



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

Analyse anatomischer und pathologischer Strukturen in OCT-Bilddaten mittels maschineller Lernverfahren

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck

Aus der Sektion Informatik/Technik

vorgelegt von
Timo Kepp
aus Minden (Westfalen)

Lübeck, 2022

1. Berichtstatter: Prof. Dr. rer. nat. habil. Heinz Handels
2. Berichtstatter: Prof. Dr.-Ing. Alfred Mertins

Tag der mündlichen Prüfung: 28.07.2022

Zum Druck genehmigt. Lübeck, den 08.08.2022

Zusammenfassung

Die optische Kohärenztomographie (OCT) ist ein nichtinvasives und kontaktloses Bildgebungsverfahren, das hochauflösende dreidimensionale Schnittbilder halbtransparenter Gewebe ermöglicht. Ihr Funktionsprinzip beruht auf der Kurzkohärenz-Interferometrie mit Nahinfrarotlicht. Mit einer Auflösung im Mikrometerbereich und einer Messtiefe von mehreren Millimetern schließt sie die Lücke zwischen Sonographie und Mikroskopie. Die OCT hat sich in den medizinischen Anwendungsbereichen der Ophthalmologie und Dermatologie etabliert, wo sie zur Untersuchung des Augenhintergrunds oder der Haut verwendet wird.

Neben der qualitativen Analyse anatomischer und pathologischer Strukturen in OCT-Bilddaten ist die Extraktion ihrer quantitativen Parameter von zentraler Bedeutung. So stellen etwa das Volumen krankheitsbedingter Flüssigkeitsansammlungen oder die Morphologie der Schichten in der Retina wichtige Biomarker dar, die zur Verbesserung von Diagnose und Therapie herangezogen werden können. Zur Berechnung dieser Parameter ist eine pixelweise Segmentierung der entsprechenden Strukturen erforderlich. Die manuelle Erstellung von Segmentierungen ist jedoch sowohl sehr aufwendig und zeitintensiv als auch fehleranfällig. Mit der Entwicklung geeigneter Algorithmen lässt sich diese Segmentierungsaufgabe automatisieren, was zu einer personellen Entlastung führt. Abhängig vom Anwendungsgebiet können u. a. starke Variationen der zu untersuchenden Gewebe oder unterschiedliche Aufnahmebedingungen durch wechselnde Scannersysteme jedoch eine automatische Analyse der OCT-Bilddaten erschweren.

In dieser Arbeit werden Methoden zur Segmentierung anatomischer und pathologischer Strukturen in OCT-Bilddaten entwickelt, mit dem Ziel, neben einer größtmöglichen Robustheit auch eine leichte Anpassungsfähigkeit und Anwendbarkeit der Algorithmen zu gewährleisten. Die hier vorgeschlagenen Ansätze beruhen auf dem maschinellen Lernen, insbesondere auf Faltungsnetzwerken, deren Einsatz anhand von drei verschiedenen Anwendungsbereichen der OCT untersucht wird.

Zuerst werden die Vorteile von Faltungsnetzwerken gegenüber traditionellen Bildverarbeitungs Pipelines bei der Segmentierung von Hautschichten der Maus im Rahmen einer präklinischen Studie demonstriert. Es konnte gezeigt werden, dass das Faltungsnetzwerk die Anzahl der erforderlichen Verarbeitungsschritte drastisch reduzieren und gleichzeitig die Genauigkeit und Robustheit steigern kann. Das entwickelte Faltungsnetzwerk dient als Grundlage für die weiteren in dieser Arbeit betrachteten Anwendungsbereiche der OCT.

Gerade in klinischen Anwendungsbereichen wie bei der Ophthalmologie ist die Robustheit automatischer Segmentierungsverfahren angesichts der unmittelbaren Auswirkungen auf den Patienten essenziell. Hier ist neben der Analyse anatomischer Strukturen die Erfassung krankheitsbedingter Biomarker entscheidend für den Verlauf einer Therapie. In einem zweiten Anwendungsfall wird das Faltungsnetzwerk zu einer Multi-Task-Learning-Architektur weiterentwickelt, um die topologische Struktur der Retinaschichten in Gegenwart von Flüssigkeitseinlagerungen zu berücksichtigen. Die Ergebnisse konnten eine Erhöhung der klinisch relevanten Robustheit der Methode bei gleichzeitiger einfacher Anwendung bestätigen.

Mit neuen kostengünstigen OCT-Systemen für die Heimanwendung kann der Krankheitsverlauf der altersbedingten Makuladegeneration (AMD) engmaschig überwacht werden, um eine rechtzeitige Therapie zu gewährleisten, die für den Erhalt der Sehkraft des Patienten entscheidend ist. Im Kontext dieses Heimanwendungsszenarios wird außer der Segmentierung eines AMD-Flüssigkeitsbiomarkers auch dessen Unsicherheit geschätzt, die als Feedback für die Zuverlässigkeit einer Vorhersage genutzt werden kann. In einer umfassenden Evaluation sind zwei Unsicherheitsarten untersucht und ihr Nutzen zur Steigerung der Zuverlässigkeit von Segmentierungsvorhersagen dargelegt worden.

Insgesamt leisten die in dieser Arbeit entwickelten Methoden des maschinellen Lernens einen wichtigen Beitrag zur robusten und zuverlässigen Analyse anatomischer sowie pathologischer Strukturen und zeigen ihren flexiblen Einsatz in verschiedenen Anwendungsbereichen der OCT.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Aufbau und wissenschaftliche Beiträge dieser Arbeit	5
2	Optische Kohärenztomographie	9
2.1	Physikalisches Grundprinzip	9
2.2	Messmethoden	11
2.3	Axiale und laterale Auflösung	13
2.4	Medizinische Anwendungsfelder	13
3	Methodische Grundlagen	19
3.1	Bildsegmentierung	19
3.1.1	Graph-Cut-Segmentierung	21
3.1.1.1	Aufbau des Graphen	23
3.1.1.2	Maximaler Fluss und minimaler Schnitt	23
3.1.1.3	Kostenfunktion	25
3.1.2	Evaluationsmetriken	26
3.2	Maschinelles Lernen	27
3.2.1	Entscheidungsbäume	28
3.2.1.1	Aufbau	29
3.2.1.2	Training und Anwendung	29
3.2.1.3	Random Forests	31
3.2.2	Faltungsnetzwerke	33
3.2.2.1	Aufbau	35
3.2.2.2	Training und Anwendung	38
3.2.2.3	Das U-Net	40
4	Segmentierung von Hautschichten der Maus in OCT-Bilddaten	43
4.1	Einleitung und Motivation	43
4.1.1	Literaturübersicht	45
4.2	Material und Methoden	49
4.2.1	Verwendete Bilddaten	49
4.2.2	Random-Forest-Segmentierung mit graphenbasierter Optimierung	50
4.2.2.1	Vorverarbeitung	50
4.2.2.2	Random-Forest-basierte Klassifikation der Hautschichten	52

4.2.2.3	Graphenbasierte Optimierung der Hautschichtgrenzen	54
4.2.3	Segmentierung mit einem Fully Convolutional Network	58
4.2.3.1	Densely Connected U-Net	58
4.2.3.2	Training und Klassifikation	60
4.3	Experimente und Ergebnisse	62
4.4	Zusammenfassung und Diskussion	66
5	Formbasierte Segmentierung von Retinaschichten und Flüssigkeiten in OCT-Bilddaten	71
5.1	Einleitung und Motivation	71
5.1.1	Literaturübersicht	73
5.2	Material und Methoden	76
5.2.1	Verwendete Bilddaten	76
5.2.2	Multi-Task-Learning zur formbasierten Segmentierung	76
5.2.2.1	Regression von Signed Distance Maps	79
5.2.2.2	Duale Aufgabenkonsistenz	81
5.2.2.3	Halb-überwachtes Training	81
5.3	Experimente und Ergebnisse	82
5.4	Zusammenfassung und Diskussion	87
6	Unsicherheitsanalyse bei der Segmentierung von PEDs in Home-OCT-Bilddaten	89
6.1	Einleitung und Motivation	89
6.2	Literaturübersicht	91
6.3	Material und Methoden	94
6.3.1	Verwendete Bilddaten	94
6.3.2	Bestimmung epistemischer und aleatorischer Unsicherheiten	95
6.3.3	Segmentierung von PEDs mit Bayesian DCU-Net	97
6.3.4	Quantifizierung und Bewertung von Unsicherheiten	98
6.4	Experimente und Ergebnisse	100
6.4.1	Anzahl der MC-Stichproben	100
6.4.2	Qualitative Auswertung	101
6.4.3	Filterung von Unsicherheiten	103
6.4.4	Analyse strukturbezogener Unsicherheiten	104
6.5	Zusammenfassung und Diskussion	107
7	Zusammenfassung und Ausblick	109
A	Ergänzende Ergebnisse zu Kapitel 6	115
A.1	Qualitative Ergebnisbeispiele von PED-Segmentierungen und ihrer Unsicherheitsschätzung	115

B Eigene Publikationen	117
Literaturverzeichnis	121

Kapitel 1

Einleitung

1.1 Motivation

Medizinische Bildgebungsverfahren ermöglichen seit mehr als einem Jahrhundert Einblicke in das Körperinnere. Heutzutage werden sie in einer Vielzahl von klinischen Fachbereichen eingesetzt und bilden eine wichtige Säule der medizinischen Grundversorgung. In diesem Zusammenhang unterstützen medizinische Bildgebungsverfahren die Erkennung und Verlaufsüberwachung von Krankheiten sowie die Planung zur Durchführung therapeutischer Eingriffe [Dössel u. a., 2014]. Die optische Kohärenztomographie (*engl.*: optical coherence tomography, kurz: OCT) stellt ein sehr junges bildgebendes Verfahren dar, das der Sonographie ähnelt, bei dem jedoch Licht anstelle von Schall verwendet wird. Bereits in den frühen 1990er-Jahren wurde ihr Prinzip zur Bildgebung verschiedener Gewebeproben demonstriert [Huang u. a., 1991]. Die OCT beruht auf der Weißlichtinterferometrie und nutzt die Interferenz, die bei der Überlagerung von zurückgestreutem Licht mit dem Referenzlicht des Interferometers entsteht, um Tiefenprofile der zu untersuchenden Gewebeprobe zu erstellen. Über die Zusammensetzung mehrerer Tiefenprofile entlang einer Linie lassen sich so hochaufgelöste Querschnittsaufnahmen generieren. Mit einer Tiefenauflösung von 1–15 μm und einer Messtiefe von 1–2 mm [Leitgeb u. a., 2004] schließt die OCT die Lücke zwischen Sonographie und Mikroskopie in der medizinischen Bildgebung (siehe Abb. 1.1). Gegenüber anderen bildgebenden Verfahren wie der Computertomographie (CT) oder der Magnetresonanztomographie (MRT) ist ihre Messtiefe zwar wesentlich geringer, bietet aber eine um den Faktor 10–100 feinere Auflösung.

Die Bildakquisition mittels OCT ist nichtinvasiv und im Unterschied zur Sonographie berührungslos, da anders als bei Schallwellen der Übergang zwischen Luft und Gewebe die Lichtausbreitung kaum behindert. Diese Eigenschaften haben die OCT vor allem interessant für die Untersuchung des Augenhintergrunds gemacht, der wegen der hohen Transparenz des Glaskörpers optisch leicht zugänglich ist. Aufgrund ihres hohen Auflösungsvermögens erlaubt die OCT eine detaillierte Darstellung der Retina, wie sie mit keinem anderen medizinischen Bildgebungsverfahren möglich ist. Diese Alleinstellungsmerkmale haben die Ophthalmologie revolutioniert, die unverändert das Hauptanwendungsgebiet der OCT ist. Zuvor ist eine detaillierte Darstellung der

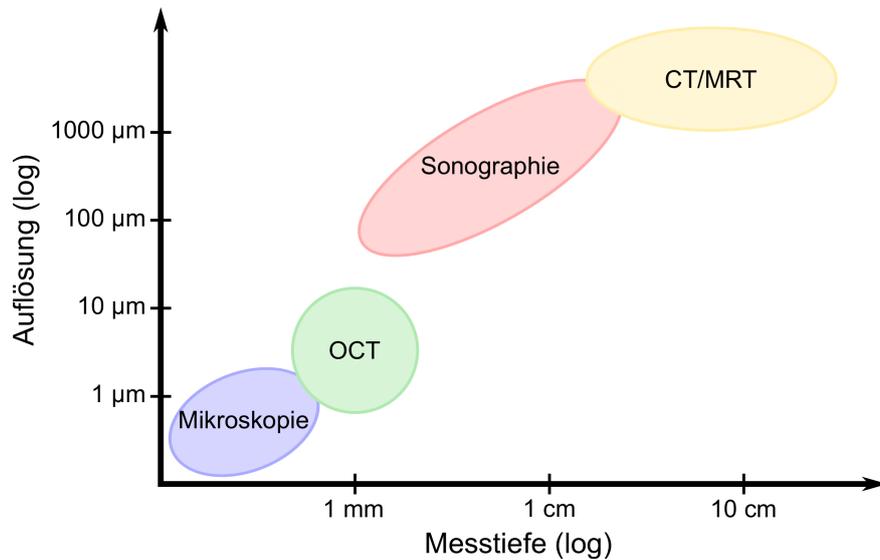


Abb. 1.1: Vergleich verschiedener medizinischer Bildgebungsverfahren hinsichtlich ihrer Auflösung und Messtiefe.

menschlichen Retina nur aus Histologien bekannt gewesen. Mittels OCT-Untersuchung werden auch Erkrankungen der Makula (*lat.*: macula lutea, *dt.*: gelber Fleck), die Stelle des schärfsten Sehens der Retina, diagnostiziert. Insbesondere bei Erkrankungen wie der altersbedingten Makuladegeneration (AMD) oder dem diabetischen Makulaödem (DMÖ) kann es in diesem Bereich zu krankheitsbedingten Flüssigkeitsansammlungen oder Veränderungen der Retinaschichten kommen, die sich mithilfe der OCT erfassen lassen.

Ein erstes OCT-System für den Einsatz in der Ophthalmologie wurde bereits 1996 von der Carl Zeiss Meditec AG unter dem Namen „OCT 1“ auf dem Markt eingeführt [Pierre-Kahn u. a., 2005]. Seitdem hat sich die OCT-Bildgebung kontinuierlich weiterentwickelt und ist stetig optimiert worden. So ermöglichen moderne OCT-Systeme heutzutage eine Bildgebung in Echtzeit und dienen als unterstützendes Instrument bei intraoperativen Eingriffen [Hattenbach u. a., 2016]. Ein neuer Entwicklungszweig beschäftigt sich mit der Konzeption kostengünstiger OCT-Geräte für die Selbstmessung zu Hause [Chopra u. a., 2021], sogenannte Home-OCT-Systeme, wodurch eine engmaschige Überwachung schnell fortschreitender Erkrankungen, wie der feuchten Verlaufsform der AMD, gewährleistet werden kann. Außer ihrer Anwendung in der Ophthalmologie wird die OCT auch in anderen medizinischen Fachgebieten verwendet. In der Dermatologie wird sie z. B. unterstützend als „optische Biopsie“ zur Differenzialdiagnose von Hauttumoren herangezogen [Welzel u. a., 2017]. Darüber hinaus wird sie in präklinischen Studien häufig zur Bildgebung in lebenden Tiermodellen eingesetzt [Salma u. a., 2018].

Die Auswertung von OCT-Bilddaten wird in der Regel von erfahrenen Experten auf ihrem jeweiligen Gebiet durchgeführt. Ergänzend zur rein qualitativen Analyse ist die Bestimmung von quantitativen Parametern aus OCT-Bilddaten von großer Bedeutung. So stellt z. B. das Volumen krankheitsbedingter Flüssigkeitseinlagerungen sowie deren Veränderung einen wichtigen Biomarker in der AMD-Diagnostik dar und kann ein direktes Behandlungskriterium sein [DOG, 2020]. Die Dicke sowie der Verlauf einzelner Schichten der Retina oder der Haut können Auskunft über morphologische Veränderungen geben. Die Extraktion der zuvor beschriebenen Parameter erfolgt auf der Basis von Segmentierungen, welche die pixelweisen Abgrenzungen einer Struktur im (OCT-)Bild beschreiben (siehe Abb. 1.2). Anhand einer Segmentierung lassen sich dann verschiedene Kenngrößen wie Fläche, Volumen oder Form ableiten. Jedoch ist die manuelle Segmentierung medizinischer (OCT-)Bilddaten sehr zeitaufwendig und fehleranfällig [Bogunović u. a., 2019]. Angesichts der zunehmenden Verfügbarkeit von OCT-Geräten und der damit einhergehenden steigenden Menge an auszuwertenden Bilddaten ist dies gerade in der klinischen Praxis nicht durchführbar. Eine Automatisierung dieser Segmentierungsaufgabe würde dementsprechend zu einer enormen Arbeitersparnis und zur personellen Entlastung führen.

Als Bestandteil klassischer Mustererkennungspipelines stellt die Entwicklung solcher automatischen Segmentierungsverfahren ein Schlüsselproblem der medizinischen Bildverarbeitung dar [Handels, 2009, S. 49 ff.]. Ihre Abfolge setzt sich aus mehreren Verarbeitungsschritten zusammen: Bildvorverarbeitung → Segmentierung → Merkmalsextraktion → Klassifikation. Zur Segmentierung von OCT-Bilddaten sind bereits verschiedene Methoden vorgeschlagen worden, welche sich an dieser Abfolge von Verarbeitungsschritten orientieren. Aufgrund des Hauptanwendungsbereichs der OCT in der Ophthalmologie beschränken sich die meisten Verfahren vorwiegend auf die Segmentierung von Retinaschichten oder Flüssigkeitsablagerungen [DeBuc, 2011; Kafieh u. a., 2013]. Insbesondere die Anwendung von maschinellem Lernen in Kombination mit graphenbasierten Ansätzen ist weitverbreitet. Zur Segmentierung von Retinaschichten nutzen Lang u. a. [2013a] einen Random-Forest-Klassifikator, welcher anhand manuell konstruierter Merkmale Schichtübergänge der Retina detektiert. Die Wahrscheinlichkeiten der Schichtübergänge werden über Graph-Cut-Methoden weiter verfeinert.

Die Flexibilität solcher klassischen Segmentierungsansätze ist jedoch häufig sehr begrenzt. Bereits ihre Verwendung zur Segmentierung in Bilddaten verschiedener OCT-Scanner stellt eine große Herausforderung dar [Bogunović u. a., 2019] und erfordert die Anpassung mehrerer Komponenten wie der Bildvorverarbeitung oder der Definition der Merkmale des Klassifikators. Gleiches gilt für die Anpassung an neue Strukturen, z. B. beim Wechsel von Retina- zu Hautschichten. Hinzu kommt, dass die Anwendbarkeit graphenbasierter Methoden durch das Auftreten von Pathologien erheblich erschwert wird, da durch sie Schichten der Retina oder der Haut teilweise unterbrochen oder verschoben werden können. Schließlich wird die Bildakquisition durch unterschiedliche Aufnahmebedingungen beeinflusst. Während der Akquisitionsprozess von Gewebepro-

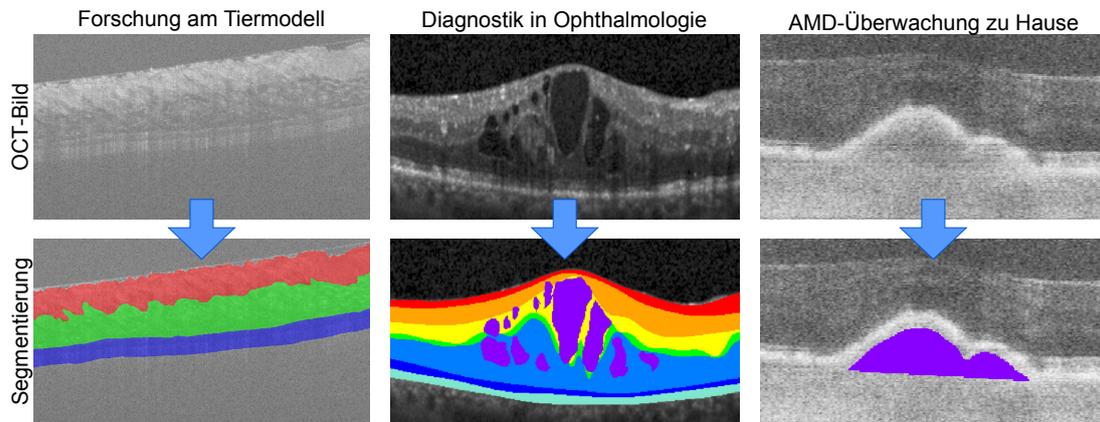


Abb. 1.2: Segmentierungsbeispiele der in dieser Arbeit bearbeiteten OCT-Anwendungsbereiche. Links: Segmentierung von Hautschichten der Maus im Rahmen einer präklinischen Studie. Mitte: Segmentierung von Retinaschichten und Flüssigkeitsablagerungen eines Patienten mit DMÖ. Rechts: Segmentierung eines AMD-bedingten Biomarkers in einem Scan eines Home-OCT-Systems für die Heimanwendung.

ben oder Tiermodellen im Rahmen von präklinischen Studien noch weitestgehend kontrolliert gestaltet werden kann, ist dieser in der Klinik oder bei einer Selbstmessung zu Hause stark vom Patienten abhängig, was zu möglichen Beeinträchtigungen der Bildqualität, z. B. durch Bewegungsartefakte, führen kann. Demnach müsste jeweils eine Anpassung der Algorithmen auf die unterschiedlichen Einflüsse erfolgen, um eine robuste Segmentierung fortlaufend zu gewährleisten.

Mit der Verfügbarkeit stetig wachsender Bilddatensätze sowie leistungsfähiger Grafikprozessoren (*engl.*: graphical processing unit, kurz: GPU) hat sich in jüngster Zeit eine rasante Entwicklung von Deep-Learning-Methoden ergeben, die eine Unterkategorie des maschinellen Lernens bilden [LeCun u. a., 2015]. Die in visuellen Anwendungsbereichen eingesetzten Faltungsnetzwerke (*engl.*: convolutional neural networks, kurz: CNNs) sind in der Lage, Merkmale auf einer rein datengetriebenen Weise zu lernen und zu extrahieren. Deep-Learning-Methoden bewirken somit einen Paradigmenwechsel [Altaf u. a., 2019], der zu einer massiven Verkürzung bisheriger Mustererkennungs- oder Segmentierungspipelines führt. Gleichzeitig kann die oben beschriebene Variabilität von OCT-Bilddaten infolge der enormen Anpassungsfähigkeit von Deep-Learning-Methoden besser aufgefangen werden.

Im Hinblick auf die Analyse medizinisch relevanter Strukturen in OCT-Bilddaten bestehen jedoch weiterhin mehrere Herausforderungen. Das Training überwachter Deep-Learning-basierter Segmentierungsansätze erfordert eine sehr hohe Anzahl annotierter Bilddaten, die als Lernvorgabe dienen. Die Verfügbarkeit solcher Annotationen ist wegen ihrer oben beschriebenen zeitaufwendigen Erstellung zumeist sehr gering. Ein weiteres Problem ist die ungleichmäßige Verteilung der zu untersuchenden Strukturen

in den OCT-Bilddaten. Vor allem Pathologien sind in der Regel deutlich unterrepräsentiert. Die geringe Anzahl von Trainingsbeispielen wirkt sich hier zusätzlich erschwerend aus. Ferner ist die Einhaltung anatomischer Beziehungen bei der Segmentierung von entscheidender Bedeutung. Die Schichten der Retina und der Haut folgen einer bestimmten Anordnung. Die Einhaltung dieser topologischen Relationen ist bei Deep-Learning-Methoden gegenüber graphenbasierten Modellen nicht automatisch gegeben. Schließlich muss für die Anwendung von Deep-Learning-Methoden in der klinischen Praxis und besonders in Kombination mit OCT-Systemen für den Heimgebrauch eine robuste und zuverlässige Arbeitsweise sichergestellt sein. Kann eine Bewertung nur mit einem geringen Maß an Sicherheit getroffen werden, könnte sich das negativ auf die Behandlungsentscheidung auswirken und folglich die Gesundheit des Patienten beeinträchtigen. Die in dieser Arbeit präsentierten Methoden des maschinellen Lernens sollen einen Beitrag zur Lösung dieser Problemstellungen bei der Analyse in OCT-Bilddaten leisten.

1.2 Aufbau und wissenschaftliche Beiträge dieser Arbeit

In der vorliegenden Arbeit werden verschiedene automatische Segmentierungsmethoden zur Analyse von OCT-Bilddaten vorgestellt, welche die zuvor beschriebenen Problemstellungen adressieren. Die dabei angewendeten Algorithmen beruhen auf Methoden des maschinellen Lernens, insbesondere auf Deep Learning. Zur besseren Veranschaulichung sind in Abbildung 1.3 die Inhalte der Methodenkapitel mit farblicher Hervorhebung ihrer Zusammenhänge dargestellt. Die darin behandelten Anwendungsbereiche der OCT umfassen die präklinische Forschung am Tiermodell und die ophthalmologische Diagnostik sowohl in der Klinik als auch beim Patienten zu Hause (siehe Abb. 1.2). Während sich die Auswertungen von OCT-Bilddaten im Rahmen präklinischer Studien noch nicht unmittelbar auf den Patienten auswirken, bilden sie eine wichtige Entscheidungsgrundlage in der medizinischen Versorgung. Dabei steht neben anatomischen Strukturen vorrangig die Analyse von Pathologien im Fokus diagnostischer Untersuchungen, deren Erkennung hingegen weitaus anspruchsvoller ist. Zudem nimmt, wie zuvor in der Motivation dargelegt, für jeden Anwendungsfall die Zahl nicht beeinflussbarer Faktoren bei der Aufnahme mit dem OCT zu, welche zu Bildartefakten führen können und damit eine Auswertung erheblich beeinträchtigen würden. Pro Kapitel und dem jeweiligen Anwendungsbereich ist demnach eine Zunahme der klinischen Relevanz von Robustheit und Zuverlässigkeit bei der Auswertung von OCT-Bilddaten erkennbar, die in Abbildung 1.3 durch einen Farbverlauf von hell- nach dunkellila gekennzeichnet ist. Im Folgenden werden die wissenschaftlichen Beiträge der Methodenkapitel mit ihren hervorgehobenen Zielen näher beschrieben.

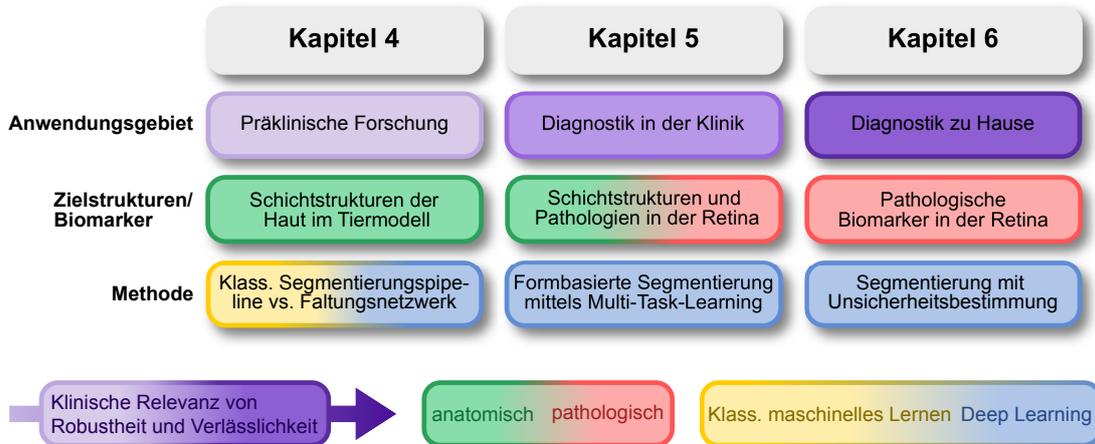


Abb. 1.3: Schematische Darstellung der in dieser Arbeit bearbeiteten Inhalte der Methodenkapitel. Die Zusammenhänge zwischen den einzelnen Kapiteln sind farblich gekennzeichnet.

Verkürzung von Segmentierungspipelines und Vereinfachung ihrer Anwendbarkeit

In Kapitel 4 werden zwei Segmentierungsansätze für die automatische Auswertung von Hautschichten der Maus im Rahmen einer präklinischen Studie vorgeschlagen, die sowohl auf einer traditionellen Segmentierungspipeline als auch auf einem Deep-Learning-Verfahren basieren. Zu den Hauptkomponenten der Segmentierungspipeline gehören zum einen ein klassisches maschinelles Lernverfahren, das anhand spezifisch definierter Merkmale eine pixelweise Klassifikation der Hautschichten vornimmt, und zum anderen ein graphenbasierter Ansatz, der eine Verfeinerung des Segmentierungsergebnisses ermöglicht. Zur drastischen Verkürzung der Prozesskette wird als zweite Methode ein CNN vorgeschlagen. In Anlehnung an [Ronneberger u. a., 2015] wird ein vollständig faltungsbasiertes Netzwerk (*engl.*: fully convolutional networks, kurz: FCN) verwendet, das einen OCT-Scan als Ganzes segmentiert. Um sicherzustellen, dass das FCN auch mit wenigen manuell annotierten Bilddaten eine hohe Generalisierbarkeit erreicht, werden beim Training mittels Augmentierungstechniken künstliche Trainingsbeispiele erzeugt. Zudem wird während des Lernprozesses die ungleichmäßige Verteilung der Zielstrukturen im OCT-Bilddatensatz berücksichtigt.

Berücksichtigung der Topologie anatomischer Strukturen bei der Segmentierung mit FCNs

Ähnlich wie die Haut ist auch die menschliche Retina in Schichten aufgebaut, deren Analyse mittels OCT Bestandteil der klinischen Routine in der Ophthalmologie ist. Gleichzeitig hat die Erkennung krankheitsbedingter Biomarker, wie z. B. DMÖ-bedingte Flüssigkeitseinlagerungen, einen hohen diagnostischen Wert, da sie einen Indikator für die Krankheitsaktivität darstellen. In Kapitel 5 wird daher eine Methode zur Segmentierung von Retinaschichten und Flüssigkeitseinlagerun-

gen in klinischen OCT-Bilddaten von DMÖ-Patienten untersucht. Um ähnlich wie bei graphenbasierten Methoden die topologische Struktur der Retinaschichten im Segmentierungsprozess zu berücksichtigen und zugleich eine einfache Anwendbarkeit des Verfahrens zu gewähren, wird das in Kapitel 4 vorgeschlagene FCN um eine Multi-Task-Learning-Strategie erweitert. Dazu wird die Segmentierung in eine formbasierte Regression der Retinaschichten und eine pixelweise Klassifikation der Flüssigkeitseinschlagerungen aufgeteilt. Auf diese Weise soll zum einen die klinisch relevante Robustheit der Methode erhöht und zum anderen aufwendige Nachbearbeitungsschritte vermieden werden. Gleichzeitig ermöglicht das Multi-Task-Learning die Nutzung nicht annotierter Bilddaten, wodurch die Anzahl verfügbarer Trainingsbeispiele erhöht wird.

Bestimmung von Unsicherheiten zur Unterstützung der Bewertung von Segmentierungen Eine rechtzeitige Erkennung krankheitsbedingter Biomarker ist bei der feuchten AMD-Verlaufsform entscheidend. Mit der Entwicklung von Home-OCT-Systemen können AMD-Patienten eine intensive Überwachung ihres Krankheitszustands selbstständig durchführen. In Anbetracht der zu erwartenden großen Menge an OCT-Bilddaten ist es sinnvoll, solche Home-OCT-Systeme mit einer automatisierten computergestützten Diagnose (*engl.*: computer aided diagnosis, kurz: CAD) zu kombinieren, die den betroffenen Patienten sowie den behandelnden Arzt über den aktuellen Krankheitszustand informieren kann. Allerdings ist die Bildqualität von Home-OCT-Systemen im Vergleich zu Klinikgeräten aufgrund ihrer kostengünstigen Bauweise geringer (siehe Abb. 1.2). Ebenso können während der Selbstmessung auftretende Bewegungsartefakte eine Auswertung der Home-OCT-Bilddaten weiter erschweren. Im Hinblick auf ein Heimanwendungsszenario der OCT nimmt die Zuverlässigkeit eines CAD-Systems folglich einen zentralen Stellenwert ein. In Kapitel 6 werden daher zwei FCN-basierte Methoden zur Erkennung und Analyse eines AMD-Biomarkers in Bilddaten eines Home-OCT-Prototyps [Sudkamp u. a., 2016] untersucht, die neben der Segmentierung eine Unsicherheitschätzung ausgeben. Die Unsicherheit kann somit als Feedback genutzt werden, wenn die Auswertung einer Aufnahme aufgrund einer unzureichenden Bildqualität nicht eindeutig ist. Als Grundlage dient hierbei ebenfalls das in Kapitel 4 vorgeschlagene FCN.

Die Kapitel dieser Arbeit sind weiterhin wie folgt gegliedert:

- Kapitel 2 gibt einen Überblick über die OCT-Bildgebung. Dieser umfasst das physikalische Grundprinzip, die unterschiedlichen Messmethoden und die klinischen Anwendungsbereiche der OCT.
- In Kapitel 3 werden die methodischen Grundlagen und Begrifflichkeiten beschrieben, auf denen die in dieser Arbeit entwickelten Verfahren beruhen.
- In Kapitel 4 werden zwei Methoden zur Segmentierung von Hautschichten der Maus in OCT-Bilddaten präsentiert und miteinander verglichen. Die erste Methode ba-

sirt auf einer Kombination aus Random-Forest-Klassifikation und graphenbasierter Optimierung, während beim zweiten Ansatz ein FCN zur Segmentierung verwendet wird. Auszüge der Untersuchungen wurden in [Kepp u. a., 2018] und in [Kepp u. a., 2019a] veröffentlicht.

- Zur FCN-basierten Segmentierung von Retinaschichten und DMÖ-bedingten Flüssigkeitsablagerungen wird in Kapitel 5 eine Multi-Task-Learning-Strategie untersucht. Eine formbasierte Regression der Retinaschichten wird angewandt, um deren topologische Struktur zu berücksichtigen. Die Segmentierung der Flüssigkeitseinlagerungen geschieht durch eine simultane Klassifikation. Die vorgeschlagene Methode baut auf der Veröffentlichung [Kepp u. a., 2019b] auf.
- Um ein Sicherheitsfeedback bei der Analyse AMD-bedingter Biomarker in Home-OCT-Bilddaten zu ermöglichen, werden in Kapitel 6 zwei FCN-basierte Methoden untersucht, die parallel zur Segmentierung auch ihre jeweilige Unsicherheit bestimmen. Die Entwicklung baut auf der Arbeit [Kepp u. a., 2020] auf. Teilergebnisse wurden außerdem in [Kepp u. a., 2022] veröffentlicht.
- Im letzten Kapitel 7 werden die in dieser Arbeit vorgeschlagenen Methoden sowie deren Ergebnisse zusammengefasst und die zentralen Schlussfolgerungen gezogen. Daneben bietet ein Ausblick Anreize für weitere zu untersuchende Problemstellungen.

Kapitel 2

Optische Kohärenztomographie

Die optische Kohärenztomographie ist ein nichtinvasives, kontaktloses Verfahren, das vorwiegend in der Medizin zur Bildgebung in streuendem Gewebe eingesetzt wird. Mit einem Auflösungsvermögen von wenigen Mikrometern und einer Messtiefe von mehreren Millimetern schließt die OCT die Lücke zwischen der Mikroskopie und der Sonographie (siehe Abb. 1.1). Das physikalische Grundprinzip beruht auf der Interferometrie, bei der die Verzögerung von zurückgestreutem Licht Rückschlüsse auf den zurückgelegten optischen Weg zulässt. Die OCT ist erstmals im Jahr 1991 von Huang u. a. [1991] in einem In-vitro-Experiment an der Retina und den Koronararterien des Menschen beschrieben worden. Eine detaillierte Übersicht der OCT ist in [Drexler u. a., 2008] und [Dössel u. a., 2014] gegeben.

Nach einer kurzen Vorstellung des physikalischen Grundprinzips (Kapitel 2.1) werden die etablierten Messmethoden der OCT (Kapitel 2.2) und ihr Auflösungsvermögen (Kapitel 2.3) beschrieben. Abschließend werden in Kapitel 2.4 klinische Anwendungen der OCT vorgestellt.

2.1 Physikalisches Grundprinzip

Aufgrund des ähnlichen Funktionsprinzips wird die OCT oft mit der Sonographie verglichen, nur dass Licht anstelle von Schall eingesetzt wird. Zur Messung von Tiefeninformationen senden beide Bildgebungsverfahren Wellen in das zu untersuchende Gewebe, wo ein bestimmter Anteil von unterschiedlichen Strukturen reflektiert wird. Über eine Analyse der zurückkommenden Wellen kann der Ort der Reflexion oder der Streuung ermittelt werden. Da sich Licht im Vergleich zum Schall weitaus schneller fortbewegt, ist eine direkte Messung nicht praktikabel. Damit die Tiefeninformationen des zurückgeworfenen Lichts dennoch erfasst werden können, wird mit polychromatischem Licht Interferometrie eingesetzt, welche das fundamentale Prinzip der OCT darstellt. In einem Interferometer (siehe Abb. 2.1, links) wird das von einer Quelle ausgesendete Licht über einen Strahlteiler in zwei Strahlengänge aufgeteilt, wobei ein Teil in den Probenarm und ein anderer Teil in den Referenzarm geleitet wird. Das zurückkommende Licht beider Strahlengänge wird im Interferometer zusammengeführt und die Intensität der kombinierten Lichtwellen kann als elektrisches Signal

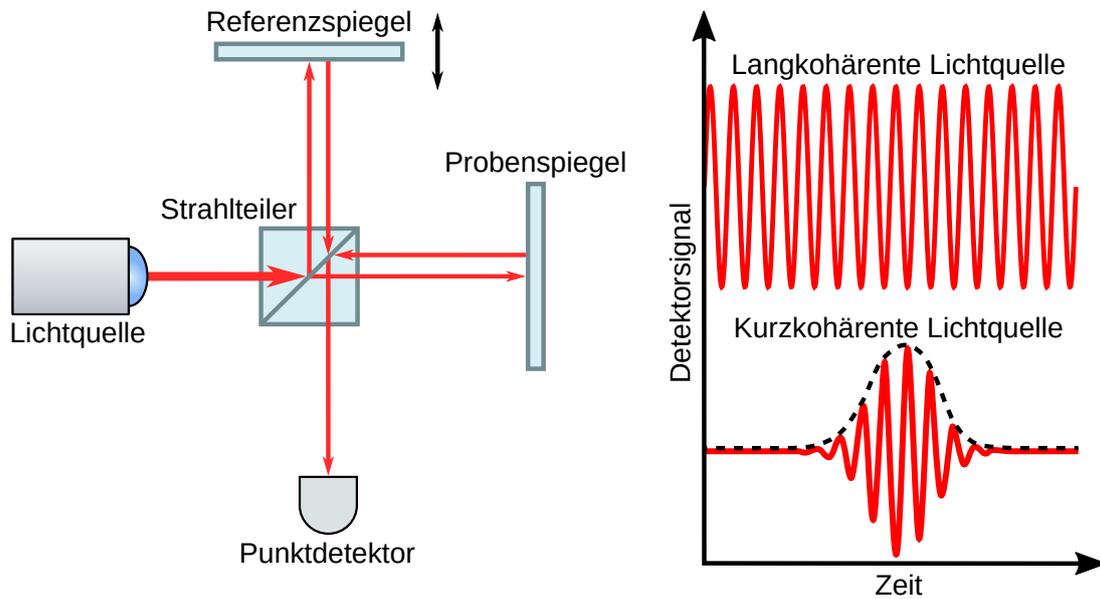


Abb. 2.1: Schematische Darstellung eines Michelson-Interferometers (links) und die Detektorsignale bei Verwendung einer lang- und kurzkohärenten Lichtquelle (rechts).

am Fotodetektor gemessen werden. Beim Zusammenführen der Strahlengänge können unter bestimmten Bedingungen Interferenzeffekte beobachtet werden. Dabei hängt die beobachtete Intensität von der Phase der Lichtwellen beider Arme ab. Ist sie höher als die Summe der Intensitäten, wird sie als konstruktive Interferenz bezeichnet, ist sie niedriger, als destruktive Interferenz. Durch das Verschieben des Referenzspiegels am Interferometer lässt sich dieser Interferenzeffekt beobachten. Bei Verwendung einer monochromatischen Lichtquelle beträgt der Abstand der Spiegelpositionen zwischen konstruktiver und destruktiver Interferenz genau eine halbe Wellenlänge.

Grundvoraussetzung für die Interferenzfähigkeit ist eine feste Phasenbeziehung der Lichtwellenzüge, die als Kohärenz bezeichnet wird. Ein Lichtwellenzug wird als zeitlich kohärent definiert, wenn eine feste Phasendifferenz zwischen verschiedenen Zeitabständen, d. h. entlang der Ausbreitungsrichtung, vorhanden ist. Lichtwellenzüge mit einer geringen zeitlichen Kohärenz weisen nur in einem begrenzten Zeitintervall eine feste Phasenbeziehung auf. Der Ausbreitungsbereich wird als Kohärenzlänge bezeichnet und ist eine wichtige Eigenschaft einer Lichtquelle. In Abbildung 2.1 sind hierzu Beispiele zweier Detektorsignale von Lichtquellen mit langer und kurzer zeitlicher Kohärenz dargestellt. Durch die Verwendung einer Lichtquelle mit kurzer Kohärenzlänge lassen sich somit Reflektionen und Streuungen einer bestimmten Tiefe zuordnen. Dabei ist eine Interferenz nur dann messbar, wenn die optischen Pfade des Referenz- und Probenarms so eingestellt sind, dass ihr Weglängenunterschied innerhalb der Kohärenzlänge der Lichtquelle liegen. Zeitliche Kohärenz wird mathematisch durch eine Autokorrelation

des Lichtwellenfeldes bestimmt und ist nach dem Wiener-Chintschin-Theorem direkt durch das Leistungsspektrum der Lichtquelle bestimmt. Bei der OCT-Bildgebung kommen daher Lichtquellen zum Einsatz, die eine große spektrale Bandbreite aufweisen, wie z. B. eine Superlumineszenzdiode (SLD), und dadurch eine geringe zeitliche Kohärenz besitzen, wohingegen monochromatisches Laserlicht eine hohe Kohärenzlänge aufweist und nicht für die OCT geeignet ist.

Ein Nachteil bei der Bildgebung mit kohärenten Wellen ist das sogenannte Speckle-Rauschen. Dieses entsteht durch die Interferenz von Wellen, die zueinander kohärent sind und von zufällig angeordneten Substrukturen unterhalb des Auflösungs-niveaus reflektiert werden. Speckle-Rauschen ist in OCT- oder Sonographie-Bilddaten als körniges Muster erkennbar, das zu einer Beeinträchtigung der Bildqualität führt.

2.2 Messmethoden

Wie zuvor dargelegt, kann die Detektion der Interferenz bei der OCT entweder als zeitliches Signal durch Bewegung des Referenzspiegels oder bei festem Referenzspiegel spektral aufgelöst erfolgen. Die erste Variante wird als Time-Domain-OCT (kurz: TD-OCT, *dt.:* Zeit-Domänen-OCT) bezeichnet. Auf ihr beruht das erste technisch umgesetzte Messsystem der OCT und ist in [Huang u. a., 1991] beschrieben. Um ein vollständiges Tiefenprofil der Probe zu erstellen, wird der Referenzspiegel über alle Messpositionen bewegt und dabei die Intensität des Interferenzsignals kontinuierlich gemessen. Aus der Einhüllenden des Messsignals (siehe Abb. 2.1, rechts unten) lässt sich der Lichtanteil ableiten, der von der Probe in der jeweiligen Tiefe zurückreflektiert oder zurückgestreut wird. Das gemessene Profil repräsentiert die einzelnen Probenreflexionen der verschiedenen Tiefen und wird als Amplituden-Scan (A-Scan) bezeichnet. Um ein Querschnittsbild der Probe zu erzeugen, wird der Lichtstrahl im Probenarm seitlich über die Probe bewegt. Die an jeder Position gemessenen A-Scans ergeben zusammen ein zweidimensionales Bild, das als B-Scan (*engl.:* brightness scan) bezeichnet wird. Mehrere B-Scans ergeben schließlich ein komplettes OCT-Volumen (siehe Abb. 2.3).

In der zweiten Variante wird anstatt die Intensität des Interferenzsignals durch mechanische Abtastung der optischen Weglänge zu messen, die Interferenz der einzelnen Spektralkomponenten erfasst. Analog zum Wiener-Chintschin-Theorem besteht ein Zusammenhang zwischen dem zeitlichen Interferenzsignal und dem Spektrum der Interferenz. Diese zweite OCT-Generation wird als Fourier-Domain-OCT (kurz: FD-OCT, *dt.:* Fourier-Domänen-OCT) bezeichnet. Doch erst mit der Verfügbarkeit schnellerer Kamera- und Computertechnik konnte die FD-OCT-Messtechnik seit 2005 umgesetzt werden. Gegenüber der TD-OCT hat die FD-OCT einen erheblichen Geschwindigkeitsvorteil, da ein komplettes Tiefenprofil ohne bewegte Elemente erfasst werden kann, was zu einer höheren Messstabilität führt [Leitgeb u. a., 2003]. Das direkte Messen im Spek-

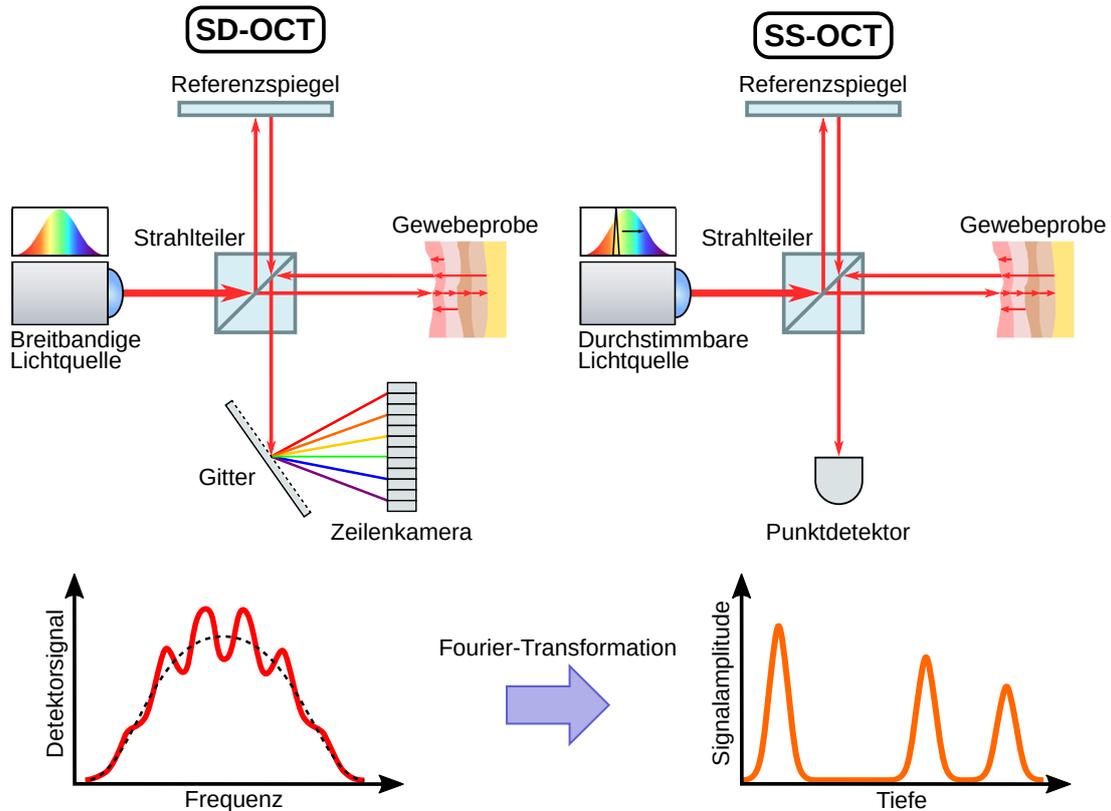


Abb. 2.2: Schematische Darstellung der SD-OCT und SS-OCT sowie dem Detektorsignal eines Interferenzspektrums, das sich mithilfe der Fourier-Transformation in ein Tiefenprofil übertragen lässt.

tralbereich führt zu einem verbesserten Signal-Rausch-Verhältnis (*engl.*: signal-to-noise ratio, kurz: SNR) [de Boer u. a., 2003]. Durch die Entwicklung der FD-OCT wurde der Grundstein für die volumetrische und später auch für die Echtzeit-Bildgebung gelegt.

Ein erstes Konzept der FD-OCT ist von Fercher u. a. [1995] vorgeschlagen worden, welches auch als Spektral-Domain-OCT (kurz: SD-OCT, *dt.*: Spektrale-Domänen-OCT) bezeichnet wird. Der Messaufbau ähnelt dem der TD-OCT, nur dass der Punktdetektor durch ein Spektrometer ersetzt wird (siehe Abb. 2.2, oben links). Über ein optisches Gitter im Spektrometer wird das Interferenzsignal in seine einzelnen Spektralkomponenten aufgeteilt und anschließend durch eine Zeilenkamera aufgenommen. Das detektierte Interferogramm enthält die gesamten Tiefeninformationen der Probe in Form einer Modulation des gemessenen Spektrums (siehe Abb. 2.2, unten links). Mithilfe der Fourier-Transformation werden die einzelnen Betragsgrößenwerte schließlich in Abhängigkeit von ihrer Tiefe in einen A-Scan umgewandelt (siehe Abb. 2.2, unten rechts).

