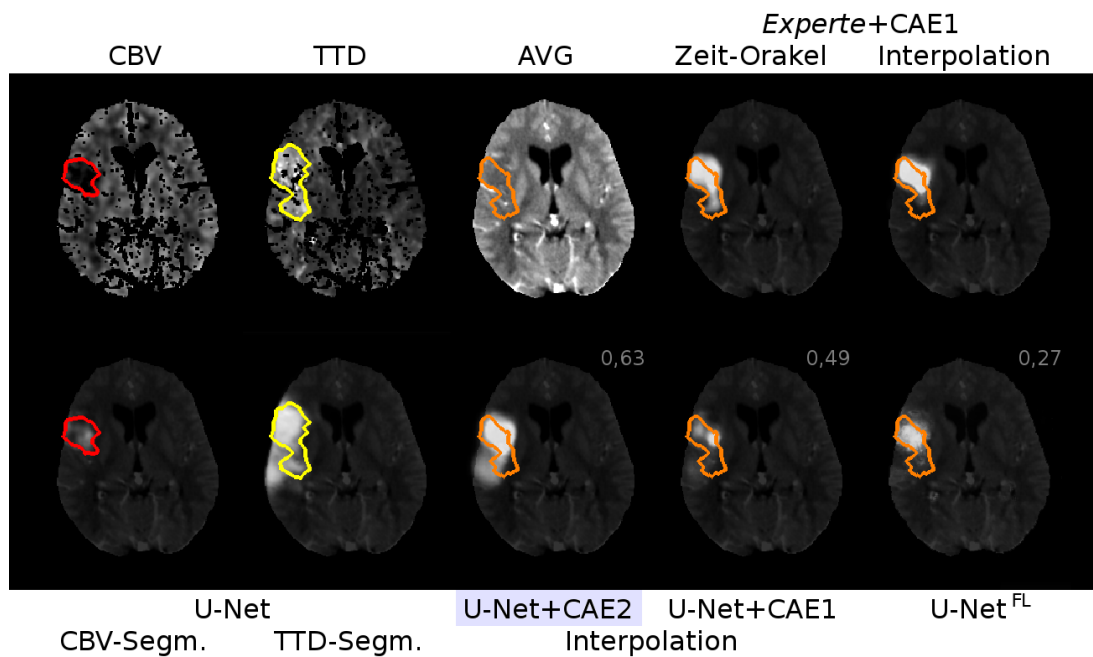


Abb. B.3: Fall 3: $t_{0 \rightarrow A} = 1 : 42$, $t_{A \rightarrow I} = 1 : 43$, $\eta = 0,21$ (Erläuterung auf Seite 125)

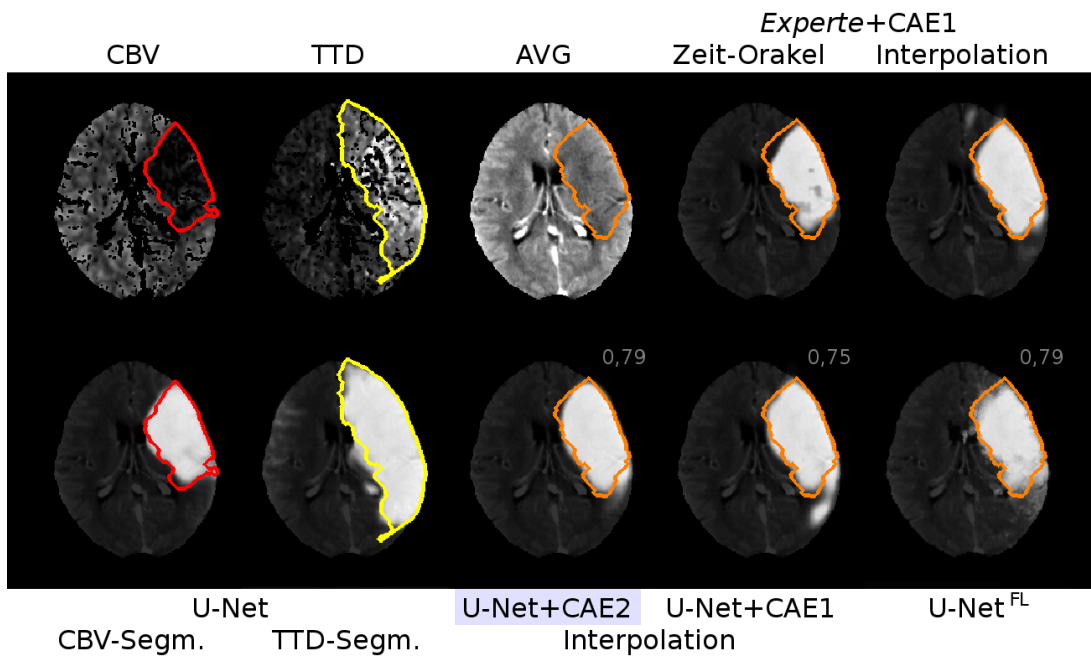


Abb. B.4: Fall 8: $t_{0 \rightarrow A} = 1 : 13$, $t_{A \rightarrow I} = 2 : 57$, $\eta = 0,34$ (Erläuterung auf Seite 125)

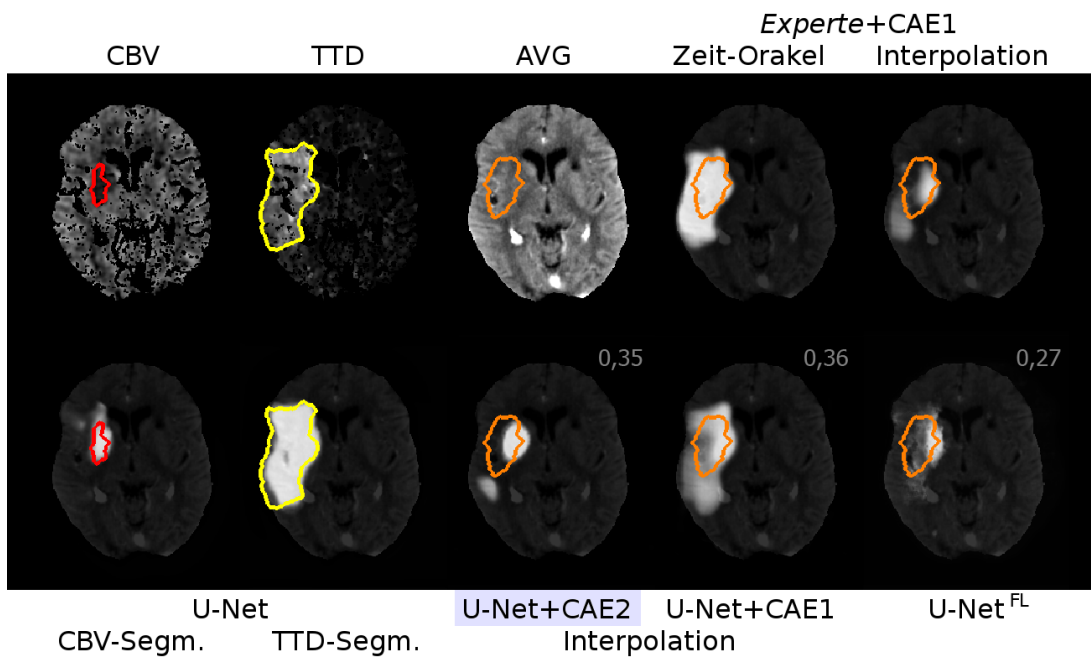


Abb. B.5: Fall 11: $t_{0 \rightarrow A} = 0 : 35$, $t_{A \rightarrow I} = 2 : 06$, $\eta = 0,22$ (Erläuterung auf Seite 125)

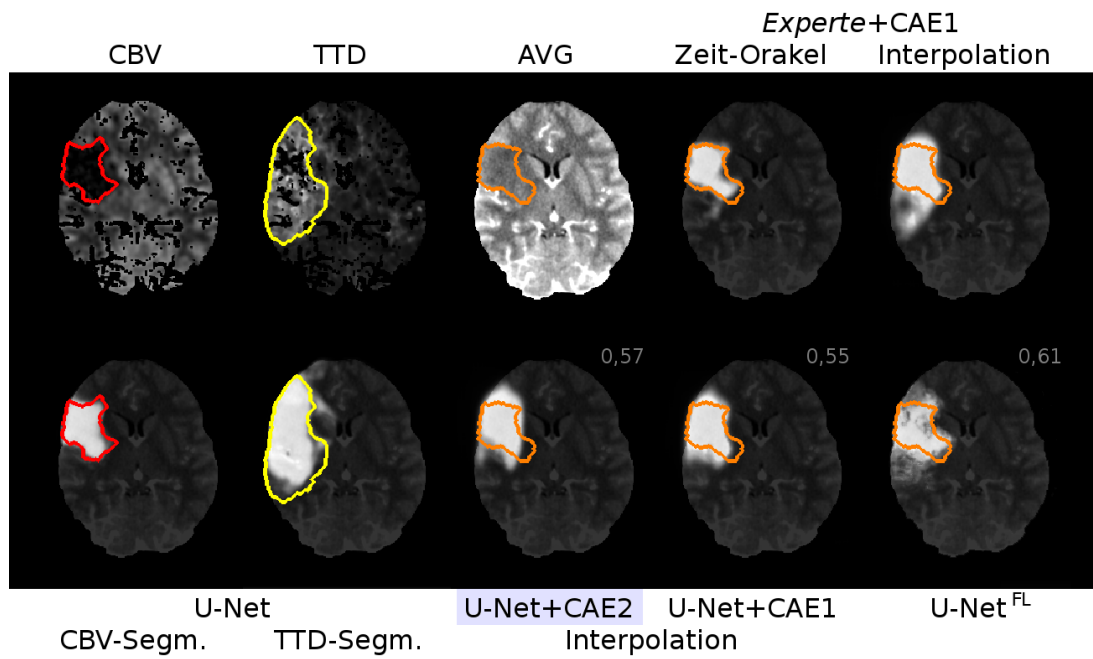


Abb. B.6: Fall 16: $t_{0 \rightarrow A} = 4 : 30$, $t_{A \rightarrow I} = 2 : 19$, $\eta = 0,42$ (Erläuterung auf Seite 125)

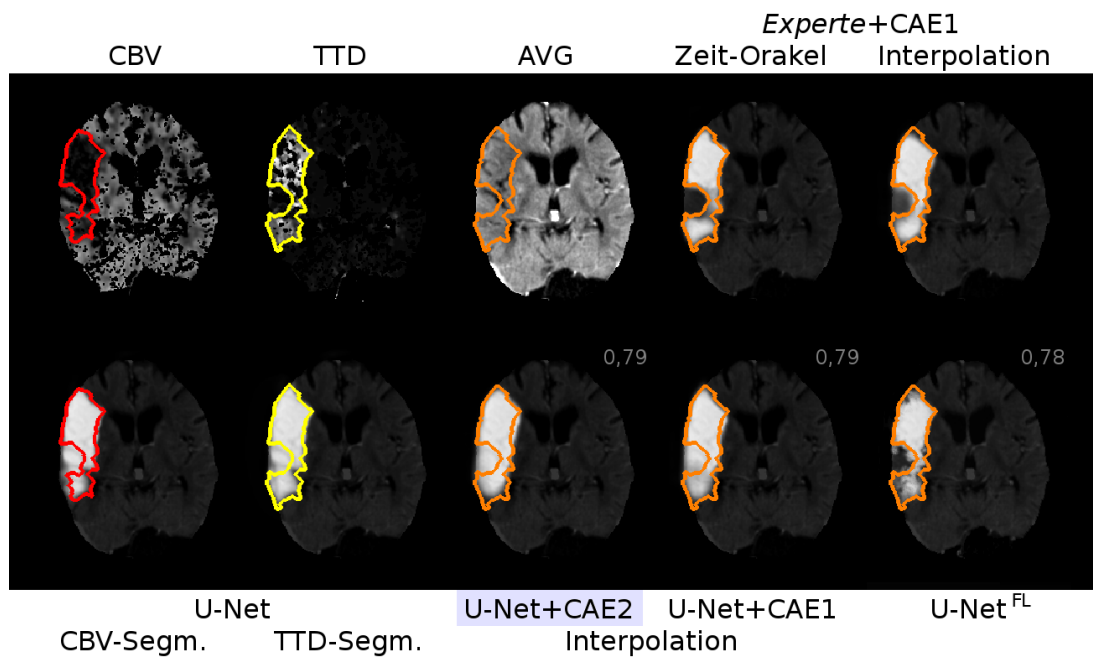


Abb. B.7: Fall 18: $t_{0 \rightarrow A} = 1 : 15$, $t_{A \rightarrow I} = 1 : 55$, $\eta = 0,22$ (Erläuterung auf Seite 125)

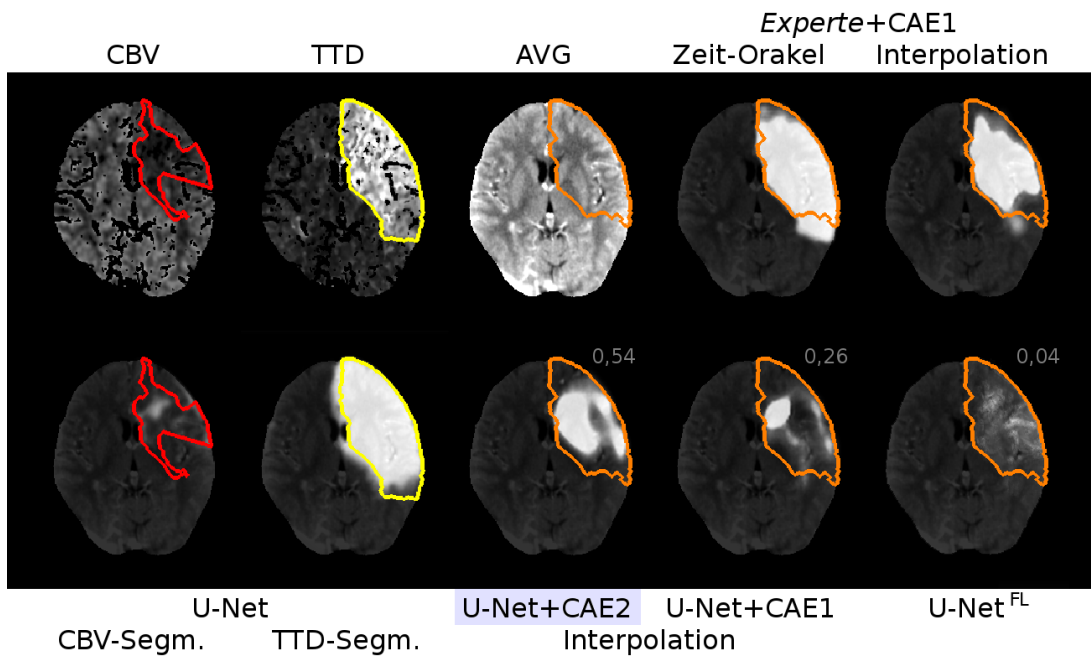


Abb. B.8: Fall 21: $t_{0 \rightarrow A} = 1 : 40$, $t_{A \rightarrow I} = 1 : 26$, $\eta = 0,16$ (Erläuterung auf Seite 125)

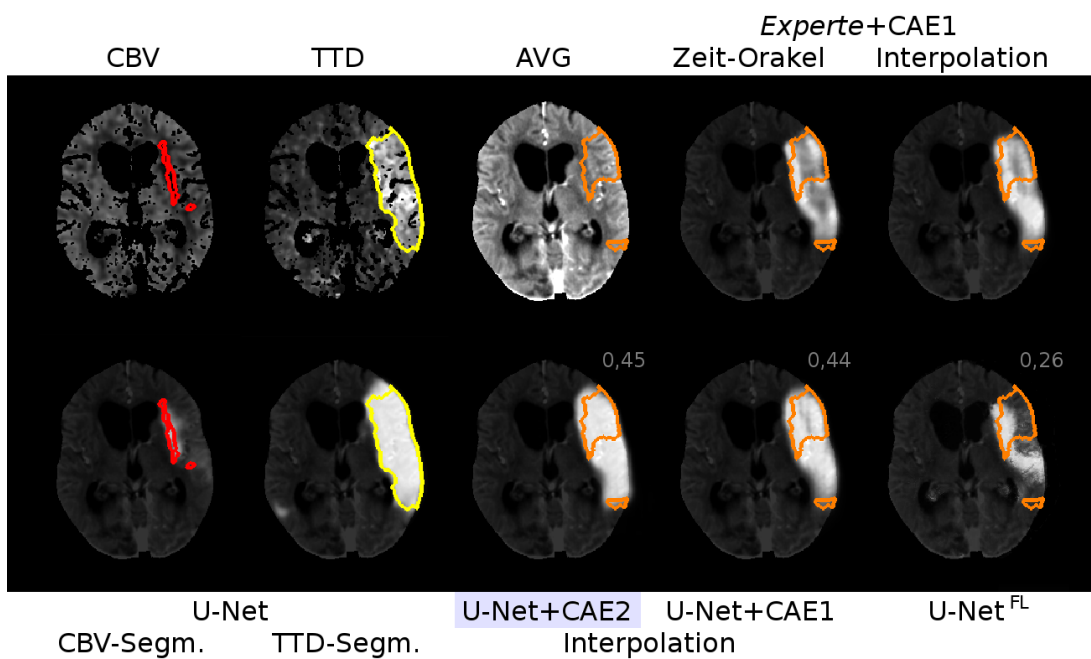


Abb. B.9: Fall 25: $t_{0 \rightarrow A} = 1 : 32$, $t_{A \rightarrow I} = 3 : 58$, $\eta = 0,47$ (Erläuterung auf Seite 125)

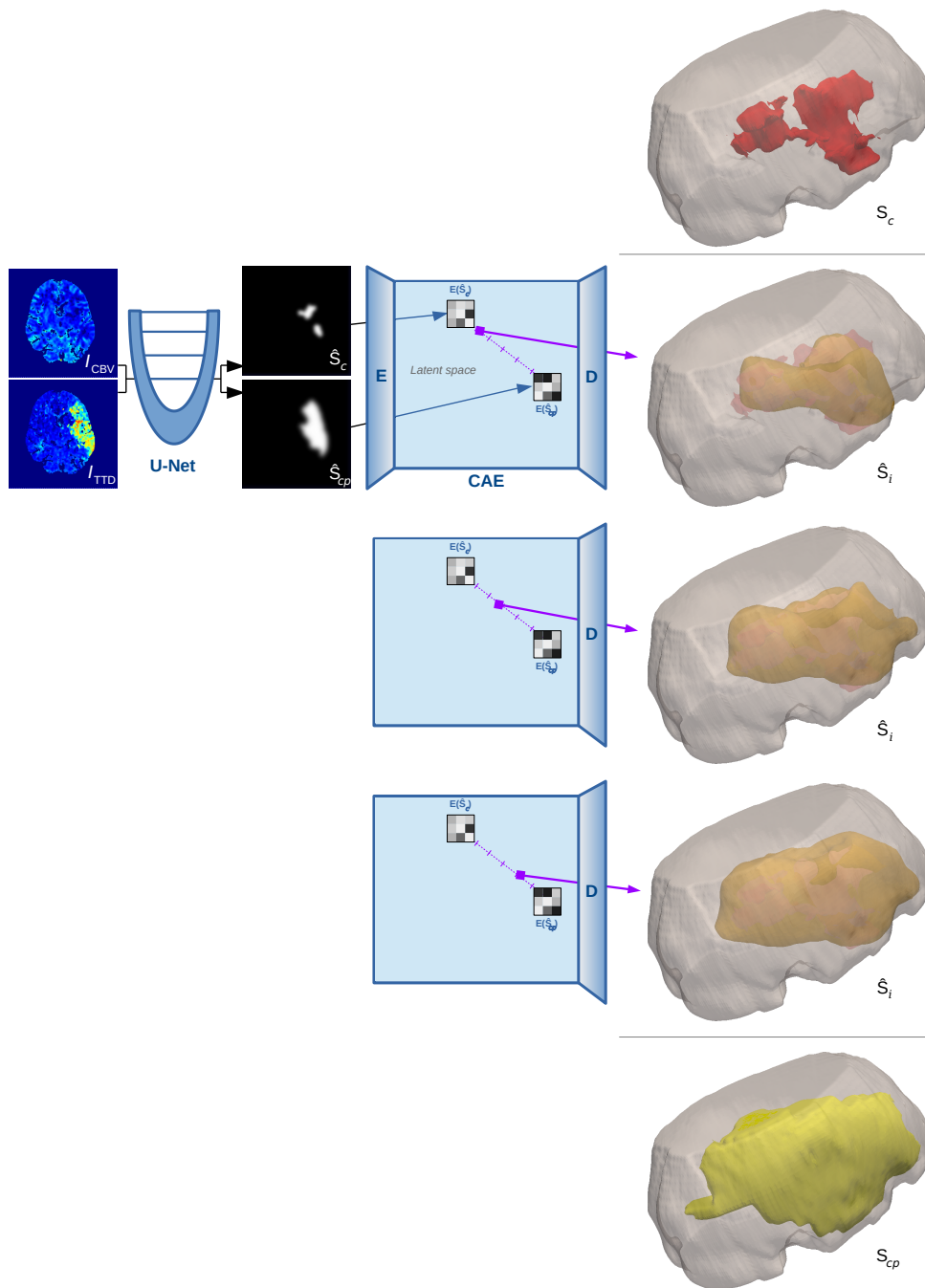


Abb. B.10: 3D-Oberflächendarstellung der binarisierten Rekonstruktion einer Interpolation im Formraum für einen Testfall aus den Experimenten aus Kapitel 6. Das oberste Bild zeigt die manuelle Segmentierung des Kerns in rot innerhalb eines Ausschnitts des Gehirnvolumens, während das unterste Bild die manuelle Segmentierung des Risikogewebes in gelb zeigt. Drei Zwischenschritte der Interpolation zwischen den Schätzungen der beiden manuellen Segmentierungen sind dazwischen dargestellt. [www.github.com/multimodallearning/stroke-prediction/blob/master/animated.gif]

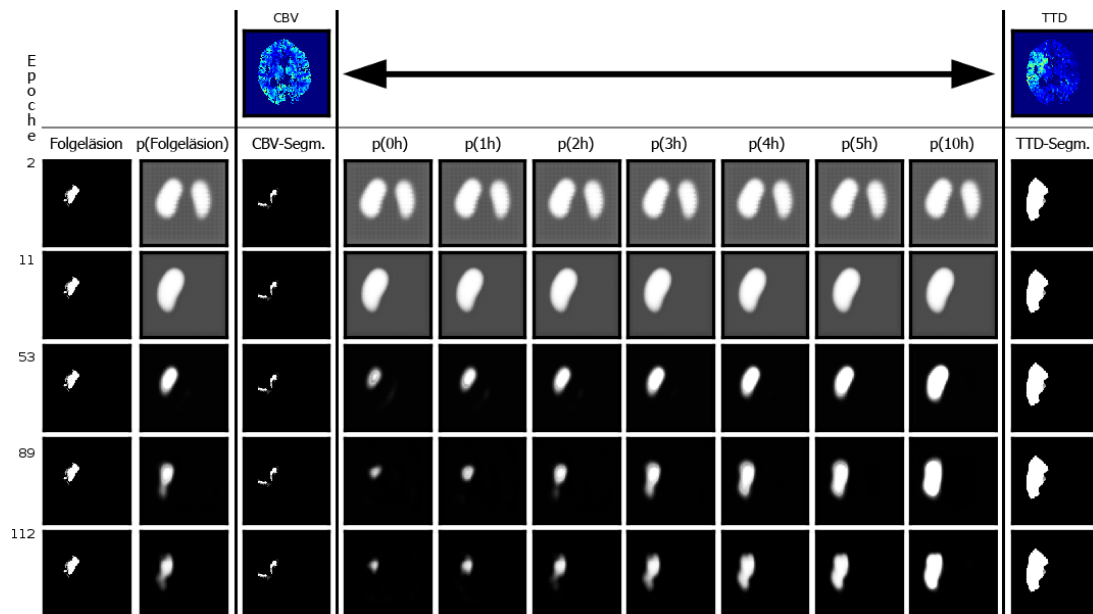


Abb. B.11: Visualisierung des Trainings vom Formraum (CAE1) anhand eines Validierungsbeispiels. Es wird deutlich, wie zuerst gelernt wird in der latenten Repräsentation die richtige Hemisphäre zu unterscheiden, danach die grobe Größenskalisierung und schließlich konkrete Läsionsformen zu abstrahieren. Parallel dazu wird möglichst eine Formrauminterpolation („p(Folgeläsion)“) nahe der Kodierung der tatsächlichen, finalen Läsionsform („Folgeläsion“) erzwungen.

Folgeläsion	Experte+CAE1	U-Net+CAE2	U-Net ^{FL}	ICC	95 %-KI
•	•			0,67	[0,41; 0,83]
•		•		0,50	[0,18; 0,73]
•			•	0,11	[-0,25; 0,45]
•	•	•		0,65	[0,45; 0,80]
	•	•		0,81	[0,63; 0,90]
	•	•	•	0,56	[0,35; 0,74]

Tabelle B.1: Interrater-Korrelation zwischen den Volumenvorhersagen in *ml* der manuellen Segmentierung des Verlaufbildes („Folgeläsion“) und den automatischen Prädiktionsmethoden. Zur Schätzung wurde das ICC(3,1)-Modell [Koo u. a., 2016] der Intraklassen-Korrelation (*engl.* Intraclass Correlation, ICC) nach Shrout u. a. [1979] verwendet, dessen Ergebnisse bei zwei Bewertern ungefähr der Korrelation in Abb. 6.9(b) auf S. 95 entsprechen. Trotz kaum vorhandener Übereinstimmung (ICC=0,11) der U-Net-Schätzungen mit den Volumen der tatsächlichen Folgeläsionen gibt es eine moderate Übereinstimmung (ICC=0,56) zwischen allen automatischen Ansätzen, wobei die Übereinstimmung zwischen den CAE-basierten Ansätzen erwartbar gut ausgeprägt ist (ICC=0,81).