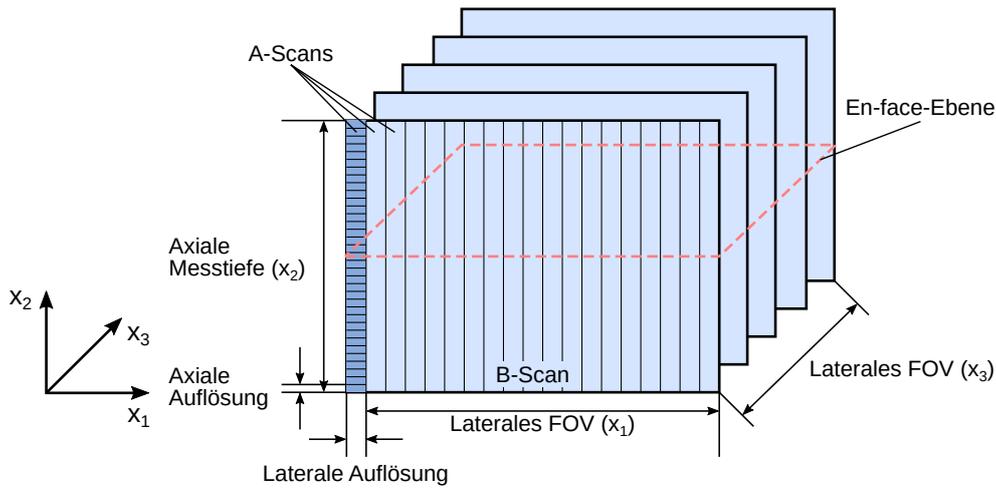
Etwas später ist von Chinn u. a. [1997] eine weitere FD-OCT-Messmethode, die Swept-Source-OCT (SS-OCT), vorgestellt worden. Auch ihr optischer Messaufbau ähnelt prinzipiell dem der TD-OCT. Bei der SS-OCT wird anstelle einer breitbandigen nun eine monochromatische Lichtquelle verwendet, deren Wellenlänge zügig über einen möglichst weiten Bereich durchgestimmt werden kann (siehe Abb. 2.2, oben rechts). Anders als bei der spektrometerbasierten SD-OCT, wird das Interferenzspektrum nicht räumlich, sondern zeitlich erfasst, wodurch ein Punktdetektor zur Messung des Signals genügt. Allerdings muss dieser auf die hohe Durchstimmrate der Lichtquelle angepasst sein. Die Transformation des Interferenzspektrums erfolgt analog zur SD-OCT durch die Fourier-Transformation.

2.3 Axiale und laterale Auflösung

Ein besonderes Merkmal der OCT ist, dass ihr axiales und laterales Auflösungsvermögen voneinander unabhängig sind. Während die axiale Auflösung von der gewählten Lichtquelle und der Signalabtastung am Detektor abhängt, wird die laterale Auflösung durch die Fokussierungsoptik im Probenarm bestimmt. Die axiale Auflösung ist umso höher, je größer die Bandbreite der Lichtquelle ist. Die axiale Messtiefe definiert denjenigen Bereich, der von einem A-Scan abgedeckt werden kann. Bei der SD-OCT bspw. wird sie durch die spektrale Auflösung begrenzt. Das laterale Auflösungsvermögen und Sichtfeld (*engl.*: field of view, kurz: FOV) hängen zum einen von der Fokussierungsoptik und zum anderen vom Abtastsystem des OCT-Systems ab. Die laterale Auflösung ist durch die Spotgröße des Lichtstrahls gegeben, die wiederum von der numerischen Apertur der Fokussierlinse abhängt. Eine schematische Darstellung einer OCT-Volumenaufnahme ist in Abbildung 2.3 illustriert.

2.4 Medizinische Anwendungsfelder

Die Ophthalmologie ist bis heute das klinische Hauptanwendungsfeld der OCT [Fujimoto u. a., 2016]. Dies liegt vorrangig an der hohen Transparenz der okulären Medien, wodurch der Augenhintergrund optisch einfach zugänglich ist. Seit ihrer Einführung in den frühen 1990er Jahren hat die OCT die Diagnose und Therapie in der Augenheilkunde revolutioniert, da sie im Gegensatz zu anderen bildgebenden Verfahren wie der Spaltlampe oder der Funduskamera die nichtinvasive Aufnahme von Querschnittsbildern des Augenhintergrunds des Menschen in Mikrometerauflösung ermöglicht. Aufgrund ihrer hohen Empfindlichkeit lassen sich feine Gewebestrukturen detailliert darstellen, was eine genaue Analyse erlaubt (siehe Abb. 2.4). Somit stellt die OCT ein bedeutendes Diagnoseinstrument für die Erkennung und Verlaufskontrolle von Augenkrankheiten dar und ermöglicht die Behandlung von Erkrankungen bereits in einem frühen Stadium. Die OCT wird vorrangig zur Unterstützung der Diagnose von Reti-



Eigene Darstellung nach [Bille, 2019, S. 67]

Abb. 2.3: Schematische Darstellung der Aufnahme eines OCT-Volumens.

naerkrankungen, wie z. B. der AMD, eingesetzt [Adhi u. a., 2013]. Mit der OCT können auch neurologische Krankheiten erkannt werden. So kann mit ihr eine Ausdünnung der Retinaschichten bei Multipler Sklerose beobachtet werden [Schippling u. a., 2015].

Neben der Abbildung der funktionellen Retina lassen sich mit der OCT Untersuchungen des Sehnervenkopfes durchführen, die eine Diagnostik in der Glaukomtherapie (grüner Star) ermöglichen [Medeiros u. a., 2004]. Hierzu werden zirkuläre Querschnittsbilder des Sehnervenkopfes erstellt, die Aufschluss über den Zustand der Nervenfaserschichten liefern. Auch der vordere Augenabschnitt lässt sich durch die OCT abbilden. Die Pachymetrie erlaubt Dickenmessungen der Hornhaut, die wichtige Kennzahlen für refraktive Eingriffe darstellen [Li u. a., 2006b]. Mit der Gonioskopie etwa können die Kammerwinkel, welche die Verbindungsstellen von Hornhaut und Iris sind, gemessen werden. Eine Veränderung der Kammerwinkel kann eine Ursache für einen erhöhten Augeninnendruck sein, der unbehandelt zum Glaukom führen kann [Sakata u. a., 2008].

Für eine optimierte Bestimmung des Behandlungszeitpunkts bei der Injektionstherapie der feuchten AMD ist eine engmaschige Überwachung des Krankheitszustands mittels OCT erforderlich. Dies hat die Entwicklung kostengünstiger und kompakter OCT-Geräte für die eigenständige Heimanwendung motiviert [Chopra u. a., 2021]. Zum einen kann dadurch die Verfügbarkeit von OCT-Systemen drastisch erhöht werden, zum anderen entlastet die Selbstmessung zu Hause den Patienten und das Klinikpersonal. Ein Prototyp eines kostengünstigen OCT-Systems [Sudkamp u. a., 2016] für die Heimanwendung zeigt Abbildung 2.5.

Weiterhin wird die OCT in der Dermatologie zur Bildgebung der oberen Hautschichten (siehe Abb. 2.6) eingesetzt, wo sie als nichtinvasive optische Biopsie fungiert [Gambichler u. a., 2007]. Im Unterschied zum Auge hat die Haut jedoch eine sehr inhomogene

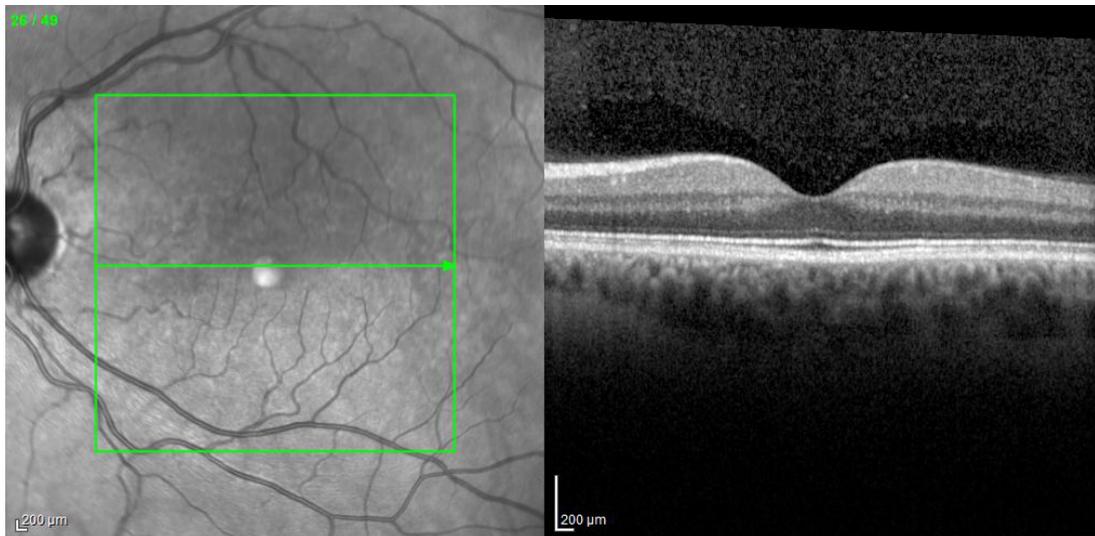


Abb. 2.4: SD-OCT des zentralen Makulabereichs der Retina (rechts). Durch eine En-face-Übersichtsaufnahme des Augenhintergrunds (links) kann das FOV (grünes Rechteck) der OCT-Aufnahme sowie die Position des B-Scans überlagert dargestellt werden. Die Aufnahme wurde mit dem Spectralis OCT-Scanner der Firma Heidelberg Engineering GmbH erstellt.

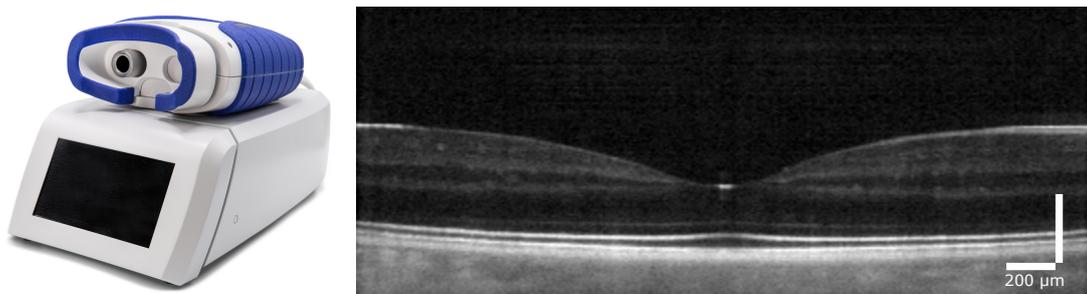


Abb. 2.5: OCT-Diagnostik zu Hause. Links: Prototyp eines Home-OCT-Systems der Firma Visotec GmbH, das eine Selbstvermessung zu Hause ermöglicht. Über ein handgehaltenes Modul kann auf einfache Weise ein volumetrisches OCT-Bild aufgenommen werden. Die einzelnen Schritte einer Messung werden an der Basisstation visuell über ein Display und per Tonausgabe veranschaulicht. Rechts: Ein B-Scan aus der volumetrischen OCT-Aufnahme einer gesunden Retina.

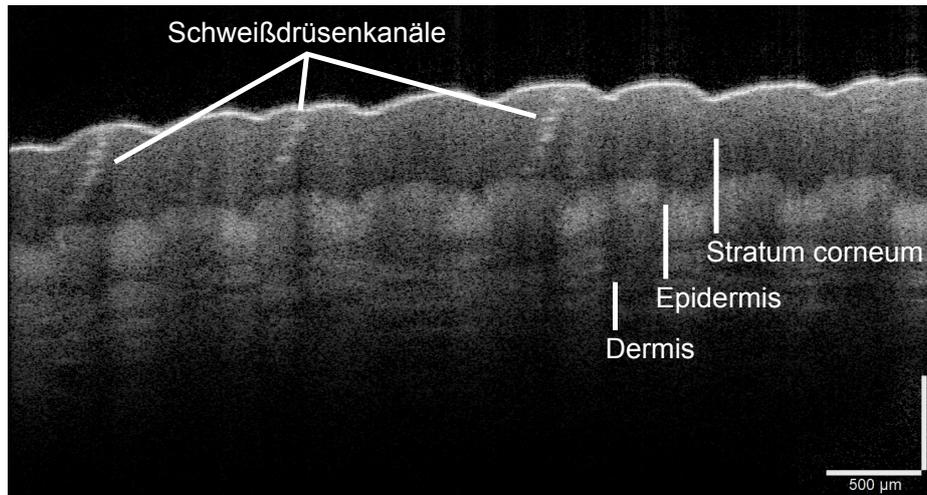


Abb. 2.6: OCT-Aufnahme der Fingerkuppe, welche mit dem OCT-System Telesto III der Firma Thorlabs GmbH erstellt wurde. Das Stratum corneum (Hornhaut), die Epidermis und die Dermis lassen sich gut voneinander unterscheiden. Es können auch mehrere spiralförmige Schweißdrüsenkanäle abgegrenzt werden.

und komplexe Beschaffenheit, die zu einer starken Streuung des Lichts führt [Welzel u. a., 1997]. Dermatologische OCT-Geräte arbeiten daher bei längeren Wellenlängen, bei denen die Gewebestreuung reduziert ist. Klinische Studien haben gezeigt, dass die OCT für die Diagnose von entzündlichen und blasenbildenden Hautkrankheiten sowie für die Differenzialdiagnose einiger Tumore hilfreich sein kann [Welzel, 2001]. Im Laufe von drei Jahrzehnten wurde die OCT-Technologie kontinuierlich optimiert und wird heute in zahlreichen dermatologischen Kliniken und Praxen routinemäßig zur Diagnose und Verlaufskontrolle von epithelialen Hauttumoren wie dem Basalzellkarzinom [Welzel u. a., 2017] eingesetzt. Überdies können mit der OCT funktionelle Parameter der Haut bestimmt werden. So lassen sich etwa Aussagen über die Rauheit oder Elastizität der Haut treffen, was für kosmetische Anwendungen interessant ist [Gambichler u. a., 2005]. Daher wird die OCT vermehrt in (prä-)klinischen Studien sowohl am Menschen als auch an Tiermodellen eingesetzt [Salma u. a., 2018]. Trotz des nachgewiesenen Nutzens werden in der klinischen Praxis die meisten Hautkrankheiten aus Kosten- und Zeitgründen mit bloßem Auge oder unter Zuhilfenahme der Auflichtmikroskopie diagnostiziert. Die konventionelle, invasive Biopsie in Kombination mit der histologischen Auswertung zur Differenzialdiagnose von malignem Gewebe stellt nach wie vor den Goldstandard für Tumorerkrankungen oder unklare Befunde dar [Heibel u. a., 2020].

Eine weitere Anwendung der OCT ist die Bildgebung von Hohlorganen. Über Glasfaserkomponenten kann ein OCT-System mit einem Endoskop kombiniert und so zur Unterstützung minimalinvasiver Verfahren eingesetzt werden. Durch die OCT ist es z. B. möglich, strukturelle Oberflächenveränderungen des Gastrointestinaltrakts zu un-

tersuchen [Jang u. a., 2001], was insbesondere für die Diagnose von Tumoren wie dem Barrett-Ösophagus bedeutsam ist [Bouma u. a., 2000]. Eine Übersicht weiterer Anwendungsfelder der OCT ist in Fercher [2010] gegeben.

Kapitel 3

Methodische Grundlagen

Die in dieser Arbeit entwickelten Methoden fußen auf einer Reihe verschiedener methodischer Grundlagen, die in diesem Kapitel vorgestellt werden. Die Analyse und Erkennung in OCT-Bilddaten beinhaltet unter anderem die Extraktion diagnostisch relevanter Bildregionen, was in der vorliegenden Arbeit durch den Einsatz von Bildsegmentierungsverfahren (Kapitel 3.1) erreicht wird. Den wesentlichen Kern dieser Verfahren bilden Methoden des maschinellen Lernens (Kapitel 3.2). Neben klassischeren Methoden des maschinellen Lernens, wie z. B. Entscheidungsbäumen (Kapitel 3.2.1), basiert der Hauptteil dieser Arbeit auf Faltungsnetzwerken, deren Funktionsweise in Kapitel 3.2.2 vorgestellt wird.

3.1 Bildsegmentierung

Durch die immer schneller fortschreitende Digitalisierung ist die Entwicklung computergestützter Bildauswertungsalgorithmen von großer Bedeutung. Die Segmentierung stellt dabei ein wichtiges Werkzeug für die automatische Erkennung und Vermessung von Objekten in Bildern dar und findet ihren Einsatz in zahlreichen Anwendungsbereichen. Besonders in der Medizin liefern Segmentierungen wertvolle Informationen über Form und Größe diagnostisch relevanter Bildobjekte, die zur quantitativen Auswertung genutzt werden können. Bezogen auf medizinische Bilddaten ist das Ziel einer Segmentierung die Abgrenzung anatomischer oder pathologischer Strukturen, welche eine Relevanz in der Diagnostik und Therapie bilden. Eine Segmentierung verschiedener Schichten der Retina ist hierzu beispielhaft in Abbildung 3.1 abgebildet.

Ein digitales Bild I besteht aus diskreten Pixeln oder Voxeln, denen ein Intensitätswert zugeordnet wird: $I : \mathbf{x} \in \Omega \mapsto \mathbb{R}$. Dabei beschreibt \mathbf{x} die Bildposition (Pixel/Voxel) einer Bilddomäne $\Omega \subseteq \mathbb{R}^d$ mit der Dimension d . Die Anzahl der Bildkanäle wird durch n (z. B. $n = 1$ bei Grauwertbildern und $n = 3$ bei Farbbildern) angegeben. Die Segmentierung eines Bildes I ist über die Abbildungsfunktion $S : \Omega \mapsto L$ definiert und weist jeder Bildposition ein Klassenlabel l zu mit $l \in L = \{0, \dots, K - 1\}$ für K unterschiedliche Klassen. Generell spricht man von einer binären Segmentierung für $K = 2$ und von einer Multi-Label-Segmentierung für $K > 2$.

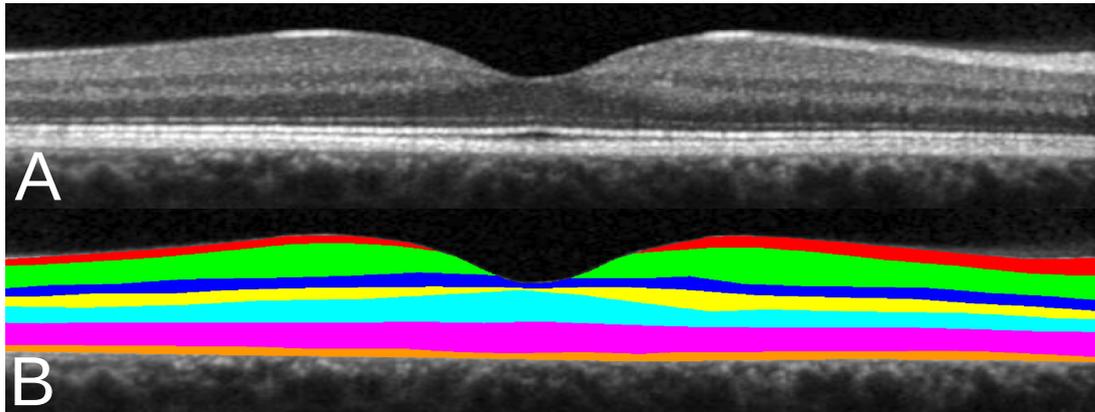


Abb. 3.1: OCT-B-Scan einer Retina (A) mit überlagertem Segmentierung (B) der einzelnen Retinaschichten.

In den vergangenen Jahrzehnten sind zahlreiche halb- und vollautomatische Segmentierungsalgorithmen entwickelt worden. Insgesamt kann zwischen punkt-, regionen-, kanten- und modellbasierten Segmentierungsverfahren unterschieden werden [Handels, 2009; Jähne, 2013], deren prinzipielle Funktionsweisen anhand kurzer Beispiele in den nachfolgenden Paragraphen vorgestellt werden. Im Anschluss daran wird die Funktionsweise der Graph-Cut-Segmentierung ausführlicher beschrieben, die Bestandteil einer in Kapitel 4 vorgestellten Methodik ist. In [Handels, 2009] wird ergänzend zwischen clusteranalytischen und atlas- sowie klassifikatorbasierten Verfahren unterschieden. Letztere bilden den Schwerpunkt der in dieser Arbeit vorgestellten Segmentierungsalgorithmen, deren methodisches Grundprinzip in Kapitel 3.2 näher beschrieben.

Punktbasierte Verfahren: Die punktbasierete Segmentierung zählt zu den einfachsten methodischen Verfahren. Bei diesen Ansätzen wird pro Bildpunkt über eine Klassenzuweisung entschieden. Das bekannteste und einfachste Verfahren stellt dabei die Schwellenwertsegmentierung dar. Dabei werden die Grauwerte anhand eines definierten Schwellenwertes im Histogramm des Bildes in zwei Bereiche unterteilt, aus denen sich die Unterteilung in Objekt- und Hintergrundklasse ergibt. Durch die Verwendung mehrerer Schwellenwerte kann einerseits die Objekt- und Hintergrundklasse präzisiert und andererseits eine Multi-Label-Segmentierung vorgenommen werden.

Regionenbasierte Verfahren: Im Gegensatz zur punktbasiereten Segmentierung berücksichtigen regionenbasierte Methoden Bildpunkte in Bezug zu ihrer Umgebung. Dies ist ein entscheidender Vorteil gegenüber punktbasiereten Verfahren, da charakteristische Merkmale eines Bildobjekts sich durch die Betrachtung zusammenhängender Bildpunkte ergeben. Als Beispiel ist hier das Regionenwachstumsverfahren (*engl.*: region growing) zu nennen, das meist zur halbautomatischen Segmentierung genutzt wird.

Ausgehend von einem manuell gesetzten Saatpunkt innerhalb des zu segmentierenden Bildobjekts werden iterativ benachbarte Pixel zum Segment hinzugefügt, sofern sie ein definiertes Homogenitätskriterium erfüllen. Bereits erfasste Pixel werden als „bearbeitet“ markiert. Können dem Segment keine weiteren Pixel hinzugefügt werden, terminiert der Algorithmus.

Kantenbasierte Verfahren: Mithilfe kantenbasierter Verfahren werden Bildobjekte segmentiert, indem ihre äußere Kontur identifiziert wird. Dazu werden diejenigen Pixel bestimmt, welche auf einer Kante im Bild liegen, z. B. durch Berechnung der Ableitung erster oder zweiter Ordnung. Wesentlicher Vorteil kantenbasierter Verfahren ist ihre höhere Robustheit gegenüber Kontrastschwankungen. Als Beispiel ist hier die Live-Wire-Segmentierung zu nennen, eine halbautomatische Methode zur interaktiven Konturierung von Bildobjekten. Beginnend von einem initialen Saatpunkt wird in Echtzeit eine verbindende Linie zum Mauszeiger generiert und angezeigt. Per Mausklick kann der Benutzer das vorgeschlagene Konturstück bestätigen und so sukzessive eine Konturierung vornehmen. Für die Berechnung der vorgeschlagenen Konturstücke kommen graphentheoretische Methoden zum Einsatz, deren Kostenfunktionen auf Kanteninformatoren basieren.

Modellbasierte Verfahren: Die bisher vorgestellten Verfahren betrachten lediglich lokale Bildinformation zu Segmentierung. Im Gegensatz dazu können mithilfe modellbasierter Verfahren spezifische Eigenschaften des Bildobjekts als Vorwissen, sogenanntes A-priori-Wissen, in die Segmentierung integriert werden, wie z. B. die geometrische Form bei Organen. So eignen sich modellbasierte Methoden gut für Segmentierungsprobleme, bei denen die Bildqualität gering ist oder Teile des Bildobjekts verdeckt sind. Zu den bekanntesten Methoden gehören statistische Formmodelle (*engl.*: statistical shape models), die vorwiegend zur Segmentierung anatomischer Strukturen genutzt werden. Zunächst werden anhand segmentierter Trainingsbilddaten Oberflächenmodelle des Bildobjekts erstellt. Anschließend werden die Oberflächenmodelle mittels Korrespondenzanalyse geometrisch aufeinander angepasst. Durch eine Hauptkomponentenanalyse kann dann ein lineares Formvariabilitätsmodell der Trainingsdaten erstellt werden. Das Modell lässt sich später mittels iterativer Optimierungsprozesse auf neue Bilddaten anpassen.

3.1.1 Graph-Cut-Segmentierung

Graph-Cut-Methoden haben ihren Ursprung in der Graphentheorie und werden z. B. zur Optimierung von Computernetzwerken oder Logistikproblemen genutzt. Sehr populär sind Graph-Cut-Methoden im Bereich der Computer Vision, wo sie vorwiegend zur Segmentierung von Bildern eingesetzt werden. Daneben finden Graph-Cut-

Methoden Anwendung bei der Lösung von Stereo-Matching-Problemen oder bei der Bildentrauschung [Boykov u. a., 2004].

Graph-Cut-Methoden bilden eine weitere Gruppe von Segmentierungsverfahren, bei denen ein Bild durch einen Graphen beschrieben wird. Dabei repräsentieren Bildpixel die Knoten des Graphen. Benachbarte Pixel werden durch Kanten im Graphen miteinander verbunden. Zusätzlich wird den Kanten eine Gewichtung zugewiesen, die durch eine Kostenfunktion definiert ist und z. B. Intensitätsunterschiede zwischen Pixeln berücksichtigt. Damit eine Segmentierung möglichst entlang der Objektkontur verläuft, erhalten Kanten mit hohem Intensitätsunterschied eine entsprechend niedrige Gewichtung durch die Kostenfunktion. Das Ziel der Graph-Cut-Segmentierung ist es, einen kostengünstigen Schnitt (*engl.*: cut) durch die Kanten des Graphen zu finden, der die Bildpixel in ein Objekt- und Hintergrundsegment aufteilt. Die Ermittlung des optimalen Schnitts entspricht der globalen Suche nach Kanten mit minimaler Kostensumme im Graphen und lässt sich mit der Max-Flow-Min-Cut-Methode von Ford und Fulkerson [Ford u. a., 1962] durchführen.

Im Vergleich zu den zuvor beschriebenen Gruppen von Segmentierungsverfahren (siehe Kapitel 3.1) bieten Graph Cuts einige Vorteile. Aufgrund der Tatsache, dass der minimale Schnitt über den gesamten Graphen gesucht wird, ist ein globales Optimum garantiert. Diese Eigenschaft macht die Graph-Cut-Segmentierung weniger anfällig für Rauschen oder andere Bildartefakte als Verfahren, die nur lokale Nachbarschaften betrachten. Zur Lösung des Segmentierungsproblems unter Verwendung von Graph Cuts wird in der Regel ein Energiefunktional minimiert, in das sich sowohl Kanten- als auch Regioneninformationen gleichermaßen einbeziehen lassen. Es können auch geometrische Eigenschaften im Graphenmodell berücksichtigt werden, um topologische Strukturbeziehungen zu erhalten [Boykov u. a., 2006b]. Ein Beispiel hierfür ist die Segmentierung von Retina- oder Hautschichten (siehe Kapitel 4 und 5), deren Anordnung einer bestimmten Topologie folgen. Ein weiterer wesentlicher Vorteil ist die Erweiterbarkeit des Graphen für höhere Dimensionen. Somit ist die Graph-Cut-Segmentierung nicht auf 2D-Bilddaten beschränkt und kann für 3D-Volumenbilddaten oder räumlich-zeitliche 4D-Bilddaten angepasst werden. Weiterhin können, ähnlich wie bei der Region-Growing-Methode, mehrere Bildpixel als Startpunkte markiert werden, die zur Bestimmung von Intensitätsverteilungen der Objekt- und der Hintergrundklasse verwendet werden und in die Kostenfunktion integriert werden können. Dabei lassen sich die Saatpunkte automatisch durch einen Algorithmus oder manuell setzen. Letzteres erlaubt ein interaktives halbautomatisches Segmentierungsverfahren, welches lediglich das Setzen mehrerer kleiner Markierungen (*engl.*: scribbles) der entsprechenden Klassen erfordert [Boykov u. a., 2001].

Eine erste Anwendung von Graph Cuts ist in der Arbeit von Greig u. a. [1989] präsentiert worden, in der sie zur Restaurierung von Binärbilddaten genutzt werden. Später haben Boykov u. a. [2004] eine weitaus effizientere Graph-Cut-Methode vorgestellt, die im Computer-Vision-Bereich durch ihre vielfältige Einsetzbarkeit sehr po-

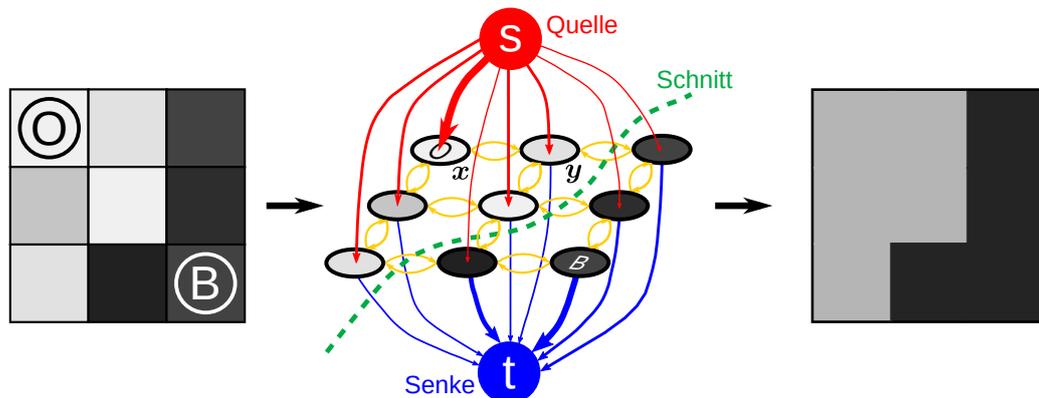
pulär geworden ist. In darauffolgenden Arbeiten konnte die Effizienz der Graph-Cut-Methoden weiter gesteigert werden. Nachfolgend werden die wesentlichen Bausteine der Graph-Cut-Segmentierung anhand eines Zwei-Klassen-Problems beschrieben. Diese setzen sich aus drei Teilen zusammen: dem Aufbau des Graphen (Kapitel 3.1.1.1), dem Max-Flow-Min-Cut-Theorem (Kapitel 3.1.1.2) sowie der Kostenfunktion (Kapitel 3.1.1.3).

3.1.1.1 Aufbau des Graphen

Die Basis der Graph-Cut-Segmentierung bildet der Graph, der anhand des zu segmentierenden Bildes konstruiert wird. Ein Graph $G = (V, E)$ besteht aus einer Menge von Knoten $v \in V$ und einer Menge von Kanten $e \in E$, die sie miteinander verbinden. Die Knoten im Graph repräsentieren die Positionen \mathbf{x} der Pixel oder Voxel des Bildes. Der Graph enthält zwei weitere Knoten $s \in V$ und $t \in V$, die auch als Terminalknoten bezeichnet werden. In der Regel wird s als Quelle (*engl.*: source) und t als Senke (*engl.*: sink) benannt, durch die eine Zuweisung der übrigen Knoten in die Objekt- oder Hintergrundklasse vorgenommen wird. Benachbarte Pixel sind im Graphen durch Kanten (\mathbf{x}, \mathbf{y}) , auch n-Links (*engl.*: neighborhood links) genannt, verbunden. Hierbei kann ein beliebiges Nachbarschaftssystem gewählt werden. Diesen Kanten werden nichtnegative Kosten oder Gewichte w_e zugewiesen, die auf dem lokalen Intensitätsgradienten, den Laplace-Nulldurchgängen, der Gradientenrichtung oder anderen Kriterien beruhen [Boykov u. a., 2004]. Die n-Links können auch durch gerichtete Kanten in einem Digraphen repräsentiert werden. In diesem Fall können sich die Kosten der umgekehrten Kante (\mathbf{y}, \mathbf{x}) von (\mathbf{x}, \mathbf{y}) unterscheiden, wie z. B. bei Verwendung der Gradientenrichtung. Boykov u. a. [2006a] konnten zeigen, dass der Gebrauch gerichteter Kanten zu besseren Segmentierungsergebnissen führt. Jeder Bildpixel wird jeweils durch Kanten, auch t-Links (*engl.*: terminal links) genannt, mit den Terminalknoten s und t verbunden. Ebenso wie den n-Links werden den t-Links nichtnegative Kosten zugewiesen. Diejenigen t-Links, welche gesetzte Saatpunkte mit dem entsprechenden Terminalknoten verbinden, erhalten unendliche Kosten. Mithilfe dieser harten Beschränkung wird sichergestellt, dass Saatpunkte durch einen Schnitt nicht von ihren Terminals getrennt werden. Die Konstruktion des Graphen ist damit abgeschlossen. Eine schematische Darstellung eines einfachen Graphenmodells ist in Abbildung 3.2 gegeben.

3.1.1.2 Maximaler Fluss und minimaler Schnitt

Um nun eine (binäre) Segmentierung erhalten zu können, wird ein Schnitt $C \subseteq E$ bestimmt, welcher den Graphen in zwei disjunkte Mengen auftrennt, welche die Objekt- und Hintergrundklasse repräsentieren. Ein Schnitt wird auch als s-t-Schnitt bezeichnet, da durch ihn die beiden Terminalknoten voneinander getrennt werden. Eine globale und effiziente Berechnung des optimalen Schnitts entspricht der Suche nach Kanten,



Eigene Darstellung nach [Boykov u. a., 2001]

Abb. 3.2: Einfaches zweidimensionales Segmentierungsbeispiel für ein 3×3 -Bild (links). Die Saatpunkte der Objekt- und Hintergrundklassen sind durch ein „O“ und ein „B“ gekennzeichnet. Der entsprechende gerichtete und gewichtete Graph ist in der Mitte dargestellt. Die Kosten jeder Kante spiegeln sich in ihrer Dicke wider. Der regionenbasierte Term und die harten Beschränkungen durch die Saatpunkte definieren die Kosten der t-Links (rot/blau). Der kantenbasierte Term bestimmt die Kosten für die n-Links (gelb). Kostengünstige Kanten stellen eine attraktive Wahl für einen minimalen Schnitt dar, durch den die endgültige Segmentierung (rechts) erzeugt wird.

deren aufsummierte Kosten $|C| = \sum_{e \in C} w_e$ minimal sind. In [Ford u. a., 1962] wird gezeigt, dass die Suche nach dem minimalen Schnitt der Bestimmung des maximalen Flusses von s zu t entspricht.

Sinnbildlich kann der maximale Fluss als eine Wassermenge interpretiert werden, die von der Quelle s zur Senke t transportiert wird. Die Kanten des Graphen können als (gerichtete) Rohre verstanden werden, deren Durchmesser dem jeweiligen Kantengewicht entspricht, woraus sich die Durchflusskapazität ergibt. Ein Fluss $f : V \times V \rightarrow \mathbb{R}$ weist jeder Kante e einen Flusswert zu und kann als eine Menge von Wasser verstanden werden, die pro Zeiteinheit fließen kann. Es gelten mehrere Regeln für einen gültigen Fluss. In einen normalen Knoten darf nicht mehr hereinfließen als hinaus, was auch als Kirchhoffsches Gesetz bekannt ist. Dagegen können Terminalknoten als Sonderfälle Flüsse erzeugen und aufnehmen, worauf auch ihre Namensgebung beruht. Des Weiteren darf ein Fluss nicht die Kapazität der Kanten überschreiten. Eine weitere wichtige Eigenschaft des Flusses ist seine Asymmetrie, nach der jeder umgekehrten Kante (y, x) der jeweilige negative Wert zugeordnet wird.

Die bereits angesprochene Methode von Ford und Fulkerson [Ford u. a., 1962] erlaubt die Lösung des Maximalfuss-Problems in polynomialer Zeit. Dabei wird ein Pfad im Graphen G zwischen s und t gesucht, ohne dabei die zuvor erwähnten Kapazitätsbeschränkungen zu verletzen. Ein solcher Pfad wird auch als augmentierender

Pfad bezeichnet, wenn dieser den Gesamtfluss vergrößert. Dieser Prozess wird iterativ wiederholt, wobei die Flüsse der einzelnen Pfade zu einem größeren Gesamtfluss zusammengefasst werden. Die Methode terminiert, sobald keine Restkapazitäten vorhanden sind und somit kein weiterer Fluss von s zu t möglich ist. Wird bei der Suche stets der kürzeste Pfad im Graphen gewählt, entspricht dies der Edmonds-Karp-Methode [Edmonds u. a., 1972]. Insgesamt gibt es zahlreiche Max-Flow-Min-Cut-Algorithmen, wie die von Goldberg u. a. [1988] vorgeschlagene Push-Relabel-Methode. Mit der Weiterentwicklung der Ford-Fulkerson-Methode durch Boykov und Kolmogorov [Boykov u. a., 2004] konnte die Laufzeit erheblich verkürzt werden und ist daher im Bereich der Computer Vision sehr beliebt geworden.

3.1.1.3 Kostenfunktion

Die Gestaltung der Kostenfunktion hat einen erheblichen Einfluss auf die Qualität der Segmentierung. Bei Graph-Cut-Methoden wird meist auf Regionen- oder Kanteninformationen zurückgegriffen. So sollen Bildregionen mit einer geringen Diskontinuität dasselbe Klassenlabel zugewiesen bekommen. Dagegen sollen Objektkonturen, die im Gegensatz eine hohe Diskontinuität aufweisen, erhalten bleiben. Bei der Graph-Cut-Segmentierung kommt meist ein Energieterm als Kostenfunktion zum Einsatz:

$$\mathcal{E}(S) = \sum_{\mathbf{x} \in \Omega} R(S(\mathbf{x})) + \sum_{\{\mathbf{x}, \mathbf{y}\} \in \mathcal{N}} B(\mathbf{x}, \mathbf{y}) \cdot |S(\mathbf{x}) - S(\mathbf{y})|. \quad (3.1)$$

Hierbei ist S eine Segmentierung, die jedem Pixel $\mathbf{x} \in \Omega$ ein Klassenlabel zuweist. Der regionenbasierte Term $R(\cdot)$ liefert die Fehlerwahrscheinlichkeit hinsichtlich der Zuweisung eines Klassenlabels für einen Pixel und lässt sich anhand von Histogrammen der Intensitätsverteilungen für Objekt und Hintergrund beschreiben, die aus den Saatpunkten abgeleitet werden können. Die regionale Fehlerabschätzung kann dann durch die negativen Logarithmuswahrscheinlichkeiten, z. B. $R(\text{„Obj“}) = -\ln \Pr(I(\mathbf{x}) | \text{„Obj“})$ für die Objektklasse bestimmt werden. Durch den kantenbasierten Term $B(\mathbf{x}, \mathbf{y})$ erfolgt eine Regularisierung des Nachbarschaftssystems \mathcal{N} , indem ähnlichen Pixeln dasselbe Klassenlabel zugewiesen wird. In [Boykov u. a., 2001] wird für den kantenbasierten Term die Funktion $B(\mathbf{x}, \mathbf{y}) \propto \exp(-(I(\mathbf{x}) - I(\mathbf{y}))^2 / 2\sigma^2) / \text{dist}(\mathbf{x}, \mathbf{y})$ vorgeschlagen, welche geringere Diskontinuitäten hart bestraft. Hohe Diskontinuitäten, die möglichen Objektkonturen entsprechen, werden dagegen weniger stark bestraft. Dementsprechend beziehen die t-Links ihre Kosten aus dem regionenbasierten Term und die n-Links ihre aus dem kantenbasierten Term. Eine Übersicht einer möglichen Kostenzuweisung für die Kanten E ist in [Boykov u. a., 2001] gegeben. Grundsätzlich lassen sich beliebige Kontextinformationen in die Kostenfunktion einbeziehen, die als sogenannte „weiche Beschränkungen“ (*engl.*: soft constraints) bezeichnet werden [Boykov u. a., 2004]. Saatpunkte hingegen werden als „harte Beschränkungen“ betrachtet, da sie eine eindeutig richtige Klassenzuordnung widerspiegeln.

3.1.2 Evaluationsmetriken

Metriken ermöglichen eine quantitative Auswertung von Segmentierungsergebnissen, anhand derer sich verschiedene Algorithmen objektiv miteinander vergleichen lassen. Eine wesentliche Voraussetzung dafür ist das Vorhandensein einer exakten Segmentierung, d. h. einer Grundwahrheit, die in der medizinischen Bildgebung häufig durch eine manuelle Segmentierung eines erfahrenen Experten erzeugt werden muss. Neben der Bewertung von Algorithmen können Metriken genutzt werden, um die Konsistenz und Zuverlässigkeit von manuell erstellten Segmentierungen eines einzelnen (Intra-Rater-Reliabilität) oder zwischen mehreren verschiedenen Experten (Inter-Rater-Reliabilität) zu untersuchen. Im Folgenden werden drei verschiedene Metriken vorgestellt, die für die quantitative Bewertung der Segmentierungsergebnisse in dieser Arbeit verwendet werden.

Eine sehr häufig verwendete Segmentierungsmetrik ist der Dice-Ähnlichkeitskoeffizient (*engl.*: Dice similarity coefficient, kurz: Dice oder DSC), welcher die Überlappung zwischen zwei Punktmengen A und B misst. Er wird definiert durch:

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

Der DSC kann Werte aus dem Bereich $[0,1]$ annehmen. Je höher die Übereinstimmung der beiden Punktmengen ist, desto weiter nähert sich der DSC dem Wert 1 an.

Die Berechnung des DSC berücksichtigt lediglich die Schnittmenge zweier Punktmengen, jedoch nicht räumliche Distanzen zwischen Oberflächen. Außerdem beeinflusst die Größe des segmentierten Objekts den DSC-Wert, da sich bspw. falsch klassifizierte Pixel auf kleinere Bereiche stärker auswirken als auf größere. Daher werden in dieser Arbeit zwei zusätzliche oberflächenbasierte Metriken zur quantitativen Auswertung herangezogen. Seien \mathcal{S}_A und \mathcal{S}_B Vektoren von Konturkoordinaten der Punktmengen A und B mit den Längen n_A und n_B . Die Oberflächendistanz (*engl.*: surface distance, kurz: SD) wird bestimmt durch:

$$\text{SD}(\mathcal{S}_A, \mathcal{S}_B) = \sum_{j=0}^{n_B-1} \min_{0 \leq i < n_A} \|\mathcal{S}_{A_i} - \mathcal{S}_{B_j}\|_2. \quad (3.3)$$

Der Oberflächenabstand hängt von der Anzahl der Konturpunkte in \mathcal{S}_B ab, weswegen der Mittelwert (*engl.*: average surface distance, kurz: ASD) gebildet wird:

$$\text{ASD}(\mathcal{S}_A, \mathcal{S}_B) = \frac{1}{n_B} \text{SD}(\mathcal{S}_A, \mathcal{S}_B). \quad (3.4)$$

Jedoch ist die ASD nicht symmetrisch, d. h. es gilt $\text{ASD}(\mathcal{S}_A, \mathcal{S}_B) \neq \text{ASD}(\mathcal{S}_B, \mathcal{S}_A)$, wodurch sie sich nicht als Metrik eignet. Deswegen wird die symmetrische ASD (ASSD) genutzt:

$$\text{ASSD} = \frac{\text{ASD}(\mathcal{S}_A, \mathcal{S}_B) + \text{ASD}(\mathcal{S}_B, \mathcal{S}_A)}{2}. \quad (3.5)$$

Schließlich wird als dritte Metrik die Hausdorff-Distanz (HD) zur Detektion von Segmentierungsausreißern angewandt. Die HD misst den maximalen Abstand zwischen zwei Segmentierungskonturen und wird definiert durch:

$$\text{HD} = \max [\text{SD}(\mathcal{S}_A, \mathcal{S}_B), \text{SD}(\mathcal{S}_B, \mathcal{S}_A)]. \quad (3.6)$$

3.2 Maschinelles Lernen

Maschinelles Lernen (*engl.*: machine learning) ist ein Teilgebiet der künstlichen Intelligenz (KI) und beschreibt Algorithmen, die Muster und Zusammenhänge in vorhandenen Datensätzen erkennen können. Im Zuge der stetig fortschreitenden Digitalisierung finden immer mehr maschinelle Lernalgorithmen ihren Weg in verschiedenste Computeranwendungen wie der Sprach- und Texterkennung oder der Produktempfehlung in Online-Warenhäusern. Durch die heutige einfache Verfügbarkeit und Generierung von Bilddaten, die durch Sensoren und dem Internet gesammelt werden, erhalten maschinelle Lernmethoden im Computer-Vision-Bereich großen Popularitätszuwachs [Khan u. a., 2020]. Beispiele hierfür können die Analyse von Straßenszenen für das autonome Fahren oder die Erkennung von Melanomen beim Hautkrebs-Screening sein.

Auf der Grundlage des erlernten Wissens sind diese Algorithmen in der Lage, Probleme zu lösen, ohne dafür explizite Anweisungen zu verwenden. Hierzu wird zunächst in einer Trainingsphase unter Verwendung eines ausgewählten Datensatzes ein statistisches Modell erstellt, das in einer anschließenden Testphase zur Auswertung neuer, ungesehener Daten verwendet wird. Eine entscheidende Rolle spielt dabei die Merkmalsextraktion, welche die Eingabedaten auf die notwendigsten Informationen reduziert. Anhand dieser Dimensionsreduktion lassen sich die Eingabedaten besser vergleichen und redundante Information wird entfernt. Je nach Wahl des Lernalgorithmus werden die Merkmale manuell oder automatisch erstellt. Zudem spielt die Generalisierung des Modells eine entscheidende Rolle. Werden einerseits die Trainingsdaten durch das Modell zu detailliert beschrieben oder gar auswendig gelernt, ist es schwierig, sie auf neue Daten zu übertragen, was als Überanpassung (*engl.*: overfitting) bezeichnet wird. Kann ein Modell andererseits die tatsächliche Komplexität der zugrunde liegenden Trainings- oder Testdaten nicht erfassen, so spricht man von einer Unteranpassung (*engl.*: underfitting).

Generell lassen sich maschinelle Lernmethoden in drei Verfahrensklassen unterscheiden: überwachtes Lernen, unüberwachtes Lernen und bestärkendes Lernen. Beim überwachten Lernen (*engl.*: supervised learning) stehen dem Lernalgorithmus Trainingsdaten zur Verfügung, bei denen die jeweilige Zielvariable bereits bekannt ist. Man spricht hierbei auch von gelabelten Daten oder einer Grundwahrheit. Während der Trainingsphase versucht der Lernalgorithmus, seine Vorhersage an die gegebene Zielvariable anzunähern, um den Prädiktionsfehler zu verringern. Dabei nimmt die Zielvariable einen kategorischen oder einen kontinuierlichen Wert an, was einem Klassifikations-

oder Regressionsproblem entspricht. Die logistische und die lineare Regression stellen Beispiele für überwachte klassifikations- und regressionsbasierte Lernalgorithmen dar. Zur gleichen Verfahrensklasse gehören Random Forests und künstliche neuronale Netze, auf denen die in dieser Arbeit vorgestellten Segmentierungsalgorithmen beruhen. Ihre Funktionsweise wird im Folgenden ausführlicher beschrieben.

Im Gegensatz dazu stehen beim unüberwachten Lernen keine gelabelten Daten zu Verfügung. Daher sucht hier der Lernalgorithmus dementsprechend eigenständig nach Mustern und Zusammenhängen in den Trainingsdaten. Daneben lässt sich wegen der fehlenden Lernvorgabe die Ausgabe des Algorithmus nicht vordefinieren. Unüberwachte Lernmethoden unterteilen sich weiter in Clusteranalyse- und Dimensionsreduktionsverfahren. Bei der Clusteranalyse wird versucht, unbekannte Muster in den Daten zu entdecken. Die gewonnenen Erkenntnisse werden dann genutzt, um die Daten zu gruppieren. Ein Beispiel für einen Clustering-Algorithmus ist die k-Means-Methode. Dagegen werden die in den Daten vorhandenen Muster durch eine Dimensionsreduktion auf wesentliche Repräsentationen reduziert, indem ähnliche Dimensionen zusammengeführt werden. Hierdurch wird die Analyse von hochkomplexen sowie mehrdimensionalen Datensätzen stark vereinfacht und kann beispielsweise mittels einer Hauptkomponentenanalyse (*engl.*: principal components analysis, kurz: PCA) durchgeführt werden. Steht neben einer größeren Menge nicht gelabelter Daten auch eine geringe Anzahl gelabelter Daten zur Verfügung, können diese für das Training kombiniert werden, was als halbüberwachtes Lernen (*engl.*: semi-supervised learning) bezeichnet wird. Des Weiteren können verrauschte, begrenzte oder ungenaue Labels als Zielvorgaben genutzt werden. Solche schwach gelabelten Daten sind wesentlich kostengünstiger und können trotzdem verwendet werden, um ein genaues Modell zu erstellen. Man spricht daher auch vom schwach überwachten Lernen (*engl.*: weakly supervised learning).

Das bestärkende Lernen (*engl.*: reinforcement learning) bildet die letzte Verfahrensklasse des maschinellen Lernens. Die Besonderheit bei dieser Lernform ist, dass der Lernalgorithmus als Agent in einem dynamischen Prozess mit seiner Umgebung interagiert und über mehrere Durchläufe eigenständig eine Strategie entwickelt. Hierzu verändert der Agent seinen Zustand durch verschiedene Aktionen, um sich dem vordefinierten Lernziel in jedem Durchlauf weiter anzunähern. Je nachdem, wie gut sich der Agent dem Lernziel angenähert hat, erhält dieser eine positive oder negative Belohnung. Das Ziel des Lernprozesses ist es also, die Belohnung zu maximieren. Bestärkende Lernalgorithmen haben einen hohen Bekanntheitsgrad erlangt, etwa durch die Beherrschung von Atari-Spielen oder AlphaGo von DeepMind (Google Inc.).