Segm.	Methode Prädiktion	Netzwerk- Parameter	Dice		
			Kern	Risikogew.	Prädiktion
<i>Experte</i>	<i>CBV Segm.</i>	–	1,00	–	0,39
<i>Experte</i>	<i>TTD Segm.</i>	–	–	1,00	0,44
<i>Experte</i>	CAE1-Orakel	$4,7 \cdot 10^6$	0,68 (R_c)	0,90 (R_{cp})	0,53 (R_i)
<i>Experte</i>	CAE1-Interp.	$4,7 \cdot 10^6$	0,68 (R_c)	0,90 (R_{cp})	0,46 (R_i)
<i>Experte</i>	CAE1-I. 24h	$4,7 \cdot 10^6$	0,70 (R_c)	0,90 (R_{cp})	0,44 (R_i)
U-Net	CAE1-Interp.	$3,6 \cdot 10^5 + 4,7 \cdot 10^6$	0,43	0,80	0,40
U-Net	CAE2-Interp.	$3,6 \cdot 10^5 + 4,7 \cdot 10^6$	0,55	0,81	0,43 (\hat{S}_i)
U-Net	Segm.-Interp.	$3,6 \cdot 10^5$	0,45 (\hat{S}_c)	0,81 (\hat{S}_{cp})	0,36
U-Net	<i>CBV Segm.</i>	$3,6 \cdot 10^5$	0,45 (\hat{S}_c)	–	0,33 (\hat{S}_c)
U-Net	<i>TTD Segm.</i>	$3,6 \cdot 10^5$	–	0,81 (\hat{S}_{cp})	0,35 (\hat{S}_{cp})
–	U-Net	$3,6 \cdot 10^5$	–	–	0,34
–	U-Net <i>Metad.</i>	$3,6 \cdot 10^5$	–	–	0,22
–	U-Net ^{FL}	$1,1 \cdot 10^7$	–	–	0,35
U-Net ^{FL}	CAE2-Interp.	$1,1 \cdot 10^7 + 4,7 \cdot 10^6$	0,47	0,83	0,41 (\hat{S}_i)
U-Net ^{FL}	<i>CBV Segm.</i>	$1,1 \cdot 10^7$	0,62 (\hat{S}_c)	–	0,36 (\hat{S}_c)
U-Net ^{FL}	<i>TTD Segm.</i>	$1,1 \cdot 10^7$	–	0,72 (\hat{S}_{cp})	0,25 (\hat{S}_{cp})

Tabelle B.2: Vollständige Dice-Resultate: Das U-Net berücksichtigt keine klinischen Metadaten, wie z. B. Zeitpunkte. Zum Vergleich ist einmal die Prädiktion mit Zeitdaten („*Metad.*“) aufgeführt. Unterschieden wird ein Standard-U-Net in 3D (Kapitel 4.2) und jenes aus Kapitel 5 (U-Net^{FL}). „*Experte*“ bezeichnet die Grundwahrheit der Segmentierungsmasken. Die Interpolation („Interp.“) ist auf einen Zeitbereich von 0–10 Stunden genormt – die Normierung auf 24h ist zum Vergleich ebenfalls gelistet. Für den CAE wurde jeweils angegeben, ob nur die erste (CAE1) oder beide (CAE2) Phasen trainiert wurden. Die in Kapitel 6.2 vorgeschlagene vollautomatische Methode (U-Net+CAE2) ist blau hinterlegt. Das Orakel optimiert den Interpolationszeitpunkt anhand der bestmöglichen Rekonstruktion der Grundwahrheit. Das beste Ergebnis für die Segmentierungsschätzungen von Kern und Penumbra sowie die Prädiktion der finalen Läsion ist jeweils **fett** markiert. Als Referenz sind außerdem die Werte der Überlappung von *CBV Segm.* und *TTD Segm.* je Segmentierungsmethode angegeben.

zu Kapitel 7

Die ersten drei Spalten zeigen die Segmentierungsgrundwahrheiten, während die letzten drei Spalten die Vorhersagen zur tatsächlichen finalen Läsion („FL Segm.“) zeigen: Prädiktion von PCA und CAE, sowie Code-Orakel der PCA.

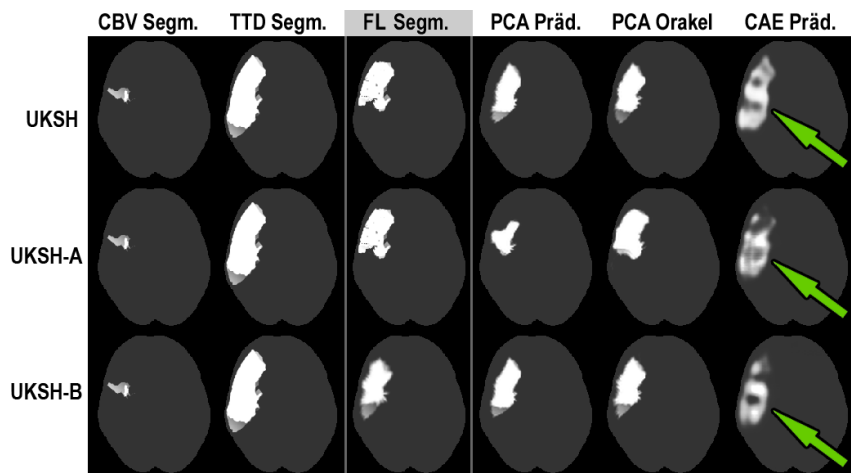


Abb. B.12: Ergebnisse für einen Beispielfall in der PCA-Deformation: Unplausible Interpolationen (grüner Pfeil) können insbesondere bei der CAE-Interpolation aus Kapitel 6 auftreten.

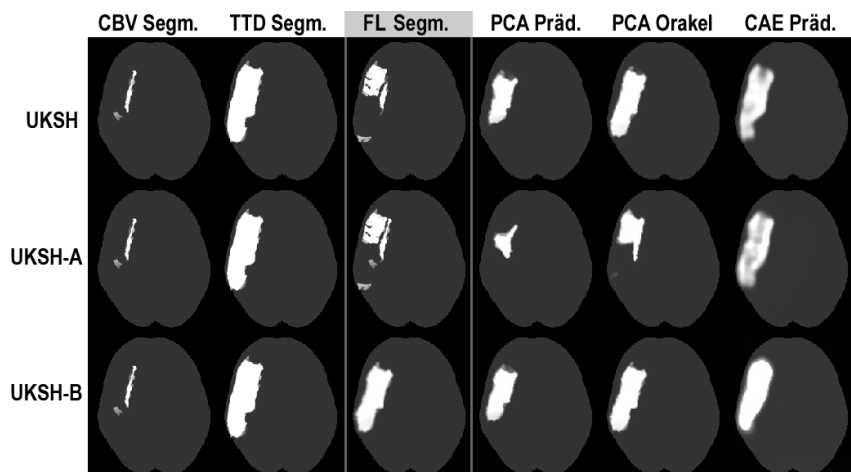


Abb. B.13: Ergebnisse für einen weiteren Beispielfall in der PCA-Deformation.

C Eigene Publikationen

Folgende Beiträge haben Teilergebnisse dieser Arbeit unter Peer-Review veröffentlicht:

Erstautorenschaft

- Lucas C, Aulmann LF, Kemmling A, Madany Mamlouk A, Heinrich MP. Abstract: Estimation of the principal ischaemic stroke growth directions for predicting tissue outcomes. In: *Bildverarbeitung für die Medizin* (BVM), 2020, S. 143. (Kapitel 7)
- Lucas C, Aulmann LF, Kemmling A, Madany Mamlouk A, Heinrich MP. Estimation of the principal ischaemic stroke growth directions for predicting tissue outcomes. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 2020, S. 69–79. (Kapitel 7)
- Lucas C, Schöttler JJ, Kemmling A, Aulmann LF, Heinrich MP. Automatic Detection and Segmentation of the Acute Vessel Thrombus in Cerebral CT. In: *Bildverarbeitung für die Medizin* (BVM), 2019, S. 74–79. (Kapitel 4.1)
- Lucas C, Kemmling A, Bouteldja N, Aulmann LF, Madany Mamlouk A, Heinrich MP. Learning to predict ischemic stroke growth on acute CT perfusion data by interpolating low-dimensional shape representations. *Frontiers in Neurology* 9(989), 2018. (Kapitel 4.2, Kapitel 6)
- Lucas C, Kemmling A, Madany Mamlouk A, Heinrich MP. Multi-scale neural network for automatic segmentation of ischemic strokes on acute perfusion images. In: *IEEE 15th International Symposium on Biomedical Imaging (ISBI)*, 2018, S. 1118–1121. (Kapitel 5)
- Lucas C, Maier O, Heinrich MP. Shallow fully-connected neural networks for ischemic stroke-lesion segmentation in MRI. In: *Bildverarbeitung für die Medizin* (BVM), 2017, S. 261–266. (Kapitel 3.4.1)

Ko-Autorenschaft

- Winzeck S, Hakim A, McKinley R, Pinto JAADSR, Alves V, Silva C, Pisov M, Krivov E, Belyaev M, Monteiro M, Oliveira A, Choi Y, Paik MC, Kwon Y, Lee H, Kim BJ, Won JH, Islam M, Ren H, Robben D, Suetens P, Gong E, Niu Y, Xu J, Pauly JM, Lucas C, Heinrich MP, Rivera LC, Castillo LS, Daza LA, Beers AL, Arbelaez P, Maier O, Chang K, Brown JM, Kalpathy-Cramer J, Zaharchuk G, Wiest R, Reyes M. ISLES 2016 and 2017-Benchmarking Ischemic Stroke Lesion Outcome Prediction Based on Multispectral MRI. *Frontiers in Neurology* 9(679), 2018. (Kapitel 5.3)

Wettbewerbsteilnahme

- Lucas C, Heinrich MP: Zweiter Platz bei der ISLES im Rahmen der MICCAI 2017 www.isles-challenge.org/ISLES2017, Stand: 20.3.2020 (Kapitel 5)

Freie Quellcodes

- Lucas C: www.github.com/multimodalllearning/stroke-prediction, Stand: 20.3.2020
3D Prädiktion des Gewebe-Outcomes bei Schlaganfall für PyTorch (Kapitel 6)

Literaturverzeichnis

- [Abulnaga u. a., 2019] Abulnaga, S. M. und Rubin, J. “Ischemic Stroke Lesion Segmentation in CT Perfusion Scans Using Pyramid Pooling and Focal Loss”. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2019, S. 352–363.
- [Alberts u. a., 2016] Alberts, E., Charpiat, G., Tarabalka, Y., Huber, T., Weber, M.-A., Bauer, J., Zimmer, C. und Menze, B. H. “A Nonparametric Growth Model for Brain Tumor Segmentation in Longitudinal MR Sequences”. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2016, S. 69–79.
- [Albers u. a., 2018] Albers, G. W., Marks, M. P., Kemp, S., Christensen, S., Tsai, J. P., Ortega-Gutierrez, S., McTaggart, R. A., Torbey, M. T., Kim-Tenser, M., Leslie-Mazwi, T., Sarraj, A., Kasner, S. E., Ansari, S. A., Yeatts, S. D., Hamilton, S., Mlynash, M., Heit, J. J., Zaharchuk, G., Kim, S., Carrozzella, J., Palesch, Y. Y., Demchuk, A. M., Bammer, R., Lavori, P. W., Broderick, J. P. und Lansberg, M. G. “Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging”. *New England Journal of Medicine* 378 (8), 2018, S. 708–718.
- [Allen u. a., 2014] Allen, L. M., Hasso, A. N., Handwerker, J. und Farid, H. “Sequence-specific MR imaging findings that are useful in dating ischemic stroke”. *Neuroradiologie Scan* 4 (1), 2014, S. 45–59.
- [AllEx, 2012] AllEx. *Alles fürs Examen. Das Kompendium für die 2. ÄP*. Georg Thieme Verlag KG, Stuttgart, 2012.
- [Angermaier u. a., 2015] Angermaier, A., Khaw, A. V., Kirsch, M., Kessler, C. und Langner, S. “Influence of Recanalization and Time of Cerebral Ischemia on Tissue Outcome after Endovascular Stroke Treatment on Computed Tomography Perfusion”. *Journal of Stroke and Cerebrovascular Diseases* 24 (10), 2015, S. 2306–2312.
- [Angermaier u. a., 2016] Angermaier, A. und Langner, S. “Recanalization and Reperfusion in Acute Stroke - More Often Different than Alike”. *Clinical Neuroradiology* 26 (3), 2016, S. 375–376.
- [Asadi u. a., 2014] Asadi, H., Dowling, R., Yan, B. und Mitchell, P. “Machine Learning for Outcome Prediction of Acute Ischemic Stroke Post Intra-Arterial Therapy”. *PLOS ONE* 9 (2), 2014, S. 1–11.

- [Badhiwala u. a., 2015] Badhiwala, J. H., Nassiri, F., Alhazzani, W., Selim, M. H., Farrokhhyar, F., Spears, J., Kulkarni, A. V., Singh, S., Alqahtani, A., Rochweg, B., Alshahrani, M., Murty, N. K., Alhazzani, A., Yarascavitch, B., Reddy, K., Zaidat, O. O. und Almenawer, S. A. “Endovascular Thrombectomy for Acute Ischemic Stroke: A Meta-analysis”. *JAMA* 314 (17), 2015, S. 1832–1843.
- [Bagher-Ebadian u. a., 2011] Bagher-Ebadian, H., Jafari-Khouzani, K., Mitsias, P. D., Lu, M., Soltanian-Zadeh, H., Chopp, M. und Ewing, J. R. “Predicting Final Extent of Ischemic Infarction Using Artificial Neural Network Analysis of Multi-Parametric MRI in Patients with Stroke”. *PLOS ONE* 6 (8), 2011, S. 1–9.
- [Barber u. a., 2000] Barber, P. A., Demchuk, A. M., Zhang, J. und Buchan, A. M. “Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy”. *The Lancet* 355 (9216), 2000, S. 1670–1674.
- [Barow u. a., 2019] Barow, E. und Thomalla, G. “Akutbehandlung des ischämischen Schlaganfalls”. *Der Nervenarzt* 90 (10), 2019, S. 979–986.
- [Bauer u. a., 2014] Bauer, S., Tessier, J., Krieter, O., Nolte, L.-P. und Reyes, M. “Integrated Spatio-Temporal Segmentation of Longitudinal Brain Tumor Imaging Studies”. In: *Medical Computer Vision. Large Data in Medical Imaging*. Springer International Publishing, 2014, S. 74–83.
- [Bentley u. a., 2014] Bentley, P., Ganesalingam, J., Jones, A. L. C., Mahady, K., Epton, S., Rinne, P., Sharma, P., Halse, O., Mehta, A. und Rueckert, D. “Prediction of stroke thrombolysis outcome using CT brain machine learning”. *NeuroImage: Clinical* 4, 2014, S. 635–640.
- [Bertels u. a., 2019] Bertels, J., Robben, D. und Vandermeulen, D. “Contra-Lateral Information CNN for Core Lesion Segmentation Based on Native CTP in Acute Stroke”. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2019, S. 263–270.
- [Berthelot u. a., 2019] Berthelot, D., Raffel, C., Roy, A. und Goodfellow, I. “Understanding and Improving Interpolation in Autoencoders via an Adversarial Regularizer”. In: *International Conference on Learning Representations (ICLR)*. 2019.
- [Bertels u. a., 2020] Bertels, J., Robben, D., Vandermeulen, D. und Suetens, P. “Optimization with Soft Dice Can Lead to a Volumetric Bias”. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2020, S. 89–97.
- [Boers u. a., 2018] Boers, A. M. M., Jansen, I. G. H., Beenen, L. F. M., Devlin, T. G., San Roman, L., Heo, J. H., Ribó, M., Brown, S., Almekhlafi, M. A., Liebeskind, D. S., Teitelbaum, J., Lingsma, H. F., Zwam, W. H. van, Cuadras, P., Mesnil de Rochemont, R. du, Beaumont, M., Brown, M. M., Yoo, A. J., Oostenbrugge, R. J. van, Menon, B. K., Donnan, G. A., Mas, J. L., Roos, Y. B. W. E. M., Oppen-

-
- heim, C., Lugt, A. van der, Dowling, R. J., Hill, M. D., Davalos, A., Moulin, T., Agrinier, N., Demchuk, A. M., Lopes, D. K., Aja Rodríguez, L., Dippel, D. W. J., Campbell, B. C. V., Mitchell, P. J., Al-Ajlan, F. S., Jovin, T. G., Madigan, J., Albers, G. W., Soize, S., Guillemin, F., Reddy, V. K., Bracard, S., Blasco, J., Muir, K. W., Nogueira, R. G., White, P. M., Goyal, M., Davis, S. M., Marquering, H. A. und Majoie, C. B. L. M. “Association of follow-up infarct volume with functional outcome in acute ischemic stroke: a pooled analysis of seven randomized trials”. *Journal of NeuroInterventional Surgery* 10 (12), 2018, S. 1137–1142.
- [Boers u. a., 2019] Boers, A. M. M., Jansen, I. G. H., Brown, S., Lingsma, H. F., Beenen, L. F. M., Devlin, T. G., Román, L. S., Heo, J.-H., Ribó, M., Almekhlafi, M. A., Liebeskind, D. S., Teitelbaum, J., Cuadras, P., Mesnil de Rochemont, R. du, Beaumont, M., Brown, M. M., Yoo, A. J., Donnan, G. A., Mas, J. L., Oppenheim, C., Dowling, R. J., Moulin, T., Agrinier, N., Lopes, D. K., Aja Rodríguez, L., Compagne, K. C. J., Al-Ajlan, F. S., Madigan, J., Albers, G. W., Soize, S., Blasco, J., Davis, S. M., Nogueira, R. G., Dávalos, A., Menon, B. K., Lugt, A. van der, Muir, K. W., Roos, Y. B. W. E. M., White, P., Mitchell, P. J., Demchuk, A. M., Zwam, W. H. van, Jovin, T. G., Oostenbrugge, R. J. van, Dippel, D. W. J., Campbell, B. C. V., Guillemin, F., Bracard, S., Hill, M. D., Goyal, M., Marquering, H. A. und Majoie, C. B. L. M. “Mediation of the Relationship Between Endovascular Therapy and Functional Outcome by Follow-up Infarct Volume in Patients With Acute Ischemic Stroke”. *JAMA Neurology* 76 (2), 2019, S. 194–202.
- [Bourcier u. a., 2019] Bourcier, R., Goyal, M., Liebeskind, D. S., Muir, K. W., Desal, H., Siddiqui, A. H., Dippel, D. W. J., Majoie, C. B., Zwam, W. H. van, Jovin, T. G., Levy, E. I., Mitchell, P. J., Berkhemer, O. A., Davis, S. M., Derraz, I., Donnan, G. A., Demchuk, A. M., Oostenbrugge, R. J. van, Kelly, M., Roos, Y. B., Jahan, R., Lugt, A. van der, Sprengers, M., Velasco, S., Lycklama Ñ Nijeholt, G. J., Ben Hassen, W., Burns, P., Brown, S., Chabert, E., Krings, T., Choe, H., Weimar, C., Campbell, B. C. V., Ford, G. A., Ribo, M., White, P., Cloud, G. C., San Roman, L., Davalos, A., Naggara, O., Hill, M. D., Bracard, S. und HERMES Trialists Collaboration, for the. “Association of Time From Stroke Onset to Groin Puncture With Quality of Reperfusion After Mechanical Thrombectomy: A Meta-analysis of Individual Patient Data From 7 Randomized Clinical Trials”. *JAMA Neurology* 76 (4), 2019, S. 405–411.
- [Breiman, 2001] Breiman, L. “Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)”. *Statistical Science* 16 (3), 2001, S. 199–231.
- [Brosch u. a., 2016] Brosch, T., Tang, L. Y. W., Yoo, Y., Li, D. K. B., Traboulsee, A. und Tam, R. “Deep 3D Convolutional Encoder Networks With Shortcuts for Multiscale Feature Integration Applied to Multiple Sclerosis Lesion Segmentation”. *IEEE Transactions on Medical Imaging* 35 (5), 2016, S. 1229–1239.

- [Broocks u. a., 2019] Broocks, G., Rajput, F., Hanning, U., Faizy, T. D., Leischner, H., Schön, G., Gellißen, S., Sporns, P., Deb-Chatterji, M., Thomalla, G., Kemmling, A., Fiehler, J. und Flottmann, F. “Highest Lesion Growth Rates in Patients With Hyperacute Stroke”. *Stroke* 50 (1), 2019, S. 189–192.
- [Caelles u. a., 2017] Caelles, S., Maninis, K.-K., Pont-Tuset, J., Leal-Taixe, L., Cremers, D. und Van Gool, L. “One-Shot Video Object Segmentation”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [Campbell, 2017] Campbell, B. C. V. “Thrombolysis and Thrombectomy for Acute Ischemic Stroke: Strengths and Synergies”. *Seminars in Thrombosis and Hemostasis* 43 (02), 2017, S. 185–190.
- [Casamitjana u. a., 2016] Casamitjana, A., Puch, S., Aduriz, A. und Vilaplana, V. “3D Convolutional Neural Networks for Brain Tumor Segmentation: A Comparison of Multi-resolution Architectures”. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2016, S. 150–161.
- [Chartrand u. a., 2017] Chartrand, G., Cheng, P. M., Vorontsov, E., Drozdal, M., Turcotte, S., Pal, C. J., Kadoury, S. und Tang, A. “Deep Learning: A Primer for Radiologists”. *RadioGraphics* 37 (7), 2017. PMID: 29131760, S. 2113–2131.
- [Chen u. a., 2016] Chen, X., Cheng, G., Cai, Y., Wen, D. und Li, H. “Semantic Segmentation with Modified Deep Residual Networks”. In: *Pattern Recognition*. Springer Singapore, 2016, S. 42–54.
- [Chen u. a., 2017] Chen, L., Bentley, P. und Rueckert, D. “Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks”. *NeuroImage: Clinical* 15, 2017, S. 633–643.
- [Cheng u. a., 2017] Cheng, J., Tsai, Y.-H., Wang, S. und Yang, M.-H. “SegFlow: Joint Learning for Video Object Segmentation and Optical Flow”. In: *The IEEE International Conference on Computer Vision (ICCV)*. 2017.
- [Chilamkurthy u. a., 2018] Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N. G., Venugopal, V. K., Mahajan, V., Rao, P. und Warier, P. “Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study”. *The Lancet* 392 (10162), 2018, S. 2388–2396.
- [Choi u. a., 2016] Choi, Y., Kwon, Y., Lee, H., Kim, B. J., Paik, M. C. und Won, J.-H. “Ensemble of Deep Convolutional Neural Networks for Prognosis of Ischemic Stroke”. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2016, S. 231–243.
- [Choi u. a., 2017] Choi, Y., Kwon, Y., Paik, M. C., Kim, B. J. und Won, J.-H. “Ischemic Stroke Lesion Segmentation with Convolutional Neural Networks for Small Data”. In: *Ischemic Stroke Lesion Segmentation – ISLES challenge 2017, held in conjunc-*

-
- tion with MICCAI 2017, Quebec, Canada. www.isles-challenge.org/ISLES2017/articles/choi.pdf. 2017.
- [Christ u. a., 2017] Christ, P. F., Ettlinger, F., Grün, F., Elshaer, M. E. A., Lipková, J., Schlecht, S., Ahmaddy, F., Tatavarty, S., Bickel, M., Bilic, P., Rempfler, M., Hofmann, F., D’Anastasi, M., Ahmadi, S.-A., Kaissis, G., Holch, J., Sommer, W. H., Braren, R., Heinemann, V. und Menze, B. H. “Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks”. *CoRR* abs/1702.05970, 2017.
- [Çiçek u. a., 2016] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. und Ronneberger, O. “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation”. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer International Publishing, 2016, S. 424–432.
- [Compagne u. a., 2019] Compagne, K. C. J., Boers, A. M. M., Marquering, H. A., Berkhemer, O. A., Yoo, A. J., Beenen, L. F. M., Oostenbrugge, R. J. van, Zwam, W. van, Roos, Y. B. W. E. M., Majoie, C. B., Es, A. C. G. M. van, Lugt, A. van der, Dippel, D. W. J., Lingsma, H. und for the MR CLEAN Investigators. “Follow-up infarct volume as a mediator of endovascular treatment effect on functional outcome in ischaemic stroke”. *European Radiology* 29 (2), 2019, S. 736–744.
- [Contin u. a., 2010] Contin, L., Beer, C., Bynevelt, M., Wittsack, H. und Garrido, G. “Semi-automatic segmentation of core and penumbra regions in acute ischemic stroke: preliminary results”. In: *17th International Conference on Systems, Signals and Image Processing (IWSSIP) 2010*. 2010, S. 178–181.
- [Copen u. a., 2011] Copen, W. A., Schaefer, P. W. und Wu, O. “MR Perfusion Imaging in Acute Ischemic Stroke”. *Neuroimaging Clinics of North America* 21 (2), 2011. Stroke Imaging, S. 259–283.
- [Couronne u. a., 2019] Couronne, R., Louis, M. und Durrleman, S. “Longitudinal autoencoder for multi-modal disease progression modelling”. Working paper or reprint: <https://hal.archives-ouvertes.fr/hal-02090886>. 2019.
- [Crimi, 2016] Crimi, A. “Brain Lesions, Introduction”. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2016, S. 1–5.
- [Dalca u. a., 2019] Dalca, A. V., Balakrishnan, G., Guttag, J. und Sabuncu, M. R. “Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces”. *Medical Image Analysis* 57, 2019, S. 226–236.
- [de Lucas u. a., 2008] de Lucas, E. M., Sánchez, E., Gutiérrez, A., Mandly, A. G., Ruiz, E., Flórez, A. F., Izquierdo, J., Arnáiz, J., Piedra, T., Valle, N., Bañales, I. und Quintana, F. “CT Protocol for Acute Stroke: Tips and Tricks for General Radiologists”. *RadioGraphics* 28 (6), 2008, S. 1673–1687.

- [De Fauw u. a., 2018] De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., Driessche, G. van den, Lakshminarayanan, B., Meyer, C., Mackinder, F., Bouton, S., Ayoub, K., Chopra, R., King, D., Karthikesalingam, A., Hughes, C. O., Raine, R., Hughes, J., Sim, D. A., Egan, C., Tufail, A., Montgomery, H., Hassabis, D., Rees, G., Back, T., Khaw, P. T., Suleyman, M., Cornebise, J., Keane, P. A. und Ronneberger, O. "Clinically applicable deep learning for diagnosis and referral in retinal disease". 24, 9 2018, S. 1342–1350.
- [d'Esterre u. a., 2015] d'Esterre, C. D., Boesen, M. E., Ahn, S. H., Pordeli, P., Najm, M., Minhas, P., Davari, P., Fainardi, E., Rubiera, M., Khaw, A. V., Zini, A., Frayne, R., Hill, M. D., Demchuk, A. M., Sajobi, T. T., Forkert, N. D., Goyal, M., Lee, T. Y. und Menon, B. K. "Time-Dependent Computed Tomographic Perfusion Thresholds for Patients With Acute Ischemic Stroke". *Stroke* 46 (12), 2015, S. 3390–3397.
- [Diener u. a., 2018] Diener, H.-C., Dichgans, M., Frank, B., Gerloff, C., Grond, M., Kleinschnitz, C., Röther, J., Thomalla, G. und Weimar, C. "New Studies in Stroke". *Neuroradiologie Scan* 8 (1), 2018, S. 61–84.
- [Doyle u. a., 2018] Doyle, S., Forbes, F., Jaillard, A., Heck, O., Detante, O. und Dojat, M. "Sub-acute and Chronic Ischemic Stroke Lesion MRI Segmentation". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2018, S. 111–122.
- [Drewer-Gutland u. a., 2015] Drewer-Gutland, F., Kemmling, A., Ligges, S., Ritter, M., Dziejwas, R., Ringelstein, E. B., Niederstadt, T. U., Heindel, W. und Heßelmann, V. "CTP-Based Tissue Outcome: Promising Tool to Prove the Beneficial Effect of Mechanical Recanalization in Acute Ischemic Stroke". *RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren* 187 (06), 2015, S. 459–466.
- [Dronne u. a., 2004] Dronne, M.-A., Boissel, J.-P., Grenier, E., Gilquin, H., Cucherat, M., Hommel, M., Barbier, E. und Bricca, G. "Mathematical Modelling of an Ischemic Stroke: An Integrative Approach". *Acta Biotheoretica* 52 (4), 2004, S. 255–272.
- [Drozdal u. a., 2016] Drozdal, M., Vorontsov, E., Chartrand, G., Kadoury, S. und Pal, C. "The Importance of Skip Connections in Biomedical Image Segmentation". In: *Deep Learning and Data Labeling for Medical Applications*. Springer International Publishing, 2016, S. 179–187.
- [Dumont u. a., 2013] Dumont, T., Duarte, M., Descombes, S., Dronne, M.-A., Massot, M. und Louvet, V. "Simulation of human ischemic stroke in realistic 3D geometry". *Communications in Nonlinear Science and Numerical Simulation* 18 (6), 2013, S. 1539–1557.