3.2.1 Entscheidungsbäume

Entscheidungsbäume (*engl.*: decision trees) sind gerichtete hierarchische Graphen, deren Aufbau der Abfolge mehrerer Entscheidungsregeln entspricht. Sie werden eingesetzt, um Daten anhand geeigneter Merkmale unter Berücksichtigung einer definier-

ten Zielgröße einzuteilen. Entscheidungsbäume bilden ein grundlegendes Instrument in der Informatik, das in verschiedenen Bereichen zum Einsatz kommt, z. B. beim Data-Mining als Such- und Sortiermethode. Anstatt die Entscheidungsbäume manuell zu konstruieren, werden sie beim maschinellen Lernen anhand der verfügbaren Trainingsdaten erlernt, um eine bestmögliche Baumstruktur zu erhalten. Entscheidungsbäume gehören zur Klasse der überwachten Lernverfahren, die zur Lösung von Klassifikations- und Regressionsproblemen eingesetzt werden können.

3.2.1.1 Aufbau

Ein Entscheidungsbaum besteht aus Knoten und Blättern. Die Knoten repräsentieren Verzweigungen, an denen sich die ankommenden Daten jeweils aufteilen. Blätter entsprechen der finalen Klassifikation und besitzen keinen weiteren Nachfolger. In dieser Arbeit werden ausschließlich binäre Entscheidungsbäume betrachtet, bei denen jeder Knoten genau zwei Nachfolger hat. Für jeden Knoten n wird eine Trennfunktion in Form einer einfachen Schwellenwertabfrage verwendet, um ankommende Daten auf einen der beiden Nachfolgeknoten aufzuteilen:

$$\psi(\mathbf{m}; m_n, t_n) = \begin{cases} 0, & \text{wenn } \mathbf{m}_{m_n} < t_n \\ 1, & \text{sonst.} \end{cases} \quad (3.7)$$

Hierbei beschreibt $\mathbf{m} \in M$ den betrachteten Merkmalsvektor aus dem Merkmalsraum M . Des Weiteren beschreiben m_n das selektierte Merkmal und t_n den gewählten Schwellenwert für den n -ten Knoten. Je nachdem an welcher Stelle im Baum ein gewisses Merkmal abgefragt wird, können sich unterschiedlich große Baumstrukturen ergeben, die jeweils dieselben Ausgaben erzeugen. Der Vorteil kleiner und kompakter Bäume ist, dass sie leichter zu interpretieren sind und sie besser generalisieren. Um eine ideale Baumkonstruktion zu ermitteln, wird diese anhand von Trainingsdaten erlernt. Dieser Lernprozess beinhaltet das Finden der optimalen Trennfunktion mit dem gewählten Merkmal m_n und dem Schwellenwert t_n . Ein Beispiel eines Entscheidungsbaums mit zugehöriger Datenklassifikation ist in Abbildung 3.3 dargestellt.

3.2.1.2 Training und Anwendung

Die Grundlage des Lernvorgangs bilden die Trainingsdaten $D_{\text{train}} = \{d_1, \dots, d_{|D_{\text{train}}|}\}$, die insgesamt $|D_{\text{train}}|$ unterschiedliche Objekte beschreiben. Für eine bessere Nachvollziehbarkeit wird D_{train} im Folgenden lediglich als D bezeichnet. Jeder Datenpunkt $d = (\mathbf{m}, l) \in D$ besteht wiederum aus einem Merkmalsvektor $\mathbf{m} = (m_1, \dots, m_{|M|}) \in M$ und einem zugehörigen Klassenlabel $l \in L = \{0, \dots, K - 1\}$. Ausgehend vom Wurzelknoten mit dem Index $n = 0$ verläuft der Aufbau knotenweise von oben nach unten bis zur Mündung in den Blattknoten. Am Knoten n wird die ankommende Teiltrainingsmenge D_n durch die Trennfunktion in die Teilmengen $D_{n,\text{links}}$ für $\psi(\mathbf{m}) = 1$ und

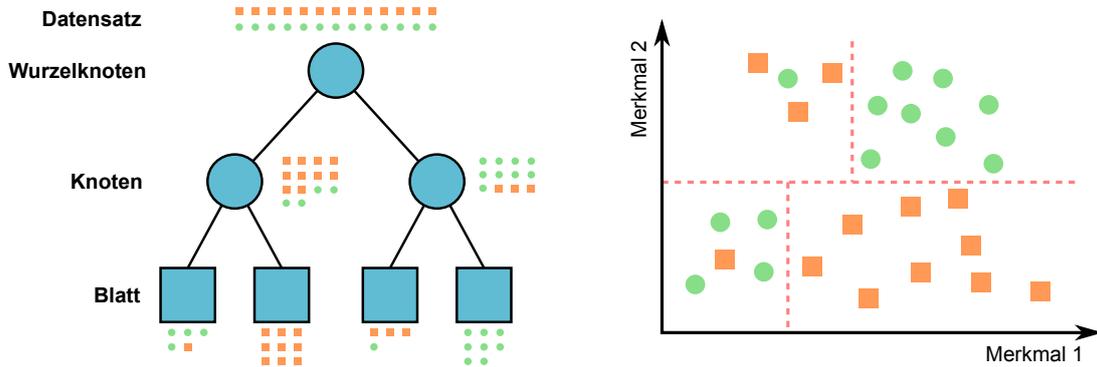


Abb. 3.3: Beispiel für das Training eines Entscheidungsbaums für ein binäres Klassifikationsproblem. Ausgehend vom Wurzelknoten wird durch eine umfangreiche Suche im Merkmalsraum eine Trennungsfunktion ermittelt, die den höchsten Informationsgewinn bietet. Auf der Grundlage der Trennfunktion werden die Trainingsdaten aufgetrennt. Dieser Prozess wiederholt sich für die nachfolgenden Knoten, bis ein Stoppkriterium erreicht ist. In diesem Fall wird das Wachstum des Baums an dieser Stelle angehalten und der entsprechende Knoten wird in ein Blatt umgewandelt, das die endgültige Klassifikation angibt. Die Klasse wird durch Mehrheitsbeschluss der im Blatt vorliegenden Trainingsdaten bestimmt.

$D_{n,\text{rechts}}$ für $\psi(\mathbf{m}) = 0$ aufgeteilt, die jeweils zum rechten oder linken Nachfolgeknoten verlaufen. Es gilt dabei $D_{n,\text{links}} \cap D_{n,\text{rechts}} = \emptyset$ sowie $D_{n,\text{links}} \cup D_{n,\text{rechts}} = D_n$. Idealerweise sollten nach jeder Aufteilung die Datenpunkte pro Teilmenge weniger unterschiedliche Klassen aufweisen als vorher. Man spricht auch von Reinheit der Daten. Als Maß für die Reinheit (*engl.*: purity measure, kurz: PM) werden bei der Klassifikation meist die Shannon-Entropie oder der Gini-Koeffizient genutzt. Die Shannon-Entropie lässt sich anhand der Klassenwahrscheinlichkeiten $P(l|D)$ der (Teil-)Menge D bestimmen:

$$H(D) = - \sum_{l=0}^{K-1} P(l|D) \log P(l|D). \quad (3.8)$$

Bei einem Zweiklassenproblem ist die Entropie maximal ($= H(D) = 1$), wenn die Klassen gleich verteilt sind. Mit zunehmendem Ungleichgewicht der Klassen sinkt der Entropiewert entsprechend. Der Gini-Koeffizient ist der Entropie ähnlich und berechnet sich folgendermaßen:

$$\text{Gini}(D) = 1 - \sum_{l=0}^{K-1} P(l|D)^2. \quad (3.9)$$

Auch der Gini-Koeffizient nimmt bei Klassengleichgewicht seinen Höchstwert an, der bei 0.5 liegt. Infolgedessen weist der Gini-Index ein flacheres Wachstum auf und ist daher gegenüber Veränderungen der Klassenhäufigkeiten etwas unempfindlicher. Mittels Entropie oder Gini-Index lässt sich nun die Zu- oder Abnahme der Reinheit infolge der

Trennung von D_n in $D_{n,\text{links}}$ sowie $D_{n,\text{rechts}}$ bestimmen, was als Informationszuwachs (*engl.*: information gain, kurz: IG) bezeichnet wird, wodurch sich die Qualität einer Datenaufteilung ermitteln lässt. Sie ist definiert durch

$$IG(D, D_{\text{links}}, D_{\text{rechts}}) = \text{PM}(D) - \frac{|D_{\text{links}}|}{|D|} \text{PM}(D_{\text{links}}) - \frac{|D_{\text{rechts}}|}{|D|} \text{PM}(D_{\text{rechts}}). \quad (3.10)$$

Für $\text{PM}(\cdot)$ kann entweder die Entropie (3.8) oder der Gini-Koeffizient (3.9) verwendet werden. Für jeden Knoten gilt es den Informationszuwachs zu maximieren

$$\psi_n(\mathbf{m}; m_n, t_n) = \arg \max_{m,t} IG(D_n, \psi(\mathbf{m}; m, t)), \quad (3.11)$$

was einer ausgiebigen Suche über alle Kombinationen von Merkmalen und Schwellenwerten entspricht und mit dem Greedy-Algorithmus gelöst wird. Ein bereits gesetzter Knoten wird im weiteren Verlauf der Baumkonstruktion nicht mehr verändert, sodass ein globales Klassifikationsoptimum nur approximiert werden kann.

Die Entscheidung, ob ein Nachfolger ein Knoten oder ein Blatt wird, kann auf der Grundlage verschiedener Stoppkriterien [Banfield u. a., 2006] getroffen werden. Andernfalls würde der Baum so lange weiterwachsen, bis jeder Datenpunkt der Trainingsmenge in einem separaten Blatt münden würde. Das wohl einfachste Stoppkriterium wäre zu überprüfen, ob die Teilmenge D_n , die am Nachfolger vorliegt, rein ist. Die alleinige Verwendung dieses Kriteriums wäre allerdings auch problematisch, da es zu großen Entscheidungsbäumen führt, die rechenintensiv sind und schlecht generalisieren. Die am häufigsten genutzten Stoppkriterien sind die maximale Baumhöhe, die minimale Anzahl von Datenpunkten in einem Nachfolger oder ein bestimmter Informationszuwachs. Schließlich wird die endgültige Klassendarstellung eines Blattes durch eine Mehrheitsentscheidung über die Labels der Trainingsdatenpunkte, die in diesem Blatt vorliegen, ermittelt.

In der Test- oder Anwendungsphase durchlaufen neue ungesehene Datenpunkte $d \in D_{\text{test}}$ am Wurzelknoten beginnend den Entscheidungsbaum, bis sie ein Blatt erreichen, durch welches das jeweilige Klassenlabel zugeordnet wird.

3.2.1.3 Random Forests

Die zuvor beschriebenen Entscheidungsbäume stellen eine einfache Klassifikationsmethode dar, da die gelernten Bäume übersichtlich und damit leicht zu interpretieren sind. Zudem haben sie einen geringen Bias, da kaum Nebenbedingungen durch die Zielfunktion formuliert werden. Allerdings ist ihre Varianz relativ hoch, sodass sie zu einer Überanpassung neigen. Außerdem sind sie anfällig für verrauschte Daten.

Durch die Kombination mehrerer Entscheidungsbäume lässt sich die Varianz verringern, wobei ein Baum-Ensemble entsteht, das als Random Forest (kurz: RF, *dt.*: Zufallswald) bezeichnet wird. Konkret werden mehrere einfache Entscheidungsbäume

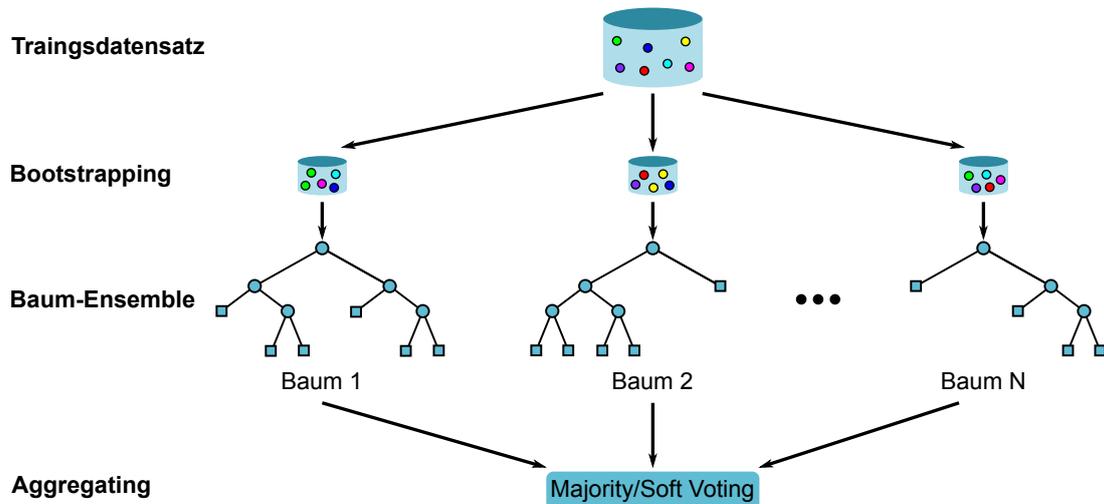


Abb. 3.4: Vereinfachte Darstellung eines Random-Forest-Klassifikators. Zunächst werden mithilfe der Bootstrapping-Technik aus dem Trainingsdatensatz mehrere unterschiedliche Teildatensätze erzeugt. Pro Datensatz wird anschließend ein individueller Entscheidungsbaum trainiert. Zusammen ergeben die trainierten Bäume ein Ensemble-Klassifikator, der z. B. neue Daten durch Mehrheitsentscheid klassifiziert.

erlernt, die in ihrer Kombination die Klassifikation vornehmen. Die Korrelation zwischen den einzelnen Bäumen sollte möglichst gering gehalten werden, da sie sonst zu einer Erhöhung der Varianz führt. Dies erfordert die Integration mehrerer Zufallsprozesse in das Training eines RF-Klassifikators. Diesbezüglich sind verschiedene Strategien vorgeschlagen worden. Ein detaillierter Überblick ist unter anderem in [Criminisi u. a., 2011] zu finden. Im Folgenden werden zwei der bekanntesten Methoden, das Bagging und die zufällige Merkmalsauswahl, vorgestellt.

Bagging ist die Kurzform für „Bootstrap Aggregating“ und ist eine von Breiman [1996] vorgeschlagene Methode zur zufälligen Aufteilung der Trainingsdaten. Dazu wird jedem Baum des Ensembles ein sogenannter Bootstrap-Datensatz, eine zufällige Teilmenge des Trainingsdatensatzes, zugewiesen und mit diesem trainiert. Die Erstellung der Bootstrap-Datensätze ist gleichbedeutend mit der wiederholten Ziehung von Stichproben mit Zurücklegen, sodass einzelne Datenpunkte mehrfach gezogen werden können. Auf diese Weise werden verschiedene Bäume erlernt, die eher klein sind und daher zu kürzeren Laufzeiten und höherer Generalisierung führen. Die Klassifikation eines neuen Datenpunktes wird durch Mehrheitsentscheid (*engl.*: majority voting) oder Mittelwertbildung (soft voting) der Einzelvorhersagen der Bäume bestimmt, was als Aggregating bezeichnet wird. Die Bagging-Strategie ist in Abbildung 3.4 illustriert.

Die zufällige Merkmalsauswahl (*engl.*: random feature selection) wird in der Arbeit von Ho [1995] vorgeschlagen und dient auch zur Variation der Eingabedaten. Anstatt beim Training für jeden Knoten über alle Merkmale des Datenpunkts zu optimieren,

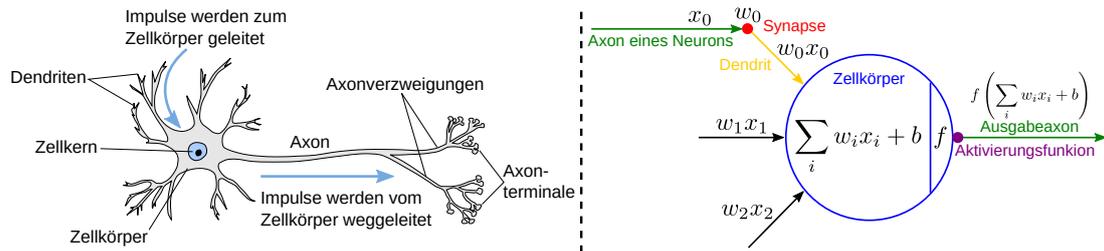
wird nur eine zufällige Teilmenge $M_n \subseteq M$ ausgewählt. Genauso wie beim Bagging führt dies zu geringeren Berechnungszeiten und erhöhter Generalisierung.

RFs stellen somit robuste Klassifikatoren dar, die in der Lage sind, gut zu generalisieren und robust gegenüber verrauschten Eingangsdaten sind. Allerdings hängt die Klassifikationsqualität von RFs maßgeblich von der Wahl der geeigneten Merkmale ab, die im Gegensatz zu modernen Methoden wie neuronalen Netzen sorgfältig manuell ausgewählt werden müssen.

3.2.2 Faltungsnetzwerke

Deep Learning (kurz: DL, *dt.*: tiefes Lernen) beschreibt eine Unterkategorie des maschinellen Lernens, die sich auf die Verwendung von künstlichen neuronalen Netzen (kurz: KNNs, *engl.*: artificial neural networks) konzentriert. Nach dem Vorbild des menschlichen visuellen Cortex stellen KNNs ein Berechnungsmodell dar, das aus einer Vielzahl von angeordneten informationsverarbeitenden Einheiten, den Neuronen, besteht. In Abbildung 3.5 ist hierzu die Analogie zwischen einem biologischen Neuron und seinem mathematischen Gegenstück dargestellt. Wie ihre biologischen Pendanten erhalten Neuronen eines KNNs Signale über verschiedene gewichtete Eingabekanäle (Dendriten). Innerhalb des Neurons (Zellkörper) werden die gewichteten Eingabesignale zusammen mit einem Bias-Wert aufsummiert. Eine Aktivierungsfunktion verarbeitet das summierte Signal weiter, was der Impulsweiterleitung über das Axon entspricht, und leitet es zum nächsten Neuron weiter. Das Summensignal wird durch eine nichtlineare Aktivierungsfunktion verwertet, die dem Schwellenwert eines Aktionspotentials entspricht, also dem Wert, der für eine Reizübertragung über das Axon überschritten werden muss. Anschließend wird die Antwort der Aktivierungsfunktion an die nächste Neuronenschicht weitergeleitet (Axon). Ein solches einschichtiges KNN wird auch als Perzeptron bezeichnet, mit dem sich bei geeigneter Wahl der Aktivierungsfunktion einfache Klassifikatoren realisieren lassen.

Die Struktur eines KNN besteht im Wesentlichen aus drei Schichttypen. Die Eingabeschicht versorgt ein KNN mit den ankommenden Signalinformationen, die dann gewichtet und an die nächste Schicht weitergeleitet werden. Hier finden weitere Gewichtungen der Eingangssignale statt. Diese Schicht wird auch als verborgene Schicht (*engl.*: hidden layer) bezeichnet, da dieser Verarbeitungsprozess für den Anwender nicht sichtbar ist. Schließlich werden die Signale zur Ausgabeschicht geleitet. Sie bildet die letzte Schicht, deren Neuronen die endgültige Entscheidung des KNNs enthalten. Die Weitergabe eines Eingabesignals wird als Vorwärtsschritt bezeichnet. Während Eingabe- und Ausgabeschicht jeweils nur einmal am Anfang und Ende vorkommen, kann es dazwischen beliebig viele versteckte Schichten geben. Die Verwendung von mehreren Schichten in Kombination mit nichtlinearen Aktivierungsfunktionen ermöglicht die Approximation hochkomplexer Funktionen. Je höher die Anzahl von Schichten ist, desto „tiefer“ ist ein KNN, woher auch die Bezeichnung „Deep Learning“ stammt.



Eigene Darstellung nach [Stanford Vision and Learning Lab, 2021a]

Abb. 3.5: Zeichnung eines biologischen Neurons (links) und sein mathematisches Modell (rechts). Die von Axonen eingehenden Signale x_i werden in den Dendriten mit w_i gewichtet, bevor sie den Zellkörper eines Neurons erreichen. Im Neuron werden die gewichteten Signale zusammen mit einem Bias-Wert aufsummiert. Die Aktivierungsfunktion $f(\cdot)$ nimmt die gewichteten Werte aus dem Summationsblock des Neurons und wandelt sie in ein Ausgangssignal um, wodurch die Größe der weitergeleiteten Signale beeinflusst wird.

Des Weiteren ist die Anzahl von Neuronen pro Schicht ebenfalls beliebig wählbar, während die Größe der Eingabe- und Ausgabeschicht auf die Eingangssignale und die erwartete Ausgabe limitiert ist. Ein dreischichtiges KNN ist links in Abbildung 3.6 dargestellt. Jedes Neuron einer Schicht ist mit jedem Neuron der vorherigen und nachfolgenden Schicht (sofern vorhanden) verbunden, was auch als vollständig verbundene Schicht (*engl.*: fully connected layer) bezeichnet wird. Da der Datenfluss nur in einer Richtung erfolgt, spricht man auch von vorwärtsgekoppelten Netzen (*engl.*: feedforward networks).

Anders als bisherige maschinelle Lernmethoden, repräsentieren KNNs einen Algorithmus, der eigenständig aus verfügbaren Trainingsdaten lernt. Hierbei werden die Netzparameter (Gewichte und Bias-Werte) der Neuronen im Training iterativ angepasst, um ein vorgegebenes Lernziel zu approximieren. Die Netzarchitektur bleibt dabei erhalten, sodass das erlernte Wissen implizit in den Netzparametern kodiert wird. Hierdurch entsteht ein Paradigmenwechsel, da die bisherige manuelle Anfertigung geeigneter Merkmale nun entfällt. KNNs machen sich den hierarchischen Aufbau vieler Signale zunutze, d. h. komplexere Merkmale können durch die Kombination einfacher Merkmale erstellt werden. In Bildern z. B. bilden die Kombinationen von Kanten oder Kreisen größere Motive, die sich wiederum zu Bildteilen zusammensetzen. Einzelne Bildteile bilden schließlich ganze Bildobjekte [LeCun u. a., 2015]. Tiefe KNNs lernen so über viele Schichten hinweg hierarchische Merkmalsrepräsentationen. Trotz der Vorgabe eines Lernziels, verbleibt die Anpassung der Netzparameter ein unüberwachter Prozess. Durch die hohe Parameteranzahl im Millionenbereich ist eine detaillierte Interpretation der gelernten Merkmale nahezu unmöglich, sodass KNNs auch als sogenannte „Black-Box-Systeme“ angesehen werden, was die explizite Extraktion oder Integration

von Wissen bislang unmöglich macht. Die hohe Parameteranzahl erfordert zudem eine sehr hohe Anzahl von Trainingsdaten, um eine Überanpassung zu vermeiden.

Obwohl wesentliche Deep-Learning-Techniken schon seit einigen Jahrzehnten bekannt sind, hat der Durchbruch erst Anfang der 2010er Jahre stattgefunden, was auf zwei wesentliche Ereignisse zurückzuführen ist. Mit der Entwicklung leistungsfähiger GPUs können nun die stark rechenintensiven Optimierungsschritte in einem Bruchteil der bisherigen Zeit durchgeführt werden, wodurch sich die Trainingszeiten drastisch reduzieren lassen. Darüber hinaus sind durch die fortschreitende Digitalisierung Daten leichter verfügbar, sodass nun eine ausreichende Anzahl von Trainingsbeispielen einer Überanpassung entgegenwirkt.

Im Computer-Vision-Bereich kommen vorwiegend Faltungsnetzwerke (*engl.*: convolutional neural networks, kurz: CNNs) zur Analyse visueller Bilddaten zum Einsatz und stellen den aktuellen Stand der Technik dar. Mit der Arbeit von Krizhevsky u. a. [2012] erlangen CNNs zum ersten Mal größere Aufmerksamkeit. Darin stellen die Autoren AlexNet vor, eine CNN-Architektur zur Klassifikation von Bildern. Im Jahr 2012 gewinnen sie mit ihrem Algorithmus die ImageNet-Challenge¹ mit einem Genauigkeitsvorsprung von über 10 % zu den anderen teilnehmenden Wettbewerbern. Der Grund für die hohe Leistung des Modells liegt in der Tiefe des Netzes, was einerseits zu einer rechenintensiven Optimierung führt, jedoch andererseits durch den Einsatz von GPUs realisierbar ist. CNNs bilden die Grundlage der meisten in dieser Arbeit vorgestellten Algorithmen. Daher wird ihr Aufbau, das Training und die Anwendung nachfolgend näher beschrieben.

3.2.2.1 Aufbau

Die zuvor beschriebenen (vorwärtsgekoppelten) KNNs bringen für die Bildanalyse einige Nachteile mit sich. Die in ihnen verwendeten vollständig verbundenen Schichten benötigen schon bei relativ kleinen Bildern eine enorme Anzahl von Trainingsparametern. Bereits bei einer Bildgröße von bspw. 100×100 Pixeln würde die Eingabeschicht 10.000 Neuronen aufweisen. Mit einer nachfolgenden verborgenen Schicht von bspw. 200 Neuronen würde das KNN bereits 2.000.000 lernbare Parameter (Bias-Werte nicht einbezogen) enthalten. CNNs sind wie die KNNs aus Neuronen aufgebaut (siehe Abb. 3.5), die ein Gewicht und einen Bias-Wert als lernbare Parameter besitzen. Da CNNs jedoch Bilder als Eingabe erhalten, ist ihre Architektur verglichen zu KNNs für diesen Zweck auf sinnvolle Weise angepasst. Ein schematischer Vergleich in der Architektur zwischen KNN und CNN ist in Abbildung 3.6 dargestellt. Die Neuronen in den einzelnen Schichten eines CNN sind dreidimensional in einem Volumen angeordnet. Ihre Höhe und Breite entsprechen der Größe des Eingabebildes und der Zwischendarstellungen, während die Tiefe die Anzahl ihrer Kanäle beschreibt.

¹ImageNet Large Scale Visual Recognition Challenge (ILSVRC)
<https://www.image-net.org/challenges/LSVRC/>

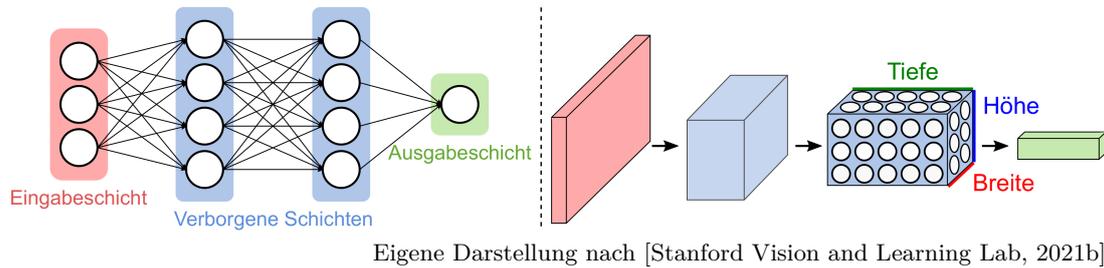


Abb. 3.6: Gegenüberstellung zwischen einem KNN (links) und einem CNN (rechts). Anders als bei KNNs sind die Neuronen in einem CNN dreidimensional in Höhe, Breite und Tiefe angeordnet. Jede Schicht transformiert ein jeweiliges Eingabevolumen in ein entsprechendes Ausgabevolumen, dessen Größe sich von der seines Vorgängers unterscheiden kann. Die Tiefe eines Volumens beschreibt dabei keine räumliche Größe, sondern die Anzahl der gelernten Filterkerne.

Wie bereits diskutiert, würde die hohe Parameteranzahl zur Überanpassung und einer langwierigen Optimierung führen. Ferner berücksichtigen vollständig verbundene Schichten nicht die Geometrie eines Bildes, was bedeutet, dass unterschiedliche Vorhersagen gemacht würden, wenn sich etwa das Objekt im Bild verschiebt. Neuronen in einer vollständig verbundenen Schicht können sich ändern, sobald sich nur eines der Neuronen aus der vorherigen Schicht ändert. Dieses Problem wird in CNNs durch die namensgebenden Faltungsschichten (*engl.*: convolutional layers) gelöst, die aus einer Reihe von lernbaren Filterkernen bestehen. Diese sind in Höhe und Breite klein, erstrecken sich aber über die gesamte Tiefe des Eingabevolumens. Der Vorteil von Faltungsschichten gegenüber vollständig verbundenen Schichten liegt darin, dass sie nur einen engeren Bereich des Eingabevolumens abdecken. Neben den Faltungsschichten werden Pooling-Schichten sowie vollständig verbundene Schichten genutzt. Die Funktionsweise der drei verschiedenen Schichttypen innerhalb eines CNN wird nachfolgend erläutert.

Während eines Vorwärtsschrittes werden die Filterkerne in Richtung der räumlichen Dimensionen über das Eingabevolumen geschoben und an jeder Position das Skalarprodukt berechnet. Dies entspricht einer Faltungsoperation, deren Ausgabe eine zweidimensionale Aktivierungskarte ist, im Folgenden auch als Merkmalskarte bezeichnet. Im Falle mehrerer gelernter Filterkerne pro Faltungsschicht werden die daraus resultierenden Merkmalskarten zu einem Volumen konkateniert. Dementsprechend stellt die Tiefe dieses Merkmalsvolumens die Anzahl der Filterkerne in der Faltungsschicht dar. Durch die Verschiebung des Filterkerns werden seine Gewichte im Raum geteilt, sodass die Anzahl der Netzparameter erheblich reduziert wird. Man spricht hier auch von geteilten Gewichten (*engl.*: shared weights). Durch die Verwendung kleiner Filterkerne werden die miteinander verbundenen Neuronen auf eine ebenso kleine Region beschränkt, wodurch die Anzahl der zu lernenden Parameter weiter reduziert wird.

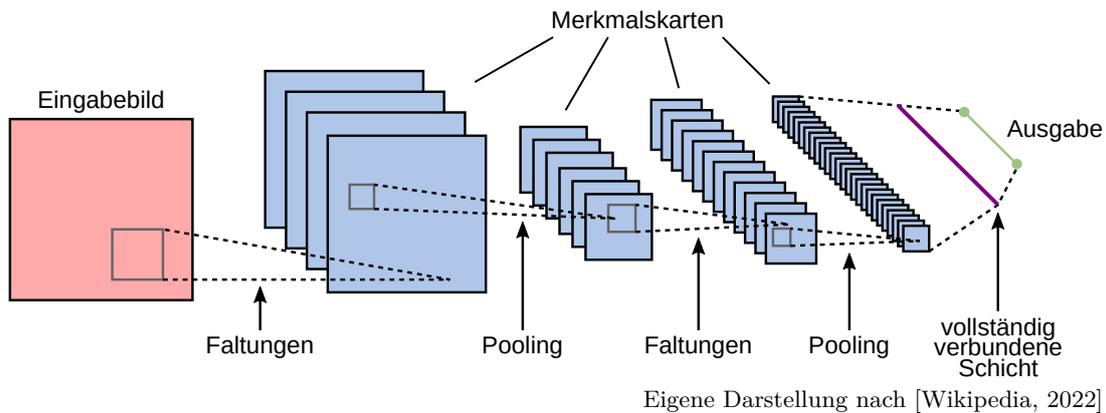


Abb. 3.7: Typische Struktur eines CNN für die Bildklassifikation. Zunächst werden Merkmalsrepräsentationen durch den abwechselnden Einsatz von Faltungs- und Pooling-Schichten über verschiedene Auflösungsstufen hinweg aus dem Eingangsbild extrahiert. Im Klassifikationsteil des CNN werden die extrahierten Merkmale schließlich in einen Vektor umgewandelt, der eine oder mehrere vollständig verbundene Schichten durchläuft, bevor die Daten in der Ausgangsschicht zusammenlaufen und eine Prädiktion bilden.

Diese räumliche Konnektivität wird als *rezeptives Feld* (*engl.*: *receptive field*) bezeichnet. Während das rezeptive Feld der Neuronen in der ersten Schicht der Größe des Filterkerns entspricht, ist es für die Neuronen aufgrund der kontinuierlichen Verkettung in den nachfolgenden Schichten größer, sodass sie einen größeren Bildkontext erfassen können. Folglich lernt das CNN einfache lokale Merkmale in früheren Schichten und kontextbezogene Merkmale in späteren Schichten. Typischerweise folgt auf eine Faltungsschicht eine nichtlineare Aktivierungsfunktion (siehe Abb. 3.5), wobei in CNNs häufig die Rectified Linear Unit (kurz: ReLU, *dt.*: lineare Gleichrichtungseinheit) genutzt wird.

Pooling-Schichten bilden eine weitere wichtige Komponente. Sie werden üblicherweise in periodischen Abständen im CNN integriert und führen eine räumliche Unterabtastung (*engl.*: *subsampling*) des Eingabevolumens unter Erhalt seiner Tiefe durch, wodurch es zu einer Verdichtung relevanter Informationen kommt, und wirken somit einer Überanpassung entgegen. Gleichzeitig sorgen Pooling-Operationen für ein schnelleres Wachstum des rezeptiven Feldes und ermöglichen somit das Erfassen größerer Kontexte im Bild. Als Pooling-Operator wird oft ein 2×2 -Kern genutzt, der mit einer Schrittweite von zwei über das Eingabevolumen verschoben wird. Zur Informationsverdichtung wird häufig das Maximum (Max-Pooling) oder der Mittelwert (Average-Pooling) verwendet. Durch den abwechselnden Einsatz von Faltungsschichten und Pooling-Operationen entsteht eine neuronale Abstraktionspyramide, in der die räumliche Bildgröße sukzessive reduziert wird bei gleichzeitiger Erhöhung der Anzahl

von Merkmalskarten pro Schicht. Eine schematische Darstellung eines CNN ist hierzu in Abbildung 3.7 illustriert.

Die Abfolge von Faltungsschichten stellt den ersten Teil eines CNN dar, in dem semantische Informationen extrahiert und gelernt werden. Im zweiten Teil findet die Klassifikation (oder Regression) der extrahierten Merkmale statt. Zu diesem Zweck werden die Informationen aus den Merkmalskarten zunächst in einen Vektor transformiert. Diese aussagekräftigen und niedrigdimensionalen Repräsentationen werden an die vollständig verbundenen Schichten weitergeleitet und stellen eine kostengünstige Möglichkeit dar, eine nichtlineare Kombination dieser Merkmale zu lernen. Im Falle einer Klassifikation werden die Informationswerte in der Ausgabeschicht über die Softmax-Funktion in Wahrscheinlichkeiten transformiert. Die Größe der Ausgabeschicht entspricht dabei der Anzahl der zu unterscheidenden Klassen. Nebst dem AlexNet sind unter anderem folgende CNN-Architekturen sehr populär geworden: die CNNs VGG-16 und VGG-19 [Simonyan u. a., 2015], das ResNet [He u. a., 2016] sowie das DenseNet [Huang u. a., 2017].

3.2.2.2 Training und Anwendung

Aufgrund des rein datengetriebenen Lernens eines CNN und dessen hoher Parameterzahl ist eine umfangreiche Menge an Bilddaten für das Training erforderlich. Deshalb werden bei geringer Datenmenge Augmentierungstechniken angewandt, mit denen die Anzahl der Trainingsbeispiele künstlich erhöht wird. Hierzu werden meist einfache Operationen wie geometrische Transformation, Hinzufügen von Rauschen oder Änderung der Bildintensitätswerte genutzt.

Das Ziel des Trainings ist die optimale Anpassung der Netzparameter, um einen Klassifikator (oder Regressor) zu erhalten, der über eine bestmögliche Genauigkeit verfügt. Während des Lernprozesses misst eine zuvor definierte Fehler- oder Verlustfunktion (*engl.*: loss function) die Abweichung zum vorgegebenen Lernziel. Eine beliebte Fehlerfunktion beim überwachten Lernen eines (Bild-)Klassifikators ist z. B. die Kreuzentropie, welche die Abweichung der geschätzten Klassenwahrscheinlichkeiten zur Grundwahrheit misst.

Die Optimierung von KNNs basiert auf der Methode des Gradientenabstiegs (*engl.*: gradient descent), die sich im Wesentlichen in drei Varianten unterscheiden lässt. Bei der Batch-Gradient-Descent-Methode wird pro Aktualisierungsschritt der gesamte Trainingsdatensatz verwendet. Diese Vorgehensweise ist zwar einfach nachzuvollziehen, bedarf jedoch für große Datenmengen ein hohes Maß an Rechen- und Speicherkapazität. Das stochastische Gradientenabstiegsverfahren (*engl.*: stochastic gradient descent, kurz: SGD) ist eine Variante des Gradientenabstiegs und führt sukzessiv eine Optimierung durch. Dabei erfolgt die Aktualisierung der Netzparameter jeweils auf einem einzelnen Trainingsbeispiel, das zufällig aus der Trainingsmenge gezogen wird, wodurch das Training im Vergleich zur Batch-Gradient-Descent-Methode wesentlicher

schneller verläuft und weniger speicher- sowie rechenintensiv ist. Allerdings führen die Fehlerberechnungen auf einzelnen Stichproben zu höheren Fluktuationen in der Fehlerfunktion. Der Mini-Batch-Gradient-Descent ist eine Variante des SGD, bei dem nicht einzelne Trainingsbeispiele, sondern „Mini-Batches“ der Trainingsmenge durch das Netz propagiert werden, was die Varianz im Aktualisierungsschritt reduziert und somit zu einem stabileren Konvergenzprozess führt. Die Zusammenstellung der Mini-Batches erfolgt zufällig und umfasst eine vordefinierte Anzahl von Trainingsbeispielen. Es sind verschiedene Erweiterungen vorgestellt worden, um die Optimierung schneller und robuster zu gestalten. Mit der Einführung von Momenten [Qian, 1999] wird in der Optimierung die Trägheit eines Objekts simuliert. Hierdurch wird die Bewegungsrichtung des vorherigen Aktualisierungsschritts zu einem gewissen Grad beibehalten. Der aktuelle Aktualisierungsschritt dient dann zur Feinabstimmung der Richtung. Auf diese Weise können durch den „Schwung“ des Moments selbst in Regionen mit geringen Gefällen größere Aktualisierungsschritte gemacht werden, was den Konvergenzprozess des Gradientenabstiegs beschleunigt. Ein sehr beliebtes Optimierungsverfahren von KNNs stellt die adaptive Momentschätzung (*engl.*: adaptive moment estimation, kurz: Adam) dar [Kingma u. a., 2014], bei dem eine individuelle adaptive Lernrate für jeden Parameter aus den Schätzungen der ersten und zweiten Momente der Gradienten berechnet wird. Die nachfolgende Beschreibung der KNN-Optimierung orientiert sich an dem Mini-Batch-Gradient-Descent-Verfahren.

Zu Beginn des Trainings werden zunächst die Netzwerkparameter initialisiert. Da eine Initialisierung mit demselben konstanten Wert zu identischen Gradienten führen würde, werden die Gewichte mit Zufallswerten nahe null initialisiert. Anschließend erhält das Netz ein Mini-Batch als Eingabe und berechnet die entsprechende Ausgabe, was auch als Vorwärtsschritt bezeichnet wird. Anhand der Fehlerfunktion wird die Abweichung zum Lernziel bestimmt. Ausgehend von der berechneten Abweichung werden nun mittels partieller Ableitung die Gradienten der einzelnen Netzparameter gebildet, was eine Fehlerrückführung [Rumelhart u. a., 1986] durch das Netz darstellt und auch als Rückwärtsschritt (*engl.*: backpropagation) bezeichnet wird. Die Gewichtaktualisierung erfolgt in Abhängigkeit einer gegebenen Lernrate, welche die Schrittweite beim Gradientenabstieg regelt und einen wesentlichen Einfluss auf den Optimierungsprozess hat. Falls sie zu niedrig gewählt wird, können lokale Minima die Optimierung vorzeitig beenden. Ist sie dagegen zu hoch angesetzt, kann ein Minimum stets übersprungen werden. Zur Vermeidung übermäßiger Gradientenaktualisierungen, die ebenso zum Überspringen lokaler Minima führen, können Regularisierungsterme wie die L1- oder L2-Norm verwendet werden. Das Wechselspiel zwischen Vorwärts- und Rückwärtsschritt wird für die gesamte Trainingsmenge wiederholt, was als Epoche bezeichnet wird. Das Training eines CNN umfasst häufig mehrere hundert Epochen, ehe das Lernziel hinreichend angenähert wird. Für gewöhnlich wird vor Beginn des Trainings eine ausreichend hohe Anzahl von Epochen festgelegt, in denen das Optimierungspro-

blem garantiert konvergiert, oder aber es wird ein Abbruchkriterium definiert, um das Training vorzeitig zu beenden.

Die Auswertung während der Anwendungsphase entspricht einem normalen Vorwärtsschritt durch das Netz ohne die Durchführung weiterer, wie oben beschriebener, Optimierungsschritte. Hierzu wird das Netz in einen Evaluationsmodus versetzt, in dem die Netzparameter fest sind.

3.2.2.3 Das U-Net

In der Arbeit von Long u. a. [2015] stellen die Autoren vollständig faltungsbasierte Netzwerke (FCNs) vor, die anstelle vollständig verbundener Schichten 1×1 -Faltungen einsetzen. Hieraus ergeben sich mehrere Vorteile. Zum einen können nun Bilddaten beliebiger Größe verwendet werden. Zum anderen wird die Anzahl der zu trainierenden Parameter weiter reduziert. Im Übrigen bietet die von Long u. a. vorgeschlagene FCN-Architektur eine pixelweise Klassifikation ganzer Bilder, weswegen sie sich besonders für die semantische Segmentierung eignet. Zur Hochskalierung der Merkmalskarten werden Entfaltungen (*engl.*: deconvolutions) genutzt, auch transponierte Faltungen genannt, die als Filterkerne gelernt werden.

Inspiziert von der Arbeit von Long u. a. schlagen Ronneberger u. a. [2015] eine Methode zur Segmentierung medizinischer Zellbilddaten vor. Darin beschreiben die Autoren das „U-Net“, eine FCN-Architektur, die sich innerhalb kürzester Zeit als Stand der Technik in der medizinischen Bildsegmentierung etabliert hat. Die originale U-Net-Architektur aus [Ronneberger u. a., 2015] wird in Abbildung 3.8 dargestellt. Die Architektur besteht aus einem kontrahierenden und einem expandierenden Pfad, die gegenüber dem FCN aus [Long u. a., 2015] symmetrisch zueinander angeordnet sind, was zu der U-förmigen und damit namensgebenden Grundarchitektur führt. Beide Pfade bestehen aus symmetrisch zueinander angeordneten Faltungsblöcken, die sich aus zwei aufeinander folgenden 3×3 -Faltungen mit jeweils einer ReLU-Aktivierung (dunkelblaue Pfeile) zusammensetzen. Das U-Net umfasst insgesamt fünf Auflösungsstufen. Im kontrahierenden Pfad, auch Encoder genannt, wird durch Max-Pooling-Operationen (rote Pfeile) jedes Mal eine neue Auflösungsstufe erreicht, wodurch die Bildgröße um den Faktor zwei reduziert wird. Gleichzeitig wird die Anzahl der Merkmalskarten pro Auflösungsstufe durch die Faltungen verdoppelt. Im sich anschließenden expandierenden Pfad, auch Decoder genannt, findet eine sukzessive Hochskalierung der gelernten Merkmalsrepräsentationen bis hin zur Ausgangsauflösung statt. Für die Hochskalierung nutzen Ronneberger u. a. ebenfalls transponierte Faltungsoperationen (grüne Pfeile). Die hochskalierten Merkmalskarten werden über sogenannte Skip Connections (*dt.*: Sprungverbindungen) mit denen der gleichen Auflösungsstufe des Encoders konkateniert, ehe sie als Eingabe der einzelnen Doppelfaltungsblöcke des Decoders weiterverarbeitet werden. Auf diese Weise lässt sich lokale Information aus dem Encoder zusammen mit Kontextinformation aus dem Decoder zu aussage-

kräftigeren Merkmalsrepräsentationen vereinen. Dabei tragen die Merkmalskarten des Encoders zur Wiederherstellung der durch die Unterabtastung verlorenen Ortsinformation bei. Da die Bildgröße aufgrund der Faltungsoperationen stetig abnimmt, werden die Merkmalskarten vor der Konkatination zugeschnitten (siehe gestrichelte Linien in Abb. 3.8). Die Anzahl der Merkmalskanäle wird im Decoder anschließend wieder graduell verringert. In der Ausgangschiicht findet schließlich die finale Klassifikation der Merkmalsrepräsentationen statt. Zu diesem Zweck wird eine 1×1 -Faltung verwendet, welche im Wesentlichen einer vollständig verbundenen Schicht eines Klassifikators entspricht.

Aufgrund ihrer hohen Performanz erfreut sich die U-Net-Architektur äußerster Beliebtheit und bildet eine fundamentale Grundlage der in dieser Arbeit entwickelten Methoden. Abgesehen von der Bildsegmentierung wird das U-Net auch in anderen Bereichen wie der Bildregistrierung oder der Bildrekonstruktion eingesetzt.

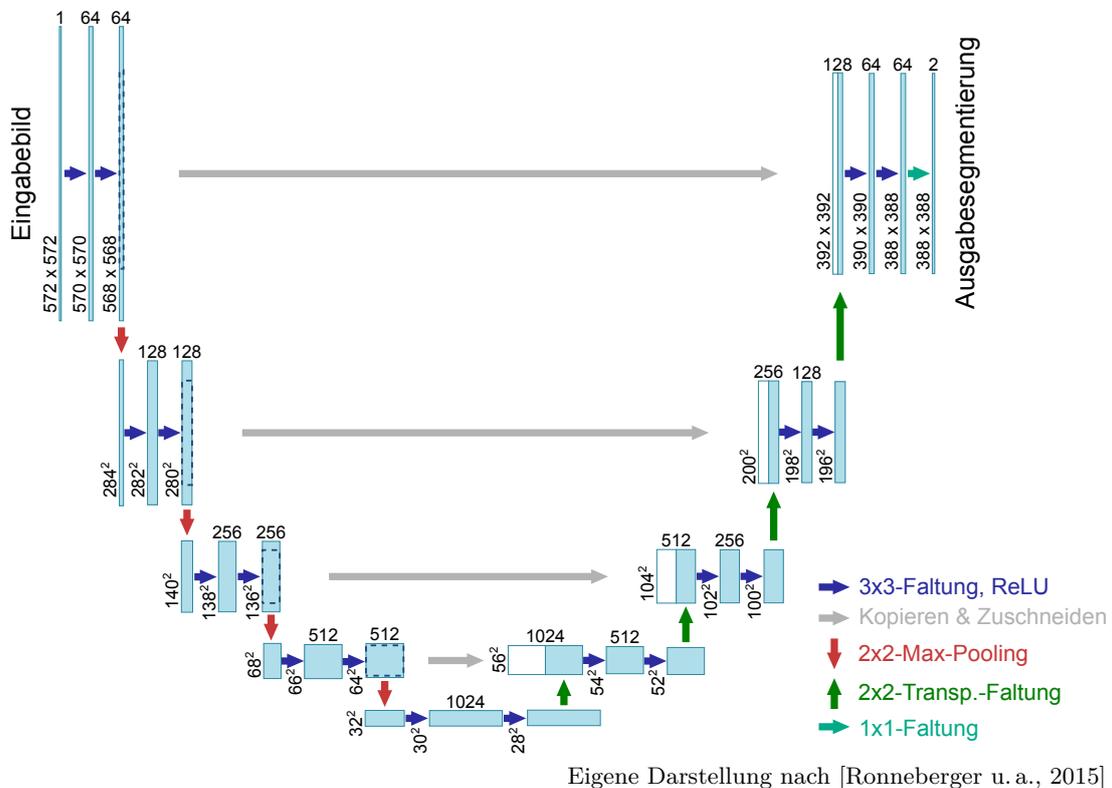


Abb. 3.8: Originale U-Net-Architektur aus [Ronneberger u. a., 2015]. Die hellblauen Rechtecke entsprechen mehrkanaligen Merkmalskarten. Die Anzahl der Kanäle ist über jeder Merkmalskarte angegeben. Die Höhe und Breite der Merkmalskarten sind unten links neben den Rechtecken zu finden. Die weißen Rechtecke im expandierenden Pfad stellen ausgeschnittene und kopierte Merkmalskarten aus dem kontrahierenden Pfad dar.

Kapitel 4

Segmentierung von Hautschichten der Maus in OCT-Bilddaten

Abgesehen von ihrem Einsatz in der Ophthalmologie wird die OCT auch im Rahmen von präklinischen Studien zur Untersuchung von Tiermodellen genutzt. Für die Evaluation des Kryolipolyseverfahrens, einer kosmetischen Methode zur nichtinvasiven subkutanen Fettreduktion, werden Mäuse als neues Tiermodell erprobt. Zur schnellen automatischen Analyse von Maushautschichten in OCT-Bilddaten werden in diesem Kapitel zwei verschiedene Segmentierungsmethoden vorgestellt. Die erste Methode basiert auf einer klassischen Segmentierungspipeline, in der eine Random-Forest-Klassifikation mit anschließender graphenbasierter Optimierung eingesetzt wird. Beim zweiten Ansatz hingegen werden Deep-Learning-Methoden angewandt, wobei ein FCN für die Segmentierung verwendet wird, um die Anzahl der erforderlichen Verarbeitungsschritte zu verringern. Im zweiten Teil des Kapitels wird ein systematischer Vergleich der beiden Methoden und der daraus resultierenden Ergebnisse vorgenommen.