-
- [Duval u. a., 2002] Duval, V., Chabaud, S., Girard, P., Cucherat, M., Hommel, M. und Boissel, J. P. “Physiologically Based Model of Acute Ischemic Stroke”. *Journal of Cerebral Blood Flow & Metabolism* 22 (8), 2002, S. 1010–1018.
- [Feigin u. a., 2015] Feigin, V., Krishnamurthi, R., Parmar, P., Norrving, B., Mensah, G., Bennett, D., Barker-Collo, S., Moran, A., Sacco, R., Truelsen, T., Davis, S., Pandian, J., Naghavi, M., Forouzanfar, M., Nguyen, G., Johnson, C., Vos, T., Meretoja, A., Murray, C. und Roth, G. “Update on the Global Burden of Ischemic and Hemorrhagic Stroke in 1990-2013: The GBD 2013 Study”. *Neuroepidemiology* 45, 2015, S. 161–176.
- [Flottmann u. a., 2017] Flottmann, F., Broocks, G., Faizy, T. D., Ernst, M., Forkert, N. D., Grosser, M., Thomalla, G., Siemonsen, S., Fiehler, J. und Kemmling, A. “CT-perfusion stroke imaging: a threshold free probabilistic approach to predict infarct volume compared to traditional ischemic thresholds”. *Nature Scientific Reports* 7, 1 2017, S. 6679.
- [Fonteijn u. a., 2011] Fonteijn, H. M., Clarkson, M. J., Modat, M., Barnes, J., Lehmann, M., Ourselin, S., Fox, N. C. und Alexander, D. C. “An Event-Based Disease Progression Model and Its Application to Familial Alzheimer’s Disease”. In: *Information Processing in Medical Imaging (IPMI)*. Springer Berlin Heidelberg, 2011, S. 748–759.
- [Foote u. a., 2019] Foote, M. D., Zimmerman, B. E., Sawant, A. und Joshi, S. C. “Real-Time 2D-3D Deformable Registration with Deep Learning and Application to Lung Radiotherapy Targeting”. In: *Information Processing in Medical Imaging (IPMI)*. Springer International Publishing, 2019, S. 265–276.
- [Gonzalez u. a., 2008] Gonzalez, R. C. und Woods, R. E. *Digital Image Processing*. 3. Aufl. Pearson Prentice Hall International, 2008.
- [Goodfellow u. a., 2016] Goodfellow, I., Bengio, Y. und Courville, A. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [Goyal u. a., 2016] Goyal, M., Menon, B. K., Zwam, W. H. van, Dippel, D. W., Mitchell, P. J., Demchuk, A. M., Dávalos, A., Majoie, C. B., Lugt, A. van der, De Miquel, M. A., Donnan, G. A., Roos, Y. B. W. E. M., Bonafe, A., Jahan, R., Diener, H.-C., Berg, L. A. van den, Levy, E. I., Berkhemer, O. A., Pereira, V. M., Rempel, J., Millán, M., Davis, S. M., Roy, D., Thornton, J., Román, L. S., Ribó, M., Beumer, D., Stouch, B., Brown, S., Campbell, B. C. V., Oostenbrugge, R. J. van, Saver, J. L., Hill, M. D. und Jovin, T. G. “Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials”. *The Lancet* 387 (10029), 2016, S. 1723–1731.
- [Greff u. a., 2016] Greff, K., Srivastava, R. K. und Schmidhuber, J. “Highway and Residual Networks learn Unrolled Iterative Estimation”. *CoRR* abs/1612.07771, 2016. arXiv: 1612.07771.

- [Hakim u. a., 2019] Hakim, A. und Wiest, R. “CT Brain Perfusion: A Clinical Perspective”. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2019, S. 15–24.
- [Havaei u. a., 2017] Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M. und Larochelle, H. “Brain tumor segmentation with Deep Neural Networks”. *Medical Image Analysis* 35, 2017, S. 18–31.
- [He u. a., 2016a] He, K., Zhang, X., Ren, S. und Sun, J. “Deep Residual Learning for Image Recognition”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.
- [He u. a., 2016b] He, K., Zhang, X., Ren, S. und Sun, J. “Identity Mappings in Deep Residual Networks”. In: *Computer Vision – ECCV 2016*. Springer International Publishing, 2016, S. 630–645.
- [He u. a., 2017] He, K., Gkioxari, G., Dollar, P. und Girshick, R. “Mask R-CNN”. In: *The IEEE International Conference on Computer Vision (ICCV)*. 2017.
- [Heimann u. a., 2009] Heimann, T., van Ginneken, B., Styner, M. A., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., Bello, F., Binnig, G., Bischof, H., Bornik, A., Cashman, P. M. M., Chi, Y., Cordova, A., Dawant, B. M., Fidrich, M., Furst, J. D., Furukawa, D., Grenacher, L., Hornegger, J., Kainmüller, D., Kitney, R. I., Kobatake, H., Lamecker, H., Lange, T., Lee, J., Lennon, B., Li, R., Li, S., Meinzer, H., Nemeth, G., Raicu, D. S., Rau, A., van Rikxoort, E. M., Rousson, M., Rusko, L., Saddi, K. A., Schmidt, G., Seghers, D., Shimizu, A., Slagmolen, P., Sorantin, E., Soza, G., Susomboon, R., Waite, J. M., Wimmer, A. und Wolf, I. “Comparison and Evaluation of Methods for Liver Segmentation From CT Datasets”. *IEEE Transactions on Medical Imaging* 28 (8), 2009, S. 1251–1265.
- [Hennerici u. a., 2017] Hennerici, M. G., Kern, R. u. a. “S1-Leitlinie Diagnostik akuter zerebrovaskulärer Erkrankungen”. In: *Leitlinien für Diagnostik und Therapie in der Neurologie*. Online: www.dgn.org/leitlinien (abgerufen am 12.09.2019). Deutsche Gesellschaft für Neurologie, 2017.
- [Heuschmann u. a., 2010] Heuschmann, P. U., Busse, O., Wagner, M., Endres, M., Villringer, A., Röther, J., Kolominsky-Rabas, P., Berger, K. u. a. “Schlaganfallhäufigkeit und Versorgung von Schlaganfallpatienten in Deutschland”. *Aktuelle Neurologie* 37 (07), 2010, S. 333–340.
- [Higashida u. a., 2003] Higashida, R. T. und Furlan, A. J. “Trial Design and Reporting Standards for Intra-Arterial Cerebral Thrombolysis for Acute Ischemic Stroke”. *Stroke* 34 (8), 2003, e109–e137.
- [Hilbert u. a., 2018] Hilbert, A., Veeling, B. S. und Marquering, H. A. “Data-efficient Convolutional Neural Networks for Treatment Decision Support in Acute Ischemic

-
- Stroke”. In: *International Conference on Medical Imaging with Deep Learning (MIDL)*. 2018.
- [Hochreiter u. a., 1997] Hochreiter, S. und Schmidhuber, J. “Long Short-Term Memory”. *Neural Computation* 9 (8), 1997, S. 1735–1780.
- [Hogea u. a., 2008] Hogea, C., Davatzikos, C. und Biros, G. “An image-driven parameter estimation problem for a reaction–diffusion glioma growth model with mass effects”. *Journal of Mathematical Biology* 56 (6), 2008, S. 793–825.
- [Hornik u. a., 1989] Hornik, K., Stinchcombe, M. und White, H. “Multilayer feedforward networks are universal approximators”. *Neural Networks* 2 (5), 1989, S. 359–366.
- [Hu u. a., 2018] Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C. M., Emberton, M., Ourselin, S., Noble, J. A., Barratt, D. C. und Vercauteren, T. “Weakly-supervised convolutional neural networks for multimodal image registration”. *Medical Image Analysis* 49, 2018, S. 1–13.
- [Hu u. a., 2019] Hu, J., Shen, L., Albanie, S., Sun, G. und Wu, E. “Squeeze-and-Excitation Networks”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
- [Huang u. a., 2010] Huang, S., Shen, Q. und Duong, T. Q. “Artificial Neural Network Prediction of Ischemic Tissue Fate in Acute Stroke Imaging”. *Journal of Cerebral Blood Flow & Metabolism* 30 (9), 2010, S. 1661–1670.
- [Huang u. a., 2011] Huang, S., Shen, Q. und Duong, T. Q. “Quantitative prediction of acute ischemic tissue fate using support vector machine”. *Brain Research* 1405, 2011, S. 77–84.
- [Huang u. a., 2017] Huang, G., Liu, Z., Maaten, L. van der und Weinberger, K. Q. “Densely Connected Convolutional Networks”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [Huang u. a., 2018] Huang, G., Chen, D., Li, T., Wu, F., Maaten, L. van der und Weinberger, K. “Multi-Scale Dense Networks for Resource Efficient Image Classification”. In: *International Conference on Learning Representations (ICLR)*. 2018.
- [Ioffe u. a., 2015] Ioffe, S. und Szegedy, C. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. In: *Proceedings of the 32nd International Conference on Machine Learning*. Bd. 37. Proceedings of Machine Learning Research. PMLR, 2015, S. 448–456.
- [Isensee u. a., 2019] Isensee, F., Petersen, J., Kohl, S. A. A., Jäger, P. F. und Maier-Hein, K. H. “nnU-Net: Breaking the Spell on Successful Medical Image Segmentation”. *CoRR* abs/1904.08128, 2019. arXiv: 1904.08128.

- [Jaderberg u. a., 2015] Jaderberg, M., Simonyan, K., Zisserman, A. und Kavukcuoglu, K. “Spatial Transformer Networks”. In: *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., 2015, S. 2017–2025.
- [James u. a., 2013] James, G., Witten, D., Hastie, T. und Tibshirani, R. *An Introduction to Statistical Learning – with Applications in R*. Springer, 2013.
- [Jenkinson u. a., 2012] Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. und Smith, S. M. “FSL”. *NeuroImage* 62 (2), 2012, S. 782–790.
- [Jha u. a., 2009] Jha, B. und Kothari, M. “Pearls & Oysters: Hyperdense or pseudo-hyperdense MCA sign”. *Neurology* 72 (23), 2009, S. 116–117.
- [Jiang u. a., 2020] Jiang, Z., Ding, C., Liu, M. und Tao, D. “Two-Stage Cascaded U-Net: 1st Place Solution to BraTS Challenge 2019 Segmentation Task”. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2020, S. 231–241.
- [Kamnitsas u. a., 2016] Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A. V., Criminisi, A., Rueckert, D. und Glocker, B. “DeepMedic for Brain Tumor Segmentation”. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2016, S. 138–149.
- [Kamnitsas u. a., 2017] Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D. und Glocker, B. “Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation”. *Medical image analysis* 36, 2017, S. 61–78.
- [Kamal u. a., 2018] Kamal, H., Lopez, V. und Sheth, S. A. “Machine Learning in Acute Ischemic Stroke Neuroimaging”. *Frontiers in Neurology* 9, 2018, S. 945.
- [Kasten, 2007] Kasten, E. *Einführung Neuropsychologie*. 1. Aufl. UTB GmbH, 2007.
- [Ke u. a., 2017] Ke, T.-W., Maire, M. und Yu, S. X. “Multigrid Neural Architectures”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [Kemmling u. a., 2013] Kemmling, A., Lev, M. H., Payabvash, S., Betensky, R. A., Qian, J., Masrur, S. und Schwamm, L. H. “Hospital Acquired Pneumonia Is Linked to Right Hemispheric Peri-Insular Stroke”. *PLOS ONE* 8 (8), 2013, S. 1–7.
- [Kemmling u. a., 2015] Kemmling, A., Flottmann, F., Forkert, N. D., Minnerup, J., Heindel, W., Thomalla, G., Eckert, B., Knauth, M., Psychogios, M., Langner, S. und Fiehler, J. “Multivariate Dynamic Prediction of Ischemic Infarction and Tissue Salvage as a Function of Time and Degree of Recanalization”. *Journal of Cerebral Blood Flow & Metabolism* 35 (9), 2015, S. 1397–1405.
- [Khan u. a., 2017] Khan, M., Baird, G. L., Goddeau, R. P., Silver, B. und Henninger, N. “Alberta Stroke Program Early CT Score Infarct Location Predicts Outcome Following M2 Occlusion”. *Frontiers in Neurology* 8, 2017, S. 98.