4.1 Einleitung und Motivation

Die Zulassung neuer Arzneimittel oder Behandlungstechniken erfordert die Durchführung von präklinischen Versuchen und Studien an spezifischen Zell- und Tiermodellen, welche biologische Eigenschaften oder Krankheiten des Menschen nachahmen. Basierend auf den in den Studien gewonnenen Erkenntnissen lassen sich Medikamentendosierungen oder Parametereinstellungen eines Behandlungsverfahrens testen und optimieren, ohne dass Menschen möglichen Risiken ausgesetzt werden. Präklinische Studien unterteilen sich in In-vitro- und In-vivo-Studien. Der lateinische Ausdruck „in vitro“ bedeutet „im Glas“ und steht für organische Prozesse, die außerhalb des lebenden Organismus ablaufen. Dabei werden bestimmte Zellen, wie z. B. Haut- oder Krebszellen, von Menschen oder Tieren entnommen und kultiviert. Mithilfe von In-vitro-Studien werden bestimmte Wirkmechanismen auf zellulärer Ebene untersucht. Der Begriff „in vivo“ bedeutet „im Lebenden“ und beschreibt Studien an lebenden Organismen. Die-

se Studien werden an Tieren wie Mäusen, Ratten oder Kaninchen durchgeführt. Je nach Studie werden bestimmte Tierarten oder (In-)Zuchtstämme verwendet, die gewünschte genetische Voraussetzungen erfüllen oder Krankheiten entwickeln. Anhand von In-vivo-Studien werden Arzneimittel nicht nur auf ihre Wirkung, sondern auch auf ihre Aufnahme, Verteilung und ihren Stoffwechsel im Tiermodell untersucht. Die Erprobung von Behandlungstechniken hingegen kann in der Regel nur an Tiermodellen erfolgen.

Die Kryolipolyse ist ein neuartiges kosmetisches Verfahren zur nichtinvasiven Reduktion von subkutanen Fettdepots. Die meisten konservativen Methoden zur Fettreduktion gehen mit einem invasiven Eingriff einher, bei dem es zu potenziellen Komplikationen im Heilungsprozess und zur unerwünschten Narbenbildung kommen kann. Die bekannteste und auch am meisten angewandte unter ihnen ist die Liposuktion (*dt.*: Fettabsaugung) [ISAPS, 2020], bei der über mehrere verteilte Hautschnitte das subkutane Fett mithilfe einer Kanüle abgesaugt wird. Im Gegensatz dazu ermöglicht die Kryolipolyse die nichtinvasive Zerstörung von subkutanem Fettgewebe durch selektive Kälteinduktion [Manstein u. a., 2008]. Bei dieser Technik werden die Fettdepots einige Zeit starker Kälte ausgesetzt. Dabei wird die Haut in einen Applikator eingesaugt und kühlt mithilfe eines Peltier-Elements das subkutane Fettgewebe herunter. Ein spezielles Vlies sowie ein Kontaktgel verhindern Erfrierungen der darüber liegenden Hautschichten. Ab einer Temperatur von ca. 4–5 °C kristallisieren die Fettzellen, wodurch ihre Apoptose (Zelltod) eingeleitet wird. Die abgestorbenen Fettzellen werden anschließend vom Körper abgebaut. Dieser Prozess dauert durchschnittlich mehrere Wochen bis Monate.

Die Kryolipolyse ist bis heute in zahlreichen klinischen Studien über vier Millionen Mal beim Menschen angewendet worden [Klein u. a., 2017]. In einer initialen Studie untersuchten Zelickson u. a. [2009] die Wirksamkeit des Kryolipolyseverfahrens an einem Tiermodell. Sie haben die Dicke der subkutanen Fettschicht von Yucatan-Schweinen (Mischweine) vor und nach einer Kryolipolysebehandlung bestimmt. Die Bestimmung der Schichtdicke wurde sowohl in den Ultraschallaufnahmen als auch in groben Sektionen der Haut nur an einzelnen Positionen bestimmt. Für fortführende klinische Studien soll ein neues Tiermodell erprobt werden, da das Schweinmodell sehr komplex und kostenintensiv ist. Daher soll die Eignung von Mäusen als Tiermodell in zukünftigen klinischen Studien zur Kryolipolyse sowie weiteren Verfahren untersucht werden. Im Vergleich zu den Yucatan-Schweinen sind Mäuse deutlich kostengünstiger und leichter in der Handhabung, was sich hauptsächlich durch den enormen Größenunterschied begründen lässt. Jedoch kann aufgrund der geringen Größe der Maus keine Ultraschallbildgebung mehr für die Dickenmessung genutzt werden.

Außer ihrem Einsatz in der klinischen Routine findet die OCT auch in präklinischen In-vivo-Studien Anwendung [Salma u. a., 2018; Singer u. a., 2007; Srinivas u. a., 2004]. Allerdings stellt die Analyse und Auswertung dreidimensionaler OCT-Bilddaten von Tiermodellen eine äußerst herausfordernde Aufgabe dar. Fehlende standardisierte

Aufnahmeroutinen, sowie variierende Anatomien der Tiere erschweren die Auswertung. Eine semantische Segmentierung im OCT-Bild ist erforderlich, um im Gegensatz zu [Zelickson u. a., 2009] eine detaillierte Dickenmessung der subkutanen Fettschicht zu erhalten, mit der sich auch lokale Änderungen über die Zeit bestimmen lassen. Jedoch geht die Erstellung manueller Segmentierungen mit einem hohen Zeitaufwand einher und ist daher in der Praxis eher unpraktikabel. Die Dickenbestimmung weiterer Hautschichten, wie z. B. der Dermis, stellt ebenfalls interessante wissenschaftliche Fragestellungen dar, sodass sich der Segmentierungsaufwand weiter vergrößert.

Für eine schnelle und präzise Dickenmessung der oberen Maushautschichten in OCT-Volumen (siehe Abb. 4.1) werden daher in diesem Kapitel zwei Segmentierungsmethoden vorgestellt. Das erste Verfahren besteht aus einer klassischen Segmentierungspipeline [Handels, 2009, S. 49 ff.], deren Kern eine Kombination aus maschinellem Lernen und Graphentheorie bildet. Im Gegensatz dazu basiert die zweite Segmentierungsmethode auf moderneren Deep-Learning-Methoden. In diesem einführenden Methodenkapitel werden in einem ausführlichen Vergleich die beiden Methoden gegenübergestellt sowie ihre Vor- und Nachteile analysiert.

Nach einer Vorstellung der Literaturübersicht wird zunächst der verwendete OCT-Bilddatensatz vorgestellt (Kapitel 4.2.1). Anschließend folgt die Beschreibung der entwickelten Segmentierungsverfahren in Kapitel 4.2.2 und 4.2.3.1. Anschließend wird in Kapitel 4.3 die Segmentierungsleistung beider Verfahren untersucht und die Ergebnisse präsentiert. Der letzte Teil des Kapitels (Kapitel 4.4) umfasst die Diskussion der Ergebnisse und die daraus gezogenen Schlussfolgerungen.

4.1.1 Literaturübersicht

Bisher sind nur wenige Arbeiten zur Segmentierung von Hautschichten in OCT-Aufnahmen publiziert worden. Im Vergleich dazu gibt es aufgrund der umfassenden Anwendung von OCT in der Ophthalmologie zahlreiche Veröffentlichungen, die sich mit der Segmentierung von Retinaschichten befassen. Eine ausführliche Übersicht ist in [DeBuc, 2011; Hu u. a., 2021; Kafieh u. a., 2013] zu finden.

Die Vielzahl der publizierten Arbeiten nutzen graphentheoretische Ansätze, die entweder auf 2D- oder 3D-Modellen aufbauen. In ihrer Arbeit stellen Chiu u. a. [2010] einen simplen graphenbasierten 2D-Ansatz zur Segmentierung von sieben verschiedenen Retinaschichten vor. Als Kosten für das Graphenmodell wird die Gradienteninformation der Intensitätswerte verwendet. Allerdings werden die Kontextinformationen benachbarter B-Scans durch das 2D-Modell nicht berücksichtigt, was zu räumlichen Inkonsistenzen im Segmentierungsergebnis führen kann. Li u. a. [2006a] schlagen in ihrer Arbeit eine 3D-Graphen-basierte Methode zur simultanen Segmentierung mehrerer Retinaschichten vor. Ihr Ansatz basiert auf einer speziellen Graphenkonstruktion, sodass der „minimale s-t-Schnitt“ (siehe Kapitel 3.1.1.2) eine Menge von Oberflächen darstellt, die den Grenzübergängen der zu segmentierenden Retinaschichten entsprechen.

Zusätzliche Glattheitsbedingungen stellen sicher, dass benachbarte Oberflächenpunkte nicht zu weit voneinander entfernt sind. Außerdem begrenzen Interaktionsbedingungen die Abstände zwischen den einzelnen Oberflächen auf einen festgelegten Bereich. Aufgrund seiner allgemeinen Formulierung ist dieser Ansatz sehr populär geworden, sodass dieser von späteren Arbeiten adaptiert und weiterentwickelt wurde [Garvin u. a., 2008; Garvin u. a., 2009; Haeker u. a., 2007]. Garvin u. a. [2009] haben in ihrer Arbeit bspw. die konstanten Beschränkungen durch eine flexible Formulierung ersetzt und führen gleichzeitig echte regionale Kosten ein. Nachteilig an den beschriebenen graphenbasierten Segmentierungsverfahren ist abgesehen von der Graphenkonstruktion die komplexe Gestaltung der Kostenfunktionen. Hierzu werden meist Kantentfilter genutzt, die jedoch für jede Oberfläche separat justiert werden müssen, was z. B. bei Aufnahmen pathologischer Retinae oder beim Wechsel zu einem anderen OCT-Scanner nötig wäre. Basierend auf dem Ansatz von Garvin u. a. werden weitere Erweiterungen vorgestellt, die texturbasierte Merkmale [Antony u. a., 2012] sowie weiche Beschränkungen umfassen.

Überdies werden Methoden des maschinellen Lernens vorgeschlagen [Karri u. a., 2016; Lang u. a., 2013b; Vermeer u. a., 2011], welche spezifische Merkmale der Retinaschichten erlernen und so eine pixelweise Klassifikation durchführen. Ein wesentlicher Vorteil ist die leichte Anpassung an neue Bilddatensätze, da als einzige Voraussetzung lediglich ein entsprechender Trainingsdatensatz benötigt wird. Allerdings ist bei den pixelweisen Klassifikationsmethoden im Gegensatz zu den graphenbasierten Modellen die Anordnung der einzelnen Retinaschichten nicht gewährleistet, sodass strenge topologische Beziehungen möglicherweise nicht eingehalten werden.

Durch die Kombination von maschinellen Lernmethoden und graphenbasierter Modelle können die Vorteile beider Ansätze kombiniert werden. Hierzu wird das Klassifikationsergebnis in die Kostenfunktion des Graphenmodells integriert. Antony u. a. [2013] nutzen einen RF-Klassifikator zur Gestaltung der Kostenfunktion ihres graphenbasierten Segmentierungsansatzes. Sie trainieren sowohl einen On-Surface- als auch einen In-Region-Klassifikator, um oberflächenbasierte und regionenbasierte Kostenterme bestimmen zu können. Srinivasan u. a. [2014] segmentieren Retinaschichten mit einem Algorithmus, der aus einer Kombination von Entrauschung, Support-Vector-Machine-Klassifikation, Graphentheorie und dynamischer Programmierung besteht. In einer ähnlichen Herangehensweise wie in [Antony u. a., 2013] segmentieren Lang u. a. [2013a] Retinaschichten sowohl von Gesunden als auch von Patienten mit Multipler Sklerose. Dabei nutzen sie ebenfalls einen RF zur Klassifikation der Grenzübergänge der Retinaschichten.

Aufgrund ihrer hohen Popularität in der Computer Vision ist das Forschungsinteresse an Deep-Learning-Methoden, insbesondere an CNNs, in der medizinischen Bildanalyse immens gestiegen. Fang u. a. [2017] schlagen ein auf Bildpatches basierendes CNN mit einer Kombination aus Graphentheorie und dynamischer Programmierung vor, um neun Retinaschichten in OCT-Bildern von Patienten mit altersbedingter Makuladegeneration zu segmentieren.

neration zu segmentieren. Darauf aufbauend untersuchen Hamwood u. a. [2018] die Auswirkungen der Patchgröße sowie der Netzwerkkomplexität auf die Gesamtleistung der Segmentierung und schlagen eine modifizierte CNN-Architektur für die Analyse von OCT-Bildern gesunder Augen vor. In einer weiteren Publikation trainieren Loo u. a. [2018] CNNs unter dem Einsatz von Transferlernen, um die ellipsoide Zone in OCT-Bildern zu segmentieren.

Die von Ronneberger u. a. [2015] vorgeschlagene U-Net-Architektur (siehe Kapitel 3.2.2.3) beschreibt ein FCN, das aus einem Multiskalen-Encoder und -Decoder besteht, die über Skip Connections verbunden sind. Im Vergleich zu den zuvor erwähnten patchbasierten Ansätzen erzeugen FCNs pixelweise Segmentierungen ihres Eingangsbildes, wodurch eine Zusammenführung der klassifizierten Bildpatches entfällt. Ronneberger u. a. haben in ihrer Arbeit weiterhin gezeigt, dass bereits eine geringe Anzahl von Trainingsbilddaten unter der ausgiebigen Anwendung erweiterter Augmentierungstechniken ausreicht, um das U-Net effizient trainieren zu können, wodurch sich das Verfahren besonders für den Einsatz in der biomedizinischen Bildanalyse eignet. Venhuizen u. a. [2017] setzen das U-Net ein, um die gesamte Retina zu segmentieren, wobei sie die Architektur um zwei weitere Auflösungsstufen ergänzen, um das rezeptive Feld zu vergrößern. Denselben Effekt erzielen Devalla u. a. [2018] in ihrer Arbeit durch den Einsatz dilatierter Faltungskerne, wodurch sich die Netzparameteranzahl signifikant reduzieren lässt. Inspiriert von der Arbeit von He u. a. [2016] werden in das vorgeschlagene DRUNET (Dilated-Residual U-Net) kurze residuale Verbindungen integriert, die einen besseren Gradientenfluss gewährleisten. Devalla u. a. haben das DRUNET zur Segmentierung von Gewebsschichten des Sehnervenkopfes in OCT-Bilddaten eingesetzt.

Wie zuvor erwähnt, sind bisher nur wenige Methoden zur Segmentierung von Hautschichten in OCT-Bilddaten veröffentlicht worden. Eine Reihe von Arbeiten befasst sich mit Ansätzen zur Segmentierung der menschlichen Epidermis in OCT-Bildern, was die Bestimmung der Hautoberfläche und der dermoepidermale Junktionszone einschließt. Weissman u. a. [2004] nutzen Gauß-ähnliche Multiskalen-Shapelets zur Segmentierung der Epidermis. Hori u. a. [2006] schlagen einen auf Pixelintensitäten basierenden Segmentierungsansatz vor, bei dem in jedem A-Scan Intensitätsspeaks detektiert werden, welche die Hautoberfläche repräsentieren. In der Arbeit von Josse u. a. [2011] ist ein Segmentierungsalgorithmus zur effizienten Bestimmung der Epidermisdicke mit sehr kurzer Berechnungszeit beschrieben. Da die Haut ähnlich wie die Retina in Schichten aufgebaut ist, sind ebenfalls graphenbasierte Ansätze zur Segmentierung vorgeschlagen worden. So nutzen Li u. a. [2015] ein Graphenmodell zur Bestimmung der Epidermis. Ähnlich wie Chiu u. a. [2010] kommt hier eine vereinfachte Version des Dijkstra-Algorithmus zum Einsatz. Die Kostenfunktion wird mithilfe vertikaler Gradienten konstruiert. Im Gegensatz dazu stellen Srivastava u. a. [2018] in ihrer Arbeit einen dreidimensionalen graphenbasierten Segmentierungsansatz zur Segmentierung der Epidermis vor, der in Anlehnung zu [Garvin u. a., 2008] kanten- und regionenba-

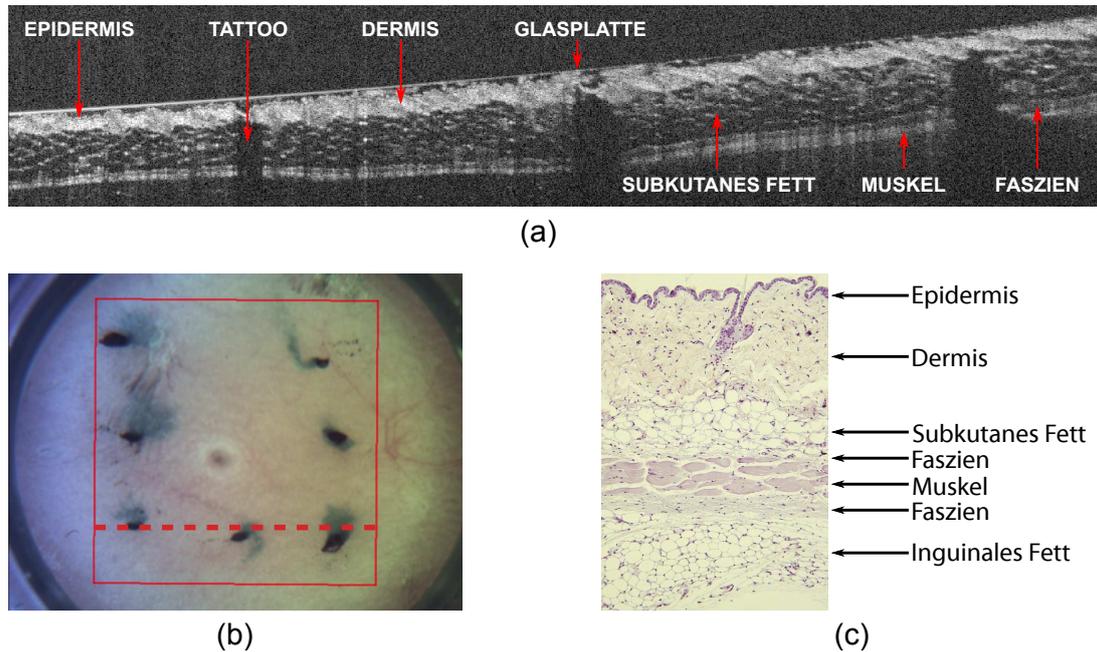


Abb. 4.1: (a) B-Scan der Mäusehaut. Das rote Rechteck im Übersichtsbild (b) stellt das FOV dar und die gestrichelte Linie zeigt die Scan-Position an. Die Tätowierungen sind U-förmig am äußeren Rand des FOV angeordnet. Ein histologischer Schnitt der Mäusehaut ist in (c) dargestellt.

sierte Informationen als Kosten nutzt. Die Segmentierungen werden genutzt, um die Rauheit der Haut bestimmen zu können. Zur Segmentierung mehrerer menschlicher Hautschichten in OCT-Bildern verwenden Calderon-Delgado u. a. [2018] eine U-Net-Variante, die eine Hochskalierung von Merkmalskarten mit Unpooling-Operationen anstelle von transponierten Faltungen durchführt. Mit einer ähnlichen Fragestellung wie in diesem Kapitel beschäftigen sich die Arbeiten von Sheet u. a. [2013, 2014]. Sie segmentieren in OCT-Aufnahmen der Mäusehaut Epidermis-, Papillar-, Dermis- und Fettgewebe. Zur Segmentierung wird ein statistisches physikalisches Modell der Photonen-Gewebe-Interaktionen in Kombination mit einem RF-Klassifikator genutzt. Fortführend wurde dieser Ansatz auch anhand von OCT-Bilddaten der menschlichen Mundschleimhaut evaluiert [Sheet u. a., 2014]. In einer nachfolgenden Forschungsarbeit haben die Autoren Denoising Autoencoder untersucht, um gewebespezifische Speckle-Präsentationen für die Klassifikation zu erlernen [Sheet u. a., 2015].

Wie einleitend erwähnt, basieren die in diesem Kapitel vorgestellten Segmentierungsmethoden einerseits auf dem Prinzip einer klassischen Segmentierungspipeline und andererseits auf modernen Deep-Learning-Methoden, deren grundlegendes Konzept in [Kepp u. a., 2018] sowie [Kepp u. a., 2019a] veröffentlicht worden ist. Die erste Methode besteht aus zwei Teilen. Zunächst werden durch einen RF-Klassifikator die Mäusehaut-

schichten in B-Scans eines OCT-Volumens segmentiert. Mit Ausnahme ihrer Anwendung in OCT-Bilddaten haben RF-Klassifikatoren auch eine hohe Präzision in anderen medizinischen Anwendungen bewiesen, z. B. bei der Segmentierung ischämischer Schlaganfallläsionen in MRT-Bilddaten [Maier u. a., 2015, 2017]. Anschließend wird die RF-Segmentierung mittels graphenbasierter Optimierung weiter verfeinert. Zu diesem Zweck wird der Ansatz von Li u. a. [2006a] verwendet, um einerseits die Genauigkeit sowie die Robustheit zu erhöhen und andererseits das Segmentierungsproblem auf den dreidimensionalen Raum zu erweitern. Die zweite Segmentierungsmethode basiert auf der Verwendung von CNNs. Genauer werden hier die Maushautschichten mithilfe eines FCNs segmentiert. Das FCN fußt auf der bekannten U-Net-Architektur und wird durch einige Modifikationen erweitert.

4.2 Material und Methoden

4.2.1 Verwendete Bilddaten

Der OCT-Bilddatensatz besteht aus 20 OCT-Volumen der Leistenregion von insgesamt vier Labormäusen des häufig verwendeten Inzuchtstammes C57BL/6. Die OCT-Bilder wurden vor, unmittelbar nach und 30 Tage nach der Kältebehandlung aufgenommen. Beim verwendeten OCT-System handelt es sich um das Telesto II SD-OCT-System (Thorlabs, Inc.) mit einer zentralen Wellenlänge von 1310 nm, einer axialen Auflösung von $3,4 \mu\text{m}$ und einer transversalen Auflösung von $6,5 \mu\text{m}$. Bei einer Bildvolumengröße von $1024 \times 1307 \times 1307$ Voxel beträgt der gesamte Abbildungsbereich $3,5 \times 8,5 \times 8,5$ mm. Für die Aufzeichnung der Daten wurde ein Probenabstandshalter (angewinkelte Glasplatte) verwendet, der direkt auf der Haut der Maus platziert wurde und den zu untersuchenden Hautbereich fixiert, um die Atembewegungen so weit wie möglich zu reduzieren. Außerdem wurde zwischen dem Probenabstandshalter und der Mäusehaut Wasser aufgebracht, was die Streuung des einfallenden Lichtstrahls verringert. Abbildung 4.1 zeigt ein Beispiel eines B-Scans mit FOV-Übersichtsbild sowie der Histologie der entsprechenden Gewebestrukturen. Um sicherzustellen, dass derselbe Hautbereich über einen längeren Zeitraum hinweg mit dem OCT abgebildet werden kann, wurden die Mäuse vor der ersten Untersuchung tätowiert, um eine Relokalisierung zu gewährleisten. Aufgrund starker Bewegungs- und Schattenartefakte mussten zwei Volumen ausgeschlossen werden, sodass der finale Datensatz insgesamt 18 OCT-Volumen umfasst.

Sowohl für das Training als auch für die Evaluation von überwachten maschinellen Lernalgorithmen wird eine Grundwahrheit benötigt. Da die manuelle Segmentierung ganzer OCT-Volumen aufgrund des hohen Aufwands nicht durchführbar war, wurden vier repräsentative B-Scans pro Volumen ausgewählt und von zwei erfahrenen Experten manuell segmentiert. Für die Annotation kam das Software-Tool ITK-Snap [Yushkevich u. a., 2006] zum Einsatz. In Ergänzung zur subkutanen Fettschicht (SFS) wurden

auch Epidermis und Dermis segmentiert, um sie für zukünftige Auswertungen nutzen zu können. Da die Unterscheidung zwischen Epidermis und Dermis an Auflagepunkten der Glasplatte des Probenabstandshalters kaum möglich (siehe Abb. 4.1 (a)), sind diese als (Epi-)Dermis (EDS) zusammengefasst worden. Aufgrund ihrer hohen Intensitätswerte lässt sich die Muskel- und Faszienschicht (MFS) optimal im B-Scan lokalisieren. Da diese direkt unter der SFS liegt, sind sie als weitere Klasse in die Segmentierung integriert worden, um die Robustheit der Segmentierung zu verbessern. Die MFS wird zur Konstruktion des Graphen genutzt, mit dem die Feinabstimmung der Hautschichtgrenzen vorgenommen wird (siehe Kapitel 4.2.2.3). Weiterhin sollen die segmentierten Tattoo-Punkte als Landmarken für die räumliche Anpassung (Registrierung) von Aufnahmen verschiedener Zeitpunkte dienen können. Demzufolge sind auch die Tätowierungen in das Klassifikationsproblem einbezogen worden. Die Regionen oberhalb und unterhalb der genannten Zielstrukturen werden als Hintergrund betrachtet. Alle Pixel jedes B-Scans sind dementsprechend einer der nachfolgenden fünf Klassen zugewiesen worden: EDS (■), SFS (■), MFS (■), Tattoo (■) und Hintergrund. Insgesamt besteht die Grundwahrheit des Datensatzes aus 72 B-Scans mit jeweils zwei individuellen Expertensegmentierungen.

4.2.2 Random-Forest-Segmentierung mit graphenbasierter Optimierung

Die Segmentierungsaufgabe des hier vorgeschlagenen Verfahrens besteht darin, jeder Bildposition $\mathbf{x} \in \Omega^{(3)}$ eines OCT-Volumens I einen Label-Wert l aus $L = \{0, \dots, K-1\}$ für $K = 5$ Klassen zuzuordnen. Die hierzu verwendete Segmentierungspipeline ist in Abbildung 4.2 dargestellt, die sich in drei Schritte unterteilen lässt. Zunächst wird ein OCT-Volumen vorverarbeitet, um einerseits die Bildqualität zu erhöhen und andererseits optimale Bedingungen für die nachfolgenden Schritte zu schaffen. Das vorverarbeitete Volumen wird dann von einem RF-Klassifikator B-Scan-weise segmentiert. Grobe Segmentierungsfehler werden durch einfache Nachbearbeitung vor der Ausführung des letzten Schritts korrigiert. Schließlich erfolgt eine räumliche Regularisierung und Verfeinerung der Grenzübergänge der EDS und SFS durch die graphenbasierte Optimierung. Das Graphenmodell berücksichtigt ebenfalls die anatomische Anordnung der Hautschichten und vermeidet so topologische Fehler. Nachfolgend werden die drei Schritte der Segmentierungspipeline im Detail vorgestellt.

4.2.2.1 Vorverarbeitung

Durch eine Vorverarbeitung werden die OCT-Volumen optimal für die nachfolgenden Schritte des Segmentierungspipeline vorbereitet. Die für gewöhnlich artefaktbehafteten OCT-Bilddaten werden so aufbereitet, dass zum einen die Aussagekraft der Merkmale des RF-Klassifikators erhöht wird und zum anderen Kanteninformationen verstärkt

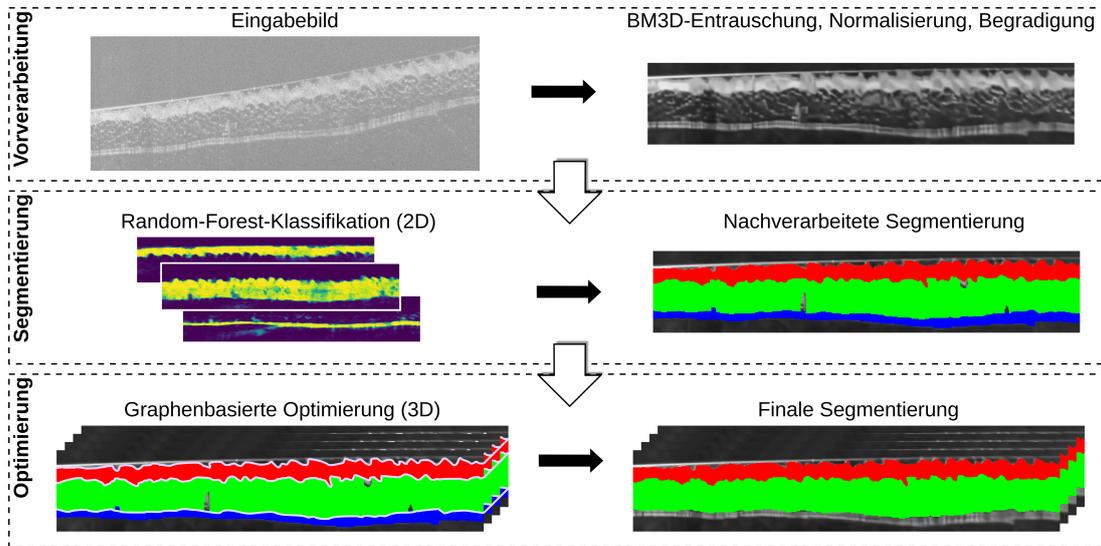


Abb. 4.2: Segmentierungspipeline des vorgeschlagenen Algorithmus.

werden, die essenziell für die Kostenfunktion des Graphenmodells sind. Daher hat die Vorverarbeitung der Bilddaten einen erheblichen Einfluss auf die Segmentierungsgüte der nachfolgenden Schritte. Im Folgenden werden die einzelnen Vorverarbeitungsschritte im Detail beschrieben.

Aufgrund ihrer sehr großen Ausgangsgröße erfordern die OCT-Volumen eine große Menge an Speicherplatz, wodurch lange Berechnungszeiten entstünden. Daher wird zuerst ein Resampling der OCT-Volumen vorgenommen, wobei deren Auflösung um die Hälfte reduziert wird. Trotz der erheblichen Größenreduzierung ist der verbleibende Detailgrad für die nachfolgenden Prozesse ausreichend.

Obwohl das Speckle-Rauschen (siehe Kapitel 2.1) grundsätzlich wichtige Informationen enthält, kann es für die automatische Analyse hinderlich sein. Da insbesondere die Merkmalsextraktion für den RF-Klassifikator durch Rauschen stark beeinträchtigt würde, wird eine OCT-Volumen glättung durchgeführt. Um den Verlust von Kanteninformationen, die für die Segmentierung wichtig sind, durch den Glättungsprozess zu verhindern, wird ein 3D-Block-Matching-Filter (*engl.* block-matching and 3D filtering, kurz: BM3D) [Dabov u. a., 2009] verwendet, der die OCT-Volumen kantenerhaltend entrauscht.

Die OCT-Volumen weisen Intensitätsschwankungen sowohl zwischen als auch innerhalb der B-Scans auf. Zur Kompensation dieser Inhomogenitäten werden die OCT-Volumen B-Scan-weise auf das Intervall $[0,1]$ skaliert. Anschließend wird ein störender Intensitätsgradient innerhalb der B-Scans über eine Ausgleichsgerade korrigiert.

Die unterschiedliche Beschaffenheit der einzelnen Hautschichten der Maus im OCT sorgen für eine erhöhte Variabilität in den Merkmalen, was zu schlechteren Klassifi-

kationsergebnissen durch den RF führen könnte. Darum wird abschließend die Bildinformationen anhand einer Referenzstruktur begradigt. Häufig wird eine Begradigung (*engl.*: flattening) als Vorverarbeitung retinaler OCT-Aufnahmen genutzt, um die patientenindividuelle Krümmung der Retina zu kompensieren. Wegen ihrer hohen Intensitätswerte wird die Dermis als Referenzstruktur verwendet. Durch die Kombination aus Mittelwertfilter und Schwellenwertverfahren kann diese grob erfasst werden. Anschließend wird ein Polynom zweiter Ordnung durch die oberen Pixel der detektierten Dermis gelegt. Zur Begradigung des Bildes wird nun jede Spalte des Bildes vertikal verschoben, sodass aus dem zuvor angepassten Polynom nun eine gerade horizontale Linie resultiert. Zur Beschleunigung wird die Begradigung nur bei jedem 20. B-Scan ausgeführt und die Verschiebungen aller übrigen B-Scans werden linear interpoliert. Überflüssige Hintergrundbereiche werden schließlich durch Beschneiden entfernt. Ein Beispiel für einen B-Scan vor und nach der Vorverarbeitung ist in Abbildung 4.2 dargestellt.

4.2.2.2 Random-Forest-basierte Klassifikation der Hautschichten

Die Segmentierung eines OCT-Volumens erfolgt B-Scan-weise durch einen RF-Klassifikator, dessen methodische Grundlagen in Kapitel 3.2.1 beschrieben sind. Die Klassifikation erfolgt anhand bildbasierter Merkmale, die jeder Bildposition $\mathbf{x} \in \Omega^{(2)}$ eines B-Scans einen Label-Wert l zuweisen. Eine B-Scan-weise 2D-Segmentierung der OCT-Volumen ist zunächst ausreichend, da eine 3D-Segmentierung durch die graphenbasierte Optimierung erfolgt. Außerdem wird durch die 2D-Segmentierung der erforderliche Speicherbedarf gering gehalten. Das Design der Merkmale spielt eine entscheidende Rolle und hat einen erheblichen Einfluss auf die Segmentierungsgüte des Klassifikators. Man spricht daher auch von sogenannten handverlesenen Merkmalen (*engl.* handcrafted features), die spezielle Struktureigenschaften im Bild beschreiben. Nachfolgend werden zunächst die verwendeten Merkmale vorgestellt. Anschließend werden sowohl das Training als auch die Anwendung des RF-Klassifikators beschrieben.

Merkmale Anstelle eines komplexen statistischen physikalischen Modells wie in [Sheet u. a., 2013, 2014] werden für den RF-Klassifikator intensitäts-, kanten- sowie descriptorbasierte Bildmerkmale verwendet. Die Extraktion einiger Merkmale findet nur in einer lokalen Nachbarschaft statt, die, sofern nicht anders angegeben, einem Bildausschnitt der Größe 9×15 Pixel entspricht. Mithilfe intensitätsbasierter Merkmale lassen sich gewebespezifische Bildrepräsentationen erfassen. Z. B. weist die EDS ein deutlich helleres Grauwertprofil auf als die SFS. Durch die Bestimmung der kumulierten Summen kann ermittelt werden, ob sich die betrachtete Position ober- oder unterhalb der EDS befindet. Des Weiteren werden Kanteninformationen genutzt, mit denen die Strukturbeschaffenheit und die Übergänge der Hautschichten modelliert werden. Zum Erlernen der Wabenstruktur der SFS wird der DoG-Filter (*engl.*: Difference

of Gaussians) verwendet, der u. a. auch als „Blob-Detektor“ eingesetzt wird. Außer der Filterantwort pro Pixel wird die mittlere Antwort des DoG-Filters in einer 15×15 -Umgebung als ergänzendes Merkmal benutzt.

Um weitere kontextbezogene Informationen in den Merkmalsraum einzubeziehen, werden ergänzend BRIEF- und Census-Deskriptoren [Calonder u. a., 2012] als Merkmale einbezogen. BRIEF steht für „Binary Robust Independent Elementary Features“ und stellt einen Descriptor dar, der die Intensitätswerte zwischen mehreren zufällig gewählten Bildpunktpaaren in einer vorgegebenen Nachbarschaft vergleicht. Positive Differenzen werden als 0 und negative als 1 gespeichert. Der Census-Descriptor ist ähnlich aufgebaut wie der BRIEF-Descriptor, nur dass der Intensitätsvergleich nicht zwischen zufälligen Bildpunktpaaren erfolgt, sondern mit dem aktuell betrachteten mittleren Bildpunkt.

Jedes Merkmal wird auf drei verschiedenen Auflösungsstufen unter Verwendung der Gauß-Pyramide berechnet. Das Ergebnis der Merkmalsextraktion ist schließlich ein Vektor mit insgesamt 282 Einträgen, der für das Training und die Anwendung an den RF-Klassifikator übergeben wird. Eine Auflistung der verwendeten Merkmale und ihrer jeweiligen Anzahl von Einträgen im Merkmalsvektor ist in Abbildung 4.3 dargestellt.

Training und Klassifikation Das Training des RF-Klassifikators erfolgt anhand eines überwachten Lernschemas, bei welchem die manuell erstellten Segmentierungen der Maushautschichten (siehe Kapitel 4.2.1) als Grundwahrheit dienen. Der RF-Klassifikator wird aus insgesamt $T = 60$ Entscheidungsbäumen konstruiert, die während des Trainings so lange wachsen, bis keine weiteren Aufteilungen mehr vorgenommen werden können oder eine maximale Baumhöhe von 64 erreicht wird. Zur Bestimmung des Informationszuwachses wird der Gini-Index als Reinheitsmaß gewählt, da sich dieser im Vergleich zur Entropie schneller berechnen lässt. Um eine höhere Generalisierung des Ensembles zu gewährleisten, werden sowohl Bagging als auch zufällige Merkmalsauswahl angewendet [Breiman, 2001]. Beim Bagging werden $5 \frac{|D|}{T}$ Datenpunkte für das Training jedes Baums ausgewählt, während die Optimierung jeder Datenaufteilung eines n -ten Knotens anhand von $\sqrt{\dim(\mathbf{m})}$ zufällig ausgewählten Merkmalen erfolgt. Die Merkmalsextraktion wird auf einem 2×2 -Gitter ausgeführt, wodurch zum einen das Training beschleunigt wird und zum anderen weniger Speicherplatz für die Merkmalsvektoren benötigt wird. Aufgrund der hohen Bildauflösung ähneln sich benachbarte Pixel sehr, sodass dieses Vorgehen kaum Einfluss auf das Segmentierungsergebnis hat.

Die Klassifikation eines neuen OCT-Volumens erfolgt B-Scan-weise. Dabei wird für jeden einzelnen Pixel eines B-Scans der Merkmalsvektor bestimmt und durch den RF-Klassifikator einer Klasse zugewiesen. Um auch diesen Prozess zu beschleunigen, wird die Klassifikation zunächst auf einem 2×2 -Gitter ausgeführt. Alle nicht klassifizierten Pixel erhalten ihr Label durch eine Nearest-Neighbor-Interpolation. Anschließend

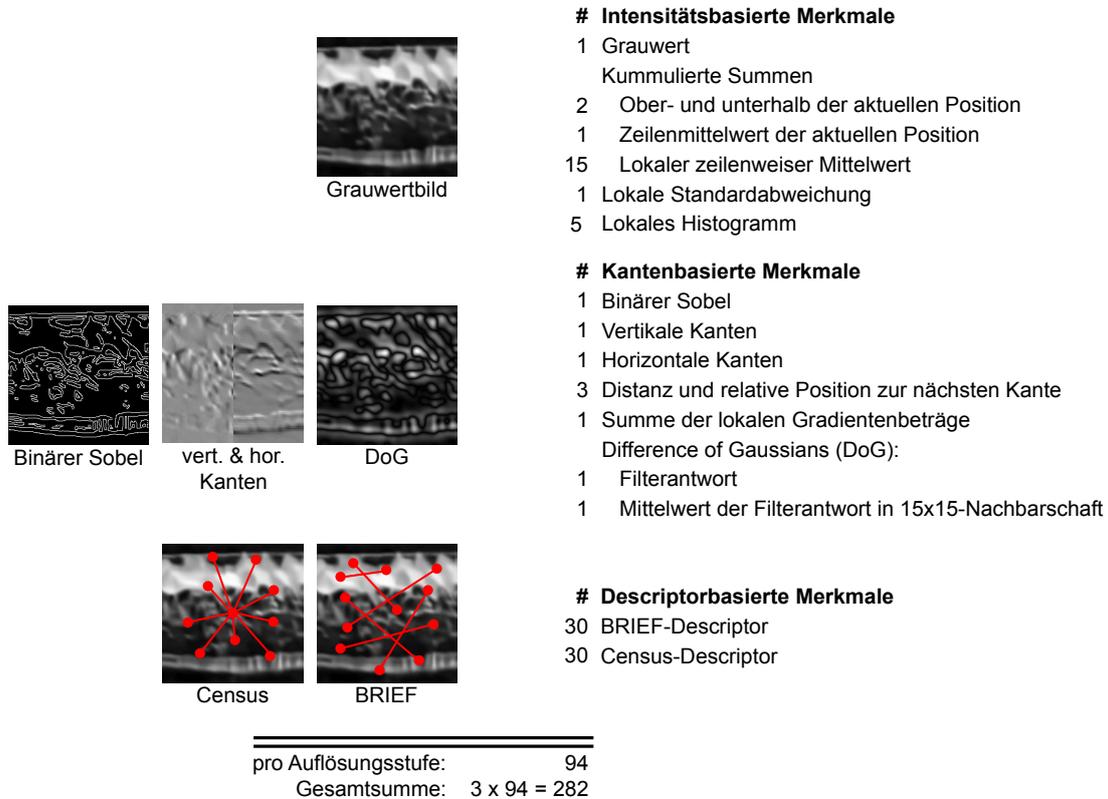


Abb. 4.3: Auflistung der verwendeten Merkmale und ihre jeweiligen Einträge im Merkmalsvektor. Sofern nicht anders angegeben werden lokale Merkmale in einer Nachbarschaft von 9×15 Pixeln berechnet.

wird eine weitere RF-Klassifikation an den Pixelpositionen durchgeführt, an denen benachbarte Pixel unterschiedliche Klassenzuordnungen erhalten haben, wie z. B. an Grenzübergängen. Anschließend wird die Segmentierung des RF-Klassifikators einer groben Nachverarbeitung unterzogen. So können falsch klassifizierte Bildpixel innerhalb einer Hautschicht oder im Hintergrund leicht erkannt und korrigiert werden, da die Hautschichten einer festen topologischen Struktur folgen.

4.2.2.3 Graphenbasierte Optimierung der Hautschichtgrenzen

Die Optimierung der segmentierten Hautschichtgrenzen stellt den letzten Schritt der Segmentierungspipeline dar und basiert im Wesentlichen auf der Arbeit Li u. a. [2006a], in der die Segmentierung von mehreren Oberflächen in Volumenbilddaten zu einem graphentheoretischen Problem (siehe Kapitel 3.1.1) umgestaltet wird. Die zentrale Neuerung der Methode basiert auf einer komplexen Graphenkonstruktion, die das Finden optimaler Oberflächensegmentierungen in die Bestimmung einer geschlossenen Men-

ge in einem knotengewichteten Digraphen überführt. Die geschlossene Menge mit den geringsten Kosten wird in dem konstruierten Graphen gefunden, indem ein minimaler s-t-Schnitt unter Anwendung des Max-Flow-Min-Cut-Algorithmus ermittelt wird.

Die Umformulierung der pixelweisen Segmentierung einzelner B-Scans in eine Detektion mehrerer Oberflächen im gesamten OCT-Volumen hat einen regulierenden Effekt auf das Segmentierungsergebnis. Hautschichten werden nun durch ihre obere und untere Grenzschicht beschrieben. Das Graphenmodell gewährleistet dabei eine strikte Berücksichtigung des topologischen Aufbaus der Haut und schränkt den Lösungsraum durch geeignete Nebenbedingungen sinnvoll ein. Auf diese Weise können auch größere Bereiche ausgebessert werden, die vom RF falsch klassifiziert wurden. Die Oberflächendetektion führt zu einer Regularisierung von Grenzverläufen innerhalb eines B-Scans, aber auch zu angrenzenden Nachbarschichten. Sie wird zur Verfeinerung der Segmentierung der EDS und SFS eingesetzt. Aufgrund ihrer morphologisch stark variierenden unteren Abgrenzung wird die MFS von der Optimierung ausgeschlossen. Lediglich die obere Abgrenzung ihrer Segmentierung wird für die Optimierung der SFS genutzt. Somit ergeben sich die drei Oberflächen bzw. Grenzübergänge: Hintergrund-EDS, EDS-SFS und SFS-MFS.

Im Folgenden wird eine Übersicht über den Ansatz aus Li u. a. [2006a] gegeben. Eine zentrale Besonderheit besteht in der nichttrivialen Konzeption des gewichteten Graphen $G = (V, E)$ mit Knoten $v \in V$, die aus Voxeln des Bildes bestehen, und Kanten $e = (v_x, v_y) \in E$, die jene miteinander verbinden. Da G ein Digraph ist, können Knoten Vor- und Nachfolger haben. Anders als in Kapitel 3.1.1 beschrieben, wird die Gewichtung zunächst anhand der Knoten und nicht der Kanten vorgenommen. Ihre Kosten werden durch

$$w(v_x) = \begin{cases} c(\mathbf{x}), & \text{wenn } x_2 = 0 \\ c(\mathbf{x}) - c(x_1, x_2 - 1, x_3), & \text{sonst} \end{cases} \quad (4.1)$$

definiert. Eine gesuchte Oberfläche wird dann durch eine geschlossene Menge C als Teilmenge von Knoten $C \subset V$ in G beschrieben, bei der sämtliche Nachfolger eines Knotens in C ebenfalls in C enthalten sind. Die Kosten der geschlossenen Menge entsprechen den aufsummierten Knotenkosten in C . Zum besseren Verständnis ist ein Koordinatensystem in Abbildung 4.4 dargestellt.

Li u. a. [2006a] nutzen verschiedene Kantenarten für die Graphenkonstruktion. Spaltenkanten verbinden Knoten innerhalb einer Bildspalte und sorgen dafür, dass alle Vorgänger eines sich in C befindenden Knoten auch in C enthalten sein müssen. Spaltenübergreifende Kanten verbinden die Knoten der einzelnen Spalten miteinander und sorgen so für den räumlichen Zusammenhang der späteren Oberfläche. Die maximale Auslenkung der Oberfläche wird durch die Parameter Δ_{x_1} und Δ_{x_3} bestimmt, wodurch sich die Glättungsbedingungen der spaltenübergreifenden Kanten regulieren lassen. Die letzte Kantenkategorie sind die Oberflächenkanten. Da im Modell die einzelnen Oberflächen durch N separate Graphen $G_i = (V_i, E_i)$ mit $i = 1, \dots, N$ dargestellt werden,

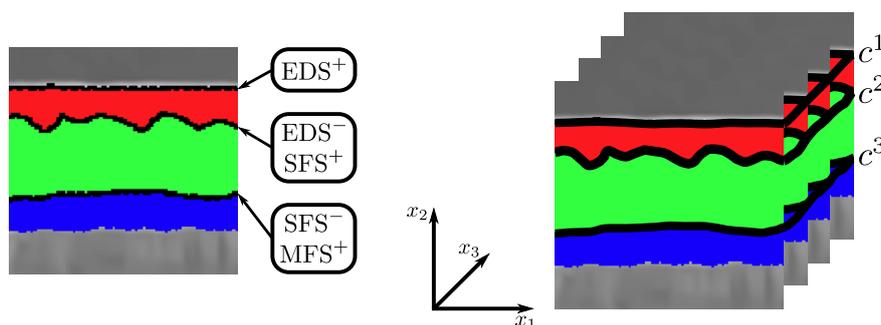


Abb. 4.4: Die RF-Segmentierungen tragen hauptsächlich für die Gestaltung der Kostenfunktionen c^{1-3} der einzelnen Oberflächen des Graphen bei (rechts). Die Voxel der Schichtübergänge werden anhand eines vertikalen Kantenfilters detektiert, der jeweils auf den binarisierten Masken der EDS, SFS und MFS angewandt wird (links).

wird ihre gegenseitige Wechselwirkung durch die Oberflächenkanten formuliert. Der minimale und maximale Abstand zwischen zwei Oberflächen wird über die Parameter δ_l und δ_u geregelt.

Ferner werden spezielle Modifikationen des Graphenmodells vorgenommen. Zum einen führen die Tattoos teilweise zu Unterbrechungen in den Oberflächen und müssen daher im Graphenmodell berücksichtigt werden. Zum anderen kann durch die Kryolipolyse die SFS an manchen Stellen komplett verschwinden. Hierzu werden zunächst alle Spalten betrachtet, in denen keine SFS erkannt wurde und anschließend alle Kanten zu diesen Spalten entfernt. Um die Tattoo-Areale noch besser modellieren zu können, werden auch schräge Kantenverläufe der Randbereiche erlaubt. Die Verwendung der zuvor beschriebenen globalen Glättungsbedingungen Δ_{x_1} und Δ_{x_3} ist dabei für Tattoos nicht sinnvoll. Besonders die Seitenränder der Tattoos verlaufen sehr steil, sodass in diesem Bereich gesonderte Glattheitsannahmen definiert werden müssen. Stattdessen werden alle spaltenübergreifenden Kanten für Tattoo-Spalten verändert, indem ein gewisser Bewegungsraum gewährt wird. Hierdurch kann sich eine Oberfläche in der Nähe eines Tattoos stärker verformen. Für die obere Grenzfläche (Hintergrund-EDS) wird die Modifikation nicht vorgenommen, da sie nicht durch Tätowierungen unterbrochen wird. Für die spätere Evaluation ist die folgende Parametrisierung empirisch ermittelt worden: $\Delta_{x_1} = 2$, $\Delta_{x_3} = 3$, $\delta_l^{\text{EDS}} = 10$, $\delta_u^{\text{EDS}} = 70$, $\delta_l^{\text{SFS}} = 2$, $\delta_u^{\text{SFS}} = 80$.

Außer der Konstruktion des Graphen spielt die Gestaltung der Kostenfunktion eine entscheidende Rolle im Hinblick auf die Segmentierungsleistung der Methode und wird nachfolgend näher dargestellt. Die Kostenfunktion basiert hauptsächlich auf den RF-Segmentierungen S_{RF} der Hautschichten sowie auf den lokalen Bildkanteninformationen und wird für jede Oberfläche separat bestimmt. Um Kanteninformationen in die Kostenfunktion zu integrieren, wird an jeder Position im Bildvolumen der normalisierte Gradientenbetrag berechnet, welcher invertiert wird, um niedrige Kosten

an Positionen mit hohem Betrag zu erhalten. Die bisher detektierten Grenzübergänge aus den RF-Segmentierungen der Hautschichten fließen in die Kostenfunktion mit ein. Hierzu wird die jeweilige Hautschicht binarisiert und mit einem vertikalen Kantenfilter gefaltet. Die daraus resultierenden positiven und negativen Filterantworten können genutzt werden, um die einzelnen Grenzübergänge oder Oberflächen zu lokalisieren. Der Übergang vom Hintergrund zur EDS entspricht den Voxeln mit positiver Filterantwort der EDS-Segmentierung und wird in der Menge EDS^+ zusammengefasst (siehe Abb. 4.4). Anschließend wird für jeden Bildvoxel die minimale Distanz zu den Voxeln in EDS^+ bestimmt. Die Kostenfunktion der ersten Oberfläche setzt sich dann wie folgt zusammen:

$$c^1(\mathbf{x}) = \min_{\mathbf{y} \in EDS^+} \|\mathbf{x} - \mathbf{y}\|_2 + \omega_G \cdot c_G(\mathbf{x}). \quad (4.2)$$

Als Distanzmaß wird die L2-Norm gewählt. Daneben beschreibt c_G die integrierte Kanteninformation, welche durch ω_G gewichtet wird. Als Nächstes werden die Kosten des EDS-SFS-Übergangs beschrieben, deren Konstruktion analog zur Kostenfunktion c^1 erfolgt. Dieser Übergang kann sowohl durch die negative Filterantwort der EDS (EDS^-) als auch durch die positive Filterantwort der SFS (SFS^+) abgebildet werden. Die Kosten der zweiten Oberfläche sind dann wie folgt definiert:

$$c^2(\mathbf{x}) = \min_{\mathbf{y} \in EDS^-} \|\mathbf{x} - \mathbf{y}\|_2 + \min_{\mathbf{y} \in SFS^+} \|\mathbf{x} - \mathbf{y}\|_2 + \omega_G \cdot c_G(\mathbf{x}). \quad (4.3)$$

Die dritte Oberfläche grenzt die SFS von der MFS ab. Dieser Übergang wird durch die Voxel der negativen Filterantwort der SFS (SFS^-) sowie der positiven Filterantwort der MFS (MFS^+) repräsentiert. Die Kostenfunktion ergibt sich daher folgendermaßen:

$$c^3(\mathbf{x}) = \min_{\mathbf{y} \in SFS^-} \|\mathbf{x} - \mathbf{y}\|_2 + \min_{\mathbf{y} \in MFS^+} \|\mathbf{x} - \mathbf{y}\|_2 + \omega_G \cdot c_G(\mathbf{x}). \quad (4.4)$$

Wie in [Garvin u. a., 2009] vorgeschlagen, wird auch regionale Information über „In-Region-Kosten“ in die Kostenfunktion integriert. Hierzu wird das Volumen über die drei Grenzübergänge vertikal in die folgenden vier Regionen unterteilt: 1. Hintergrund, 2. EDS, 3. SFS, 4. MFS/Hintergrund. Die In-Region-Kosten der drei Oberflächen $i = 1, 2, 3$ sind dann durch

$$w_{IR}^i(v_{\mathbf{x}}) := \begin{cases} \omega_R, & \text{wenn } S_{RF}(\mathbf{x}) = \text{Region}_i \\ -\omega_R, & \text{wenn } S_{RF}(\mathbf{x}) = \text{Region}_{i+1} \\ 0, & \text{sonst} \end{cases} \quad (4.5)$$

gegeben. Diese Kosten sind negativ für die Region_{i+1} , die sich unterhalb der i -ten Oberfläche befinden sollte, und positiv für die darüber liegende Region_i . Da in der minimalen geschlossenen Menge alle Knoten enthalten sind, die unterhalb der i -ten Oberfläche liegen, wird diese vorzugsweise wegen der niedrigen Kosten oberhalb der

Region $_{i+1}$ und unterhalb der Region $_i$ verlaufen, die hohe Kosten aufweist. Die Gesamtkosten w_i jeder i -ten Oberfläche ergeben sich schließlich aus der Kombination von w_{IR}^i und Gleichung 4.1 zu

$$w^i(v_{\mathbf{x}}) = \begin{cases} -1, & \text{wenn } x_2 = 0 \\ c^i(\mathbf{x}) - c^i(x_1, x_2 - 1, x_3) + w_{\text{IR}}^i, & \text{sonst.} \end{cases} \quad (4.6)$$

Ferner werden die Kosten für $x_2 = 0$ auf -1 abgeändert, um eine leere geschlossene Menge zu vermeiden. Die Gewichte ω_G sowie ω_R werden jeweils auf 10 und 1 gesetzt.

Zur effizienten Bestimmung der minimalen geschlossenen Mengen kommt der Max-Flow-Min-Cut-Algorithmus zum Einsatz. Der Graph wird hierzu um die Terminalknoten s und t erweitert (siehe Kapitel 3.1.1.2). Die Quelle s ist über Kanten mit allen Knoten mit positiven Kosten (V^+) verbunden, während die Senke t mit Knoten mit negativen Kosten (V^-) verbunden ist. Der neu entstandene Graph $G_{s,t}$ ist nun kantengewichtet. Dazu erhalten die n-Links und die t-Links die betragsmäßigen Kosten des Knotens, mit dem sie verbunden sind. Über einen s-t-Schnitt kann nun die minimale geschlossene Menge ermittelt werden. Die Oberfläche lässt sich schließlich aus den maximalen x_2 -Koordinaten jeder Bildspalte bestimmen. Dieser Prozess wird für jede i -te Oberfläche wiederholt. Zum Schluss werden die vom RF-Klassifikator segmentierten Tattoo-Areale in die Segmentierung wieder integriert.

4.2.3 Segmentierung mit einem Fully Convolutional Network

Der im Folgenden beschriebene Ansatz nutzt Deep-Learning-Methoden unter Verwendung eines FCN, um die Maushautschichten in den OCT-Bilddaten zu segmentieren. Ähnlich wie im vorangegangenen Kapitel 4.2.2 wird ein OCT-Volumen durch das FCN B-Scan-weise segmentiert. Die FCN-basierte Segmentierung bietet mehrere Vorteile gegenüber dem RF-basierten Ansatz. Durch das automatische Lernen hochkomplexer Merkmalsrepräsentationen entfällt die Notwendigkeit, geeignete Merkmale zu gestalten. Infolgedessen werden B-Scans in ihrer Rohfassung durch das FCN verarbeitet, was die Anzahl der Verarbeitungsschritte vergleichsweise erheblich reduziert (siehe Abb. 4.5). Das FCN gibt eine vollständige Segmentierung des Eingangsbildes aus, sodass die Klassifikation jedes einzelnen Pixels im B-Scan vermieden wird. Im Folgenden wird die verwendete FCN-Architektur sowie die Trainings- und Klassifikationsphasen vorgestellt.

4.2.3.1 Densely Connected U-Net

Die vorgeschlagene FCN-Architektur ist vom U-Net inspiriert (siehe Kapitel 3.2.2.3). Dahin gehend werden nachfolgend nur die von der originalen U-Net-Architektur abweichenden Modifikationen beschrieben. Zur verbesserten Veranschaulichung ist in

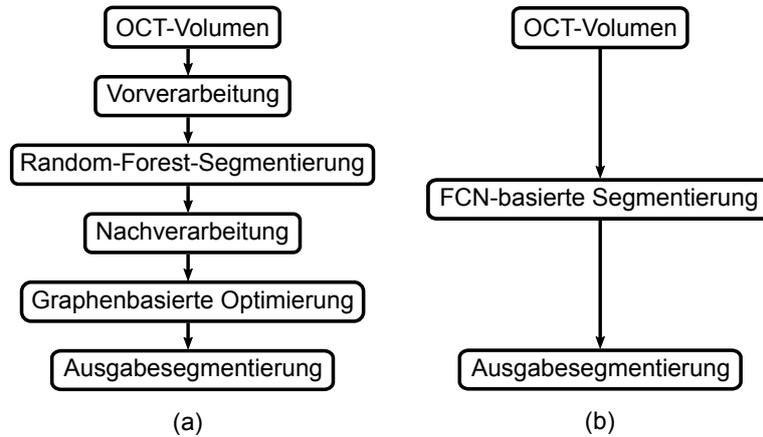


Abb. 4.5: Vergleich der Segmentierungspipelines des (a) RF-Klassifikators mit graphenbasierter Optimierung (siehe Kapitel 4.2.2) und (b) des hier vorgestellten FCN.

Abbildung 4.6 ein schematischer Aufbau der FCN-Architektur dargestellt. Die Doppelfaltungsblöcke der originalen U-Net-Architektur werden um kurze Skip Connections ergänzt. Anstelle der Verwendung von Residual-Blöcken [He u. a., 2016] werden die Eingangskanäle jedes einzelnen Doppelfaltungsblocks mit seiner Ausgabe konkatiniert. Die Anzahl der Ausgangskanäle der zweiten 3×3 -Faltung und einer weiteren 1×1 -Faltung in der Skip Connection werden jeweils auf $f/2$ gesetzt, wobei f die Anzahl der ausgehenden Merkmalskanäle des vorangegangenen Doppelfaltungsblocks angibt. Hierdurch wird sichergestellt, dass f sich durch die Konkatination nicht ändert. Die Verwendung solcher dicht verbundenen (*engl.*: densely connected, kurz: DC) DC-Blöcke hat mehrere Vorteile. Sie verringern die Problematik kleiner werdender oder verschwindender Gradienten (*engl.*: vanishing gradients), verbessern und fördern die Wiederverwendung von Merkmalen im Netz [Huang u. a., 2017]. Wegen der DC-Blöcke wird die FCN-Architektur auch als „DCU-Net“ bezeichnet. Weiterhin wird ein Zero-Padding vor jeder Faltung ausgeführt. Hierdurch wird verhindert, dass sich die Bildgröße durch die Faltungsoperation verringert. Eine Batch-Normalisierung [Ioffe u. a., 2015] nach jeder 3×3 -Faltung sorgt überdies für eine schnellere Konvergenz während des Trainings. Die verschiedenen Auflösungsstufen werden im Encoder über ein 2×2 -Average-Pooling mit einer Kernelschrittweite von zwei realisiert, wodurch sich die Auflösung zu jeder Stufe halbiert. Wie bei der originalen U-Net-Architektur wird die Anzahl der Merkmalskanäle auf jeder Auflösungsstufe verdoppelt. Beginnend bei $f=32$ Merkmalskanälen ergibt sich so eine Anzahl von $f=512$ Merkmalskanälen in der untersten Encoderschicht. Insgesamt besteht das DCU-Net aus fünf Auflösungsbenen und weist eine Anzahl von insgesamt 610228 trainierbaren Parametern auf. Im Decoder werden die transponierten Faltungen jeweils durch eine bilineare Interpolation ersetzt, durch welche dann die Hochskalierung der Merkmalskarten erfolgt. Hierdurch

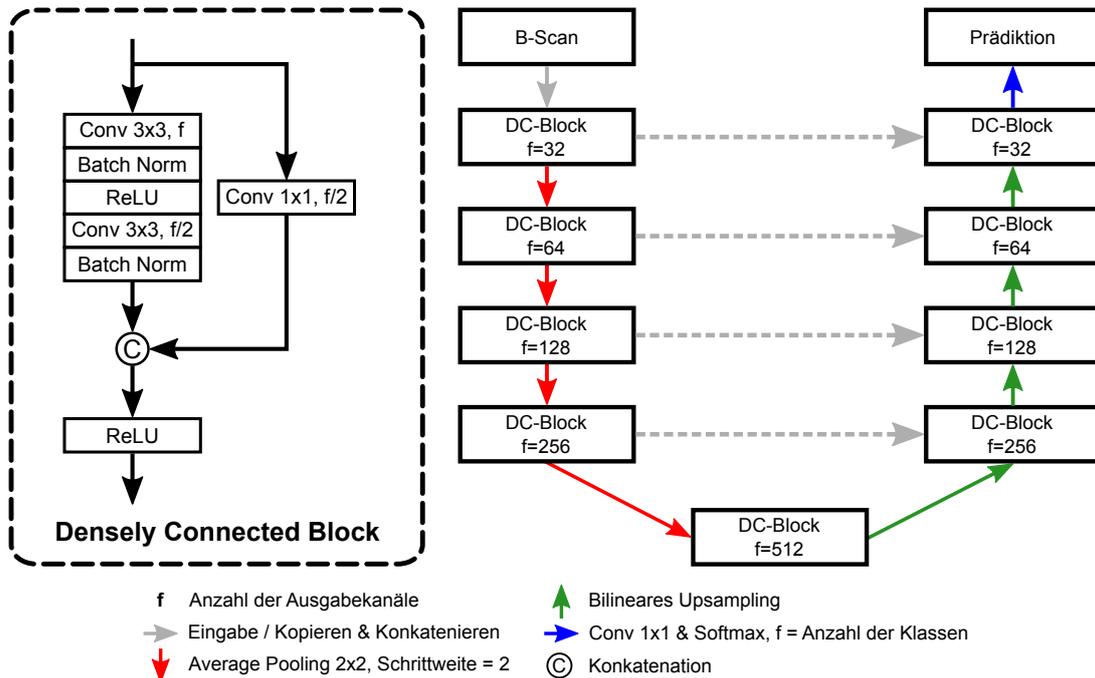


Abb. 4.6: Netzwerkarchitektur des DCU-Nets.

wird zum einen die Anzahl der Netzparameter reduziert und zum anderen werden Checkerboard-Artefakte in der finalen Segmentierung vermieden [Odena u. a., 2016].

4.2.3.2 Training und Klassifikation

Das DCU-Net wird wie der RF-Klassifikator (siehe Kapitel 4.2.2.2) mithilfe eines überwachten Lernverfahrens trainiert, wobei jedoch die unveränderten B-Scans der OCT-Volumen benutzt werden. Die Bilddaten werden lediglich in ihrer Größe reduziert, um die Recheneffizienz des Trainings zu steigern. Zu diesem Zweck werden unnötige Hintergrundbereiche sowohl im oberen als auch im unteren Bildbereich der B-Scans entfernt und die Auflösung um die Hälfte reduziert. Schließlich wird die Differenz zur endgültigen Bildgröße von 256×656 Pixeln durch ein Zero-Padding aufgefüllt.

Zur Bestimmung des Prädiktionsfehlers kommt der von Sudre u. a. [2017] vorgeschlagene Generalized Dice Loss (kurz: GDL *dt.*: generalisierte Dice-Fehlerfunktion) zum Einsatz. Der GDL unterscheidet sich von dem in Kapitel 3.1.2 vorgestellten DSC durch einen Gewichtungsfaktor, der Invarianz für jede einzelne Klasse gewährleistet und damit effektiv dem Klassenungleichgewicht entgegenwirken kann. Dies ist wichtig, da die Hintergrundklasse deutlich überrepräsentiert im OCT-Datensatz vorliegt. Sehr kleine und/oder selten vorkommende Strukturen, wie hier die Tätowierungen, können

durch den Gewichtungsfaktor stärker in der Fehlerfunktion berücksichtigt werden. Der GDL wird wie folgt definiert:

$$\mathcal{L}_{\text{GDL}} = 1 - \frac{2 \sum_{l=0}^{K-1} \alpha_l \sum_{\mathbf{x}} P_l(\mathbf{x}) S_l(\mathbf{x}) + \varepsilon}{\sum_{l=0}^{K-1} \alpha_l \sum_{\mathbf{x}} P_l^2(\mathbf{x}) + \sum_{\mathbf{x}} S_l^2(\mathbf{x}) + \varepsilon}. \quad (4.7)$$

Hier beschreiben $P_l(\mathbf{x})$ und $S_l(\mathbf{x})$ jeweils die Bildwerte der prädizierten Klassenwahrscheinlichkeit und die Grundwahrheit der Klasse l an der Pixelposition \mathbf{x} . Für den Gewichtungsfaktor α_l wird die normierte inverse Häufigkeitsverteilung der einzelnen Klassen verwendet, wodurch die bekannte Korrelation zwischen Regionsgröße einer Klasse und dem DSC reduziert wird. Der Parameter ε sorgt für numerische Stabilität, welcher auf 10^{-5} gesetzt wird. In der Zwischenzeit sind in der Literatur viele weitere Fehlerfunktionen vorgeschlagen worden, mit dem Ziel, dem Ungleichgewicht der Klassen während des Trainings entgegenzuwirken. Eine umfassende Übersicht ist hierzu in der Arbeit von Ma u. a. [2021] gegeben.

Vor Trainingsbeginn werden die Netzparameter nach der Methode von Glorot u. a. [2010] initialisiert. Zur Anpassung der Parameter während des Trainings wird der Adam-Optimierer (siehe Kapitel 3.2.2.2) genutzt. Seine anfängliche Lernrate wird auf 0.001 festgelegt, die mit fortschreitendem Training exponentiell abnimmt. Zur Vermeidung übermäßiger Gradientenaktualisierungen wird ein L2-Regularisierungsterm benutzt. Das Training verläuft über insgesamt 300 Epochen. Im Anschluss wird diejenige Netzwerkkonfiguration für die nachfolgende Testphase verwendet, welche den geringsten Validierungsfehler aufweist. Das Training wird auf einer NVidia GeForce 1080 TI mit einer Mini-Batch-Größe von fünf durchgeführt, um die Ausnutzung des GPU-Speichers weitestgehend zu maximieren.

Um mit der begrenzten Menge an verfügbaren Trainingsbeispielen auszukommen, wird eine Datenaugmentierung durchgeführt, um eine Überanpassung der Netzparameter zu verhindern und die Generalisierbarkeit zu erhöhen. Für diese Arbeit wird die Datenaugmentierung aus [Kepp u. a., 2019a] erweitert und als Standard für die weiteren Kapitel verwendet. Die Augmentierung besteht aus folgenden Komponenten, deren Parametrisierung zufällig erfolgt: horizontale Spiegelungen, Rotationen ($\pm 7^\circ$), Translationen (± 50 Pixel in hor./vert. Richtung), Skalierungen (Faktor ± 0.25) und nichtlineare Intensitätsverschiebungen nach [Devalla u. a., 2018], die durch

$$\bar{I}(\mathbf{x}) = -a + (a + b + 1) \cdot I^c(\mathbf{x}) \quad (4.8)$$

definiert wird. Die Faktoren a und b werden aus $[0, 0.1]$ und der Exponent c aus $[0.3, 2.5]$ zufällig gezogen. In Abbildung 4.7 werden mehrere Augmentierungsbeispiele eines B-Scans dargestellt.

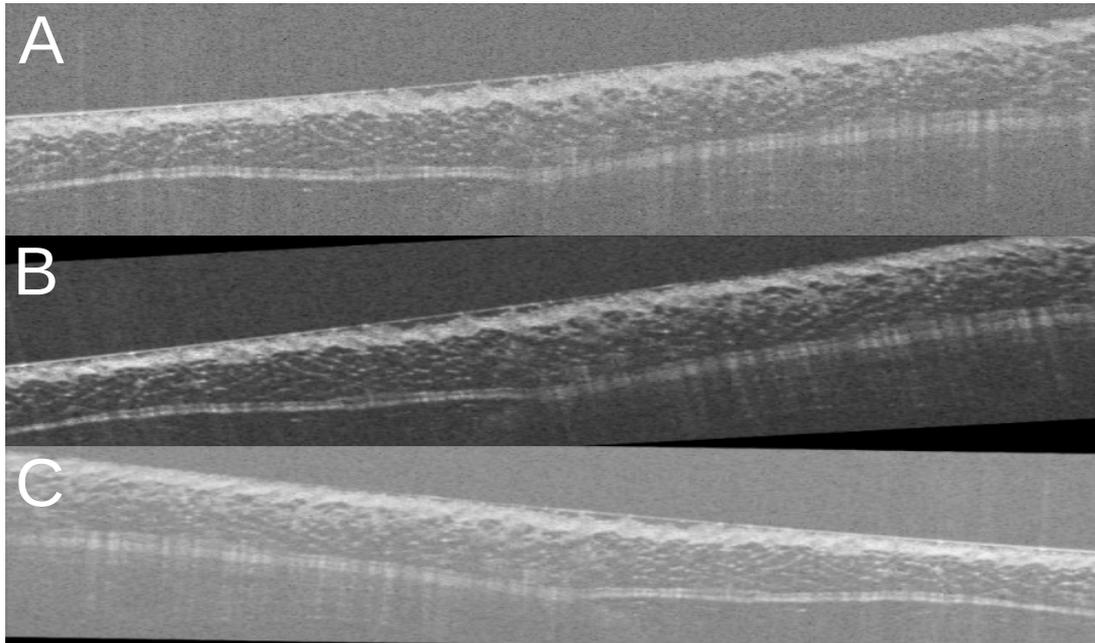


Abb. 4.7: Visualisierung eines B-Scans (A) mit zwei verschiedenen Augmentierungsbeispielen (B und C).

4.3 Experimente und Ergebnisse

Im Folgenden werden die beiden zuvor vorgestellten Segmentierungsverfahren aus Kapitel 4.2.2 und 4.2.3 hinsichtlich ihrer Genauigkeit und Robustheit evaluiert und miteinander verglichen. Dabei werden die Erweiterungen und Modifikationen der Methoden hinsichtlich ihrer Beeinflussung auf das Segmentierungsergebnis untersucht. Um die Auswirkung der graphenbasierten Optimierung (GO) auf das Segmentierungsergebnis untersuchen zu können, wird die RF-Segmentierung sowohl allein (RF) als auch in Kombination mit dem Optimierungsschritt (RF+GO) evaluiert. Des Weiteren wird der Vorteil der DC-Blöcke (DCU-Net) gegenüber den einfachen Doppelfaltungsblöcken (U-Net) analysiert. Schließlich wird die graphenbasierte Optimierung auch auf die Segmentierungen des (DC)U-Nets angewandt ((DC)U-Net+GO), um zu überprüfen, ob sich die Robustheit des Verfahrens noch weiter steigern lässt. Die Evaluation wird mittels 18-facher Kreuzvalidierung durchgeführt. Zu diesem Zweck werden die B-Scans der 18 OCT-Volumen in nicht überlappende Trainings-, Validierungs- und Testdatensätze im Verhältnis von 13–4–1 aufgeteilt. Insgesamt wird die Kreuzvalidierung jeweils mithilfe der Grundwahrheit jedes Experten ausgeführt, was bedeutet, dass pro Grundwahrheit 18 verschiedene FCNs oder RF-Klassifikatoren trainiert werden. Die Ergebnisse der Durchläufe werden anschließend für die Auswertung zusammengeführt. Die quantitative Auswertung basiert auf den Metriken DSC, ASSD und HD

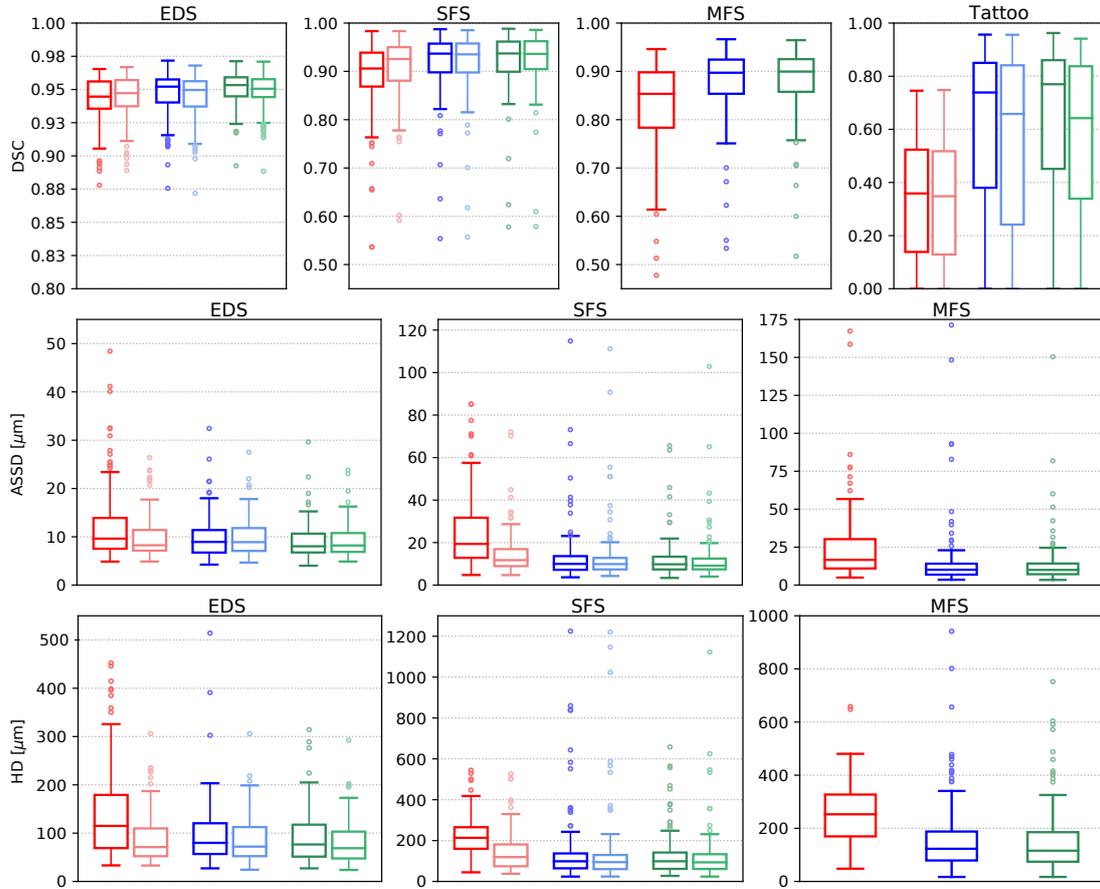


Abb. 4.8: Boxplot-Darstellung der DSC-, ASSD- und HD-Werte beider Kreuzvalidierungen. Zur Unterscheidung der Methoden wird folgende Farbkodierung genutzt: RF ●, RF+GO ●, U-Net ●, U-Net+GO ●, DCU-Net ●, DCU-Net+GO ●.

(siehe Kapitel 3.1.2), die jeweils zwischen den Segmentierungen der Methoden und den Grundwahrheiten bestimmt werden. Ergänzend wird die Inter-Rater-Reliabilität (IRR) zwischen den manuellen Annotationen der beiden Experten für alle Metriken berechnet. Für jede Methode und Struktur sind die gemittelten Metriken in Tabelle 4.1 dargestellt. Boxplots in Abbildung 4.8 veranschaulichen zudem die Ergebnisverteilungen im Detail. Es ist zu beachten, dass die MFS-Klasse nicht durch die graphenbasierte Optimierung bestimmt wird. Um signifikante Unterschiede identifizieren zu können, wird der Wilcoxon-Rangsummentest mit einem Signifikanzniveau von 5 % angewandt. Die Implementierung des RF+GO-Verfahrens wurde in Matlab R2015 mithilfe der Toolbox aus [Dollár, 2022] realisiert. Die Netzwerkarchitektur sowie die Trainings- und Testroutinen der FCNs wurden in dem Python-basierten Framework PyTorch [Paszke u. a., 2017] umgesetzt.

Tabelle 4.1: Gemittelte Ergebnisse der beiden Kreuzvalidierungen, durchgeführt mit Annotationen des jeweiligen Experten. Die besten Ergebnisse sind fett gedruckt. Die Abstände für ASSD und HD sind in μm angegeben (Pixelhöhe und -breite entsprechen $6.8\ \mu\text{m}$ bzw. $13\ \mu\text{m}$). Es ist zu beachten, dass die MFS-Klasse nicht bei der graphenbasierten Optimierung (GO) berücksichtigt wird.

	Strukturen \rightarrow	EDS	SFS	MFS	Tattoo
	Methoden \downarrow	●	●	●	●
DSC	RF	0.942	0.891	0.830	0.340
	RF+GO	0.944	0.909	–	0.339
	U-Net	0.947	0.920	0.877	0.593
	U-Net+GO	0.945	0.921	–	0.543
	DCU-Net	0.951	0.924	0.880	0.616
	DCU-Net+GO	0.949	0.925	–	0.558
	IRR	0.949	0.937	0.881	0.583
ASSD	RF	12.14	24.79	24.62	–
	RF+GO	9.95	14.69	–	–
	U-Net	9.83	13.08	15.22	–
	U-Net+GO	9.95	12.89	–	–
	DCU-Net	9.03	11.82	14.55	–
	DCU-Net+GO	9.15	11.78	–	–
	IRR	9.05	9.96	11.97	–
HD	RF	140.03	219.93	257.75	–
	RF+GO	87.12	140.96	–	–
	U-Net	96.27	141.50	162.21	–
	U-Net+GO	87.89	132.85	–	–
	DCU-Net	90.59	124.55	162.64	–
	DCU-Net+GO	81.71	117.98	–	–
	IRR	87.82	95.10	128.66	–

Insgesamt werden die Gewebeschichten von allen Verfahren mit einer hohen Genauigkeit segmentiert. Beginnend mit der EDS über die SFS bis hin zur MFS nimmt die Segmentierungsgenauigkeit etwas ab, was mit der immer komplexer werdenden Strukturbeschaffenheit einhergeht. Für EDS und SFS liegen 50% aller DSC-Werte über 0.85, für das (DC)U-Net sogar über 0.9. Die MFS wird mit einer etwas niedrigeren Genauigkeit segmentiert. Im Vergleich dazu ist die Segmentierung der Tattoos deutlich anspruchsvoller, was sich in den stärker gestreuten DSC-Werten ausdrückt. Der RF erreicht gegenüber dem (DC)U-Net nur geringere DSC-Werte für diese Klasse. Ähnlich zum DSC spiegeln auch die ASSD- und HD-Werte die allgemeine hohe Segmentierungsleistung wider. Allerdings fällt auf, dass der RF allein höhere Distanzen erzielt. Besonders im Hinblick auf die SFS und MFS unterscheiden sich die ASSD- und HD-Werte des RF von denen der U-Net-Varianten. Qualitativ lässt sich der Un-

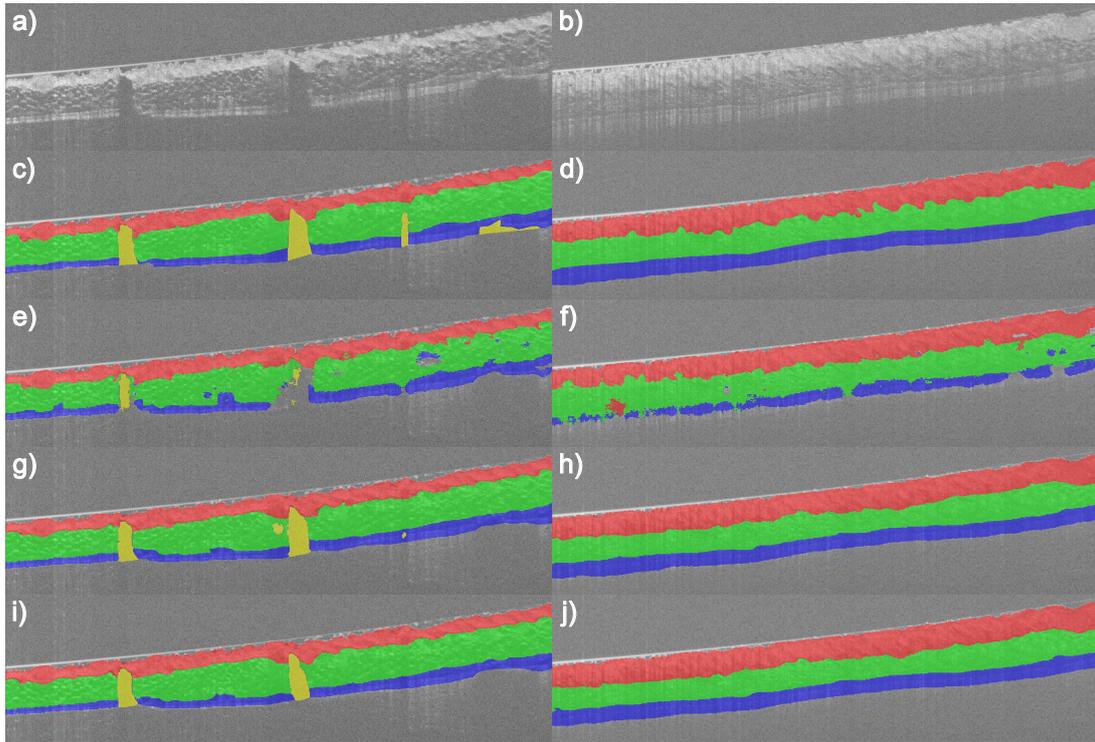


Abb. 4.9: Zwei beispielhafte Darstellungen von Segmentierungen der EDS ■, SFS ■, MFS ■ und Tattoos ■: a)/b) B-Scan; c)/d) Expertensegmentierung; e)/f) RF; g)/h) U-Net; i)/j) DCU-Net.

terschied zwischen RF- und FCN-basierter Segmentierung ebenfalls beobachten. In Abbildung 4.9 sind Segmentierungsergebnisse zwei verschiedener B-Scans der Methoden RF, U-Net und DCU-Net dargestellt. Die Segmentierungen des RF werden zu diesem Zweck wieder auf den ursprünglichen B-Scan übertragen, d. h. die Transformation der Begradigung erfolgte invertiert. Anhand der Darstellung fällt der Unterschied zwischen RF und (DC)U-Net noch deutlicher auf. Tattoos und von Haaren verursachte Schattenwürfe führen beim RF eher zu Fehlklassifikationen als bei den FCNs. Trotz der hohen Heterogenität der OCT-Scans sowie der Verwendung von nicht vorverarbeiteten Bilddaten weisen die Segmentierungen der U-Net-Varianten eine hohe Qualität auf.

Zwischen allen Methoden erreicht das DCU-Net durchschnittlich die höchste Segmentierungsgüte. Bezogen auf die EDS und der Tattoos übertrifft es sogar die IRR. Gegenüber der RF-basierten Segmentierung lässt sich, neben den geringfügigen Verbesserungen für die EDS und SFS, ein Anstieg des mittleren DSC-Wertes um ca. 81 % der Tattoo-Klasse beobachten. Für die Klassen EDS, SFS und MFS konnten die ASSD-Werte um ca. 26 %, 52 % und 41 % und die HD-Werte um ca. 35 %, 43 % und

37% reduziert werden. Das DCU-Net segmentiert im Vergleich zum Standard-U-Net die EDS, SFS und Tattoos präziser. Die Ergebnisse der beiden Netzarchitekturen unterscheiden sich signifikant, mit Ausnahme der ASSD- und HD-Werte für die SFS und aller Metriken für die MFS. Abgesehen von der Tattoo-Klasse ist der Zugewinn an Genauigkeit allerdings nur gering. Ebenso lässt sich qualitativ (siehe Abb. 4.9) nur ein geringer Unterschied zwischen den beiden U-Net-Varianten feststellen. Dagegen zeigen die Boxplots des DCU-Nets in Abbildung 4.8 kleinere Interquartilabstände sowie eine geringere Anzahl von Ausreißern bezüglich der HD-Werte, was auf eine erhöhte Robustheit gegenüber dem Standard-U-Net hinweist.

Durch die Kombination des RF mit der graphenbasierten Optimierung (RF+GO) steigert sich die Segmentierungsleistung deutlich, was sich durch die regularisierende Eigenschaft des Verfahrens begründen lässt. So verbessern sich jeweils für die EDS und SFS die ASSD-Werte um ca. 18% und 41% und die HD-Werte um ca. 38% und 36%. Dahingegen lässt sich nur eine kleinere Verbesserung der DSC-Werte für die EDS und SFS feststellen. Die kombinierte RF+GO-Methode erreicht sogar hinsichtlich der ASSD- und HD-Werte der EDS dasselbe Niveau wie das Standard-U-Net, d.h. es kann kein signifikanter Unterschied festgestellt werden. In Abbildung 4.10 wird beispielhaft dargestellt, welchen Einfluss die graphenbasierte Optimierung auf das RF-Segmentierungsergebnis ausübt. Trotz der eher schwachen Segmentierung des RF-Klassifikators konnte durch den Optimierungsschritt die Genauigkeit bedeutend gesteigert werden. Für das in Abbildung 4.10 dargestellte Beispiel konnte für die EDS und die SFS der DSC um ca. 4% und 19% gesteigert und jeweils die ASSD um ca. 74% und 36% sowie die HD um ca. 58% und 28% verringert werden.

Darüber hinaus wird die graphenbasierte Optimierung als optionale Nachverarbeitung für die FCN-basierten Methoden untersucht. Der Zuwachs an Genauigkeit ist im Vergleich zum RF etwas geringer. Allerdings lässt sich eine signifikante Veränderung der Ergebniswerte, mit Ausnahme der ASSD-Werte des DCU-Nets für die EDS, feststellen. Der DSC nimmt zwar für die EDS etwas ab, aber erhöht sich für die SFS. Demgegenüber verringern sich die DSC-Werte für die Tattoo-Klasse erheblich stärker, was mit der sehr variablen Ausprägung der Tattoo-Areale zu begründen ist, die sich nur schwer durch den Graphen modellieren lassen. Daher senkt der Optimierungsschritt die Segmentierungsgüte der FCNs in Bezug auf die Tattoo-Klasse. Ähnlich zum DSC wirkt sich die graphenbasierte Optimierung kaum auf die ASSD-Werte aus. Dagegen konnten die HD-Werte des (DC)U-Nets um 10% und 9% für die EDS und um 5% und 6% für die SFS reduziert werden.

4.4 Zusammenfassung und Diskussion

In diesem Kapitel wurden zwei verschiedene Ansätze des maschinellen Lernens zur Segmentierung mehrerer Hautschichten der Maus sowie von Tätowierungen in OCT-Bild-

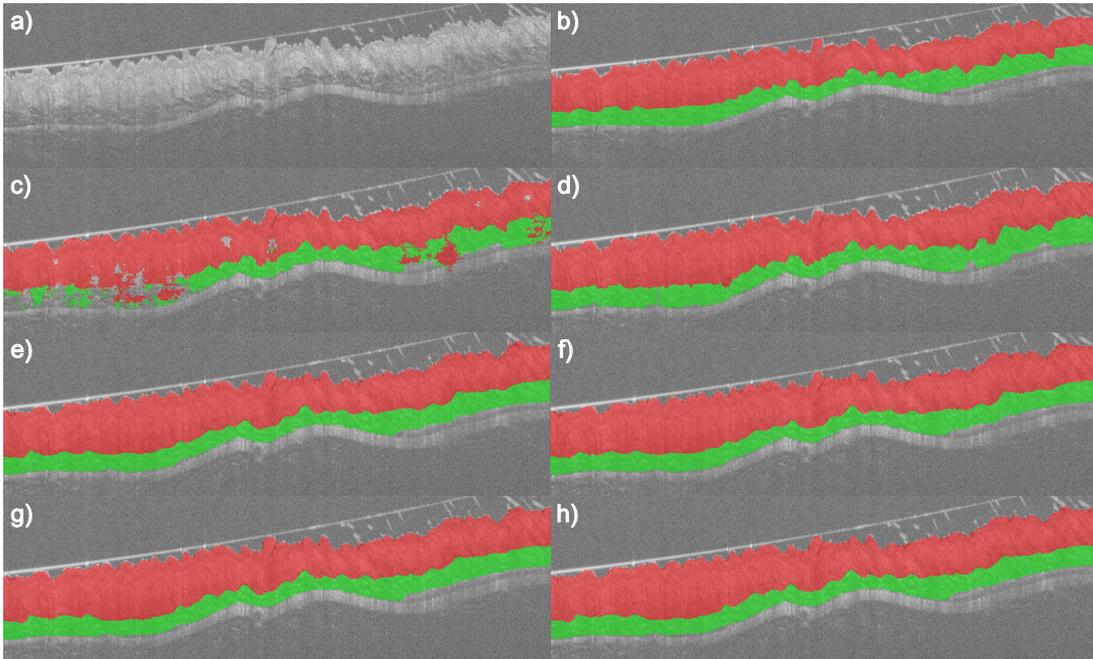


Abb. 4.10: Beispielhafte Darstellung von Segmentierungen der EDS ● und SFS ● vor und nach der graphenbasierten Optimierung: a) B-Scan; b) Grundwahrheit; c) RF; d) RF+GO; e) U-Net; f) U-Net+GO; g) DCU-Net; h) DCU-Net+GO.

daten vorgestellt. Die erste Methode folgt einer klassischen Segmentierungspipeline, die aus einer Vorverarbeitung, einer RF-Segmentierung und einem graphenbasierten Optimierungsschritt besteht. Im Gegensatz dazu basiert der zweite Segmentierungsansatz auf dem DCU-Net, das auf der populären U-Net-Architektur beruht und neben geringfügigen Modifikationen dicht verbundene Faltungsblöcke verwendet.

Die Ergebnisse der Evaluation haben gezeigt, dass sowohl das RF+GO-Verfahren als auch das DCU-Net mit einer hohen Genauigkeit segmentieren. Im direkten Vergleich erzielte das DCU-Net sowohl gegenüber dem alleinigen RF als auch in Kombination mit der graphenbasierten Optimierung deutlich bessere Ergebnisse.

Dies ist auf den autonomen Lerncharakter eines FCN und seine wesentlich höhere Anzahl von Parametern zurückzuführen, wodurch sowohl aussagekräftige Merkmale als auch globale Kontextinformationen selbst in unbearbeiteten OCT-Bildern erzeugt werden. Durch den Einsatz geeigneter Augmentierungstechniken konnte zum einen eine erhöhte Generalisierbarkeit trotz einer geringen Anzahl von Trainingsbeispielen erreicht und zum anderen komplexe und stark variierende Strukturen wie die SFS oder MFS sowie Tattoo-Bereiche besser erfasst werden. Dagegen erwies sich die Modellierung geeigneter Merkmale für den RF entschieden aufwendiger. Somit konnte gezeigt

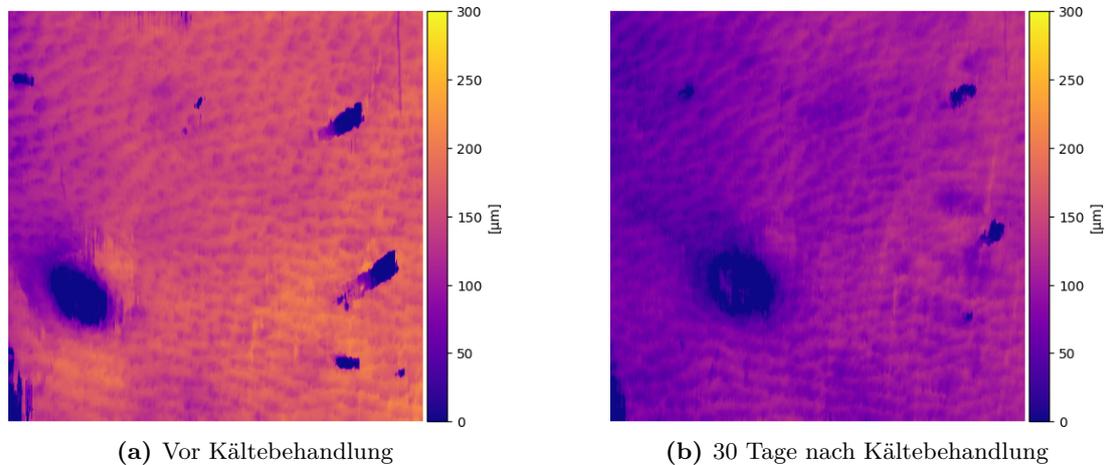


Abb. 4.11: Visualisierung der segmentierten SFS als Dickenkarte. Die Abbildungen zeigen jeweils die SFS zweier OCT-Volumen, welche direkt vor (a) und 30 Tage nach der Kältebehandlung (b) aufgenommen wurden. Anhand der Farbskala lässt sich eine Reduktion der SFS erkennen.

werden, dass mithilfe des (DC)U-Nets der Segmentierungsprozess erheblich vereinfacht und gleichzeitig die Genauigkeit erhöht werden kann.

Die Verwendung der DC-Blöcke gegenüber den Doppelfaltungsblöcken des U-Nets zeigte, dass sich die Ergebnisse signifikant verbessern lassen. Allerdings ist der Zuegewinn an Genauigkeit und Robustheit gering. Eine ähnliche Beobachtung haben Isensee u. a. [2021] gemacht, indem sie ein selbstkonfigurierendes U-Net anhand zahlreicher frei verfügbarer medizinischer Bilddatensätze evaluiert haben. Ihre Ergebnisse haben gezeigt, dass im Wesentlichen die Wahl geeigneter Hyperparameter sowie der Einsatz umfassender Augmentierungstechniken zu einer hohen Segmentierungsgüte führen. Verglichen mit [Kepp u. a., 2019a] konnten niedrigere ASSD- und HD-Werte für das (DC)U-Net erreicht werden. Dabei ließ sich eine Annäherung der Ergebniswerte zwischen DCU-Net und U-Net beobachten. Da bis auf die erweiterte Augmentierung die Trainingsprozedur gegenüber [Kepp u. a., 2019a] unverändert blieb, konnten somit die Beobachtungen aus [Isensee u. a., 2021] bestätigt werden.

In Kombination mit der graphenbasierten Optimierung erreicht der RF-Klassifikator eine deutlich höhere Genauigkeit und Präzision, die der des Standard-U-Nets nahekommt. Anspruchsvolle Bildbereiche, die vom RF falsch segmentiert werden, lassen sich auf diese Weise korrigieren (siehe Abb. 4.10). Das Graphenmodell berücksichtigt überdies benachbarte B-Scans, was eine zusätzliche Regularisierung ermöglicht. Dies konnte jedoch im Rahmen dieser Arbeit nur qualitativ verifiziert werden, da hierfür keine Grundwahrheit eines kompletten OCT-Volumens vorliegt.

Die Verwendung der graphenbasierten Optimierung als optionale Nachverarbeitung der (DC)U-Net-Segmentierungen zeigte, dass die Robustheit erhöht und die Anzahl der Ausreißer deutlich verringert werden kann. Ebenso konnten durch das Graphenmodell topologische Segmentierungsfehler korrigiert werden. Gleichzeitig hatte dieser Nachverarbeitungsschritt nur eine geringe Auswirkung auf die Segmentierungsgenauigkeit der Hautschichten, was für die bereits hohe Präzision des (DC)U-Nets spricht. Die Segmentierungsgenauigkeit der Tattoo-Klasse nimmt sogar merklich durch die Nachverarbeitung ab. Trotz der Berücksichtigung von Tattoos im Graphenmodell verbleibt es anspruchsvoll, sämtliche Variationen dieser Klasse durch eine geeignete Parameterwahl zu berücksichtigen. Ein weiterer Nachteil ist die schlechte Adaptierbarkeit des Graphenmodells auf neue Segmentierungsaufgaben, wie der Anpassung an ein neues Tiermodell.

Obwohl sich beide hier vorgestellten Methoden aufgrund ihrer sehr hohen Segmentierungsleistung für die quantitative Analyse von OCT-Bildern der Mäusehaut eignen, bietet das DCU-Net mehrere Vorteile gegenüber dem RF-GO-Ansatz. Dazu gehören neben der hohen Genauigkeit und Robustheit vorrangig die erhebliche Reduzierung der Segmentierungspipeline (siehe Abb. 4.5) sowie die leichte Anpassbarkeit des Verfahrens. Diese Eigenschaften des DCU-Nets sind entscheidend für den Einsatz in Forschungsstudien, z. B. zur Bewertung des Kryolipolyse-Verfahrens [Salma u. a., 2018]. In Abbildung 4.11 ist hierzu die Dickenmessung der SFS vor und 30 Tage nach einer Kältebehandlung dargestellt.

Um die Robustheit der Segmentierung ohne weitere Nachbearbeitungsschritte zu erhöhen, wird das DCU-Net im folgenden Kapitel um eine Multi-Task-Lernstrategie zur Auswertung von Retinaschichten und DMÖ-bedingten Flüssigkeitseinlagerungen erweitert.

Kapitel 5

Formbasierte Segmentierung von Retinaschichten und Flüssigkeiten in OCT-Bilddaten

Die Hauptanwendung der OCT liegt in der Ophthalmologie, wo sie zur Diagnose verschiedener Augenkrankheiten eingesetzt wird. Die automatische Segmentierung der einzelnen Retinaschichten sowie pathologischer Strukturen ist hilfreich für die klinische Untersuchung. Wie bei der Haut folgen auch die Schichten der Retina einer bestimmten Anordnung, die von aktuellen Deep-Learning-Methoden oft nicht berücksichtigt wird. Die Ergebnisse des vorangegangenen Kapitels zeigen, dass sich graphenbasierte Methoden für eine optionale Nachbearbeitung zur Korrektur von Topologiefehlern eignen. Ihre Anwendbarkeit ist jedoch aufwendig und vor allem in Gegenwart von Pathologien sehr komplex. In diesem Kapitel wird daher eine Segmentierungsmethode vorgeschlagen, die ergänzende Forminformationen der Retinaschichten nutzt, um eine verbesserte Topologieerhaltung zu gewährleisten. Zu diesem Zweck wird das bereits vorgestellte DCU-Net so erweitert, dass neben der pixelweisen Klassifikation auch eine Regression der Forminformationen der Retinaschichten möglich ist. Hierdurch wird eine räumliche Regularisierung eingeführt und es können plausible Segmentierungen erzeugt werden. Weiterhin erlaubt eine duale Aufgabenkonsistenz ein halb-überwachtes Training. Eine umfassende Evaluation der Segmentierungsmethode wird anhand von OCT-Bilddaten von Patienten mit diabetischem Makulaödem durchgeführt.