-
- [Kingma u. a., 2015] Kingma, D. P. und Ba, J. “Adam: A Method for Stochastic Optimization”. In: *International Conference on Learning Representations (ICLR)*. 2015.
- [Kolominsky-Rabas u. a., 2002] Kolominsky-Rabas, P. L. und Heuschmann, P. U. “Inzidenz, Ätiologie und Langzeitprognose des Schlaganfalls”. *Fortschritte der Neurologie · Psychiatrie* 70 (12), 2002, S. 657–662.
- [Koo u. a., 2016] Koo, T. K. und Li, M. Y. “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research”. *Journal of Chiropractic Medicine* 15 (2), 2016, S. 155–163.
- [Krizhevsky u. a., 2012] Krizhevsky, A., Sutskever, I. und Hinton, G. E. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems 25*. 2012, S. 1097–1105.
- [Krüger u. a., 2015] Krüger, J., Ehrhardt, J. und Handels, H. “Probabilistic Appearance Models for Segmentation and Classification”. In: *The IEEE International Conference on Computer Vision (ICCV)*. 2015.
- [Lansberg u. a., 2012] Lansberg, M. G., Straka, M., Kemp, S., Mlynash, M., Wechsler, L. R., Jovin, T. G., Wilder, M. J., Lutsep, H. L., Czartoski, T. J., Bernstein, R. A., Chang, C. W., Warach, S., Fazekas, F., Inoue, M., Tipirneni, A., Hamilton, S. A., Zaharchuk, G., Marks, M. P., Bammer, R. und Albers, G. W. “MRI profile and response to endovascular reperfusion after stroke (DEFUSE 2): a prospective cohort study”. *The Lancet Neurology* 11 (10), 2012, S. 860–867.
- [Larrazabal u. a., 2019] Larrazabal, A. J., Martinez, C. und Ferrante, E. “Anatomical Priors for Image Segmentation via Post-processing with Denoising Autoencoders”. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer International Publishing, 2019, S. 585–593.
- [LeCun u. a., 1998] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. u. a. “Gradient-based learning applied to document recognition”. *Proceedings of the IEEE* 86 (11), 1998, S. 2278–2324.
- [Li u. a., 2016] Li, S., Jiao, J., Han, Y. und Weissman, T. “Demystifying ResNet”. *CoRR* abs/1611.01186, 2016. arXiv: 1611.01186.
- [Lin u. a., 2014] Lin, M., Chen, Q. und Yan, S. “Network In Network”. In: *International Conference on Learning Representations (ICLR)*. 2014.
- [Lin u. a., 2017] Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B. und Belongie, S. “Feature Pyramid Networks for Object Detection”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [Lisowska u. a., 2017] Lisowska, A., O’Neil, A., Dilys, V., Daykin, M., Beveridge, E., Muir, K., McLaughlin, S. und Poole, I. “Context-Aware Convolutional Neural

- Networks for Stroke Sign Detection in Non-contrast CT Scans”. In: *Medical Image Understanding and Analysis*. Springer International Publishing, 2017, S. 494–505.
- [Lisowska u. a., 2018] Lisowska, A., Beveridge, E., O’Neil, A., Dilys, V., Muir, K. und Poole, I. “Evaluation of Dense Vessel Detection in NCCT Scans”. In: *Biomedical Engineering Systems and Technologies*. Springer International Publishing, 2018, S. 134–145.
- [Livne u. a., 2018] Livne, M., Boldsen, J. K., Mikkelsen, I. K., Fiebach, J. B., Sobesky, J. und Mouridsen, K. “Boosted Tree Model Reforms Multimodal Magnetic Resonance Imaging Infarct Prediction in Acute Stroke”. *Stroke* 49 (4), 2018, S. 912–918.
- [Livne u. a., 2019] Livne, M., Rieger, J., Aydin, O. U., Taha, A. A., Akay, E. M., Kossen, T., Sobesky, J., Kelleher, J. D., Hildebrand, K., Frey, D. und Madai, V. I. “A U-Net Deep Learning Framework for High Performance Vessel Segmentation in Patients With Cerebrovascular Disease”. *Frontiers in Neuroscience* 13, 2019, S. 97.
- [Long u. a., 2015] Long, J., Shelhamer, E. und Darrell, T. “Fully Convolutional Networks for Semantic Segmentation”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015.
- [Louis u. a., 2019] Louis, M., Couronné, R., Koval, I., Charlier, B. und Durrleman, S. “Riemannian Geometry Learning for Disease Progression Modelling”. In: *Information Processing in Medical Imaging (IPMI)*. Springer International Publishing, 2019, S. 542–553.
- [Lucas u. a., 2017] Lucas, C., Maier, O. und Heinrich, M. P. “Shallow Fully-Connected Neural Networks for Ischemic Stroke-Lesion Segmentation in MRI”. In: *Bildverarbeitung für die Medizin 2017*. Springer Berlin Heidelberg, 2017, S. 261–266.
- [Lucas u. a., 2018a] Lucas, C., Kemmling, A., Mamlouk, A. M. und Heinrich, M. P. “Multi-scale neural network for automatic segmentation of ischemic strokes on acute perfusion images”. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. 2018, S. 1118–1121.
- [Lucas u. a., 2018b] Lucas, C., Kemmling, A., Bouteldja, N., Aulmann, L. F., Madany Mamlouk, A. und Heinrich, M. P. “Learning to Predict Ischemic Stroke Growth on Acute CT Perfusion Data by Interpolating Low-Dimensional Shape Representations”. *Frontiers in Neurology* 9, 2018, S. 989.
- [Lucas u. a., 2019] Lucas, C., Schöttler, J. J., Kemmling, A., Aulmann, L. F. und Heinrich, M. P. “Automatic Detection and Segmentation of the Acute Vessel Thrombus in Cerebral CT”. In: *Bildverarbeitung für die Medizin 2019*. Springer Fachmedien Wiesbaden, 2019, S. 74–79.
- [Lucas u. a., 2020a] Lucas, C., Aulmann, L. F., Kemmling, A., Madany Mamlouk, A. und Heinrich, M. P. “Estimation of the Principal Ischaemic Stroke Growth

-
- Directions for Predicting Tissue Outcomes”. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2020, S. 69–79.
- [Lucas u. a., 2020b] Lucas, C., Aulmann, L. F., Kemmling, A., Mamlouk, A. M. und Heinrich, M. P. “Abstract: Estimation of the Principal Ischaemic Stroke Growth Directions for Predicting Tissue Outcomes”. In: *Bildverarbeitung für die Medizin 2020*. Springer Fachmedien Wiesbaden, 2020, S. 143.
- [Maas u. a., 2009] Maas, M. B., Lev, M. H., Ay, H., Singhal, A. B., Greer, D. M., Smith, W. S., Harris, G. J., Halpern, E., Kemmling, A., Koroshetz, W. J. und Furie, K. L. “Collateral Vessels on CT Angiography Predict Outcome in Acute Ischemic Stroke”. *Stroke* 40 (9), 2009, S. 3001–3005.
- [Machner u. a., 2009] Machner, B., Sprenger, A., Kömpf, D., Sander, T., Heide, W., Kimmig, H. und Helmchen, C. “Visual search disorders beyond pure sensory failure in patients with acute homonymous visual field defects”. *Neuropsychologia* 47 (13), 2009, S. 2704–2711.
- [Machner u. a., 2012] Machner, B., Dorr, M., Sprenger, A., Gablentz, J. von der, Heide, W., Barth, E. und Helmchen, C. “Impact of dynamic bottom-up features and top-down control on the visual exploration of moving real-world scenes in hemispatial neglect”. *Neuropsychologia* 50 (10), 2012, S. 2415–2425.
- [Maier u. a., 2015a] Maier, O., Schröder, C., Forkert, N. D., Martinetz, T. und Handels, H. “Classifiers for Ischemic Stroke Lesion Segmentation: A Comparison Study”. *PLOS ONE* 10 (12), 2015, S. 1–16.
- [Maier u. a., 2015b] Maier, O., Wilms, M., Gablentz, J. von der, Krämer, U. M., Münte, T. F. und Handels, H. “Extra Tree forests for sub-acute ischemic stroke lesion segmentation in MR sequences”. *Journal of Neuroscience Methods* 240, 2015, S. 89–100.
- [Maier u. a., 2016] Maier, O. und Handels, H. “Predicting Stroke Lesion and Clinical Outcome with Random Forests”. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2016, S. 219–230.
- [Maier u. a., 2017] Maier, O., Menze, B. H., Gablentz, J. von der, Häni, L., Heinrich, M. P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., Christiaens, D., Dutil, F., Egger, K., Feng, C., Glocker, B., Götz, M., Haeck, T., Halme, H.-L., Havaei, M., Iftekharuddin, K. M., Jodoin, P.-M., Kamnitsas, K., Kellner, E., Korvenoja, A., Larochelle, H., Ledig, C., Lee, J.-H., Maes, F., Mahmood, Q., Maier-Hein, K. H., McKinley, R., Muschelli, J., Pal, C., Pei, L., Rangarajan, J. R., Reza, S. M., Robben, D., Rueckert, D., Salli, E., Suetens, P., Wang, C.-W., Wilms, M., Kirschke, J. S., Krämer, U. M., Münte, T. F., Schramm, P., Wiest, R., Handels, H. und Reyes, M. “ISLES 2015 - A public evaluation benchmark for ischemic stroke

- lesion segmentation from multispectral MRI”. *Medical Image Analysis* 35, 2017, S. 250–269.
- [Maier, 2017] Maier, O. “Decision forest variants for brain lesion segmentation”. Diss. Universität zu Lübeck, 2017.
- [Maldjian u. a., 2001] Maldjian, J. A., Chalela, J., Kasner, S. E., Liebeskind, D. und Detre, J. A. “Automated CT Segmentation and Analysis for Acute Middle Cerebral Artery Stroke”. 22 (6), 2001, S. 1050–1055.
- [Masci u. a., 2011] Masci, J., Meier, U., Cireşan, D. und Schmidhuber, J. “Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction”. In: *Artificial Neural Networks and Machine Learning – ICANN 2011*. Springer Berlin Heidelberg, 2011, S. 52–59.
- [Menze u. a., 2011] Menze, B. H., Van Leemput, K., Honkela, A., Konukoglu, E., Weber, M.-A., Ayache, N. und Golland, P. “A Generative Approach for Image-Based Modeling of Tumor Growth”. In: *Information Processing in Medical Imaging (IPMI)*. Springer Berlin Heidelberg, 2011, S. 735–747.
- [Meyer u. a., 2009] Meyer, B. C. und Lyden, P. D. “The Modified National Institutes of Health Stroke Scale: its Time has Come”. *International Journal of Stroke* 4 (4), 2009, S. 267–273.
- [Milletari u. a., 2016] Milletari, F., Navab, N. und Ahmadi, S. “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation”. In: *2016 Fourth International Conference on 3D Vision (3DV)*. 2016, S. 565–571.
- [Milletari u. a., 2017] Milletari, F., Rothberg, A., Jia, J. und Sofka, M. “Integrating Statistical Prior Knowledge into Convolutional Neural Networks”. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer International Publishing, 2017, S. 161–168.
- [Minnerup u. a., 2016] Minnerup, J., Wersching, H., Teuber, A., Wellmann, J., Eyding, J., Weber, R., Reimann, G., Weber, W., Krause, L. U., Kurth, T., Berger, K., null, null, Homberg, V., Petrovitch, A., Heuser, L., Mönnigs, P., Krogias, C., Wallner, B., Hennigs, S., Ahlers, A., Sahl, H., Ranft, A., Dobis, C., Brassel, F., Nolden-Koch, M., Schmitt, H., Chapot, R., Nordmeyer, H., Schlamann, M., Weimar, C., Busch, F., Busch, E., Sigges, E., Ruf, H., Wohlfahrt, K., Karatschai, R., Klein, B., Höhle, T., Haass, A., Nasreldein, A., Büchele, B., Gahn, G., Sterker, M., Hantel, T., Krämer, C., Henningsen, H., Adelt, I., König, M., Schmidt, C., Hofmann, A., Niederstadt, T., Unrath, M., Rehfeldt, T., Fauser, B., Pfeiffer, A., Lowens, S., Stögbauer, F., Staudacher, T., Erdmann, P., Grottemeyer, K., Spüntrup, E., Bücke, P., Wienecke, P., Faiss, J., Wolzik-Großmann, M., Brune, N., Isenmann, S., Thomas, C. und Mucha, D. “Outcome After Thrombectomy and Intravenous Thrombolysis in Patients With Acute Ischemic Stroke”. *Stroke* 47 (6), 2016, S. 1584–1592.