5.1 Einleitung und Motivation

Das diabetische Makulaödem gilt als die häufigste Ursache für eine Sehbeeinträchtigung bei Patienten mit Diabetes mellitus [Bhagat u. a., 2009]. Das DMÖ stellt eine Komplikation der diabetischen Retinopathie (DR) dar, bei der es infolge eines hohen Blutzuckerspiegels zur Bildung abnormaler, undichter Blutgefäße im Bereich der Ma-

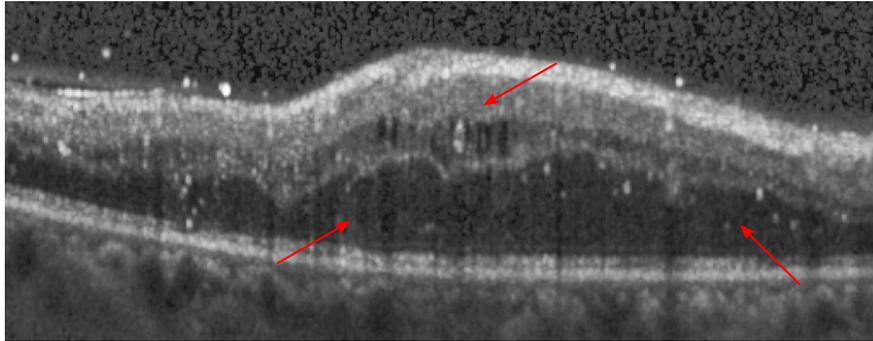


Abb. 5.1: Darstellung eines OCT-B-Scans einer Retina mit DMÖ. Durch Einblutungen entstehen Flüssigkeitseinlagerungen (rote Pfeile), die zu morphologischen Veränderungen führen.

kula kommt. Mit dem Fortschreiten der Krankheit kommt es zu Ödembildungen, die den Bereich innerhalb der Makula anschwellen lässt und das scharfe zentrale Sehen erheblich beeinträchtigt.

Im Frühstadium der Krankheit bleiben Retinaschäden zunächst unbemerkt, weshalb für Diabetiker regelmäßige Kontrolluntersuchungen beim Augenarzt unerlässlich sind. Erste pathologische Veränderungen am Augenhintergrund können vom Ophthalmologen durch eine Funduskopie mit der Spaltlampe festgestellt werden. Daneben wird bei Vorerkrankungen sowie zur Überwachung des therapeutischen Verlaufs die OCT als bildgebendes Verfahren eingesetzt [Spital, 2018]. Ihre hochauflösende Schnittbild-darstellung ermöglicht im Vergleich zur Funduskopie eine detaillierte Visualisierung krankheitsspezifischer Veränderungen und eine quantitative Auswertung [Markomiche-lakis u. a., 2004]. Abbildung 5.1 zeigt einen OCT-B-Scan einer Retina mit DMÖ. Durch Einblutungen entsteht vermehrt intraretinale Flüssigkeit (*engl.*: intraretinal fluid, kurz: IRF) im Bereich der inneren nukleären sowie äußeren plexiformen Schicht. Die durch das IRF verursachte Schwellung hat außerdem zur Folge, dass sich die Retinaschichten morphologisch stark verformen und teilweise sogar durchtrennt werden. Sowohl das IRF als auch die Beschaffenheit der Retinaschichten stellen wichtige Biomarker dar und dienen als funktionelle Prognosefaktoren bei der DMÖ-Therapie [Zur u. a., 2018].

Um Biomarker in OCT-Bilddaten präzise erfassen zu können, sind Segmentierungen der entsprechenden Strukturen erforderlich. Wie im vorangegangenen Kapitel 4 darge-legt, ist eine manuelle Annotation sehr zeitaufwendig und besonders in der klinischen Praxis nicht durchführbar. Um sowohl Zeit als auch personelle Ressourcen zu sparen, ist die Entwicklung automatischer Segmentierungsmethoden vorangetrieben worden.

Ebenso wie die Haut besteht auch die Retina des menschlichen Auges aus mehre-ren Schichten, die einer strikten Hierarchie folgen (siehe Abb. 5.1). Aus diesem Grund nutzen nach dem heutigen Stand der Technik viele Segmentierungsverfahren Graphen-

modelle (siehe Kapitel 4.2.2.3) zur Nachverarbeitung, um eine korrekte Topologie der Retinaschichten zu gewährleisten. Jedoch ist die manuelle Konzeption eines Graphenmodells mit unterschiedlichen Beschränkungen für die individuellen Retinaschichten und Pathologien sehr komplex und reduziert die Flexibilität des Verfahrens (siehe Kapitel 4.4).

Aufgrund ihrer Vorteile basieren aktuelle Segmentierungsverfahren vorwiegend auf FCNs. Die Hierarchie der Retinaschichten wird jedoch von den meisten dieser Methoden nicht berücksichtigt, sodass es zu topologischen Fehlern im Segmentierungsergebnis kommen kann. In diesem Kapitel wird daher eine Methodik vorgeschlagen, welche die Forminformation der einzelnen Retinaschichten zur Regularisierung der Segmentierung nutzt, wodurch eine korrekte Topologie eingehalten werden soll. Neben den Retinaschichten werden DMÖ-bedingte Flüssigkeitseinlagerungen segmentiert, die einen wichtigen Biomarker für das Krankheitsbild darstellen.

Der Aufbau dieses Kapitel wird im Folgenden kurz beschrieben. Nach der Präsentation der Literaturübersicht (Kapitel 5.1.1) wird zunächst der verwendete Bilddatensatz vorgestellt (Kapitel 5.2.1), auf dem die Entwicklung der hier vorgestellten Methode fußt. Anschließend erfolgt die Präsentation der Methoden des vorgeschlagenen Segmentierungsverfahrens in Kapitel 5.2.2. Die Evaluation und der Vergleich mit anderen Methoden sowie die daraus resultierenden Ergebnisse werden in Kapitel 5.3 beschrieben. Abschließend werden die Ergebnisse in Kapitel 5.4 zusammengefasst und die gewonnenen Erkenntnisse diskutiert.

5.1.1 Literaturübersicht

In den vergangenen Jahren sind mehrere Methoden zur kombinierten Segmentierung von Retinaschichten und Flüssigkeitseinlagerungen in OCT-Bilddaten vorgestellt worden. Wie aus der Literaturübersicht des vorangegangenen Kapitels 4.1.1 zu entnehmen ist, verfolgen die meisten Stand-der-Technik-Verfahren eine der folgenden drei Strategien: pixelweise Segmentierung, graphenbasierte Segmentierung und eine Kombination aus beidem.

Typische Vertreter der pixelweisen Segmentierungsverfahren nutzen meist einen trainierten Klassifikator, wie z. B. einen RF oder ein CNN, die eine Klassenwahrscheinlichkeit für jeden Pixel ausgeben. Jedoch werden topologische Zusammenhänge vom Klassifikator nicht berücksichtigt, was für die Segmentierung von hierarchisch aufgebauten Strukturen wie der Retina nachteilig ist und einen weiteren Nachverarbeitungsschritt erfordert. So schlagen Chiu u. a. [2015] in ihrer Arbeit eine Kernregression zur Klassifikation von acht Retinaschichten und Flüssigkeiten vor. Ein Graphenmodell in Verbindung mit dynamischer Programmierung dient als Nachverarbeitung. In einer ähnlichen Vorgehensweise segmentieren Montuoro u. a. [2017] elf verschiedene Retinaschichten zusammen mit verschiedenen Flüssigkeitstypen. Für eine initiale pixelweise Klassifikation setzen sie einen unüberwachten RF-Klassifikator ein. Anschließend wird

die Segmentierung durch ein Graphenmodell verfeinert. Um die komplexe Interaktion der einzelnen Retinastrukturen zu berücksichtigen, wird die Segmentierung iterativ mithilfe einer speziellen Autokontext-Methodik ergänzt.

Eine detaillierte Übersicht graphen- und modellbasierter Segmentierungsmethoden wurde bereits im vorherigen Kapitel vorgestellt. Diese Modelle basieren dabei direkt auf der topologischen Gegebenheit der Zielstruktur, wodurch ihre Konzeption nicht trivial ist. Novosel u. a. [2017] stellen in ihrer Arbeit einen Ansatz zur gemeinsamen Segmentierung von Retinaschichten und Läsionen unter Verwendung von locker verknüpften Level Sets vor. Dabei werden die Läsionen als zusätzliche veränderbare Ebene modelliert, die durch Hilfsflächen abgegrenzt wird.

Roy u. a. [2017] schlagen das ReLayNet vor, ein auf der U-Net-Architektur aufbauendes FCN, mit dem sie sowohl Retinaschichten also auch Flüssigkeiten segmentieren. Liu u. a. [2018] verwenden eine halb-überwachte Segmentierungsmethode, bei der ein FCN mittels adversereller Lernstrategie trainiert wird. Ungeachtet ihrer Fähigkeit höchst komplexe Kontextinformation anhand ihrer Trainingsbilddaten zu erlernen, können Faltungsnetzwerke sowie die zuvor genannten pixelweisen Klassifikationsmethoden topologische Beziehungen nicht zuverlässig garantieren.

Infolgedessen wurden mehrere FCN-basierte Methoden entwickelt, die flexibel Form- und Topologieinformationen in den Segmentierungsprozess integrieren, um diese Problematik zu adressieren. Ravishankar u. a. [2017] nutzen einen Autoencoder, um Forminformation zu lernen. BenTaieb u. a. [2016] entwickelten eine spezielle Fehlerfunktion, um geometrisches und topologisches Vorwissen in die Segmentierung einzubauen. Mithilfe einer Adjazenzmatrix modellieren Ganaye u. a. [2019] Nachbarschaftsbeziehungen zwischen anatomischen Strukturen, welche in die Fehlerfunktion integriert werden. Auf diese Weise lassen sich unplausible Strukturbeziehungen bestrafen. Um Retinaschichten in Gegenwart von Flüssigkeiten topologisch korrekt zu segmentieren, schlagen Wei u. a. [2020] die Mutex-Dice-Fehlerfunktion vor, welche die gegenseitigen Beziehungen der Strukturen berücksichtigt. Mithilfe einer Kaskade aus zwei U-Nets segmentieren He u. a. [2019] Retinaoberflächen sowie mikrozystische Makulaödeme. Dabei führt das erste Netz eine pixelweise Klassifikation aus, während das zweite zur Nachverarbeitung eingesetzt wird, um eine fehlerfreie Topologie der Retinaoberflächen sicherzustellen. Hierzu statten He u. a. das Netzwerk mit vollständig verbundenen Schichten an der Ausgabe aus, um direkt die Abstände der Retinaoberflächen zu regressieren. In einer sich anschließenden Arbeit kombinieren He u. a. [2021] diese beiden Schritte zu einem einheitlichen Framework, indem sie direkt die Verteilung der Oberflächenpositionen modellieren.

Weitere Arbeiten nutzen Forminformationen zur Regularisierung der Segmentierung, wobei vermehrt Distanzkarten zum Einsatz kommen. Anders als bei (binären) Segmentierungsmasken, deren Pixel in Vorder- und Hintergrund eingeteilt werden, entsprechen die Intensitäten einer Distanzkarte den jeweiligen Abständen zur nächsten Objektkan-

te. Oft finden vorzeichenbehaftete Distanzkarten (*engl.*: signed distance map, kurz: SDM) Anwendung, die negative Distanzen innerhalb der Objektkontur aufweisen.

Glocker u. a. [2012] stellen eine Kombination aus pixelweiser Klassifikation und Formregression unter Verwendung von RFs zur Multiorgansegmentierung in CT-Bildern des Thorax vor. Die Forminformation wird implizit durch SDMs realisiert, was zu robusteren Segmentierungsergebnissen führt. Mithilfe einer CNN-basierten Regression segmentieren Heinrich u. a. [2018] synaptische Verbindungen in elektronenmikroskopischen Aufnahmen und haben den ersten Platz in der CREMI-Challenge² erzielt. Kepp u. a. [2019b] regressieren SDMs zur gemeinsamen Mehrklassensegmentierung von Retinaschichten und Flüssigkeiten in OCT-Bilddaten. Sie zeigen, dass der Einsatz von SDMs topologische Fehler reduziert. Die Evaluation wurde auf dem Datensatz aus [Chiu u. a., 2015] ausgeführt.

Aufgrund ihrer flexiblen Anwendbarkeit in FCN-basierten Frameworks haben SDMs jüngst erhöhte Aufmerksamkeit erhalten. In einer detaillierten Benchmark-Studie fassen [Ma u. a., 2020b] aktuelle Methoden zusammen, die anhand frei verfügbarer Bilddatensätze ausgewertet wurden. Ma u. a. teilen dabei die Methoden in zwei Klassen ein: zum einen in die Entwicklung neuer Fehlerfunktionen unter Verwendung von SDMs [Kervadec u. a., 2019; Xue u. a., 2020] und zum anderen in die Hinzunahme der Regression von SDMs als unterstützende Hilfsaufgabe (*engl.*: auxiliary tasks) [Navarro u. a., 2019].

Inspiziert durch die Arbeit von Kepp u. a. [2019b] nutzen weitere Arbeiten SDMs zur Segmentierung von Retinaschichten. Cao u. a. [2020] verwenden in ihrer Arbeit ein Multi-Task-Framework, das Retinaschichten in Form von SDMs ergänzend zu ihrer pixelweisen Klassifikation regressiert. Aus derselben Forschungsgruppe schlagen Liu u. a. [2020] eine ähnliche Methode zur Segmentierung von Retinaschichten vor. Ähnlich zu ihrer Vorarbeit wird eine adverserielle Lernstrategie angewandt, um das FCN auf schwierige OCT-Regionen zu fokussieren. Anders als in [Kepp u. a., 2019b] nutzen beide erwähnten Ansätze SDMs ausschließlich zur Regression der einzelnen Retinaoberflächen. Zwar zeigt die SDM-basierte Regression aus [Kepp u. a., 2019b] eine verbesserte Berücksichtigung der Topologie, allerdings erzeugt das Verfahren eine eher diffuse, unscharfe Segmentierung von Flüssigkeitsregionen. Dies liegt mitunter an der Tatsache, dass anders als die Retinaschichten Flüssigkeiten keiner bestimmten Topologie folgen. Krankheitsbedingte lokale Verformungen der Retinaschichten werden vom Verfahren nur bedingt erfasst.

Aufbauend auf [Kepp u. a., 2019b] wird in diesem Kapitel ein Multi-Task-Framework vorgeschlagen, bei dem parallel zur SDM-Regression der Retinaschichten ein weiterer Klassifikationspfad im FCN verwendet wird, um eine schärfere Abgrenzung der Flüssigkeitsregionen zu gewährleisten. Um die Konsistenz zwischen Regression und pixel-

²Circuit Reconstruction from Electron Microscopy Images (CREMI)
<https://cremi.org/>

weiser Klassifikation zu erhöhen, wird die duale Aufgabenkonsistenz (*engl.*: dual task consistency, kurz: DTC) aus [Luo u. a., 2020] adaptiert. Die DTC erlaubt eine halbüberwachte Lernstrategie, wodurch die Anzahl verfügbarer Trainingsbilddaten deutlich erhöht wird. Um Stabilitätsprobleme während des Trainings zu vermeiden, schlagen Xue u. a. [2020] hierfür eine spezielle Fehlerfunktion vor, welche ebenfalls im hier vorgeschlagenen Framework Anwendung findet.

5.2 Material und Methoden

5.2.1 Verwendete Bilddaten

Die in diesem Kapitel vorgestellte Methodik wurde anhand des in [Chiu u. a., 2015] vorgestellten und öffentlich zugänglichen Duke-OCT-Datensatzes³ entwickelt und evaluiert. Dieser Datensatz besteht aus insgesamt zehn OCT-Scans von Patienten mit DMÖ, die mit einem Heidelberg Spectralis SD-OCT-Scanner (Heidelberg Engineering GmbH) aufgenommen wurden. Jede Aufnahme besteht aus insgesamt 61 B-Scans, die jeweils eine Größe von 512×740 Pixel aufweisen. Pro OCT-Aufnahme haben zwei medizinische Experten manuell jeweils Retinaschichten sowie Flüssigkeiten in elf an der Fovea zentrierten B-Scans (Abstand zum fovealen B-Scan: 0, ± 2 , ± 5 , ± 10 , ± 15 , ± 20) annotiert. Hieraus ergeben sich insgesamt 110 annotierte B-Scans pro Experte. Abgesehen von den Flüssigkeiten (■) wurden folgende Retinaschichten segmentiert: retinale Nervenfaserschicht (*engl.*: retinal nerve fibre layer, kurz NFL ■), Ganglienzellschicht (*engl.*: ganglion cell layer, kurz: GCL) und innere plexiforme Schicht (*engl.*: inner plexiform layer, kurz: IPL) zusammengefasst zu GCL-IPL (■), innere nukleäre Schicht (*engl.*: inner nuclear layer, kurz: INL ■), äußere plexiforme Schicht (*engl.*: outer plexiform layer, kurz: OPL ■), äußere nukleäre Schicht (*engl.*: outer nuclear layer, kurz: ONL) und myoide Zone der Photorezeptorinnensegmente (*engl.*: inner segment myeloid, kurz: ISM) zusammengefasst zu ONL-ISM (■), ellipsoide Zone der Photorezeptorinnensegmente (*engl.*: inner segment ellipsoid, kurz: ISE ■), Außensegmente der Photorezeptorzellen (*engl.*: outer segments of the photoreceptor cells, kurz: OS) bis retinales Pigmentepithel (*engl.*: retinal pigment epithelium, kurz: RPE) zusammengefasst zu OS-RPE (■). Gemeinsam mit dem Hintergrund besteht die Grundwahrheit somit aus $K = 9$ Klassen. In Abbildung 5.2 wird beispielhaft ein B-Scan mit den manuellen Segmentierungen beider Experten dargestellt. Weitere Details zum Datensatz können aus [Chiu u. a., 2015] entnommen werden.

5.2.2 Multi-Task-Learning zur formbasierten Segmentierung

Das grundlegende Ziel des Multi-Task-Learnings (MTL) besteht darin, bereits in den ersten Schichten des Netzes gemeinsame Merkmalsrepräsentationen zu erhalten, wäh-

³Duke-OCT-Datensätze: <http://people.duke.edu/~sf59/software.html>

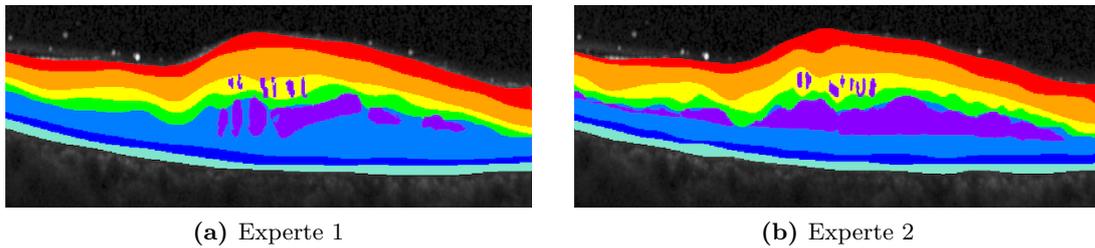


Abb. 5.2: Segmentierungsbeispiel eines B-Scans durch die beiden Experten.

rend die verschiedenen Aufgaben in späteren Netzabschnitten in ihren jeweiligen separaten Pfaden erlernt werden. Gegenüber dem Single-Task-Learning (STL) hat das MTL mehrere Vorteile. Durch die gemeinsame Nutzung von Netzwerkschichten wird der erforderliche Speicherbedarf erheblich reduziert. Außerdem zeigen die MTL-Netze höhere Inferenzgeschwindigkeiten, da die Mehrfachberechnung von Merkmalen für jede Aufgabe in den gemeinsam genutzten Schichten explizit vermieden wird. Gerade MTL hat das Potenzial, die Leistung eines Netzes zu verbessern, falls die entsprechenden Aufgaben komplementäre Eigenschaften aufweisen oder eine regulierende Wirkung haben [Vandenhende u. a., 2021]. MTL-Methoden, die SDMs zur Regularisierung der Segmentierung integrieren, basieren meist auf FCNs, die aus einer Encoder-Decoder-Architektur (z. B. U-Net) bestehen, deren aufgabenspezifische Pfade entweder aus separaten Decodern (*engl.* multi-reconstruction branches), auch „Soft-Parameter-Sharing“ (SPS) genannt, oder aus separaten Ausgabeschichten (*engl.* multi-head, kurz: MH) bestehen, was als „Hard-Parameter-Sharing“ (HPS) bezeichnet wird [Ma u. a., 2020b].

Um sowohl die Anzahl der gemeinsamen Merkmalsrepräsentationen zu maximieren als auch die Summe der lernbaren Netzwerkparameter gering zu halten, wird für das hier vorgeschlagene MTL-Segmentierungsframework eine Multi-Head-FCN-Architektur gewählt. Ein schematischer Überblick der Methode ist in Abbildung 5.3 dargestellt. Als Basis-FCN wird das in Kapitel 4.2.3.1 vorgestellte DCU-Net verwendet. Um sicherzustellen, dass selbst Retinae mit stark geschwollener Makula komplett vom rezeptiven Feld erfasst werden, wird die Architektur des DCU-Nets auf fünf Auflösungsstufen erweitert. Die Aufgabenpfade werden durch zwei entsprechende 1×1 -Faltungsschichten am Netzwerkkende realisiert.

Während der Trainingsphase wird das Segmentierungsframework als MTL-Modell bestehend aus Klassifikation und Regression formuliert. Dabei wird für jeden Eingabe-B-Scan zum einen eine pixelweise Klassifikation über alle Klassen erzeugt und zum anderen eine SDM für jede Retinaschicht regressiert. Die pixelweise Klassifikation erfolgt analog zu Kapitel 4.2.3, wobei jedem Bildpixel \mathbf{x} des Eingangs-B-Scans eine von insgesamt $K = 9$ Klassen, bestehend aus sieben Retinaschichten, Flüssigkeiten und Hintergrund, zugeordnet wird. Für das Training des Klassifikationspfades kommt der

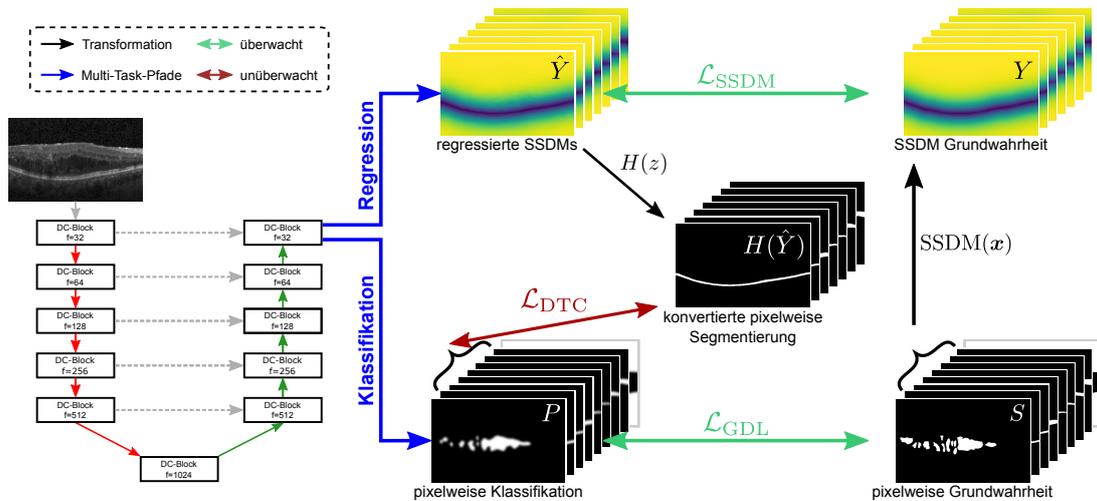


Abb. 5.3: Schematische Übersicht der vorgeschlagenen MTL-Segmentierungsmethode. Zusätzlich zu einer pixelweisen Klassifikation wird eine Regression der Retinaschichten in Form von skalierten SDMs (SSDMs) vorgenommen. Die DTC sorgt für Konsistenz zwischen den einzelnen Aufgaben und ermöglicht ein halb-überwachtes Training.

GDL (\mathcal{L}_{GDL}) als Fehlerfunktion (siehe Kapitel 4.2.3.2) zum Einsatz, welcher zwischen der Softmax-Wahrscheinlichkeit P und der Grundwahrheit S berechnet wird. Über den Regressionspfad werden skalierte SDMs (SSDMs) der Retinaschichten geschätzt ($\mathcal{L}_{\text{SSDM}}$), was sich regularisierend auf die Segmentierung auswirkt und zu einer besseren Einhaltung der topologischen Beziehungen führen soll. Hierzu werden die binären Masken der Retinaschichten in der Grundwahrheit selektiert (siehe geschweifte Klammern in Abb. 5.3) und in SSDMs Y überführt. Die Flüssigkeits- und Hintergrundklasse werden nicht durch die SDMs regressiert, da sie keiner bestimmten Topologie folgen. Beide Aufgabenpfade sind über die DTC (\mathcal{L}_{DTC}) verbunden, die für Konsistenz zwischen den klassifizierten und regressierten Retinaschichten beider Aufgabenpfade sorgt. Die DTC erlaubt außerdem die Nutzung nicht gelabelter B-Scans über eine halb-überwachte Trainingsstrategie. Während der Testphase wird lediglich der Klassifikationspfad zur Erzeugung der finalen Segmentierung S genutzt, die durch die Anwendung des $\arg \max$ -Operators auf die pixelweisen Klassenwahrscheinlichkeiten bestimmt wird.

Im Folgenden wird die vorgeschlagene Methode in mehreren Unterkapiteln vorgestellt, darunter die Berechnung und Regression von SDMs, das halb-überwachte Lernen mithilfe dualer Aufgabenkonsistenz sowie die Trainingsprozedur.

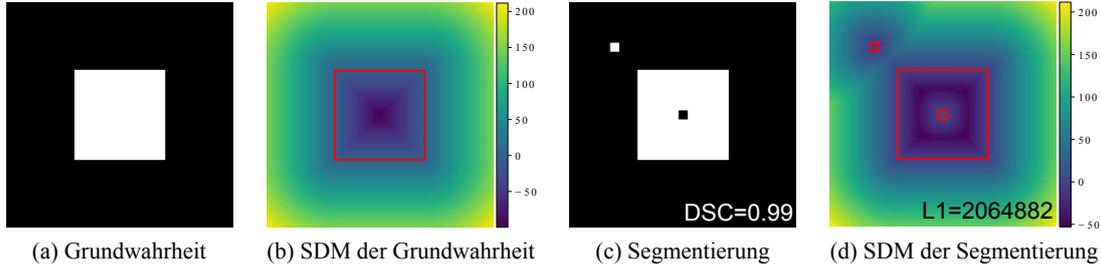


Abb. 5.4: Vergleich zwischen einer binären Segmentierungsmaske und ihrer SDM. Die in (a) und (c) gezeigten binären Masken stellen beispielhaft eine Grundwahrheit und die Segmentierung dar. Die zugehörigen SDMs sind in (b) und (d) dargestellt. Der Nullpunkt der SDM ist durch eine rote Kontur gekennzeichnet. Dieser Vergleich zeigt, dass bei der Regression von SDMs selbst kleinste Segmentierungsfehler im Gegensatz zur pixelweisen Klassifikation stark bestraft werden, was der Vergleich zwischen DSC- und L1-Wert zeigt.

5.2.2.1 Regression von Signed Distance Maps

Die Regression von SDMs erfordert zunächst die Erstellung der entsprechenden Grundwahrheit, welche sich einfach anhand der Grundwahrheitsmasken S_l berechnen lassen. Für jedes Klassenlabel l kann die entsprechende SDM unter Verwendung der binären One-Hot-Kodierung folgendermaßen berechnet werden:

$$\text{SDM}(\mathbf{x}) = \begin{cases} -\inf_{\mathbf{y} \in \mathcal{S}} \|\mathbf{x} - \mathbf{y}\|_2, & \mathbf{x} \in \Omega_{\text{in}} \\ 0, & \mathbf{x} \in \mathcal{S} \\ \inf_{\mathbf{y} \in \mathcal{S}} \|\mathbf{x} - \mathbf{y}\|_2, & \mathbf{x} \in \Omega_{\text{out}} \end{cases} \quad (5.1)$$

wobei Ω_{in} und Ω_{out} jeweils den Bereich innerhalb oder außerhalb des Objekts bezeichnen und \mathcal{S} repräsentiert die Objektkontur. Daneben gibt $\|\mathbf{x} - \mathbf{y}\|_2$ die euklidische Distanz zwischen den Bildpunkten \mathbf{x} und \mathbf{y} an. Für eine effiziente Berechnung der SDM wird der Algorithmus von Maurer u. a. [2003] verwendet.

Mithilfe der SDMs können Segmentierungsmasken in einen höherdimensionalen Raum kodiert werden, welcher komplexere Informationen über die Objektstruktur enthält. So kennzeichnen bspw. negative Werte Bildbereiche innerhalb des Objekts, während positive Werte Hintergrundbereiche darstellen. Die Werte nahe oder gleich null bezeichnen Positionen nahe oder auf der Objektkontur. Kleine Änderungen wirken sich bei der pixelweisen Klassifikation nur auf die entsprechenden lokalen Punkte aus, was sich kaum auf den DSC-Wert überträgt (siehe Abb. 5.4, Unterbild (c)). Dagegen führen bereits kleine Änderungen zu einer erheblichen Beeinflussung der SDM-Werte und damit zu deutlich größeren Fehlern (siehe Abb. 5.4, Unterbild (d)). Infolgedessen sind SDMs im Vergleich zu (binären) Segmentierungsmasken wesentlich empfindlicher gegenüber Fehlern und bestrafen insbesondere Ausreißer härter [Ma u. a., 2020a]. Da große Entfernungen zur jeweiligen Retinaschicht nicht aussagekräftig sind, werden,

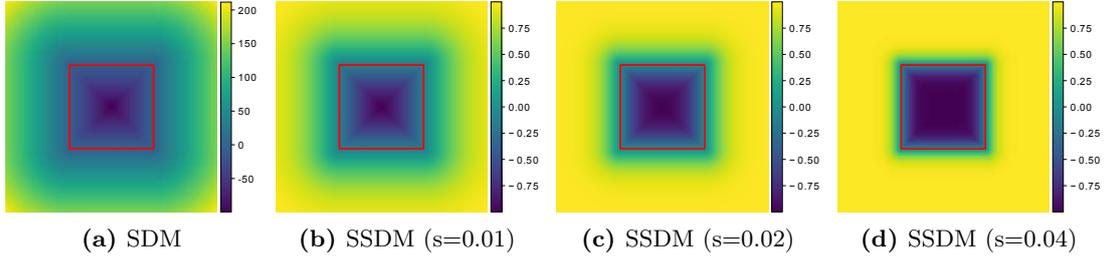


Abb. 5.5: Vergleich zwischen SDM und SSDM. Über die Skalierung werden die Distanzwerte des SSDM auf den Wertebereich $[-1, 1]$ skaliert und auf eine bestimmte Distanz begrenzt, die über den Parameter s gesteuert wird.

wie von [Heinrich u. a., 2018] vorgeschlagen, SSMDs verwendet, die auf den Bereich $[-1, 1]$ begrenzt sind:

$$\text{SSDM}(\mathbf{x}) = \tanh(\text{SDM}(\mathbf{x}) \cdot s). \quad (5.2)$$

Mithilfe des Skalierungsfaktors s wird festgelegt, bei welchem maximalen Abstand von der Objektkontur die SSDM bereits gegen -1 oder 1 konvergiert. Somit werden durch die Skalierung weiter entfernte Bereiche weniger berücksichtigt und die Regressionsaufgabe wird vereinfacht. In Abbildung 5.5 wird beispielhaft eine SDM mit verschiedenen Skalierungen dargestellt. Für die nachfolgenden Experimente wird $s = 0.02$ gesetzt, was empirisch ermittelt wurde.

Die Regression der SSDMs der Retinaschichten erfolgt über den entsprechenden Pfad im FCN-Modell (siehe Abb. 5.3). Für jede Retinaschicht wird eine zugehörige SSDM gelernt. Als nichtlineare Aktivierungsfunktion wird die tanh-Funktion genutzt, welche die SSDM-Werte auf den Bereich $[-1, 1]$ beschränkt und somit gleichzeitig zu einer Beschleunigung des Lernprozesses führt. Zur Regression der SSDMs wird die L1-Norm als Fehlerfunktion eingesetzt, da sie gegenüber der L2-Norm weniger sensitiv auf Ausreißer reagiert. Sie ist wie folgt definiert:

$$\mathcal{L}_{L1} = \sum_{\mathbf{x}} |\hat{Y}(\mathbf{x}) - Y(\mathbf{x})|. \quad (5.3)$$

Dabei beschreibt Y die Grundwahrheit und \hat{Y} die Prädiktion des Netzes. Xue u. a. [2020] zeigen in ihrer Arbeit, dass beim Regressieren mehrerer Strukturen mit der L1-Fehlerfunktion ein instabiles Training auftreten kann. Um dieses Problem zu lösen, schlagen Xue u. a. eine ergänzende produktbasierte Fehlerfunktion vor:

$$\mathcal{L}_{\text{Produkt}} = - \sum_{l=1}^{K-2} \frac{\sum_{\mathbf{x}} \hat{Y}_l(\mathbf{x}) Y_l(\mathbf{x}) + \varepsilon}{\sum_{\mathbf{x}} \hat{Y}_l(\mathbf{x}) Y_l(\mathbf{x}) + \sum_{\mathbf{x}} \hat{Y}_l^2(\mathbf{x}) + \sum_{\mathbf{x}} Y_l^2(\mathbf{x}) + \varepsilon}. \quad (5.4)$$

Durch das Produkt wird ein falsches Vorzeichen in der regressierten SSDM bestraft. Es ist zu beachten, dass $\mathcal{L}_{\text{Produkt}}$ nur für die Klassenlabels der Retinaschichten $l \in$

$\{1, \dots, K-2\}$ und nicht für den Hintergrund ($l = 0$) oder die Flüssigkeiten ($l = K-1$) berechnet wird. Für die SSDM-Regression wird die Kombination beider Fehlerfunktionen genutzt: $\mathcal{L}_{\text{SSDM}} = \mathcal{L}_{\text{Produkt}} + \mathcal{L}_{\text{L1}}$. Die Kombination aus beiden Fehlerfunktionen weist größere Gradientenbeträge auf, sodass zum einen das Regressionsergebnis verbessert und zum anderen das Training robuster gestaltet werden kann [Xue u. a., 2020].

5.2.2.2 Duale Aufgabenkonsistenz

Da die Retinaschichten einerseits auf Pixelebene (Klassifikation) und andererseits auch als höherdimensionale SSDM (Regression) über die beiden Aufgabenpfade ausgegeben werden, ist es notwendig, dass beide Prädiktionen zueinander konsistent sind. Luo u. a. [2020] schlagen hierfür eine duale Aufgabenkonsistenz (DTC) vor. Zu diesem Zweck werden die Werte der regressierten SSDMs mithilfe der Heaviside-Funktion, die durch eine differenzierbare Approximation angenähert werden kann, in Wahrscheinlichkeiten umgewandelt:

$$H(z) = \frac{1}{1 + e^{-z \cdot k}}. \quad (5.5)$$

Hierbei repräsentiert z den SSDM-Wert an der Pixelposition \mathbf{x} und k bestimmt die Steilheit der Funktion. Analog zu Xue u. a. [2020] wird der Wert von k in den nachfolgenden Experimenten auf einen Wert von 1500 gesetzt. Die DTC wird schließlich mithilfe einer auf der L1-Norm basierenden Fehlerfunktion (\mathcal{L}_{DTC}) umgesetzt und wird zwischen $H(\hat{Y}_l(\mathbf{x}))$ und $P_l(\mathbf{x})$ berechnet. Ein wesentlicher Vorteil durch die DTC ist, dass keine Grundwahrheit benötigt wird, denn \mathcal{L}_{DTC} wird zwischen den (transformierten) Prädiktionen der Retinaschichten des Regressions- und Klassifikationspfades bestimmt (siehe Abb. 5.3). Dies ermöglicht eine halb-überwachte Trainingsprozedur, die im Folgenden vorgestellt wird.

5.2.2.3 Halb-überwachtes Training

Mithilfe der DTC wird ein halb-überwachtes Training ermöglicht (siehe Abb. 5.3). Da die DTC-Fehlerfunktion keine Grundwahrheit erfordert, können zuzüglich zu den elf gelabelten B-Scans auch die 50 nicht gelabelten B-Scans pro OCT-Scan für das Training verwendet werden. Die pixelweise Klassifikation (\mathcal{L}_{GDL}) sowie die Regression der SSDMs ($\mathcal{L}_{\text{SSDM}}$) folgt einem überwachten Training, wohingegen die DTC (\mathcal{L}_{DTC}) vollständig unüberwacht abläuft. Die gesamte Fehlerfunktion setzt sich wie folgt zusammen:

$$\mathcal{L} = \underbrace{\mathcal{L}_{\text{GDL}} + \beta \mathcal{L}_{\text{SSDM}}}_{\text{überwacht}} + \underbrace{\gamma_t \mathcal{L}_{\text{DTC}}}_{\text{unüberwacht}} \quad (5.6)$$

Dabei beschreiben β und γ_t Gewichtungsfaktoren. Für β wird ein Wert von fünf in den Experimenten gewählt. Hinter γ_t verbirgt sich eine von Luo u. a. [2020] vorgeschlagene zeitabhängige Gewichtungsfunktion, mit der das Gleichgewicht zwischen den

überwachten und unüberwachten Fehlerfunktionen gesteuert wird und folgendermaßen definiert ist:

$$\gamma_t(t) = e^{\left(-5\left(1-\frac{t}{t_{\max}}\right)^2\right)}. \quad (5.7)$$

Hier geben t die aktuelle und t_{\max} die maximale Trainingsepoche an. Durch $\gamma_t(t)$ wird das Netzwerk in den ersten Epochen zunächst überwacht trainiert. Im späteren Verlauf des Trainings erhält die unüberwachte Komponente der Fehlerfunktion \mathcal{L}_{DTC} schrittweise mehr Gewichtung. Die Verwendung von $\gamma_t(t)$ ist essenziell für den Gebrauch der DTC, da sie anderenfalls zu Beginn des Trainings das Lernen der Teilaufgaben behindern würde.

5.3 Experimente und Ergebnisse

Die vorgeschlagene Methodik wird anhand mehrerer Vergleichsverfahren evaluiert. Um die Vorteile von MTL zu untersuchen, werden die einzelnen Teilaufgaben des vorgestellten Frameworks als STL-Version realisiert. Hierzu wird aus dem jeweiligen Aufgabenpfad ein separates STL-Framework gebildet, das einerseits auf pixelweiser Klassifikation (\mathcal{L}_{GDL}) und andererseits auf der Regression von SSDMs ($\mathcal{L}_{\text{SSDM}}$) beruht. Klassifikations- und Regressionsframework werden nachfolgend jeweils als $\text{STL}_{\text{Klass.}}$ und $\text{STL}_{\text{Regr.}}$ beschrieben. Da sich SSDMs mithilfe der ableitbaren Heaviside-Funktion $H(z)$ in pixelweise Klassifikationen überführen lassen, liegt der Gedanke nahe, Regression und Klassifikation nacheinander in einem MTL-Framework durchzuführen, was der Methodik von Xue u. a. [2020] entspricht. Ein Vorteil dieses Ansatzes ist, dass keine aufgabenspezifischen Ausgabepfade erforderlich sind (striktes HPS) und die Aufgabenkonsistenz bereits implizit durch die aufeinander aufbauende Reihenfolge sichergestellt ist. Die Methodik wird im weiteren Verlauf als MTL_{Xue} bezeichnet. Schließlich wird die DTC und die damit verbundene halb-überwachte Lernstrategie mithilfe einer Ablation dieser Komponente untersucht. Folglich wird das vorgeschlagene Multi-Head-FCN ohne (MTL_{MH}) und mit der DTC ($\text{MTL}_{\text{MH}_{\text{DTC}}}$) evaluiert.

Die Evaluation der fünf zuvor beschriebenen Verfahren ($\text{STL}_{\text{Klass.}}$, $\text{STL}_{\text{Regr.}}$, MTL_{Xue} , MTL_{MH} , $\text{MTL}_{\text{MH}_{\text{DTC}}}$) wird mittels Leave-One-Patient-Out-Kreuzvalidierung durchgeführt. Als Bilddatensatz wird der in Kapitel 5.2.1 vorgestellte Bilddatensatz genutzt. Dadurch, dass zwei separate Expertenannotationen existieren, wird eine Kreuzvalidierung zweimal mit jeweils einer anderen Grundwahrheit durchlaufen. Als Vorbereitung für die Kreuzvalidierung werden alle B-Scans sowohl in der Höhe als auch in der Breite zugeschnitten, um einerseits irrelevante Bildbereiche auszuschließen und andererseits die Bilddaten an die manuellen Annotationen anzupassen, die einen kleineren Bildbereich einnehmen. Gleichzeitig wird durch den Zuschnitt Speicher eingespart, da bereits durch die weitere Auflösungsstufe des DCU-Nets ein erhöhter Speicherverbrauch berücksichtigt werden muss. Pro Kreuzvalidierungsdurchlauf werden die OCT-Scans des Datensatzes in nicht überlappende Trainings-, Validierungs- und Testdatensätze

Tabelle 5.1: Gemittelte Ergebnisse der beiden Kreuzvalidierungsdurchläufe, durchgeführt mit Annotationen des jeweiligen Experten. Die besten Ergebnisse sind fett gedruckt. Die Abstände für ASSD und HD sind in Pixel angegeben. Ergänzend wird die IRR zwischen beiden Experten angegeben.

Strukturen → Methoden ↓	NFL	GCL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE	Flüssigkeit	
									
DSC	STL _{Klass.}	0.861	0.896	0.782	0.740	0.872	0.859	0.842	0.553
	STL _{Regr.}	0.844	0.888	0.779	0.739	0.876	0.857	0.848	0.519
	MTL _{Xue}	0.856	0.896	0.784	0.745	0.877	0.859	0.845	0.520
	MTL _{MH}	0.869	0.900	0.791	0.758	0.882	0.863	0.852	0.530
	MTL _{MH_{DTC}}	0.871	0.902	0.796	0.746	0.877	0.866	0.853	0.522
	IRR	0.846	0.888	0.770	0.724	0.868	0.847	0.842	0.522
ASSD	STL _{Klass.}	1.46	1.64	1.96	2.07	2.36	1.13	1.18	–
	STL _{Regr.}	2.70	1.80	2.27	2.16	2.63	1.16	1.16	–
	MTL _{Xue}	2.49	1.64	2.06	2.09	2.43	1.17	1.20	–
	MTL _{MH}	1.58	1.56	2.12	1.99	2.52	1.10	1.13	–
	MTL _{MH_{DTC}}	1.40	1.53	1.93	1.91	2.32	1.06	1.14	–
	IRR	1.70	1.76	1.92	1.87	2.17	1.17	1.19	–
HD	STL _{Klass.}	13.80	12.13	15.34	15.83	19.04	5.97	5.86	–
	STL _{Regr.}	21.00	11.00	14.39	15.91	20.20	5.34	4.67	–
	MTL _{Xue}	17.98	9.71	13.74	15.25	19.23	6.29	5.40	–
	MTL _{MH}	13.16	9.20	13.29	14.84	20.10	5.05	5.27	–
	MTL _{MH_{DTC}}	11.83	8.55	12.22	12.98	17.80	4.62	4.30	–
	IRR	13.60	9.33	11.78	11.20	17.83	4.48	4.38	–

im Verhältnis von 7–2–1 aufgeteilt. Um das jeweilige Netzwerk weniger anfällig für atypische Morphologien zu gestalten, wie z. B. in den Regionen mit DMÖ-bedingten Flüssigkeiten, werden die Augmentierungstechniken aus Kapitel 4.2.3.2 durch elastische Transformationen [Simard u. a., 2003] ergänzt. Abgesehen von der erweiterten Augmentierung ist das Trainingsverfahren dasselbe wie im Kapitel zuvor. Zur quantitativen Analyse werden der DSC, die ASSD sowie die HD als Metriken (siehe Kapitel 3.1.2) zwischen den berechneten Segmentierungen und der korrespondierenden Grundwahrheit bestimmt. Die Mittelwerte dieser Metriken sind für jede Struktur in Tabelle 5.1 aufgeführt. Zusätzlich wird die IRR zwischen den Grundwahrheiten beider Experten für jede Metrik berechnet. Die Verteilung der Ergebnisse wird durch Boxplots veranschaulicht (siehe Abb. 5.6). Darüber hinaus werden drei qualitative Segmentierungsbeispiele in Abbildung 5.7 dargestellt.

Alle Verfahren segmentieren die einzelnen Retinaschichten mit einer hohen Genauigkeit, die sich in Bezug auf DSC und ASSD sich auf IRR-Niveau befindet. Aufgrund ihrer sehr dünnen Beschaffenheit, vor allem im Bereich der Fovea, erweist sich die Segmentierung der NFL als schwierig, was qualitativ zu beobachten ist (siehe Abb. 5.7).

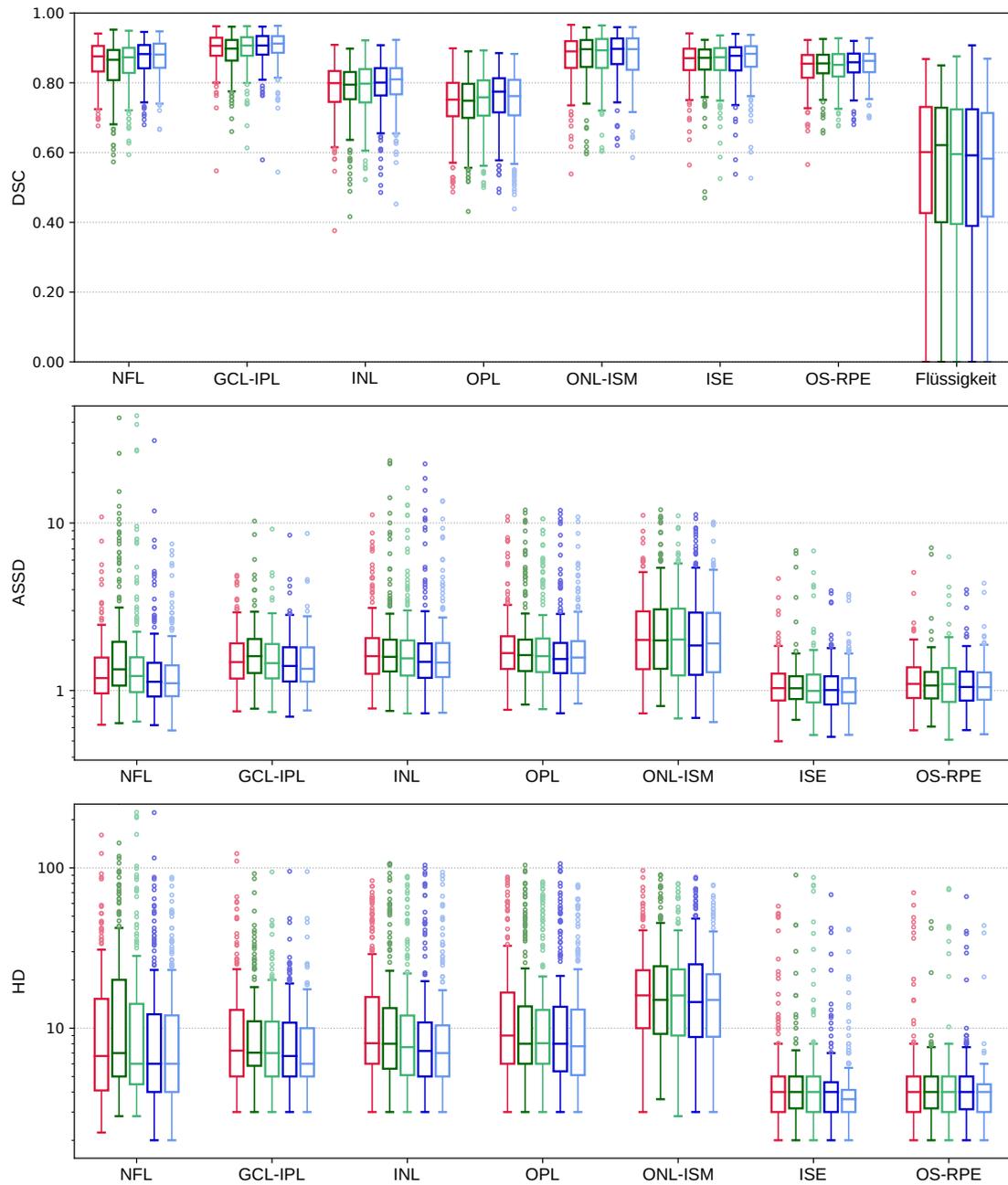


Abb. 5.6: Boxplot-Darstellung der DSC-, ASSD- und HD-Werte beider Kreuzvalidierungsdurchläufe. Zur Unterscheidung der Methoden wird folgende Farbkodierung genutzt: STL_{Klass}. (■), STL_{Reg}. (■), MTL_{Xue} (■), MTL_{MH} (■), MTL_{MH_{DTc}} (■).

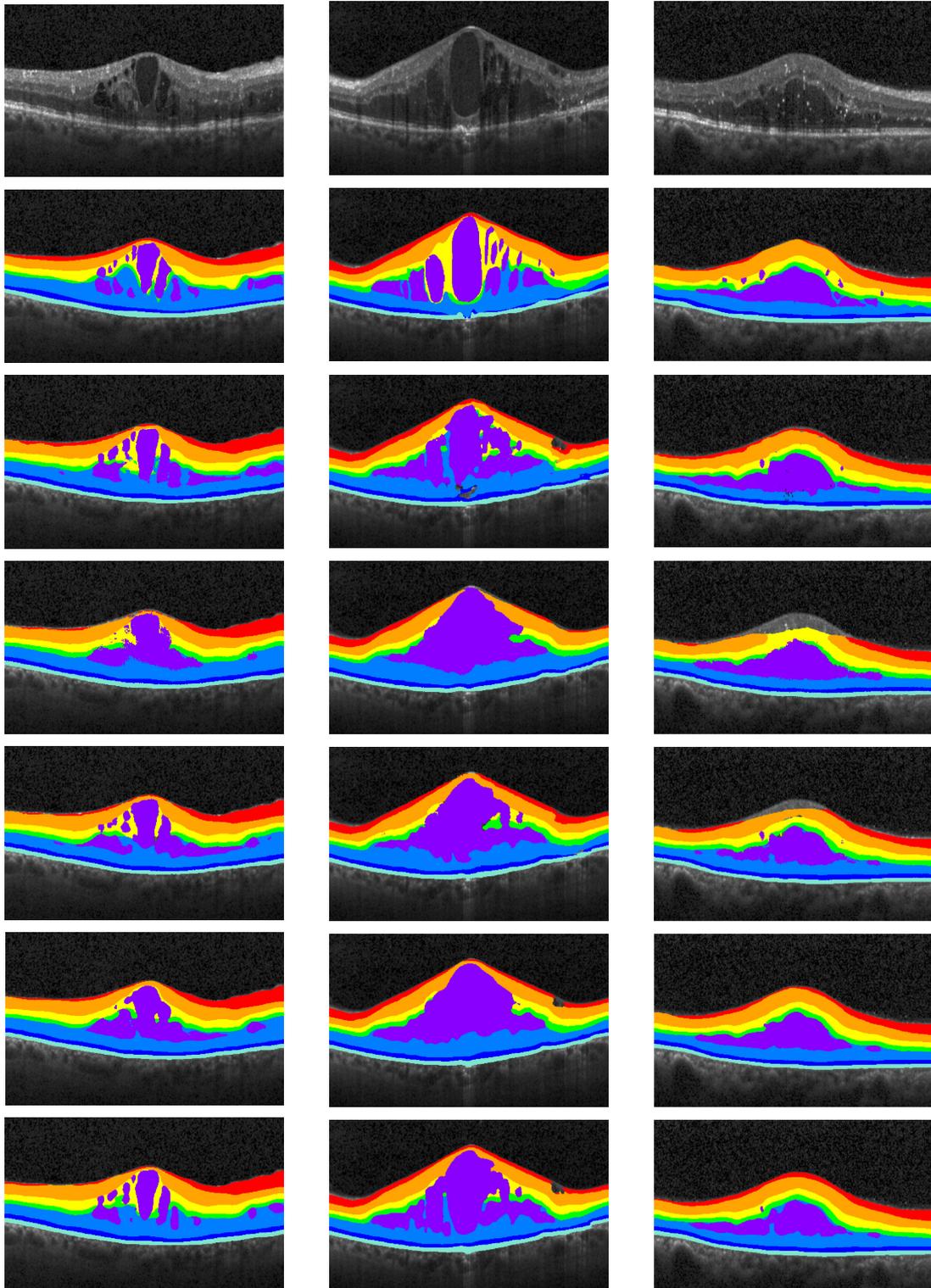


Abb. 5.7: Darstellung von drei beispielhaften Segmentierungen. Von oben nach unten: B-Scan, Grundwahrheit, $STL_{Klass.}$, $STL_{Regr.}$, MTL_{Xue} , MTL_{MH} , $MTL_{MH_{DTC}}$.

Noch herausfordernder stellt sich die Segmentierung der INL, OPL und ONL-ISM heraus, da sich in ihrer Region die DMÖ-bedingten Flüssigkeiten akkumulieren. Zwar erreichen alle Methoden einen hohen DSC für die ONL-ISM, was jedoch an der relativ großen Fläche dieser Schicht liegt. Zieht man ASSD und HD hinzu, so fällt auf, dass die mittleren Ergebniswerte gegenüber den restlichen Retinaschichten deutlich größere Distanzen aufweisen. Im Verhältnis zu ihrer Schichtdicke lassen sich sowohl die ISE als auch die OS-RPE am robustesten segmentieren, was auch an den sehr kurzen Interquartilabständen der entsprechenden Boxplots zu erkennen ist. Im Kontrast zu den Retinaschichten werden Flüssigkeiten von allen Methoden mit einer deutlich geringeren Genauigkeit segmentiert. Das $STL_{\text{Klass.}}$ -Verfahren erreicht hier den höchsten mittleren DSC-Wert von 0.553. Die DSC-Werte der übrigen Methoden liegen im Durchschnitt bei 0.522, was dem IRR-Niveau entspricht.

Gegenüber der Klassifikation ($STL_{\text{Klass.}}$) liefert die alleinige Regression ($STL_{\text{Regr.}}$) zwar schlechtere Ergebniswerte, jedoch wird die topologische Reihenfolge der Retinaschichten berücksichtigt, was sich in den beiden qualitativen Segmentierungsbeispielen in Abbildung 5.7 gut erkennen lässt. Andererseits werden kleinere Variationen im Verlauf der Retinaschichten weniger durch die Regression berücksichtigt, was sich in den höheren ASSDs widerspiegelt. Sehr schwierig ist die Regression der NFL und GCL-IPL, welche zum Teil in Bereichen mit starker Retinanschwellung nicht segmentiert werden. Flüssigkeiten lassen sich nur schwer regressieren und sind meist diffus segmentiert, sodass vereinzelte Flüssigkeitseinschlüsse nicht abgegrenzt werden können. Dies gelingt mithilfe der alleinigen Klassifikation ($STL_{\text{Klass.}}$) besser und es werden mit ihr im Mittel die höchsten DSC-Werte erreicht. Allerdings treten beim $STL_{\text{Klass.}}$ -Verfahren in Gegenwart von größeren Flüssigkeitsbereichen (siehe Abb. 5.7, mittleres Beispiel) vermehrt Topologiefehler auf.

Durch die nachgeschaltete Klassifikation (MTL_{Xue}) kann die Segmentierungsgenauigkeit bezüglich der NFL, GCL-IPL und INL im Vergleich zur alleinigen Regression ($STL_{\text{Regr.}}$) signifikant verbessert werden. Gegenüber der alleinigen Klassifikation lassen sich signifikante Verbesserungen hinsichtlich der HD-Werte für die GCL-IPL, INL und OPL feststellen. Qualitativ kann auch gezeigt werden, dass sich Flüssigkeiten besser abgrenzen lassen und weniger diffus segmentiert werden bei gleichzeitiger Berücksichtigung der topologischen Struktur. Mit Bezug auf die Retinaschichten erzielen die MTL_{MH} -Verfahren (MTL_{MH} und MTL_{MHDTC}) die besten Segmentierungsergebnisse. Bis auf wenige Ausnahmen unterscheiden sich die Ergebniswerte zu denen der STL-Methoden signifikant voneinander. Durch den Einsatz der DTC konnten die HD-Werte verringert werden, welche sich bis auf die NFL und die INL signifikant unterscheiden. Die MTL_{MHDTC} -Methode zeigt gegenüber allen STL-Verfahren signifikant kleinere HD-Werte, was für die MTL_{MH} -Variante lediglich nur auf die ersten zwei bis vier Retinaschichten (beginnend bei der NFL) zutrifft. Im Gegensatz dazu unterscheiden sich zwar die ASSD-Werte beider MTL_{MH} -Varianten nicht, wohl aber die DSC-Werte der Retinaschichten OPL, ONL-ISM sowie ISE. Bezogen auf die DSC-Werte in Tabel-

le 5.1 werden Flüssigkeiten von den MTL_{MH} -Methoden auf einem ähnlichen Niveau segmentiert wie die regressionsbasierten STL-Verfahren. Die qualitativen Beispieldarstellungen in Abbildung 5.7 zeigen andererseits, dass durch die Verwendung der DTC Flüssigkeiten schärfer abgegrenzt werden als ohne sie. Trotz der Aufspaltung in zwei unterschiedliche Aufgabenpfade werden topologische Fehler weitestgehend vermieden. Zudem werden durch die DTC kleinere Interquartilsabstände der DSC-Werte erreicht.

5.4 Zusammenfassung und Diskussion

Aufbauend auf der DCU-Net-Architektur aus Kapitel 4.2.3.1 wurde in diesem Kapitel ein Segmentierungsframework präsentiert, das mithilfe separater Aufgabenpfade sowohl Klassifikation als auch Regression miteinander kombiniert. Durch die Regression der einzelnen Retinaschichten als SSDMs wird die Forminformation implizit erlernt und damit die topologische Struktur der Retina berücksichtigt. Außerdem wird ein halb-überwachtes Trainingsschema durch die Integration der DTC ermöglicht, welche Konsistenz zwischen Klassifikation und Regression etabliert.

Die Auswertungen zeigten, dass durch die Verwendung von SSDMs die Topologie der Retinaschichten gegenüber der pixelweisen Klassifikation ($STL_{Klass.}$) verbessert berücksichtigt wird, allerdings reduziert sich durch die alleinige Regression ($STL_{Regr.}$) die Segmentierungsgüte. Gerade bei Retinae mit stärkeren Schwellungen treten vermehrt Segmentierungsfehler in den oberen Schichten auf. Flüssigkeiten lassen sich durch SSDMs nur sehr ungenau segmentieren, sodass fein abgegrenzte Einschlüsse als diffus zusammenhängendes Areal erkannt werden. Eine naheliegende Erklärung hierfür ist, dass Flüssigkeiten, anders als die Retinaschichten, keiner bestimmten Form folgen und je nach Lage anders ausgeprägt sind.

Eine verbesserte Flüssigkeitsabgrenzung ließ sich durch das Anfügen einer weiteren Klassifikationsaufgabe ($MTL_{X_{ue}}$) demonstrieren. Auch die Segmentierungsgüte der Retinaschichten konnte durch diese Methode optimiert werden. Jedoch scheint die direkte Kopplung an die SSDMs mittels approximierter Heaviside-Funktion zu derselben Problematik zu führen, da ebenfalls vermehrt die oberen Retinaschichten unvollständig segmentiert werden. Die MTL_{MH} -Methoden erzielen hinsichtlich der Retinaschichtsegmentierung im Mittel die besten Resultate. Durch das Lernen mehrerer Aufgaben über separate Aufgabenpfade wird eine höhere Generalisierbarkeit erreicht, was einer Überanpassung der Netzwerkparameter entgegenwirkt. Gleichzeitig werden so Merkmalsrepräsentationen gelernt, die von beiden Aufgaben präferiert werden. Gerade bei einer nur geringen Anzahl verfügbarer Trainingsbilddaten, die aufgrund der DMÖ-Erkrankung zudem große Variationen aufweisen, wird hierdurch der Fokus auf relevante Merkmale gelenkt.

Mithilfe der DTC wird eine Konsistenz zwischen beiden Aufgaben erreicht und eine gegenseitige Unterstützung gewährleistet. Diese Kopplung ist dabei nicht so streng wie

bei der $MTL_{X_{ue}}$ -Methode, was gewisse Freiheiten der einzelnen Aufgaben erlaubt. Die Anwendung der DTC zeigte, dass durch die Nutzung weiterer nicht gelabelter Bilddaten sich die Robustheit steigern lässt und auch die Flüssigkeitsabgrenzung verfeinert werden kann. Die Segmentierungsergebnisse der $MTL_{MH_{DTC}}$ -Methode weisen am wenigsten Topologiefehler auf, die anderenfalls eine mühsame Nachverarbeitung nach sich ziehen, besonders in Gegenwart von Flüssigkeiten.

Im Vergleich zur Nachverarbeitung mittels graphenbasierter Optimierung (Kapitel 4.2.2.3) bietet das vorgeschlagene $MTL_{MH_{DTC}}$ -Framework einige Vorteile. Da kein komplexes Graphenmodell benötigt wird, lässt sich die Methode leicht auf andere Fragestellungen anpassen. Die Prozesskette bleibt nach wie vor kurz (siehe Abb. 4.5). Der einzige Nachteil ist, dass Topologiefehler zwar signifikant reduziert, aber nicht vollständig vermieden werden. Insbesondere nur vereinzelt auftretende Variationen im Bilddatensatz können zu Unsicherheiten und damit zu Fehlern führen, wie bspw. kleine Flüssigkeitseinlagerungen im Bereich der NFL (siehe Abb. 5.7, mittlere Spalte).

Generell ist die Segmentierung von Flüssigkeiten in OCT-Bilddaten durch FCNs eine herausfordernde Aufgabe. Während Probleme wie ein starkes Klassenungleichgewicht durch spezielle Fehlerfunktionen [Ma u. a., 2021] oder Sampling-Strategien angegangen werden können, verbleibt eine Restunsicherheit in der Abgrenzung der einzelnen Flüssigkeitsregionen. Aufgrund des schwachen Bildkontrasts in diesen Regionen ist selbst eine manuelle Identifikation oftmals nicht eindeutig. Dies wird unter anderem durch die sehr unterschiedlich segmentierten Flüssigkeiten der beiden Experten (siehe Abb. 5.2) und der niedrigen IRR bestätigt. Dieselben Erkenntnisse sind in [Bogunović u. a., 2019] gemacht worden.

Die Bestimmung von Unsicherheiten eines Segmentierungsverfahrens oder der zugrunde liegenden Bilddaten würde somit eine sinnvolle Ergänzung darstellen. Für die Anwendbarkeit in der klinischen Praxis können unsichere Bereiche in einer Segmentierung kenntlich gemacht werden und dienen somit als ergänzendes Feedback für die Interpretation. Das nachfolgende Kapitel beschäftigt sich mit der Bestimmung und Analyse von Unsicherheiten bei der Segmentierung eines OCT-basierten Flüssigkeitsbiomarkers der AMD.

Kapitel 6

Unsicherheitsanalyse bei der Segmentierung von PEDs in Home-OCT-Bilddaten

Neue innovative und kostengünstige OCT-Systeme ermöglichen eine flexible Überwachung der altersbedingten Makuladegeneration zu Hause. In Kombination mit aktuellen FCN-basierten Segmentierungsalgorithmen kann die Bewertung AMD-bedingter Flüssigkeitsbiomarker, wie z. B. der Pigmentepithelabhebung, unterstützt werden. Allerdings kann die begrenzte Verfügbarkeit medizinischer Bilddaten sowie verrauschte Grundwahrheiten zu einer geringen Generalisierung der FCN-Modelle führen. Um die Zuverlässigkeit der Modellvorhersagen zu beurteilen, können Schätzungen der segmentierungsbezogenen Unsicherheiten verwendet werden. In diesem Kapitel werden zwei Unsicherheitsarten bei der Segmentierung von Pigmentepithelabhebungen in Bilddaten eines OCT-Geräts zur Heimanwendung analysiert. Epistemische und aleatorische Unsicherheiten werden durch Dropout und Augmentierung zum Testzeitpunkt bestimmt. Die Auswertungen werden sowohl mit pixelweisen als auch mit strukturbezogenen Unsicherheitsmetriken durchgeführt.

6.1 Einleitung und Motivation

Die altersbedingte Makuladegeneration gehört zu den Hauptgründen für Erblindung in der westlichen Welt [Steinmetz u. a., 2021]. Ähnlich wie beim DMÖ verursacht die AMD Schäden im Bereich der Makula, wodurch das scharfe zentrale Sehen erheblich beeinträchtigt wird und alltägliche Aufgaben wie das Lesen und die Gesichtserkennung nicht mehr möglich sind. Die AMD verläuft in verschiedenen Stadien. Hierbei wird zwischen einem frühen, einem mittleren und einem späten Stadium unterschieden. Letzteres unterteilt sich nochmals in eine trockene und eine feuchte Form. Im höheren Alter lagern sich vermehrt Abbauprodukte des Stoffwechsels in der Retina zwischen dem retinalen Pigmentepithel (RPE) und der Bruch-Membran (BM) ab, die

als Drusen bezeichnet werden und zu Pigmentepithelabhebungen in den betroffenen Bereichen führen. Im frühen und mittleren Stadium der AMD nimmt die Anzahl sowie Größe der Drusen zu, was sich aber in den meisten Fällen kaum auf den Seheindruck auswirkt. Etwa 15 % der Patienten erreichen das späte AMD-Stadium, in dem ca. 5 % die trockene und ca. 10 % die feuchte Form ausbilden [Brandl u. a., 2016]. Bei der späten trockenen AMD kommt es zu einem langsam fortschreitenden, großflächigen Absterben der Photorezeptorzellen, das mit einer zunehmend stärkeren Sehbehinderung einhergeht. Sie wird daher auch als geographische Atrophie (GA) bezeichnet. Trotz ihres relativ geringen Vorkommens ist die feuchte Form für 90 % aller AMD-bedingten Erblindungen verantwortlich [Yonekawa u. a., 2015]. Die feuchte AMD schreitet rasant voran, sodass es bereits in einem kurzen Zeitraum zu einer gravierenden Minderung des Sehvermögens kommen kann. Durch die Nährstoffunterversorgung der Photorezeptorzellen werden Botenstoffe gebildet, primär der Vascular Endothelial Growth Factor (VEGF), die das Einwachsen abnormaler Blutgefäße in Richtung Retina fördern. Allerdings verdrängen die neuen Blutgefäße unkontrolliert umliegende Strukturen und führen zu unregelmäßigen Flüssigkeitsaustritten und deren Einlagerungen in der Retina, wodurch diese Form auch als exsudative oder neovaskuläre AMD bezeichnet wird. Diese sogenannten Exsudationen und nachfolgende Vernarbungsprozesse verursachen bei den Patienten eine rasche Sehverschlechterung bis hin zur Erblindung. Wohingegen es derzeit noch keine zugelassene Behandlung für die GA gibt, so lässt sich das Blutgefäßwachstum der AMD durch VEGF-Antikörper (Anti-VEGF) reduzieren und teilweise unterdrücken, was zu einer deutlichen Verbesserung des Sehvermögens führt oder eine fortlaufende Verschlechterung aufhält [Rosenfeld u. a., 2006].

Umfassende Studien haben ergeben, dass es jedoch nicht ausreicht, die Indikation zur Behandlung allein auf den subjektiven Seheindruck des Patienten zu stützen, da dieser Veränderungen oft erst zu spät wahrnimmt [Schmidt-Erfurth u. a., 2014]. Durch hochauflösende OCT-Aufnahmen der Retina kann bereits lange vor den subjektiven Sehveränderungen eine Krankheitsaktivität festgestellt werden. Die charakteristischen Retinaveränderungen der AMD können im OCT als Biomarker genutzt werden. Zu den prominentesten gehören die intraretinale Flüssigkeit (IRF), die subretinale Flüssigkeit (*engl.*: subretinal fluid, kurz: SRF) und die Pigmentepithelabhebung (*engl.*: pigment epithelial detachment, kurz: PED), welche als direkte Behandlungsindikatoren verwendet werden [Schmidt-Erfurth u. a., 2016]. Der Behandlungszeitpunkt sollte deshalb idealerweise zum frühestmöglichen Zeitpunkt im Krankheitsverlauf festgelegt werden, um irreversible Schäden zu vermeiden [Lim u. a., 2012]. Das interindividuell sehr unterschiedliche Therapieansprechen und Rezidivverhalten stellt eine große Herausforderung für die behandelnden Ärzte dar. Alle modernen und gleichzeitig erfolgreichen Behandlungsstrategien basieren auf regelmäßigen Kontrollen der Retina mittels OCT. Dabei hat die Untersuchungsfrequenz einen entscheidenden Einfluss auf den Erfolg der AMD-Behandlung. Allerdings sind die Untersuchungsintervalle derzeitiger Behandlungsschemata oft zu lang und unflexibel, sodass eine frühzeitige Diagnose

pathologischer Veränderungen nicht gewährleistet ist. Zur Optimierung der Therapie müsste die Untersuchungsfrequenz signifikant erhöht werden, z. B. auf ein wöchentliches oder sogar tägliches Intervall. Kürzere Untersuchungsintervalle führen hingegen zu einer starken Belastung für den Patienten und gleichzeitig zu hohen Personalkosten. Demzufolge lässt sich eine solche angestrebte Überwachungsfrequenz in der Praxis nur durch eine Heimlösung realisieren, z. B. durch ein OCT-System zur Selbstuntersuchung im Haus des Patienten. Ein neuartiges Konzept eines kostengünstigen Home-OCT zur Selbstuntersuchung ist kürzlich als Prototyp von der Firma Visotec GmbH⁴ entwickelt worden und wird derzeit in klinischen Studien getestet [Sudkamp u. a., 2016, 2018]. Jedoch ist eine manuelle Auswertung aufgrund der, durch den täglichen Einsatz zu erwartenden, großen Menge an Bilddaten nicht praktikabel. Daher ist eine wesentliche Anforderung an das Home-OCT für die AMD-Überwachung eine automatisierte computergestützte Auswertung (CAD) der Bilddaten, um den Verlauf der Erkrankung verfolgen zu können.

In diesem Kapitel wird daher eine Methodik zur Segmentierung von AMD-Biomarkern in Home-OCT-Bilddaten vorgeschlagen, um dem Patienten zu Hause ein direktes Feedback geben zu können, ob eine therapeutische Behandlung in der Klinik erforderlich ist. Als Biomarker werden PEDs verwendet, da diese aufgrund der hohen Reflektivität des RPE gut detektiert werden können. Weitere Gründe für die Auswahl dieses Biomarkers werden in Kapitel 6.3.1 gegeben. Gängige CAD-Systeme bieten kaum eine Einschätzung, wie sicher sie ihre Entscheidung getroffen haben. Besonders für den Monitoring Einsatz beim Patienten zu Hause wäre ein derartiges Unsicherheitsfeedback von großer Bedeutung. Beispielsweise könnte so das CAD-System den Patienten warnen, falls eine Auswertung nur unter einer geringen Sicherheit getroffen wird, z. B. falls die Bildqualität unzureichend war. Angesichts dessen wird eine Unsicherheitsschätzung mittels Bayes'scher Inferenz in die vorgeschlagene Segmentierungsmethode integriert.

Das Kapitel ist folgendermaßen strukturiert. Zuerst wird der aktuelle Stand der Technik vorgestellt. Danach wird im Material- und Methodenteil des Kapitels zunächst der verwendete Bilddatensatz (Kapitel 6.3.1) des Home-OCT beschrieben. Die Bestimmung von Unsicherheiten (Kapitel 6.3.2) bei der Segmentierung von PED (Kapitel 6.3.4) wird im Anschluss vorgestellt. Zur Analyse der Unsicherheiten werden in Kapitel 6.3.4 verschiedene Metriken präsentiert. Die Durchführung der Experimente und die daraus resultierenden Ergebnisse werden in Kapitel 6.4 vorgestellt, die zum Schluss des Kapitels (Kapitel 6.5) zusammengefasst und diskutiert werden.

6.2 Literaturübersicht

Sowohl das Auftreten als auch die Volumenzunahme der eingangs beschriebenen AMD-Biomarker stellen ein direktes Behandlungskriterium dar. Für eine praktikable klini-

⁴<https://visotec.health/>

sche Auswertung wurden in den letzten Jahren zahlreiche Methoden zur semantischen Segmentierung solcher OCT-basierter Flüssigkeitsbiomarker für die AMD-Diagnose entwickelt. Da in diesem Kapitel eine Methodik zur PED-Segmentierung vorgeschlagen wird, fokussiert sich die nachfolgende Literaturübersicht vorwiegend auf entsprechende Arbeiten. Eine systematische Übersicht von Algorithmen zur automatischen Analyse verschiedener AMD-Biomarkern ist in [Wintergerst u. a., 2017] gegeben.

PED wird als Ablösung des RPE von der inneren Kollagenschicht der BM definiert. Der durch diese Trennung entstandene Raum wird von Blut, serösem Exsudat (eiweißhaltige Flüssigkeit), drusenoidem Material, fibrovaskulärem Gewebe oder einer Kombination davon eingenommen [Mrejen u. a., 2013]. Aufgrund der klar definierten Lage des PED nutzen frühe Arbeiten primär graphenbasierte Segmentierungsmodelle [Chiu u. a., 2012; Dufour u. a., 2012]. Chen u. a. [2012] schlagen ein zweistufiges Verfahren zur Segmentierung von SRF und PED vor. Nach einer Klassifikation der Retinaschichten werden mithilfe eines Graphenmodells die Schichtübergänge segmentiert, wobei SRF- und PED-Bereiche in die Kostenfunktion integriert werden. Chen u. a. [2013] stellen in ihrer Arbeit eine Methode zur Segmentierung von Drusen vor, die ebenfalls zu Abhebungen des RPE führen, und nutzen die spezielle Morphologie der Retina als A-priori-Wissen für ihren Algorithmus aus.

Aktuellere Arbeiten verwenden überwiegend Deep-Learning-Verfahren zur PED-Segmentierung, von denen ein Großteil auf Encoder-Decoder-Architekturen basiert. In ihrer Arbeit stellen Zadeh u. a. [2017] drei verschiedene Ansätze zur Drusensegmentierung vor, bei der das U-Net zum Einsatz kommt. Asgari u. a. [2019] schlagen einen MTL-Ansatz zur Segmentierung von RPE, Drusen und BM vor, wobei für jede Klasse ein separater Decoder verwendet wird. Ferner sind im Rahmen der RETOUCH-Challenge⁵ Arbeiten zur Segmentierung von IRF, SRF und PED in Bilddaten drei verschiedener OCT-Scanner vorgestellt worden. Auch hier fußt die Mehrzahl der Beiträge auf dem U-Net, das unter anderem in seiner Architektur erheblich erweitert [Apostolopoulos u. a., 2017], durch adverserielle Lernstrategien optimiert [Tennakoon u. a., 2017] oder durch weitere Nachverarbeitungen ergänzt worden ist, um Falschklassifikationen zu eliminieren [Lu u. a., 2019]. Eine umfassende Analyse und Zusammenfassung der Challenge wird in [Bogunović u. a., 2019] beschrieben.

Eine erste Methode zur Segmentierung von PED sowie der gesamten Retina in Home-OCT-Bilddaten schlagen Kepp u. a. [2020] vor. Sie nutzen eine U-Net-Architektur in Kombination mit einem Convolutional Denoising Autoencoder, welcher genutzt wird, um Fehler der Retinasegmentierung zu korrigieren, die bspw. durch Bewegungsartefakte verursacht werden.

Trotz des hohen Niveaus aktueller Forschungsarbeiten, bleibt es stets herausfordernd, zuverlässige Segmentierungsergebnisse zu erzielen, vorrangig in medizinischen Bilddaten. Gründe dafür können eine unzureichende Bildqualität, pathologisch be-

⁵Retinal OCT Fluid Challenge (RETOUCH); <https://retouch.grand-challenge.org/>

dingte Inhomogenitäten und verschiedene Bildmodalitäten sein. Hinzu kommt, dass die für das Training benötigten manuellen Annotationen meist nur in geringer Anzahl vorhanden sind und in ihrer Erstellung sehr zeitintensiv sind, da sie einen medizinischen Experten benötigen. Jedoch können auch Expertensegmentierungen in schwierigen Fällen fehlerbehaftet sein, was sich wiederum negativ auf das Training des Modells auswirkt. Infolgedessen ist die Unsicherheitsbestimmung entscheidend zur Bewertung der Zuverlässigkeit von Segmentierungsmethoden. Mehrere Arbeiten haben die Bestimmung von Unsicherheiten untersucht. Ein populärer Ansatz ist das Bayesian Deep Learning von Gal u. a. [2016]. Gal u. a. nutzen das Prinzip des Monte-Carlo-Samplings (MC) durch Dropout zum Testzeitpunkt (*engl.*: test time dropout, kurz: TTD), wodurch sich Stichproben aus der Posterior-Verteilung approximieren lassen. Mehrere generierte MC-Stichproben können dann genutzt werden, um die epistemische Unsicherheit zu bestimmen, welche die Unsicherheit der Netzwerkparameter aufgrund von unvollständigem Wissen beschreibt. In [Kendall u. a., 2017a,b] wird diese Methodik verwendet, um Unsicherheiten von semantischen Segmentierungen zu berechnen. Anstelle des MC-Dropouts schlagen Lakshminarayanan u. a. [2017] Deep Ensembles zur Unsicherheitsbestimmung vor, wobei mehrere verschieden initialisierte Netzwerkmodelle trainiert werden. Die Modelle erzeugen für jede Eingabe verschiedene Ausgaben, die zur Bestimmung der Unsicherheiten eingesetzt werden können. [Wang u. a., 2019] zeigen in ihrer Arbeit, dass sich durch verschiedene Augmentierungen derselben Eingabe zur Testzeit (*engl.*: test time augmentation, kurz: TTA) die aleatorische Unsicherheit berechnen lässt. Im Gegensatz zur epistemischen Unsicherheit definiert die aleatorische Unsicherheit die Existenz von Störungen der Eingabedaten. Kendall u. a. [2017b] demonstrieren in ihrer Arbeit, dass sich die aleatorische Unsicherheit ebenso durch Hinzufügen einer zusätzlichen Ausgabe schätzen lässt.

Ferner sind Unsicherheiten bei der Segmentierung von OCT-Bilddaten durch mehrere Forschergruppen analysiert worden. So werden sowohl epistemische als auch aleatorische Unsicherheiten bei der Segmentierung von Photorezeptoren [Orlando u. a., 2019] oder von Retinaschichten [Sedai u. a., 2018] untersucht. Letzteres integrieren Sedai u. a. [2019] in ein schwach überwachtes Teacher-Student-Modell, wobei die Unsicherheiten zur Verbesserung der Genauigkeit und Zuverlässigkeit beitragen. Darüber hinaus können Unsicherheiten zur unüberwachten Anomaliedetektion genutzt werden. Hierzu trainieren Seeböck u. a. [2019] ein Bayesian U-Net mit Pseudo-Grundwahrheiten von Retinaschichten gesunder Patienten, die durch ein graphenbasiertes Verfahren erstellt werden. Das trainierte Netzwerk erhält dann zur Testzeit OCT-Scans pathologischer Retinae. Aufgrund fehlender Information über pathologische Strukturen können auf diese Weise Pathologien erkannt werden. Des Weiteren können Unsicherheiten Feedback für interaktive Annotationstools bieten [Zadeh u. a., 2019].

Die Schätzungen von epistemischen und aleatorischen Unsicherheiten bei der Segmentierung von Drusen in Klinik-OCT-Bilddaten mittels U-Net wird in der Arbeit von Joy u. a. [2021] untersucht. In Anlehnung an die Untersuchungen von Joy u. a.

[2021] werden in diesem Kapitel sowohl epistemische als auch aleatorische Unsicherheiten bei der zweidimensionalen PED-Segmentierung in Home-OCT-Bilddaten analysiert. Im Vergleich zu [Joy u. a., 2021] dient anstelle des Standard-U-Nets hierzu das in Kapitel 4.2.3.1 vorgestellte DCU-Net. Weiterhin wird die Analyse der geschätzten Unsicherheiten durch weitere Untersuchungen ergänzt. Zum einen wird durch eine Filterung unsicherer Bildpunkte die Qualität der Unsicherheiten bewertet. Mithilfe eines sich daraus ableitenden Bewertungsmaßes lassen sich die ermittelten Unsicherheiten quantitativ vergleichen. Zum anderen werden im Vergleich zu Joy u. a. [2021] weitere strukturbezogene Unsicherheitsmetriken untersucht. Die Methodik baut dabei auf den gewonnenen Erkenntnissen aus [Kepp u. a., 2020] auf. Teile der in diesem Kapitel präsentierten Ergebnisse wurden in [Kepp u. a., 2022] veröffentlicht.

6.3 Material und Methoden

6.3.1 Verwendete Bilddaten

Das kostengünstige Home-OCT-System der Firma Visotec GmbH soll dem Patienten ermöglichen, den Verlauf der eigenen AMD-Erkrankung selbstständig zu Hause zu untersuchen, ohne dass ein Arzt anwesend sein muss. Anstatt Querschnittsbilder in axialer Richtung, wie bei kommerziell erhältlichen OCT-Systemen üblich, setzt sich ein Scan des Home-OCT aus transversalen Bildern verschiedener Tiefen zusammen. Mithilfe moderner Kamertechnik und einer einfachen breitbandigen Lichtquelle können kostengünstige und kompakte OCT-Systeme konstruiert werden, die für den Patienten leicht zugänglich und einfach zu bedienen sind. Aufgrund der kostengünstigen Bauweise weisen die Scans des Home-OCT im Vergleich zu kommerziellen Klinik-OCT-Systemen eine geringere Qualität auf und haben ein deutlich reduziertes SNR. Außerdem können Bewegungsartefakte die Bildinterpretation beeinträchtigen. Das FOV des Home-OCT ist aufgrund der schnellen Akquisitionszeit (~ 0.9 Sekunden pro Volumen) kleiner als bei klinischen OCT-Systemen, genügt aber für die klinische Diagnose gängiger AMD-Biomarker [von der Burchard u. a., 2017]. Ein qualitativer Vergleich zwischen Scans des Home-OCT und eines in der Klinik genutzten SD-OCT ist in Abbildung 6.1 dargestellt.

Aus einer Studie [von der Burchard u. a., 2021] stehen Bilddaten des Home-OCT von insgesamt 39 Patienten mit einer AMD-Erkrankung zur Verfügung. Insgesamt umfasst der Bilddatensatz 68 Augen, von denen 90 % eine AMD und 4 % eine andere Erkrankung haben. Die restlichen 6 % der Augen sind gesund. Um möglichst viele Bilddaten für das Training nutzen zu können, werden bei 29 Patienten ergänzend zum Studienauge auch die Aufnahmen des anderen Auges verwendet. Alle OCT-Scans haben eine axiale Auflösung von $5.8 \mu\text{m}$ und eine laterale Auflösung von $3 \mu\text{m}$. Die Volumengröße beträgt $250 \times 968 \times 608$ px mit einem Abbildungsbereich von $1.45 \times 2.9 \times 1.8$ mm.

Alle OCT-Scans wurden durch einen medizinischen Experten manuell annotiert. Im Gegensatz zum PED zeichnen sich IRF und SRF aufgrund des geringen SNR weniger

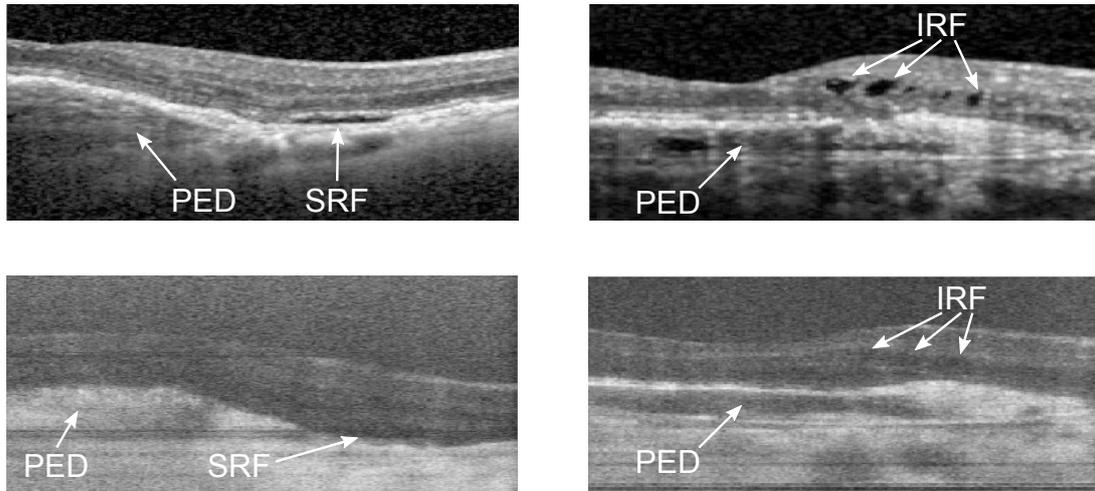


Abb. 6.1: Darstellung von B-Scans zweier Retinae mit verschiedenen AMD-Biomarkern, die sowohl mit einem Klinik-OCT-System (Heidelberg Spectralis, obere Reihe) als auch mit dem Home-OCT (untere Reihe) aufgenommen wurden. Aufgrund unterschiedlicher FOVs sind jeweils nur Ausschnitte der B-Scans dargestellt.

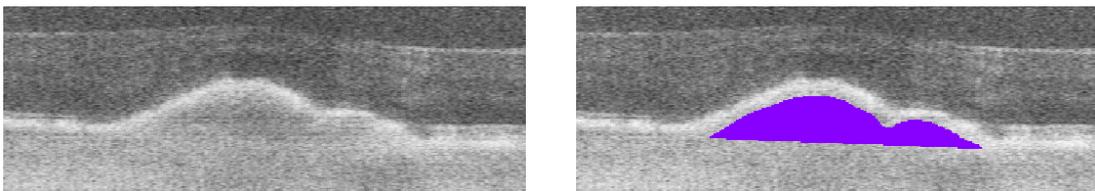


Abb. 6.2: Ausschnitt eines Home-OCT-B-Scans ohne und mit überlagelter PED-Grundwahrheit.

deutlich in den Scans des Home-OCT ab (siehe Abb. 6.1) und kommen im Gegensatz zum PED äußerst selten im gesamten Datensatz vor. Demzufolge wird das Segmentierungsproblem auf PED als Biomarker beschränkt. PED wird oberhalb durch das RPE und unterhalb durch die BM begrenzt. Ein Beispiel für eine manuelle Segmentierung ist in Abbildung 6.2 dargestellt. Um die Annotationsarbeit in einem überschaubaren Rahmen zu halten, wurden pro Home-OCT-Volumen 32 äquidistante B-Scans manuell segmentiert. Somit umfasst der gesamte Bilddatensatz 2176 gelabelte B-Scans.

6.3.2 Bestimmung epistemischer und aleatorischer Unsicherheiten

Vorhersagen von KNNs werden stets mit einer gewissen Konfidenz getroffen. Für den binären Klassifikationsfall sind nur die Ausgaben „wahr“ und „falsch“ möglich. Das Modell ist nicht in der Lage, die Information zu geben: „Ich bin mir nicht sicher“, was gerade in Anwendungsbereichen wie der Medizin oder dem autonomen Fahren

problematisch sein kann. Die Integration von Unsicherheitsbestimmungen in CAD-Systeme ist daher essenziell.

Insgesamt kann zwischen zwei Gruppen von Unsicherheiten unterschieden werden: die epistemische und die aleatorische Unsicherheit. Die mangelnde Kenntnis eines Systems wird als epistemische Unsicherheit (*griech.*: episteme, *dt.*: Wissen) bezeichnet. Bezogen auf KNNs beschreibt sie, was ein Modell aufgrund mangelnder Trainingsdaten nicht oder nur unzureichend gelernt hat und gibt somit die Unsicherheit über die Modellparameter wieder. Durch das Hinzufügen weiterer Trainingsdaten kann die epistemische Unsicherheit reduziert werden. Als aleatorische Unsicherheit (*lat.*: alea, *dt.*: Würfel) wird die durch probabilistische Variationen eines zufälligen Ereignisses verursachte Unsicherheit bezeichnet und ist nicht reduzierbar, da die zugrundeliegenden Variablen immer Schwankungen unterliegen werden. Ein Beispiel hierfür ist der Würfelwurf, dessen inhärente Unsicherheit nicht reduzierbar ist und die Wahrscheinlichkeit eines Ergebnisses stets dieselbe bleibt. Im Folgenden wird die Bestimmung der epistemischen und aleatorischen Unsicherheit mit Bezug zur FCN-basierten Segmentierung vorgestellt.

Die epistemische Unsicherheit lässt sich mithilfe von Bayesian Deep Learning ermitteln. Eine umfassende Darstellung ist hierzu in [Gal u. a., 2016] gegeben. Für eine betrachtete Menge von Trainingsbilddaten $\mathbf{I} = \{I_1, \dots, I_{|\mathbf{I}|}\}$ und entsprechenden Grundwahrheiten $\mathbf{S} = \{S_1, \dots, S_{|\mathbf{S}|}\}$ mit $|\mathbf{I}| = |\mathbf{S}|$ wird ein probabilistisches Modell $F_{\text{seg}} : I \rightarrow S$ gelernt. Zur Testzeit erzeugt dieses Modell die Segmentierung S_{test} für das Eingabebild I_{test} mit der Wahrscheinlichkeit

$$p(S_{\text{test}}|I_{\text{test}}, \mathbf{I}, \mathbf{S}) = \int p(S_{\text{test}}|I_{\text{test}}, \mathbf{w})p(\mathbf{w}|\mathbf{I}, \mathbf{S})d\mathbf{w}. \quad (6.1)$$

Hierbei beschreibt \mathbf{w} die Netzparameter von $F_{\text{seg}}(\cdot)$. Zudem gibt $p(S_{\text{test}}|I_{\text{test}}, \mathbf{w})$ die Softmax-Wahrscheinlichkeit von S_{test} sowie $p(\mathbf{w}|\mathbf{I}, \mathbf{S})$ die Wahrscheinlichkeit der Netzparameter in Abhängigkeit von den Trainingsbilddaten \mathbf{I} und \mathbf{S} an. Aufgrund der hohen Dimensionalität von $p(\mathbf{w}|\mathbf{I}, \mathbf{S})$ ist die genaue Bestimmung der Posteriori-Verteilung aus Gleichung 6.1 kaum möglich und wird deshalb mittels Variationsinferenz angenähert [Graves, 2011]. Hierzu wird eine Variationsverteilung der Netzwerkparameter $q(\mathbf{w})$ durch die Minimierung der Kullback-Leibler-Divergenz $\text{KL}(q(\mathbf{w})||p(\mathbf{w}|\mathbf{I}, \mathbf{S}))$ gelernt, welche ein Maß für die Ähnlichkeit zweier Verteilungen ist. Gal u. a. [2016] zeigen, dass das Training von CNNs mit einem Dropout nach jeder Faltungsschicht die Minimierung der Kullback-Leibler-Divergenz annähert. Auf diese Weise kann die prädiktive Verteilung

$$\begin{aligned} q(S_{\text{test}}|I_{\text{test}}, \mathbf{I}, \mathbf{S}) &= \int p(S_{\text{test}}|I_{\text{test}}, \mathbf{w})q(\mathbf{w})d\mathbf{w} \\ &\approx \frac{1}{N} \sum_{i=1}^N p(S_{\text{test}}|I_{\text{test}}, \hat{\mathbf{w}}_i) =: \frac{1}{N} \sum_{i=1}^N P_i \end{aligned} \quad (6.2)$$

näherungsweise ermittelt werden. Durch die Summe mehrerer aus $\hat{\boldsymbol{w}} \sim q(\boldsymbol{w})$ gezogener MC-Stichproben kann so das Integral aus Gleichung 6.2 geschätzt werden. Gal u. a. [2015, 2016] zeigen in ihrer Arbeit, dass das Ziehen aus $q(\boldsymbol{w}_i)$ durch einen Dropout zur Testzeit angenähert werden kann. Generell sorgen Dropout-Schichten beim Training für eine bessere Generalisierung, indem sie eine vorher spezifizierte Anzahl von Neuronen pro Berechnungsschritt zufällig deaktivieren [Srivastava u. a., 2014]. Beim Bayesian Deep Learning wird ein Dropout verwendet, um die Verteilung der Netzgewichte zu modellieren. Anschließend werden die Dropout-Schichten zur Testzeit (TTD) aktiviert, um MC-Stichproben zu generieren, wobei dasselbe Eingabebild N -mal vom Bayesian FCN verarbeitet wird. Über den Mittelwert aller MC-Stichproben ergibt sich die finale Segmentierung. Unsicherheiten lassen sich durch verschiedene Kenngrößen ermitteln wie bspw. Varianz, Standardabweichung oder Entropie.

Die aleatorische Unsicherheit kann ähnlich zum MC-Dropout über stichprobenbasierte Augmentierung zur Testzeit (TTA) ermittelt werden [Wang u. a., 2019]. Generell werden Augmentierungstechniken während der Trainingsphase von KNNs genutzt, um die Anzahl verfügbarer Daten künstlich zu erhöhen und so eine bessere Generalisierung zu gewährleisten. Hier wird nun die Augmentierung dazu genutzt, um zur Testzeit N verschiedene Versionen desselben Bilds zu erzeugen. Diese erhält das FCN jeweils als Eingabe, um für jede Variation eine Prädiktion zu erzeugen. Die finale Segmentierung sowie ihre aleatorische Unsicherheit wird analog zum MC-Dropout aus den erzeugten Stichproben ermittelt. Sofern geometrische Transformationen wie Spiegelung, Rotation oder Translation Bestandteil der Augmentierung sind, muss vor Berechnung der Segmentierung und der Unsicherheit die entsprechende inverse Transformation angewandt werden.

6.3.3 Segmentierung von PEDs mit Bayesian DCU-Net

Zur Segmentierung von PEDs in den Home-OCT-Bilddaten wird das in Kapitel 4.2.3.1 vorgestellte DCU-Net genutzt. Für das Bayesian DCU-Net (BDCU-Net) wird hinter jedem DC-Block eine Dropout-Schicht integriert. Durch Dropout wird beim Training eine bessere Generalisierung gewährleistet, während gleichzeitig die Posterior-Verteilung der Netzgewichte modelliert wird. Anschließend können durch den TTD oder die TTA MC-Stichproben generiert werden. Hierbei erhält das BDCU-Net N mal das gleiche Bild (TTD) oder N verschiedene Augmentierungen dieses Bildes (TTA) als Eingabe, sodass bei jedem Durchlauf eine unterschiedliche Ausgabe erzeugt wird. Um die finale Segmentierung zu erhalten, werden zunächst die Softmax-Wahrscheinlichkeiten P_i via MC-Integration zusammengeführt. Die Übertragung in binäre Klassenlabels ($K = 2$) erfolgt dann schließlich über den $\arg \max$ -Operator. Die Variation der einzelnen Ausgaben spiegelt die Unsicherheit des Modells wider. Die Quantifikation der Unsicherheiten wird nachfolgend in Kapitel 6.3.4 beschrieben. Anstatt nur gelabelte B-Scans als Netzwerkinput zu nutzen, werden ebenso benachbarte, nicht gelabelte B-Scans be-

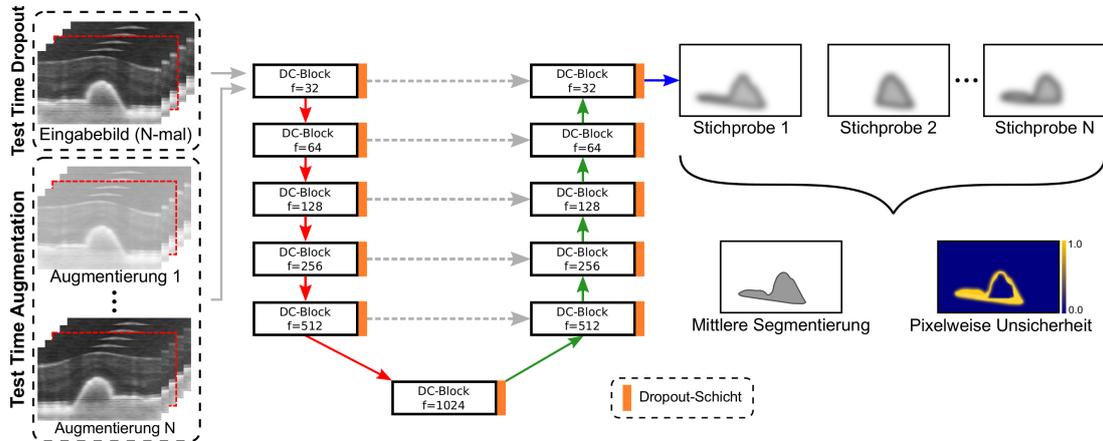


Abb. 6.3: Schematische Darstellung der vorgeschlagenen Methodik. Um MC-Stichproben zu erzeugen, wird zum einen das TTD-Verfahren genutzt. Hierbei werden Dropout-Schichten hinter jeden DC-Block des DCU-Nets integriert und während der Testzeit aktiviert. Zum anderen wird die TTA-Methode genutzt, bei der mehrere verschieden augmentierte Versionen desselben Bildes die MC-Stichproben erzeugen.

rücksichtigt. Jeweils zwei B-Scans vor und nach einem gelabelten B-Scan werden zu einem 2D-Multikanalbild zusammengesetzt, das als Pseudo-3D-Eingabe dient und dem Segmentierungsnetzwerk weitere Nachbarschaftsinformation bereitstellt. Eine schematische Übersicht des BDCU-Nets ist in Abbildung 6.3 dargestellt.

6.3.4 Quantifizierung und Bewertung von Unsicherheiten

Zur Quantifizierung der Segmentierungsunsicherheiten werden im Folgenden verschiedene Metriken vorgestellt. Wie bereits in Kapitel 6.3.2 erwähnt, eignen sich bspw. Standardabweichung, Varianz oder Entropie zur Schätzung der Diversität der Segmentierungsstichproben. In dieser Arbeit wird die Entropie zur pixelweisen Unsicherheitsquantifizierung verwendet. Sie wird definiert durch die mittlere Entropie der Softmax-Wahrscheinlichkeiten P_i aller N MC-Stichproben über alle Klassen K :

$$U = \sum_{l=0}^{K-1} -\frac{1}{N} \sum_{i=1}^N P_i \cdot \log(P_i). \quad (6.3)$$

Für nachfolgende Untersuchungen wird die Entropie auf den Bereich $[0,1]$ normiert.