-
- [Mitra u. a., 2014] Mitra, J., Bourgeat, P., Fripp, J., Ghose, S., Rose, S., Salvado, O., Connelly, A., Campbell, B., Palmer, S., Sharma, G., Christensen, S. und Carey, L. “Lesion segmentation from multimodal MRI using random forest following ischemic stroke”. *NeuroImage* 98, 2014, S. 324–335.
- [Mok u. a., 2017] Mok, T. C. W. und Chung, A. C. S. “Deep Adversarial Networks for Stroke Lesion Segmentation”. In: *Ischemic Stroke Lesion Segmentation – ISLES challenge 2017, held in conjunction with MICCAI 2017, Quebec, Canada*. www.isles-challenge.org/ISLES2017/articles/mok.pdf. 2017.
- [Monteiro u. a., 2017] Monteiro, M. und Oliveira, A. L. “Fully Convolutional Neural Network for 3D Stroke Lesion Segmentation”. In: *Ischemic Stroke Lesion Segmentation – ISLES challenge 2017, held in conjunction with MICCAI 2017, Quebec, Canada*. www.isles-challenge.org/ISLES2017/articles/monteiro.pdf. 2017.
- [Mould, 2012] Mould, D. R. “Models for Disease Progression: New Approaches and Uses”. *Clinical Pharmacology & Therapeutics* 92 (1), 2012, S. 125–131.
- [Newell u. a., 2016] Newell, A., Yang, K. und Deng, J. “Stacked Hourglass Networks for Human Pose Estimation”. In: *Computer Vision – ECCV 2016*. Springer International Publishing, 2016, S. 483–499.
- [Newton u. a., 2016] Newton, A. J. und Lytton, W. W. “Computer modeling of ischemic stroke”. *Drug Discovery Today: Disease Models* 19, 2016. Computational Models of Neurological Disorder, S. 77–83.
- [Nielsen u. a., 2018] Nielsen, A., Hansen, M. B., Tietze, A. und Mouridsen, K. “Prediction of Tissue Outcome and Assessment of Treatment Effect in Acute Ischemic Stroke Using Deep Learning”. *Stroke* 49 (6), 2018, S. 1394–1401.
- [Nogueira u. a., 2009] Nogueira, R. G., Liebeskind, D. S., Sung, G., Duckwiler, G. und Smith, W. S. “Predictors of Good Clinical Outcomes, Mortality, and Successful Revascularization in Patients With Acute Ischemic Stroke Undergoing Thrombectomy”. *Stroke* 40 (12), 2009, S. 3777–3783.
- [Noh u. a., 2015] Noh, H., Hong, S. und Han, B. “Learning Deconvolution Network for Semantic Segmentation”. In: *The IEEE International Conference on Computer Vision (ICCV)*. 2015.
- [Noh u. a., 2019] Noh, J., Bae, W., Lee, W., Seo, J. und Kim, G. “Better to Follow, Follow to Be Better: Towards Precise Supervision of Feature Super-Resolution for Small Object Detection”. In: *The IEEE International Conference on Computer Vision (ICCV)*. 2019.
- [O’Donnell u. a., 2016] O’Donnell, M. J., Chin, S. L., Rangarajan, S., Xavier, D., Liu, L., Zhang, H., Rao-Melacini, P., Zhang, X., Pais, P., Agapay, S. u. a. “Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study”. *The Lancet* 388 (10046), 2016, S. 761–775.

- [Oktay u. a., 2018] Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S. A., de Marvao, A., Dawes, T., O'Regan, D. P., Kainz, B., Glocker, B. und Rueckert, D. "Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation". *IEEE Transactions on Medical Imaging* 37 (2), 2018, S. 384–395.
- [Olive-Gadea u. a., 2019] Olive-Gadea, M., Martins, N., Boned, S., Carvajal, J., Moreno, M. J., Muchada, M., Molina, C. A., Tomasello, A., Ribo, M. und Rubiera, M. "Baseline ASPECTS and e-ASPECTS Correlation with Infarct Volume and Functional Outcome in Patients Undergoing Mechanical Thrombectomy". *Journal of Neuroimaging* 29 (2), 2019, S. 198–202.
- [Otton u. a., 2013] Otton, J., Morton, G., Schuster, A., Bigalke, B., Marano, R., Olivotti, L., Nagel, E. und Chiribiri, A. "A direct comparison of the sensitivity of CT and MR cardiac perfusion using a myocardial perfusion phantom". *Journal of Cardiovascular Computed Tomography* 7 (2), 2013, S. 117–124.
- [Pace u. a., 2018] Pace, D. F., Dalca, A. V., Brosch, T., Geva, T., Powell, A. J., Weese, J., Moghari, M. H. und Golland, P. "Iterative Segmentation from Limited Training Data: Applications to Congenital Heart Disease". In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer International Publishing, 2018, S. 334–342.
- [Paszke u. a., 2017] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L. und Lerer, A. "Automatic differentiation in PyTorch". In: *Conference on Neural Information Processing Systems (NIPS) – Autodiff workshop*. 2017.
- [Perazzi u. a., 2017] Perazzi, F., Khoreva, A., Benenson, R., Schiele, B. und Sorkine-Hornung, A. "Learning Video Object Segmentation From Static Images". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [Petersen u. a., 2019] Petersen, J., Jäger, P. F., Isensee, F., Kohl, S. A. A., Neuberger, U., Wick, W., Debus, J., Heiland, S., Bendszus, M., Kickingereder, P. und Maier-Hein, K. H. "Deep Probabilistic Modeling of Glioma Growth". In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer International Publishing, 2019, S. 806–814.
- [Pexman u. a., 2001] Pexman, J. H. W., Barber, P. A., Hill, M. D., Sevick, R. J., Demchuk, A. M., Hudon, M. E., Hu, W. Y. und Buchan, A. M. "Use of the Alberta Stroke Program Early CT Score (ASPECTS) for Assessing CT Scans in Patients with Acute Stroke". *American Journal of Neuroradiology* 22 (8), 2001, S. 1534–1542.
- [Pham u. a., 2019] Pham, D. D., Dovletov, G., Warwas, S., Landgraeber, S., Jäger, M. und Pauli, J. "Deep Learning With Anatomical Priors: Imitating Enhanced Autoencoders In Latent Space For Improved Pelvic Bone Segmentation In MRI".

In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. 2019, S. 1166–1169.

- [Pierot u. a., 2019] Pierot, L., Jayaraman, M. V., Szikora, I., Hirsch, J. A., Baxter, B., Miyachi, S., Mahadevan, J., Chong, W., Mitchell, P. J., Coulthard, A., Rowley, H. A., Sanelli, P. C., Tampieri, D., Brouwer, P. A., Fiehler, J., Kocer, N., Vilela, P., Rovira, A., Fischer, U., Caso, V., Worp, B. van der, Sakai, N., Matsumaru, Y., Yoshimura, S.-i., Anxionnat, R., Desal, H., Biscoito, L., Pumar, J. M., Diaz, O., Fraser, J. F., Linfante, I., Liebeskind, D. S., Nogueira, R. G., Hacke, W., Brainin, M., Yan, B., Soderman, M., Taylor, A., Pongpech, S., Tanaka, M., Karel, T. u. a. “Standards of practice in acute ischemic stroke intervention: International recommendations”. *Interventional Neuroradiology* 25 (1), 2019, S. 31–37.
- [Pinto u. a., 2018] Pinto, A., Mckinley, R., Alves, V., Wiest, R., Silva, C. A. und Reyes, M. “Stroke Lesion Outcome Prediction Based on MRI Imaging Combined With Clinical Information”. *Frontiers in Neurology* 9, 2018, S. 1060.
- [Pohlen u. a., 2017] Pohlen, T., Hermans, A., Mathias, M. und Leibe, B. “Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [Poudel u. a., 2017] Poudel, R. P. K., Lamata, P. und Montana, G. “Recurrent Fully Convolutional Neural Networks for Multi-slice MRI Cardiac Segmentation”. In: *Reconstruction, Segmentation, and Analysis of Medical Images*. Springer International Publishing, 2017, S. 83–94.
- [Qin u. a., 2018] Qin, C., Bai, W., Schlemper, J., Petersen, S. E., Piechnik, S. K., Neubauer, S. und Rueckert, D. “Joint Learning of Motion Estimation and Segmentation for Cardiac MR Image Sequences”. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Springer International Publishing, 2018, S. 472–480.
- [Qin u. a., 2019] Qin, C., Shi, B., Liao, R., Mansi, T., Rueckert, D. und Kamen, A. “Unsupervised Deformable Registration for Multi-modal Images via Disentangled Representations”. In: *Information Processing in Medical Imaging*. Springer International Publishing, 2019, S. 249–261.
- [Rangaraju u. a., 2015] Rangaraju, S., Streib, C., Aghaebrahim, A., Jadhav, A., Frankel, M. und Jovin, T. G. “Relationship Between Lesion Topology and Clinical Outcome in Anterior Circulation Large Vessel Occlusions”. *Stroke* 46 (7), 2015, S. 1787–1792.
- [Ravi u. a., 2017] Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B. und Yang, G. “Deep Learning for Health Informatics”. *IEEE Journal of Biomedical and Health Informatics* 21 (1), 2017, S. 4–21.

- [Ravishankar u. a., 2017] Ravishankar, H., Venkataramani, R., Thiruvankadam, S., Sudhakar, P. und Vaidya, V. “Learning and Incorporating Shape Models for Semantic Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer International Publishing, 2017, S. 203–211.
- [Redmon u. a., 2016] Redmon, J., Divvala, S., Girshick, R. und Farhadi, A. “You Only Look Once: Unified, Real-Time Object Detection”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.
- [Rekik u. a., 2012] Rekik, I., Allasonnière, S., Carpenter, T. K. und Wardlaw, J. M. “Medical image analysis methods in MR/CT-imaged acute-subacute ischemic stroke lesion: Segmentation, prediction and insights into dynamic evolution simulation models. A critical appraisal”. *NeuroImage: Clinical* 1 (1), 2012, S. 164–178.
- [Robben u. a., 2017] Robben, D. und Suetens, P. “Dual-scale fully convolutional neural network for final infarct prediction”. In: *Ischemic Stroke Lesion Segmentation – ISLES challenge 2017, held in conjunction with MICCAI 2017, Quebec, Canada*. www.isles-challenge.org/ISLES2017/articles/robben.pdf. 2017.
- [Robben u. a., 2020] Robben, D., Boers, A. M., Marquering, H. A., Langezaal, L. L., Roos, Y. B., Oostenbrugge, R. J. van, Zwam, W. H. van, Dippel, D. W., Majoie, C. B., Lugt, A. van der, Lemmens, R. und Suetens, P. “Prediction of final infarct volume from native CT perfusion and treatment parameters using deep learning”. *Medical Image Analysis* 59, 2020.
- [Rohé u. a., 2017] Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M. und Pennec, X. “SVF-Net: Learning Deformable Image Registration Using Shape Matching”. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Springer International Publishing, 2017, S. 266–274.
- [Ronneberger u. a., 2015] Ronneberger, O., Fischer, P. und Brox, T. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer International Publishing, 2015, S. 234–241.
- [Rosso u. a., 2019] Rosso, C., Blanc, R., Ly, J., Samson, Y., Lehericy, S., Gory, B., Marnat, G., Mazighi, M., Consoli, A., Labreuche, J., Saleme, S., Costalat, V., Bracard, S., Desal, H., Piotin, M. und Lapergue, B. “Impact of infarct location on functional outcome following endovascular therapy for stroke”. *Journal of Neurology, Neurosurgery & Psychiatry* 90 (3), 2019, S. 313–319.
- [Roth u. a., 2018] Roth, H. R., Oda, H., Zhou, X., Shimizu, N., Yang, Y., Hayashi, Y., Oda, M., Fujiwara, M., Misawa, K. und Mori, K. “An application of cascaded 3D fully convolutional networks for medical image segmentation”. *Computerized Medical Imaging and Graphics* 66, 2018, S. 90–99.
- [Rumelhart u. a., 1986] Rumelhart, D. E., Hinton, G. E. und Williams, R. J. “Learning representations by back-propagating errors”. *Nature* 323 (6088), 1986, S. 533–536.

-
- [Saake u. a., 2012] Saake, M., Goelitz, P., Struffert, T., Breuer, L., Volbers, B., Doerfler, A. und Kloska, S. “Comparison of Conventional CTA and Volume Perfusion CTA in Evaluation of Cerebral Arterial Vasculature in Acute Stroke”. *American Journal of Neuroradiology* 33 (11), 2012, S. 2068–2073.
- [Sakamoto u. a., 2017] Sakamoto, M., Nakano, H., Zhao, K. und Sekiyama, T. “Multi-stage Neural Networks with Single-Sided Classifiers for False Positive Reduction and Its Evaluation Using Lung X-Ray CT Images”. In: *Image Analysis and Processing - ICIAP 2017*. Springer International Publishing, 2017, S. 370–379.
- [Scalzo u. a., 2012] Scalzo, F., Hao, Q., Alger, J. R., Hu, X. und Liebeskind, D. S. “Regional Prediction of Tissue Fate in Acute Ischemic Stroke”. *Annals of Biomedical Engineering* 40 (10), 2012, S. 2177–2187.
- [Schiratti u. a., 2015] Schiratti, J.-B., Allasonniere, S., Colliot, O. und Durrleman, S. “Learning spatiotemporal trajectories from manifold-valued longitudinal data”. In: *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., 2015, S. 2404–2412.
- [Schmidt-Pogoda u. a., 2019] Schmidt-Pogoda, A. und Minnerup, J. “Seltene Schlaganfallursachen – im Alltag gar nicht so selten”. *Der Nervenarzt* 90 (10), 2019, S. 1013–1020.
- [Schmidhuber, 2015] Schmidhuber, J. “Deep learning in neural networks: An overview”. *Neural Networks* 61, 2015, S. 85–117.
- [Seidenstein u. a., 2015] Seidenstein, A. H., Barone, F. C. und Lytton, W. W. “Computer modeling of ischemic stroke”. *Scholarpedia* 10 (3), 2015. revision #150806, S. 32015.
- [Setio u. a., 2016] Setio, A. A. A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., van Riel, S. J., Wille, M. M. W., Naqibullah, M., Sánchez, C. I. und van Ginneken, B. “Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks”. *IEEE Transactions on Medical Imaging* 35 (5), 2016, S. 1160–1169.
- [Shibata u. a., 2019] Shibata, K., Hashimoto, T., Miyazaki, T., Miyazaki, A. und Nobe, K. “Thrombolytic Therapy for Acute Ischemic Stroke: Past and Future”. *Current Pharmaceutical Design* 25 (3), 2019, S. 242–250.
- [Shrout u. a., 1979] Shrout, P. E. und Fleiss, J. L. “Intraclass correlations: uses in assessing rater reliability.” *Psychological bulletin* 86 (2), 1979, S. 420.
- [Simonyan u. a., 2015] Simonyan, K. und Zisserman, A. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: *International Conference on Learning Representations (ICLR)*. 2015.
- [Song u. a., 2019] Song, T. und Huang, N. “Integrated Extractor, Generator and Segmentor for Ischemic Stroke Lesion Segmentation”. In: *Brainlesion: Glioma*,

- Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2019, S. 310–318.
- [Sporns u. a., 2018] Sporns, P. B., Kemmling, A., Schwake, M., Minnerup, J., Nawabi, J., Broocks, G., Wildgruber, M., Fiehler, J., Heindel, W. und Hanning, U. “Triage of 5 Noncontrast Computed Tomography Markers and Spot Sign for Outcome Prediction After Intracerebral Hemorrhage”. *Stroke* 49 (10), 2018, S. 2317–2322.
- [Srivastava u. a., 2014] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. und Salakhutdinov, R. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. *Journal of Machine Learning Research* 15, 2014, S. 1929–1958.
- [Srivastava u. a., 2015] Srivastava, R. K., Greff, K. und Schmidhuber, J. “Training Very Deep Networks”. In: *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., 2015, S. 2377–2385.
- [Stahmeyer u. a., 2019] Stahmeyer, J. T., Stubenrauch, S., Geyer, S., Weissenborn, K. und Eberhard, S. “Häufigkeit und Zeitpunkt von Rezidiven nach inzidentem Schlaganfall”. *Deutsches Ärzteblatt* 116 (42), 2019, S. 711–717.
- [Stier u. a., 2015] Stier, N., Vincent, N., Liebeskind, D. und Scalzo, F. “Deep learning of tissue fate features in acute ischemic stroke”. In: *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2015, S. 1316–1321.
- [Straka u. a., 2010] Straka, M., Albers, G. W. und Bammer, R. “Real-time diffusion-perfusion mismatch analysis in acute stroke”. *Journal of Magnetic Resonance Imaging* 32 (5), 2010, S. 1024–1037.
- [Sukkar u. a., 2012] Sukkar, R., Katz, E., Zhang, Y., Raunig, D. und Wyman, B. T. “Disease progression modeling using Hidden Markov Models”. In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2012, S. 2845–2848.
- [Suntrup u. a., 2015] Suntrup, S., Kemmling, A., Warnecke, T., Hamacher, C., Oelenberg, S., Niederstadt, T., Heindel, W., Wiendl, H. und Dziewas, R. “The impact of lesion location on dysphagia incidence, pattern and complications in acute stroke. Part 1: dysphagia incidence, severity and aspiration”. *European Journal of Neurology* 22 (5), 2015, S. 832–838.
- [Suntrup-Krueger u. a., 2017] Suntrup-Krueger, S., Kemmling, A., Warnecke, T., Hamacher, C., Oelenberg, S., Niederstadt, T., Heindel, W., Wiendl, H. und Dziewas, R. “The impact of lesion location on dysphagia incidence, pattern and complications in acute stroke. Part 2: Oropharyngeal residue, swallow and cough response, and pneumonia”. *European Journal of Neurology* 24 (6), 2017, S. 867–874.
- [Swieten u. a., 1988] Swieten, J. C. van, Koudstaal, P. J., Visser, M. C., Schouten, H. J. und Gijn, J. van. “Interobserver agreement for the assessment of handicap in stroke patients.” *Stroke* 19 (5), 1988, S. 604–607.

-
- [Szegedy u. a., 2015] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. und Rabinovich, A. “Going Deeper With Convolutions”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015.
- [Szegedy u. a., 2016] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. und Wojna, Z. “Rethinking the Inception Architecture for Computer Vision”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.
- [Takahashi u. a., 2014] Takahashi, N., Lee, Y., Tsai, D.-Y., Matsuyama, E., Kinoshita, T. und Ishii, K. “An automated detection method for the MCA dot sign of acute stroke in unenhanced CT”. *Radiological Physics and Technology* 7 (1), 2014, S. 79–88.
- [Thomalla u. a., 2009] Thomalla, G., Audebert, H. J., Berger, K., Fiebach, J. B., Fiehler, J., Kaps, M., Neumann-Haefelin, T., Schellinger, P. D., Siebler, M., Sobesky, J., Villringer, A., Witte, O. W. und Röther, J. “Bildgebung beim Schlaganfall – eine Übersicht und Empfehlungen des Kompetenznetzes Schlaganfall”. *Aktuelle Neurologie* 36 (7), 2009, S. 354–367.
- [Trepel, 2017] Trepel, M. *Neuroanatomie*. 7. Aufl. Urban & Fischer Verlag, 2017.
- [Tsai u. a., 2018] Tsai, J. P., Mlynash, M., Christensen, S., Kemp, S., Kim, S., Mishra, N. K., Federau, C., Nogueira, R. G., Jovin, T. G., Devlin, T. G., Akhtar, N., Yavagal, D. R., Bammer, R., Straka, M., Zaharchuk, G., Marks, M. P., Albers, G. W. und Lansberg, M. G. “Time From Imaging to Endovascular Reperfusion Predicts Outcome in Acute Stroke”. *Stroke* 49 (4), 2018, S. 952–957.
- [Turowski u. a., 2015] Turowski, B. und Schramm, P. “An Appeal to Standardize CT- and MR-Perfusion”. *Clinical Neuroradiology* 25 (2), 2015, S. 205–210.
- [Ušinskas u. a., 2004] Ušinskas, A., Dobrovolskis, R. A. und Tomandl, B. F. “Ischemic stroke segmentation on CT images using joint features”. *Informatica* 15 (2), 2004, S. 283–290.
- [Veit u. a., 2016] Veit, A., Wilber, M. J. und Belongie, S. “Residual Networks Behave Like Ensembles of Relatively Shallow Networks”. In: *Advances in Neural Information Processing Systems 29*. Curran Associates, Inc., 2016, S. 550–558.
- [Vincent u. a., 2008] Vincent, P., Larochelle, H., Bengio, Y. und Manzagol, P.-A. “Extracting and composing robust features with denoising autoencoders”. In: *Proceedings of the 25th International Conference on Machine Learning*. Proceedings of Machine Learning Research. ACM, 2008, S. 1096–1103.
- [Vymazal u. a., 2012] Vymazal, J., Rulseh, A. M., Keller, J. und Janouskova, L. “Comparison of CT and MR imaging in ischemic stroke”. *Insights into Imaging* 3 (6), 2012, S. 619–627.

- [Wahl u. a., 2018] Wahl, B., Cossy-Gantner, A., Germann, S. und Schwalbe, N. R. “Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings?” *BMJ Global Health* 3 (4), 2018.
- [Wall u. a., 2012] Wall, D. P., Kosmicki, J., DeLuca, T. F., Harstad, E. und Fusaro, V. A. “Use of machine learning to shorten observation-based screening and diagnosis of autism”. *Translational Psychiatry* 2 (4), 2012, e100–e100.
- [Wang u. a., 2014] Wang, X., Sontag, D. und Wang, F. “Unsupervised Learning of Disease Progression Models”. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '14. ACM, 2014, S. 85–94.
- [Wang u. a., 2016] Wang, L., Wang, L., Lu, H., Zhang, P. und Ruan, X. “Saliency Detection with Recurrent Fully Convolutional Networks”. In: *Computer Vision – ECCV 2016*. Springer International Publishing, 2016, S. 825–841.
- [Wardlaw u. a., 2005] Wardlaw, J. M. und Mielke, O. “Early Signs of Brain Infarction at CT: Observer Reliability and Outcome after Thrombolytic Treatment—Systematic Review”. *Radiology* 235 (2), 2005, S. 444–453.
- [Wheeler u. a., 2015] Wheeler, H. M., Mlynash, M., Inoue, M., Tipiririni, A., Liggins, J., Bammer, R., Lansberg, M. G., Kemp, S., Zaharchuk, G., Straka, M., Albers, G. W. und Investigators, D. 2. “The growth rate of early DWI lesions is highly variable and associated with penumbral salvage and clinical outcomes following endovascular reperfusion”. *International Journal of Stroke* 10 (5), 2015, S. 723–729.
- [Winzeck u. a., 2018] Winzeck, S., Hakim, A., McKinley, R., Pinto, J. A. A. D. S. R., Alves, V., Silva, C., Pisov, M., Krivov, E., Belyaev, M., Monteiro, M., Oliveira, A., Choi, Y., Paik, M. C., Kwon, Y., Lee, H., Kim, B. J., Won, J.-H., Islam, M., Ren, H., Robben, D., Suetens, P., Gong, E., Niu, Y., Xu, J., Pauly, J. M., Lucas, C., Heinrich, M. P., Rivera, L. C., Castillo, L. S., Daza, L. A., Beers, A. L., Arbelaezs, P., Maier, O., Chang, K., Brown, J. M., Kalpathy-Cramer, J., Zaharchuk, G., Wiest, R. und Reyes, M. “ISLES 2016 and 2017-Benchmarking Ischemic Stroke Lesion Outcome Prediction Based on Multispectral MRI”. *Frontiers in Neurology* 9, 2018, S. 679.
- [Xie u. a., 2017] Xie, S., Girshick, R., Dollar, P., Tu, Z. und He, K. “Aggregated Residual Transformations for Deep Neural Networks”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [Xie u. a., 2019] Xie, Y., Jiang, B., Gong, E., Li, Y., Zhu, G., Michel, P., Wintermark, M. und Zaharchuk, G. “Use of Gradient Boosting Machine Learning to Predict Patient Outcome in Acute Ischemic Stroke on the Basis of Imaging, Demographic, and Clinical Information”. *American Journal of Roentgenology* 212 (1), 2019, S. 44–51.

-
- [Yoo u. a., 2012] Yoo, A. J., Chaudhry, Z. A., Nogueira, R. G., Lev, M. H., Schaefer, P. W., Schwamm, L. H., Hirsch, J. A. und González, R. G. “Infarct Volume Is a Pivotal Biomarker After Intra-Arterial Stroke Therapy”. *Stroke* 43 (5), 2012, S. 1323–1330.
- [Yu u. a., 2016] Yu, X., Zhou, F. und Chandraker, M. “Deep Deformation Network for Object Landmark Localization”. In: *Computer Vision – ECCV 2016*. Springer International Publishing, 2016, S. 52–70.
- [Yu u. a., 2017] Yu, F., Koltun, V. und Funkhouser, T. “Dilated Residual Networks”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [Zaidi u. a., 2012] Zaidi, S. F., Aghaebrahim, A., Urra, X., Jumaa, M. A., Jankowitz, B., Hammer, M., Nogueira, R., Horowitz, M., Reddy, V. und Jovin, T. G. “Final Infarct Volume Is a Stronger Predictor of Outcome Than Recanalization in Patients With Proximal Middle Cerebral Artery Occlusion Treated With Endovascular Therapy”. *Stroke* 43 (12), 2012, S. 3238–3244.
- [Zhang u. a., 2018a] Zhang, Q.-s. und Zhu, S.-c. “Visual interpretability for deep learning: a survey”. *Frontiers of Information Technology & Electronic Engineering* 19 (1), 2018, S. 27–39.
- [Zhang u. a., 2018b] Zhang, Z., Yang, L. und Zheng, Y. “Translating and Segmenting Multimodal Medical Volumes With Cycle- and Shape-Consistency Generative Adversarial Network”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018.
- [Zhang u. a., 2019] Zhang, L., Lu, L., Wang, X., Zhu, R. M., Bagheri, M., Summers, R. M. und Yao, J. “Spatio-Temporal Convolutional LSTMs for Tumor Growth Prediction by Learning 4D Longitudinal Patient Data”. *IEEE Transactions on Medical Imaging*, 2019, S. 1–1.
- [Zhou u. a., 2018] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N. und Liang, J. “UNet++: A Nested U-Net Architecture for Medical Image Segmentation”. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA)*. Springer International Publishing, 2018, S. 3–11.
- [Zhu u. a., 2016] Zhu, W., Xiao, L., Lin, M., Liu, X. und Yan, B. “Large-Vessel Occlusion Is Associated with Poor Outcome in Stroke Patients Aged 80 Years or Older Who Underwent Intravenous Thrombolysis”. *Journal of Stroke and Cerebrovascular Diseases* 25 (11), 2016, S. 2712–2716.

Danksagung

An dieser Stelle möchte ich denjenigen Personen meinen Dank aussprechen, die mich während der Promotion unterstützt haben.

Dieser gilt insbesondere Prof. Dr. Mattias Heinrich für seine persönliche wie auch umfassende Betreuung und Unterstützung bei der Durchführung der gesamten Arbeit. Ihm und Priv.-Doz. Dr. med. André Kemmling danke ich außerdem für die Überlassung des Themas. Dr. Kemmling hat darüber hinaus die Arbeit mit Blick auf die klinische Relevanz stetig hilfreich begleitet.

Besonders danken möchte ich auch Dr. med. Linda Rudolf für die Unterstützung bei der Annotation der Datensätze, ohne die die Methoden nicht hätten trainiert und evaluiert werden können. Darüber hinaus danke ich PD Dr. rer. nat. Amir Madany Mamlouk für seine Hinweise zur Planung und Umsetzung der Arbeit, sowie Prof. Dr. rer. nat. Thomas Martinetz für die Begutachtung der Dissertation.

Nicht zuletzt gilt mein Dank den Kolleginnen und Kollegen des Instituts für Medizinische Informatik für den regelmäßigen Austausch in offener Atmosphäre und die zahlreichen Korrekturhinweise.

Meiner Frau und meinen Eltern danke ich für ihre von Beginn an vorbehaltlose Unterstützung und Ermutigung zur Anfertigung dieser Arbeit.