Die berechnete Entropiekarte sollte geringe Werte an Stellen korrekter Vorhersagen des Segmentierungsmodells und hohe Werte im entgegengesetzten Fall aufweisen. Zur Qualitätsbewertung von Unsicherheiten schlagen Mehta u. a. [2020] hierzu eine Evaluationsmethode vor, die während der BraTS2019-Challenge⁶ zum Einsatz kam. Dabei

⁶Multimodal Brain Tumor Segmentation Challenge 2019 (BraTS2019)

<https://www.med.upenn.edu/cbica/brats-2019/>

werden unsichere Bildpixel anhand von vordefinierten Unsicherheitsschwellenwerten T gefiltert und für nachfolgende Auswertungen nicht weiter berücksichtigt. Beispielsweise werden bei einem Schwellenwert von $T = 0.75$ alle Bildpixel mit einer Unsicherheit von ≥ 0.75 als unsicher markiert und die verbleibenden Pixel zur Auswertung verwendet. Auf diese Weise werden diejenigen Segmentierungsmodelle belohnt, die ihre falschen Klassifikationen (falsch-positiv oder -negativ, *engl.*: false positive/negative, kurz: FP/FN) als unsicher und ihre richtigen Klassifikationen (richtig-positiv oder -negativ, *engl.*: true positive/negative, kurz: TP/TN) als sicher einstufen. Mit einer steigenden Anzahl von gefilterten unsicheren Bildpunkten sollten Metriken wie der DSC somit steigen. Um auch untersuchen zu können, wie viele richtig klassifizierte Bildpunkte gefiltert werden, wird ihr gefilterter Anteil ebenfalls berechnet. Der Anteil gefilterter TPs (FTPs) bei verschiedenen Schwellenwerten T ist gegeben durch $\text{FTP} = (\text{TP}_{1.00} - \text{TP}_T) / \text{TP}_{1.00}$. Die Berechnung für den gefilterten Anteil von TNs (FTNs) erfolgt hierzu analog. Anhand eines Gesamt-Scores U_{score} können verschiedene Unsicherheitsschätzungen miteinander verglichen werden. Dieser vereint die Flächen unter der Kurve (*engl.*: areas under the curves, kurz: AUCs) der drei Metriken DSC vs. T (AUC_{DSC}), FTP vs. T (AUC_{FTP}) und FTN vs. T (AUC_{FTN}) für verschiedene T-Werte. Der Gesamt-Score wird wie folgt berechnet:

$$U_{\text{score}} = \frac{\text{AUC}_{\text{DSC}} + (1 - \text{AUC}_{\text{FTP}}) + (1 - \text{AUC}_{\text{FTN}})}{3}. \quad (6.4)$$

In Anlehnung an [Roy u. a., 2019] werden weitere Metriken für die strukturbezogene Unsicherheitsanalyse verwendet, mit denen sich gezielt Unsicherheiten einzelner Klassenlabel l untersuchen lassen:

Mittlere Strukturunsicherheit Mittlere pixelweise Unsicherheit über diejenigen Pixel, die als Klasse l gelabelt worden sind:

$$\bar{U}_l = E \left[\{U(\mathbf{x})\}_{\mathbf{x} \in \{S=l\}} \right]. \quad (6.5)$$

Paarweiser DSC Mittlerer DSC über alle möglichen Paarungen der MC-Stichproben:

$$\text{pwDSC}_l = E \left[\{\text{DSC}((S_i = l), (S_j = l))\}_{i < j} \right]. \quad (6.6)$$

IoU aller Stichproben Intersection over Union (IoU) aller MC-Stichproben:

$$\text{IoU}_l = \frac{|(S_1 = l) \cap (S_2 = l) \cap \dots \cap (S_N = l)|}{|(S_1 = l) \cup (S_2 = l) \cup \dots \cup (S_N = l)|}. \quad (6.7)$$

Variationskoeffizient Volumenvariation aller MC-Stichproben. Über den Mittelwert μ_l und die Standardabweichung σ_l aller Volumina lässt sich der Variationskoeffizient für die Klasse l wie folgt berechnen:

$$\text{CV}_l = \frac{\sigma_l}{\mu_l}. \quad (6.8)$$

6.4 Experimente und Ergebnisse

Im Folgenden werden mehrere Experimente vorgestellt, deren Evaluation Aufschluss über die Güte und Qualität der Segmentierungen sowie deren Unsicherheitsschätzungen der TTD- und TTA-Methode geben soll. Die Auswertungen basieren auf einer vierfachen Kreuzvalidierung. Pro Durchlauf werden die OCT-Bilddaten der 68 Patientenaugen entsprechend in Trainings- und Testdatensatz (51-17) aufgeteilt. Die OCT-Bilddaten werden durch eine Vorverarbeitung entsprechend angepasst. Alle OCT-Scans werden hierzu durch einen adaptiven Histogrammausgleich [Yadav u. a., 2014] sowie einer Gauß-Glättung der lateralen Dimensionen (Fenstergröße=21, $\sigma=10$) vorverarbeitet, um die Bildqualität hinsichtlich des SNR zu verbessern. Des Weiteren werden die OCT-Scans in axialer Richtung auf 192 Pixel zugeschnitten und die erste laterale Dimension um das Vierfache reduziert, um sie auf das FCN-Training vorzubereiten. Wie bereits in Kapitel 5 angedeutet, ist die Segmentierung krankheitsspezifischer Flüssigkeitsbiomarker der Retina äußerst anspruchsvoll. Insbesondere ihr geringes Vorkommen, ihre pathologische Inhomogenität sowie die geringe Anzahl verfügbarer Trainingsdaten stellen eine große Herausforderung dar. Für die Lösung dieser Probleme werden mehrere Schritte unternommen, um ein robustes Training auf den Home-OCT-Bilddaten zu gewährleisten. Da PEDs nicht in jedem der selektierten B-Scans vorkommen (59%), wird bei der Aufteilung der Bilddaten in Trainings-, Validierungs- und Testdatensatz darauf geachtet, dass jeweils dieselbe Verteilung von Klassen existiert. Zusätzlich wird während des Trainings durch eine entsprechende Gewichtung dafür gesorgt, dass B-Scans, die PEDs enthalten, im Verhältnis von 3:1 häufiger gezogen werden. Auch der GDL wirkt als Fehlerfunktion mit einer entsprechenden Gewichtung dem Klassenungleichgewicht entgegen. Für die Experimente wird eine empirisch ermittelte Dropout-Rate von 0.3 festgelegt. Während der Testphase werden durch den TTD und die TTA 20 MC-Stichproben erzeugt (siehe Kapitel 6.4.1), anhand derer die mittlere Segmentierung und pixelweise Entropie ermittelt werden. Für die TTA kommt dieselbe Augmentierung zum Einsatz, die auch beim Training verwendet wird. Das Trainingsschema einschließlich der Augmentierung stimmt ansonsten mit dem in Kapitel 4.2.3.2 überein. Die einzelnen Experimente sowie deren Ergebnisse sind in den nachfolgenden Unterkapiteln weiter beschrieben.

6.4.1 Anzahl der MC-Stichproben

Zunächst wird die Anzahl der erforderlichen MC-Stichproben für TTD und TTA untersucht. Wie in [Roy u. a., 2019] wird die Anzahl der MC-Stichproben nach zwei Kriterien gewählt. Die Genauigkeit der mittleren MC-Segmentierung sollte mindestens auf dem Niveau des DCU-Nets liegen. Gleichzeitig muss die Entropieschätzung robust sein, sodass das Hinzufügen weiterer MC-Stichproben nicht zu großen Ver-

Tabelle 6.1: Mittlere DSC-Werte ohne und mit TTD und TTA bei unterschiedlicher Anzahl von MC-Stichproben.

Stichprobenanzahl	DSC
Kein(e) TTD/TTA	0.600
1	0.594
5	0.609
10	0.613
15	0.614
20	0.614
25	0.613
30	0.613

Tabelle 6.2: Mittlere absolute Abweichung der pixelweisen Entropien U_i und U_j , berechnet mit jeweils i und j MC-Stichproben.

$i \rightarrow j$	$E [U_i - U_j] \cdot 10^{-3}$
1 \rightarrow 5	4.759
5 \rightarrow 10	1.749
10 \rightarrow 15	1.017
15 \rightarrow 20	0.719
20 \rightarrow 25	0.557
25 \rightarrow 30	0.455

änderungen führt. Hierzu ist die Kreuzvalidierung mit verschiedenen Anzahlen $N = \{1, 5, 10, 15, 20, 25, 30\}$ von MC-Stichproben für TTD und TTA durchgeführt worden.

Zunächst wird die Auswirkung der MC-Stichprobenanzahl auf die Segmentierungsgüte untersucht. Die gemittelten DSC-Werte ohne und mit TTD oder TTA sind in Tabelle 6.1 aufgelistet. Generell ist der Unterschied zwischen allen DSC-Werten mit einer Standardabweichung von 0.007 sehr gering. Ab einer Anzahl von fünf MC-Stichproben wird mit TTD und TTA ein leicht höherer DSC-Wert erreicht als ohne. Ab zehn MC-Stichproben bleibt der DSC-Wert nahezu konstant. Die höchsten DSC-Werte werden für eine Anzahl von 15 und 20 MC-Stichproben erzielt.

Um zu ermitteln, wie viele MC-Stichproben eine verlässliche Unsicherheitsschätzung gewährleisten, wird der Unterschied zwischen den Entropien U_i und U_j untersucht, welche jeweils aus i und j MC-Stichproben berechnet werden. Als Maß dient hierzu die mittlere absolute Differenz $E [|U_i - U_j|]$. Die Ergebnisse dieser Auswertung sind in Tabelle 6.2 dargestellt. Mit steigender Anzahl verwendeter MC-Stichproben nehmen die Unterschiede in den Übergängen stetig ab. Die geringste mittlere absolute Differenz wird beim Übergang von 25 zu 30 MC-Stichproben ($0.455 \cdot 10^{-3}$) erreicht. Da eine höhere Anzahl von MC-Stichproben mit einem ebenfalls höheren Rechenaufwand einhergeht, wird im Hinblick auf die Ergebnisse aus Tabelle 6.1 eine MC-Stichprobenanzahl von 20 für die nachfolgenden Untersuchungen festgelegt.

6.4.2 Qualitative Auswertung

Im Folgenden werden die Ergebnisse der Kreuzvalidierung qualitativ untersucht. Hierzu sind in Abbildung 6.4 zwei Segmentierungsbeispiele mit entsprechender Grundwahrheit

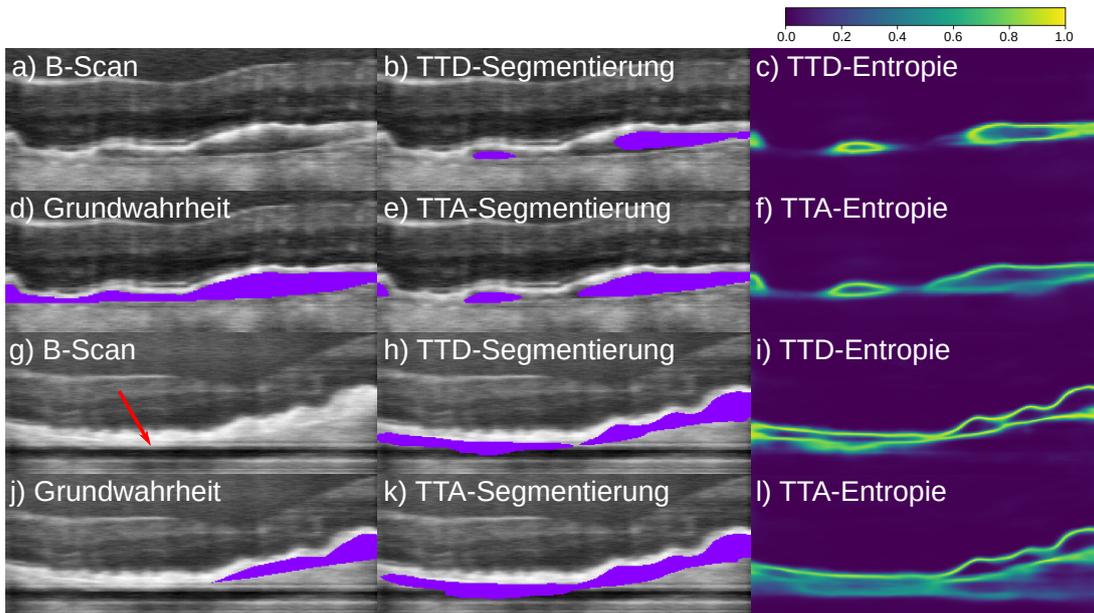


Abb. 6.4: Beispielhafte Darstellung von zwei Segmentierungen und den jeweiligen epistemischen (TTD) und aleatorischen (TTA) Unsicherheiten. Das zweite Beispiel weist ein stärkeres Bewegungsartefakt auf, das mit einem roten Pfeil markiert ist.

dargestellt. Die geschätzten epistemischen (TTD) und aleatorischen (TTA) Unsicherheiten werden in Form von pixelweisen Entropien abgebildet.

Insgesamt fällt auf, dass mittelgroße PEDs mit einer höheren Übereinstimmung zur Grundwahrheit segmentiert werden. Dagegen werden kleinere PED-Bereiche eher ungenau oder nicht segmentiert. Dies ist bei drusenoidem und fibrovaskulärem PED der Fall, das im Gegensatz zu serösem PED hell im OCT-Scan abgebildet wird (siehe Abb. 6.4, Unterbilder a) und g)), da es sich durch die fehlende dunkle Abgrenzung kaum vom RPE differenzieren lässt. Selbst für den medizinischen Experten stellen die unterschiedlichen PED-Varianten, besonders in kleineren Ausprägungen, eine Herausforderung dar, was sich mitunter in nicht konsistenten manuellen Annotationen widerspiegelt. Die Unterschiede zwischen den TTD- und TTA-Segmentierungen sind eher gering. Bezogen auf das erste Beispiel aus Abbildung 6.4 haben die TTA-Segmentierungen eine etwas höhere Übereinstimmung mit der Grundwahrheit als die TTD-Segmentierungen.

Verglichen zu den Segmentierungen lassen sich deutlichere Unterschiede zwischen den epistemischen (TTD) und aleatorischen (TTA) Unsicherheitsschätzungen erkennen. Wie erwartet zeichnen sich die höchsten Entropiewerte in Kantenregionen ab. Dabei sind die Entropiewerte der epistemischen Unsicherheiten deutlich höher und fallen rascher mit steigender Distanz ab. Dagegen sind die Entropiewerte der aleatorischen Unsicherheiten geringer und sind in Kantenregionen nicht gleichmäßig „hell“. In schwer differenzierbaren Regionen, wie kleinen PEDs oder diffusem RPE, sind die

Entropiewerte aleatorischer Unsicherheiten höher als in klar abgrenzbaren Bereichen z. B. beim Wechsel von serösem PED zum Choroid (siehe Abb. 6.4, Unterbild f)). Außer Kantenregionen zeigen die aleatorischen Unsicherheiten auch in anderen herausfordernden Bereichen höhere Entropiewerte. Dies lässt sich z. B. gut am zweiten qualitativen Beispiel darstellen (siehe Abb. 6.4, Unterbilder g) bis k)). Ein starkes Bewegungsartefakt (siehe Abb. 6.4, roter Pfeil in Unterbild g)) nahe der RPE-Schicht sorgt für höhere Entropiewerte sowohl in der epistemischen als auch in der aleatorischen Unsicherheit. Durch seine dunkle Erscheinung wird das Bewegungsartefakt fälschlicherweise in Teilbereichen als PED segmentiert. Anders als die epistemische hat die aleatorische Unsicherheitsschätzung in diesem Bereich deutlich höhere Werte.

6.4.3 Filterung von Unsicherheiten

Als Nächstes wird die Qualität der geschätzten Unsicherheiten mithilfe der von Mehta u. a. [2020] vorgeschlagenen Methodik (siehe Kapitel 6.3.4) untersucht. Abbildung 6.5 zeigt hierzu zwei qualitative Beispiele. Zur Darstellung sind die pixelweisen Unsicherheiten anhand vier unterschiedlicher Schwellenwerte $T = \{1.00, 0.75, 0.50, 0.25\}$ gefiltert worden. Ohne Filterung ($T = 1.00$) erscheinen die PED-Segmentierungsbeispiele des TTD und der TTA relativ ähnlich hinsichtlich ihres Anteils an falsch-positiven (FP) und falsch-negativen (FN) Pixeln. Mit abnehmendem Schwellenwert T ist eine deutliche Zunahme gefilterter und damit als unsicher markierter Pixel, die sich zunächst in Kantenregionen ausbreiten, zu beobachten. Ab einem Schwellenwert von $T = 0.50$ ist zu erkennen, dass die durch die TTA geschätzte aleatorische Unsicherheit vergleichsweise mehr FPs und FNs filtert als die epistemische. Für $T = 0.25$ verschwinden nahezu alle FPs und FNs durch die Filterung in den TTA-Segmentierungen. Allerdings steigt ab diesem Schwellenwert der Anteil gefilterter TPs deutlich an und ist im Gegensatz zur TTD-Segmentierung etwas höher.

Die Beobachtungen aus Abbildung 6.5 bestätigen sich in der quantitativen Auswertung, welche in Tabelle 6.3 für dieselben Schwellenwerte durchgeführt worden ist. Außerdem wird in Abbildung 6.6 die Veränderung von DSC, FTP und FTN für eine kleinschrittige Filterung dargestellt, wobei Schwellenwerte aus dem Bereich $[0.01, 1.00]$ mit einer Schrittweite von 0.01 gewählt worden sind. Ebenso sind die AUC-Werte (siehe Gleichung 6.4) für jede Auswertung angegeben. Ohne Filterung erzielt die TTA-Methode einen signifikant ($p=3.934 \cdot 10^{-10}$) höheren DSC-Wert von 0.621. Mit steigender Anzahl gefilterter Bildpixel steigt der DSC sowohl für den TTD auch für die TTA, was bedeutet, dass vermehrt falsche Klassifikationen beider Verfahren nicht mehr für die Auswertung betrachtet werden. Dabei nähern sich die DSC-Werte kurzzeitig an ($T = 0.50$), scheren danach aber wieder auseinander. Für $T > 0.5$ werden mehr TPs der TTD-Segmentierungen gefiltert als bei denen der TTA, wobei ein Wechsel bei $T = 0.50$ stattfindet. Ein ähnliches Verhalten lässt sich ebenfalls für die Änderung der FTNs beobachten, nur dass sich die Verläufe weniger stark unterscheiden. Die einzelnen

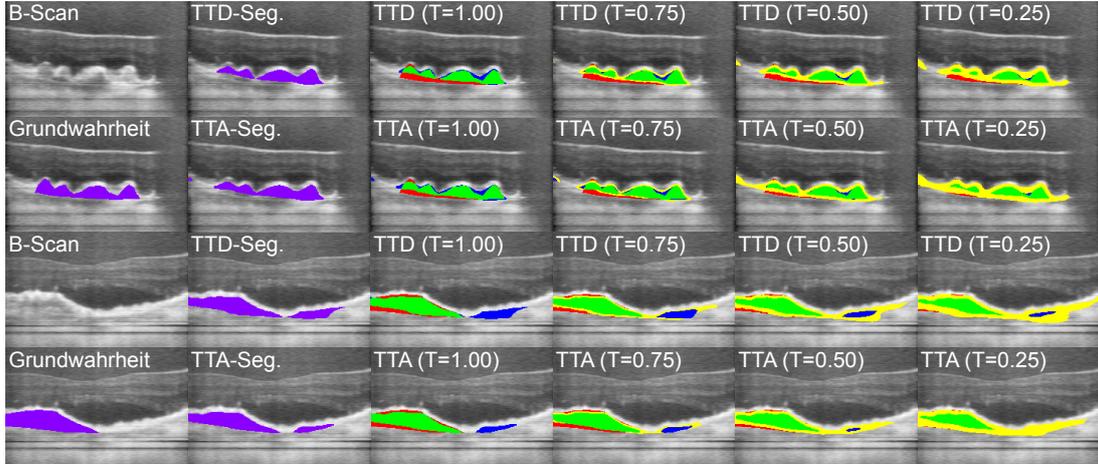


Abb. 6.5: Auswirkung unterschiedlicher Unsicherheitsschwellenwerte auf zwei exemplarische PED-Segmentierungen (jeweils die beiden ersten und letzten Zeilen). Die Farbkodierung ist wie folgt: Segmentierung (■), TP (■), FP (■), FN (■), unsichere Bildpunkte (■).

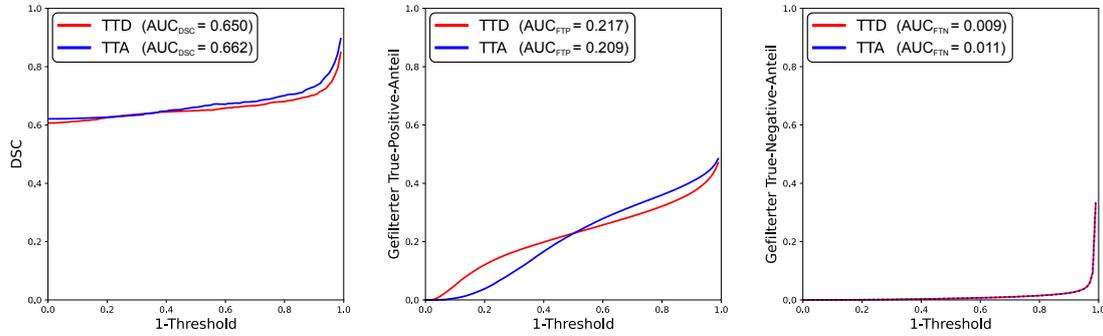
AUCs beider Verfahren werden mithilfe von Gleichung 6.4 im finalen U_{score} vereint. Dabei werden die aleatorischen Unsicherheiten der TTA mit einem mittleren U_{score} von 0.814 etwas höher bewertet als die epistemischen des TTD ($U_{\text{score}} = 0.808$).

6.4.4 Analyse strukturbezogener Unsicherheiten

In einem abschließenden Experiment wird untersucht, inwiefern die von Roy u. a. [2019] vorgeschlagenen strukturbezogenen Unsicherheitsmetriken genutzt werden können, um die Qualität der PED-Segmentierungen zu beurteilen. Zu diesem Zweck zeigt Abbildung 6.7 Streudiagramme der vier Unsicherheitsmetriken in Bezug auf den DSC. Jeder Datenpunkt steht für einen B-Scan, der PED enthält. Die PED-Größe wird in den Streudiagrammen anhand der Kantenlänge farblich kodiert. Um die Zuverlässigkeit der Qualitätskontrolle quantifizieren zu können, wird der Korrelationskoeffizient nach Pearson zwischen den Unsicherheitsmetriken und dem DSC berechnet. Die Korrelationskoeffizienten r sowie die zugehörigen p-Werte sind in den Streudiagrammen dargestellt. Generell werden die höchsten Korrelationskoeffizienten für die aleatorischen Unsicherheitsmetriken erzielt. Im Gegensatz zu den Streudiagrammen des TTD (epistemische Unsicherheit) zeigen die Punktwolken der TTA (aleatorische Unsicherheit) einen deutlich kompakteren Zusammenhang mit einem klar erkennbaren linearen Verlauf. Obwohl sich für die Streudiagramme der mittleren Strukturunsicherheit \bar{U}_{PED} eine klare Heteroskedastizität beobachten lässt, so ist der r -Wert sowohl für das TTD- als auch das TTA-Verfahren nahe null. Die Ursache hierfür liegt in der großen Anzahl sehr kleiner PED-Segmentierungen, deren DSC-Wert zur Grundwahrheit annähernd

Tabelle 6.3: Einfluss ausgewählter Unsicherheitsschwellenwerte auf die Metriken DSC, FTP und FTN.

Metriken ↓	Schwellenwerte → Unsicherheiten ↓	1.00	0.75	0.50	0.25
DSC	epistemisch	0.607	0.632	0.649	0.676
	aleatorisch	0.621	0.630	0.661	0.689
FTP	epistemisch	0.00	0.145	0.228	0.303
	aleatorisch	0.00	0.067	0.228	0.341
FTN	epistemisch	0.00	0.002	0.005	0.009
	aleatorisch	0.00	0.001	0.005	0.012

**Abb. 6.6:** Auswirkung der Änderung von Unsicherheitsschwellenwerten auf DSC, FTP und FTN. Die AUCs der Kurvenverläufe sind der jeweiligen Legende angegeben.

Null ist bei gleichzeitig geringer Unsicherheit. Dieselbe Beobachtung kann für IoUP_{PED} gemacht werden. Demgegenüber weisen die Metriken $\text{pwDSC}_{\text{PED}}$ und CV_{PED} für solche Fälle eher höhere Unsicherheitswerte zu, wodurch sich die Datenpunkte entlang der y-Achse auffächern, was auch zu höheren r-Werten führt. Die höchsten r-Werte werden für $\text{pwDSC}_{\text{PED}}$ erreicht. Die Farbkodierung der Datenpunkte kann verwendet werden, um den Einfluss der PED-Größe im B-Scan auf den DSC und die strukturbezogenen Unsicherheitsmetriken zu untersuchen. Dementsprechend werden mittelgroße PEDs (hellblau bis orange) am genauesten und zuverlässigsten segmentiert. Die Datenpunkte kleinerer PEDs (dunkelblau bis hellblau) befinden sich dagegen überwiegend in den oberen Regionen der einzelnen Punktwolken, was auf eine unzuverlässigere Segmentierungsschätzung hindeutet. Bei der Segmentierung von (sehr) großen PEDs (orange bis dunkelrot) besteht nur eine geringe Übereinstimmung mit der Grundwahrheit. Die Vorhersagen werden mit einem mittleren bis geringen Grad an Sicherheit getroffen.

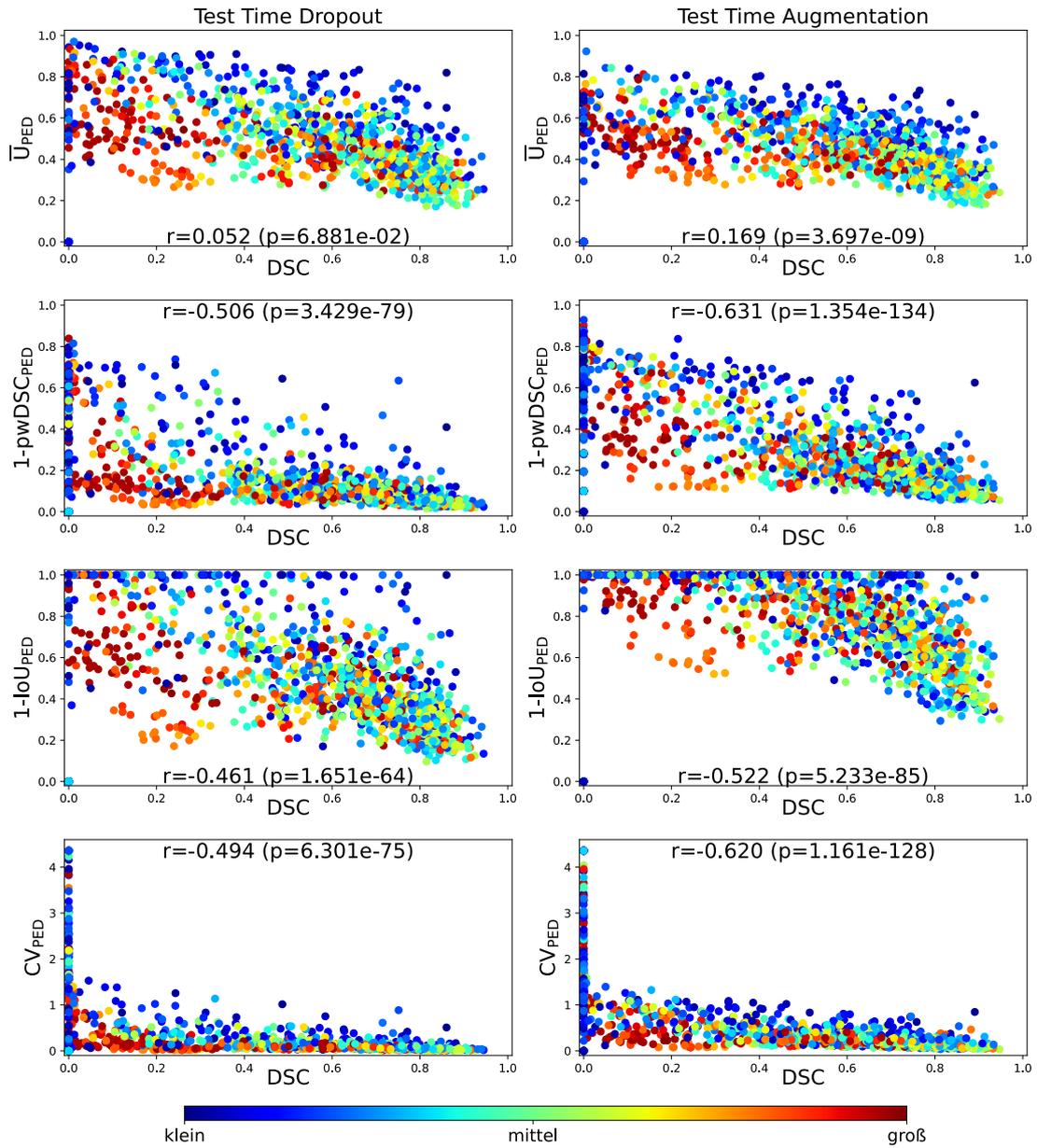


Abb. 6.7: Streudiagramme der vier strukturellen Unsicherheitsmetriken und dem DSC, die anhand von B-Scans mit PED für den TTD (links) und die TTA (rechts) bestimmt werden. Die Datenpunkte sind anhand der jeweiligen PED-Größe farblich kodiert.

6.5 Zusammenfassung und Diskussion

Die Behandlung der AMD stellt stets eine große Herausforderung dar. Durch die Kombination handlicher OCT-Systeme für den Heimgebrauch und KI-basierter CAD-Systeme können pathologische Veränderungen früher erkannt und Therapiemaßnahmen individuell auf den Patienten zugeschnitten werden. Um die Patientensicherheit zu erhöhen, ist die Integration von Unsicherheitsabschätzungen sinnvoll. Hierzu wurden in diesem Kapitel epistemische und aleatorische Unsicherheiten bei der Segmentierung von PEDs in Home-OCT-Bilddaten im Detail untersucht. Zur Schätzung der Unsicherheiten wurden zwei stichprobenbasierte Methoden vorgestellt, die auf dem Prinzip des TTD und der TTA basieren. Im Rahmen einer vierfachen Kreuzvalidierung wurden Güte und Qualität der Segmentierungen wie auch ihre Unsicherheiten von beiden Ansätzen ausführlich untersucht.

Die Evaluation der Experimente zeigte, dass die Segmentierungsgenauigkeit der TTA gegenüber der TTD-Variante höher ist (siehe Tab. 6.3) ($DSC = 0.6$). Außerdem wurde eine höhere Genauigkeit im Vergleich zu einer früheren Arbeit [Kepp u. a., 2020] erzielt, in der bereits PEDs als Biomarker in Home-OCT-Bilddaten untersucht worden ist. Allerdings sind die Ergebnisse nur bedingt mit anderen Stand-der-Technik-Methoden zu vergleichen, da sich die Home-OCT-Scans in ihrer Bildqualität signifikant von öffentlich verfügbaren Bilddatensätzen [Bogunović u. a., 2019] handelsüblicher OCT-Systeme unterscheiden. Ferner weist der Home-OCT-Bilddatensatz eine hohe Heterogenität AMD-relevanter Flüssigkeitsbiomarker wie PEDs auf. Zusätzlich zu ihrer unterschiedlichen Beschaffenheit (serös, drusenoid, fibrovaskulär) variieren die einzelnen PEDs mitunter erheblich in ihrer Größe, was das Lernen repräsentativer Merkmale erschwert. Die Heterogenität der Bilddaten führt darüber hinaus zu einer inkonsistenten Grundwahrheit (siehe Abb. 6.4), die sich ebenfalls negativ auf den Lernprozess auswirkte.

Neben der Segmentierungsgenauigkeit wurde in den Experimenten gezeigt, dass mithilfe der TTA vertrauenswürdiger Unsicherheiten geschätzt werden können, als mit dem TTD. Die deutlich höheren Entropiewerte in den pixelweisen epistemischen Unsicherheiten der TTD-Methode bewirken bereits frühzeitig ($T > 0.5$) eine verstärkte Filterung von TPs und FPs (siehe Abb. 6.6), was sich im Vergleich zu den aleatorischen Unsicherheiten der TTA in einem niedrigeren Gesamt-Score U_{score} widerspiegelt.

Da aleatorische Unsicherheiten inhärente Schwankungen der Bilddaten abbilden, werden durch sie auch schwierigere Bereiche wie bspw. Bewegungsartefakte berücksichtigt (siehe Abb. 6.4). Daher erweisen sich die aleatorischen Unsicherheiten bei der Aufnahme mit dem Home-OCT als sehr hilfreich und können einen Hinweis darauf liefern, ob eine Auswertung aufgrund von Bewegungsartefakten möglicherweise nicht zuverlässig ist. Im Unterschied dazu stellen epistemische Unsicherheiten fehlendes Wissen des trainierten Modells dar (siehe Kapitel 6.3.2). Sie können insbesondere in der Entwicklungsphase als Indikator verwendet werden, um zu beurteilen, ob die Generalisierung des Segmentierungsmodells zufriedenstellend ist. Anderenfalls könnte sie durch

eine Anpassung der Trainingsroutine oder durch Hinzufügen weiterer Trainingsbeispiele verringert werden, sofern dies erforderlich ist.

Die Ergebnisse der Untersuchungen anhand strukturbezogener Unsicherheitsmetriken zeigten, dass mittelgroße PEDs durchschnittlich sowohl am genauesten als auch am sichersten segmentiert werden. Dagegen sind die geschätzten Segmentierungen von kleinen PEDs in der Regel unsicherer, was in einigen Fällen darauf zurückzuführen ist, dass der Anteil schwer differenzierbarer Kantenregionen im Verhältnis zur Fläche groß ist. Ähnliches konnte jedoch auch für den Variationskoeffizienten CV_{PED} beobachtet werden, der unabhängig von der Strukturgröße ist. Die geringen Übereinstimmungen mit der Grundwahrheit bei großen PEDs können auf eine starke Unterrepräsentation im Bilddatensatz zurückgeführt werden. So werden nur einzelne Randsegmente des RPE durch das Segmentierungsmodell erfasst, auf die auch die pixelweisen Unsicherheits-schätzungen beschränkt sind. Hierdurch weisen (sehr) große PED-Segmentierungen mit geringem DSC-Wert eine verhältnismäßig geringe Unsicherheit auf. Ähnliches gilt auch für zahlreiche (sehr) kleine PED-Segmentierungen, deren DSC-Wert nahe null liegt, die jedoch als sicher eingestuft werden. Weitere qualitative Beispiele sind hierzu in Anhang A gegeben.

Insgesamt haben die Ergebnisse der umfassenden Untersuchungen gezeigt, dass durch die Bestimmung von Unsicherheiten ein verbessertes Feedback über die Zuverlässigkeit einer Segmentierungsschätzung möglich ist.

Kapitel 7

Zusammenfassung und Ausblick

Die in der vorliegenden Arbeit behandelten Anwendungsbereiche der OCT demonstrieren die vielseitigen Einsatzmöglichkeiten dieser Technologie und zugleich die damit verbundenen Herausforderungen bei der automatisierten Auswertung und Beurteilung ihrer Bilddaten. Von der präklinischen Forschung am Tiermodell (Kapitel 4) über die Diagnostik in der ophthalmologischen Klinik (Kapitel 5) bis hin zur Überwachung der Krankheitsaktivität von Augenerkrankungen beim Patienten zu Hause (Kapitel 6) nehmen die Anforderungen an die Auswertung von OCT-Bilddaten stetig zu. Zur Unterstützung der Analyse von OCT-Bilddaten aus diesen Anwendungsbereichen wurden in dieser Arbeit verschiedene Segmentierungsalgorithmen unter Einsatz von maschinellem Lernen, insbesondere Deep Learning, vorgeschlagen und untersucht. Dabei stand die notwendige Robustheit und Verlässlichkeit des jeweiligen Anwendungsbereiches im Vordergrund, was einerseits die Beurteilung sowohl anatomischer als auch pathologischer Strukturen sowie andererseits die unterschiedlichen OCT-Aufnahmebedingungen und -techniken einschließt. Die Hauptbeiträge der Methodenkapitel 4 bis 6 werden nachfolgend unter Berücksichtigung der in dieser Arbeit formulierten Ziele zusammengefasst.

Verkürzung von Segmentierungspipelines und Vereinfachung ihrer Anwendbarkeit Zur automatisierten Analyse einzelner Hautschichten der Maus im Rahmen einer präklinischen Studie wurden in Kapitel 4 zwei Segmentierungsverfahren entwickelt und systematisch miteinander verglichen. Die erste Methode umfasst eine RF-Klassifikation in Verbindung mit einer graphenbasierten Optimierung, wohingegen die zweite Methode auf einem FCN, dem DCU-Net, beruht. Die Ergebnisse der Evaluation zeigten, dass der Einsatz des DCU-Nets zu einer höheren Segmentierungsgüte gegenüber dem RF-Klassifikator führt. Aufgrund seines selbstlernenden Charakters ist das DCU-Net in der Lage, komplexe Merkmale sowie deren Zusammenhänge bereits auf unverarbeiteten OCT-Bilddaten zu erlernen und anzuwenden, was generell die Anzahl von Prozessierungsschritten signifikant reduziert und somit die Anpassungsfähigkeit an andere Anwendungsbereiche erhöht. Um trotz der begrenzten Menge an Trainingsbeispielen eine hohe Generalisierbarkeit des DCU-Nets zu erreichen, wurden einerseits Augmentierungstechniken und andererseits eine gewichtete Fehlerfunk-

tion zum Ausgleich des Klassenungleichgewichts realisiert. Weiterhin konnte durch die graphenbasierte Optimierung die Anzahl der Ausreißer in den Segmentierungen des RF-Klassifikators sowie des DCU-Nets verringert und dadurch die Robustheit erhöht werden. Die Adaptierung des Graphenmodells an neue Anwendungsbereiche, z. B. der Segmentierung von Retinaschichten, ist jedoch aufgrund seiner komplexen Konstruktion sowie der umfangreichen Parametrisierung äußerst aufwendig. Gleiches gilt für die Gestaltung von Merkmalen des RF-Klassifikators. Ein zusätzlicher Vorteil des DCU-Nets ist seine schnelle Inferenz, die bei Verwendung einer GPU zu sehr kurzen Berechnungszeiten für eine Segmentierung führt. FCN-basierte Segmentierungsverfahren wie das DCU-Net eignen sich daher hervorragend für die Auswertung von großen OCT-Bilddatenmengen präklinischer Studien, wie z. B. der Erprobung eines neuen Tiermodells zur Evaluation des Kryolipolyseverfahrens. Hinsichtlich seiner Vorzüge wurde das DCU-Net als Grundlage für die nachfolgenden Methodenkapitel dieser Arbeit herangezogen.

Berücksichtigung der Topologie anatomischer Strukturen bei der Segmentierung mit FCNs Um die Robustheit des DCU-Nets ohne zusätzliche Nachbearbeitungsschritte zu erhöhen, wurde in Kapitel 5 eine Multi-Task-Learning-Strategie zur Segmentierung von Retinaschichten und Flüssigkeitseinlagerungen in OCT-Bilddaten von DMÖ-Patienten entwickelt. Zu diesem Zweck wurde das DCU-Net zu einer Multi-Head-FCN-Architektur modifiziert, bei der sich die Ausgabeschicht jeweils in einen separaten Klassifikations- und Regressionspfad aufteilt. Dadurch wurde eine formbasierte Regression von Retinaschichten als SSDMs neben der pixelweisen Klassifikation als Hilfsaufgabe formuliert, was zu einer verstärkten Berücksichtigung ihrer topologischen Beziehungen beiträgt. Zudem bewirkt die ausschließliche Klassifikation von Flüssigkeiten eine schärfere Abgrenzung im Vergleich zur Regression aller Klassen wie in [Kepp u. a., 2019b]. Die vorgeschlagene Multi-Head-FCN-Architektur lässt sich leicht erweitern, sodass die Segmentierung durch weitere Hilfsaufgaben unterstützt werden kann. Einen weiteren Vorteil bietet die Kopplung beider Aufgabenpfade durch die duple Aufgabenkonsistenz (DTC), die eine Nutzung nicht gelabelter Bilddaten zulässt und somit ein halb-überwachtes Training erlaubt. Die Ergebnisse der Evaluation zeigten, dass die vorgeschlagene Methode im Allgemeinen sowohl zu einer höheren Segmentierungsgenauigkeit der Retinaschichten als auch zu einer besseren Berücksichtigung ihrer Topologie führt. Trotz ihrer verbesserten Abgrenzung gestaltet sich die automatische Segmentierung krankheitsbedingter Flüssigkeiten wegen ihrer hohen Variabilität sowie ihres geringen Auftretens jedoch als schwierig. Selbst bei einer manuellen Annotation zwischen Experten führt dies zu erheblichen Abweichungen.

Bestimmung von Unsicherheiten zur Unterstützung der Bewertung von Segmentierungen Mit Ausnahme des DMÖ sind krankheitsbedingte Flüssigkeiten

wichtige Biomarker für die Diagnose der AMD, deren Veränderungen bei der feuchten Verlaufsform unbehandelt in kürzester Zeit zur Erblindung des Patienten führen können. Neuartige OCT-Systeme für den Heimgebrauch ermöglichen es dem Patienten, seinen Krankheitszustand zu Hause selbstständig zu erfassen. Die Ergänzung solcher OCT-Systeme durch KI-basierte CAD-Systeme könnte die Diagnose unterstützen und die rechtzeitige Einleitung von therapeutischen Maßnahmen sicherstellen. In diesem Zusammenhang wurden in Kapitel 6 Methoden zur Analyse von PEDs als AMD-bedingte Flüssigkeitsbiomarker in Bilddaten eines Home-OCT-Prototyps untersucht. Abgesehen von der in Kapitel 5 beschriebenen Problematik bei der Erkennung krankheitsbedingter Flüssigkeiten, führen auch die geringere Bildqualität der Home-OCT-Scans sowie das Auftreten von Bewegungsartefakten durch die Selbstmessung des Patienten zu Unsicherheiten in der Auswertung. Damit diese verbleibende Restunsicherheit als Feedback zur Verfügung gestellt werden kann, wurden ergänzend zur PED-Segmentierung ihre pixelweisen Unsicherheiten ermittelt. Untersucht wurden sowohl epistemische als auch aleatorische Unsicherheiten, die durch einen Dropout (TTD) oder eine Augmentierung (TTA) zur Testzeit bestimmt werden. Diesbezüglich wurde das DCU-Net um Dropout-Schichten (BDCU-Net) erweitert.

In einer umfassenden Bewertung hat sich gezeigt, dass mittels TTA sowohl genauere Segmentierungen als auch zuverlässigere Unsicherheiten geschätzt werden können als mit dem TTD. Die Qualität der geschätzten Unsicherheiten wurde dabei ergänzend zu einer rein qualitativen Beurteilung auch quantitativ über ein Filterverfahren bewertet. Weiterhin wurde die Zuverlässigkeit der Segmentierungen in Bezug auf die PED-Größe anhand strukturbezogener Unsicherheitsmetriken untersucht. Es hat sich gezeigt, dass mittelgroße PEDs in Bilddaten des Home-OCT am genauesten und zuverlässigsten segmentiert werden. Im Gegensatz dazu ist die Erkennung von kleinen PEDs schwieriger und unsicherer. Außerdem weisen große PEDs trotz geringer Segmentierungsgenauigkeit eine ebenso geringe Unsicherheit auf, was wiederum die Zuverlässigkeit beeinträchtigen kann. Insgesamt zeigten die Untersuchungen, dass die Bestimmung epistemischer und aleatorischer Unsicherheiten einen wichtigen Aspekt bei der Entwicklung und Anwendung FCN-basierter CAD-Systeme für handliche Home-OCT-Geräte darstellt. Aufgrund der Tatsache, dass aleatorische Unsicherheiten die inhärente Zufälligkeit von Daten beschreiben, können sie verwendet werden, um die Auswertbarkeit einer Messung mit dem Home-OCT in Bezug auf die Bildqualität zu beurteilen. Hingegen können epistemische Unsicherheiten zur Untersuchung der Generalisierbarkeit von FCN herangezogen werden, da sie Hinweise auf fehlendes Wissen aufzeigen.

Ungeachtet ihrer aufgeführten Vorteile haben sich auch Einschränkungen der in dieser Arbeit entwickelten Methoden gezeigt, die zu weiteren Forschungsfragen führen.

- Zur Gewährleistung einer optimalen Generalisierung von FCN-basierten Segmentierungsverfahren ist eine große Menge gelabelter Bilddaten für ihr überwachtes

Training erforderlich. Im Unterschied zu anderen Bereichen der Computer Vision ist die Verfügbarkeit solcher Trainingsdatensätze in der medizinischen Bildverarbeitung jedoch sehr begrenzt, da für ihre Erstellung in der Regel ein Experte auf dem jeweiligen Gebiet herangezogen werden muss. Obwohl die Anzahl der Trainingsbeispiele durch Augmentierungstechniken künstlich erhöht werden kann, geht dies nicht mit einer Steigerung der anatomischen und pathologischen Variationen einher. Abgesehen von einer zeitaufwendigen manuellen Erstellung weiterer Trainingsbeispiele wäre es daher interessant, sich verstärkt mit Methoden zu beschäftigen, die ähnlich wie in Kapitel 5 die Verwendung von nicht oder nur schwach gelabelten Bilddaten erlauben. Insbesondere eröffnen generative Deep-Learning-Techniken eine interessante Möglichkeit, die Variationen der Trainingsbilddaten zu erhöhen. Hierzu wird in [Tajbakhsh u. a., 2020] ein Überblick verschiedener Methoden gegeben, die diese Problemstellung adressieren.

- Die Evaluationen der Kapitel 5 und 6 haben die Problematik der Segmentierung von krankheitsbedingten Flüssigkeitsbiomarkern des DMÖ und der AMD verdeutlicht. Im Gegensatz zu anatomischen Strukturen wie den Retinaschichten sind Pathologien viel seltener in OCT-Bilddatensätzen vertreten. Selbst mit anspruchsvollen Sampling-Strategien und der Verwendung gewichteter Fehlerfunktionen ist es nach wie vor schwierig, dem Klassenungleichgewicht entgegenzuwirken. Ebenso führt die zuvor geschilderte Vielfalt der Flüssigkeitsbiomarker zu ihrer geringeren Segmentierungsgenauigkeit durch die hier entwickelten Methoden. Um sicherzustellen, dass diese Vielfalt an Variationen von Flüssigkeitsbiomarkern durch eine ausreichend umfangreiche Menge von Trainingsbeispielen abgedeckt wird, wären Untersuchungen generativer Methoden sinnvoll, die wie in [Uzunova u. a., 2020] eine gezielte Synthese von Pathologien erlauben.
- Alle in dieser Arbeit entwickelten Versionen des DCU-Nets sind bisher nur für die Verarbeitung von 2D-OCT-Bilddaten geeignet. Dabei hat der Verfeinerungsschritt durch die graphenbasierte Optimierung gezeigt, dass die Erweiterung auf den dreidimensionalen Raum zu einer verbesserten Regularisierung führt. Zudem bieten OCT-Systeme wie das Telesto II SD-OCT-System oder das Home-OCT von Visotec eine dichte räumliche Abtastung, sodass die Erweiterung des DCU-Nets zur Verarbeitung von 3D-OCT-Bilddaten eine wertvolle Ergänzung zur Erfassung einer größeren räumlichen Umgebung des untersuchten Gewebes darstellt.
- Obwohl in Kapitel 5 gezeigt werden konnte, dass die Regression von SSDMs der Retinaschichten zu einer verbesserten Berücksichtigung ihrer Topologie bei der Segmentierung beiträgt, garantiert sie nicht die strikte Einhaltung der Schichtreihenfolge wie bei der graphenbasierten Optimierungsmethode. Mehrere Lösungsansätze sind für diese Problematik möglich. So könnten anstelle der gesamten Retinaschichten nur deren Übergänge mit den SSDMs regressiert werden. Dadurch ließe sich

aus den detektierten Kanten ein kontinuierliches Oberflächenmodell ähnlich dem Graphenmodell generieren. Zur Vermeidung von Überlappungen der einzelnen Flächen könnten unterstützende Randbedingungen wie in [He u. a., 2021] herangezogen werden.

- Die Bestimmung epistemischer und aleatorischer Segmentierungsunsicherheiten bietet ein hilfreiches Feedback für die Bewertung der Zuverlässigkeit von Modellvorhersagen. Die Auswertung der strukturellen Unsicherheitsmetriken hat demgegenüber gezeigt, dass die geschätzten Unsicherheiten nicht immer der eigentlichen Modellzuverlässigkeit entsprechen. Um die Unsicherheitsschätzung an das Konfidenzniveau des Segmentierungsmodells anzupassen, sind in der Literatur verschiedene Kalibrierungstechniken [Mehrtaš u. a., 2020] vorgeschlagen worden, die mögliche Anknüpfungspunkte für zukünftige Untersuchungen darstellen.
- Wegen des niedrigen SNR des Home-OCT ist die Erkennung anderer Flüssigkeitsbiomarker wie IRF und SRF neben PED nicht möglich. Abgesehen von der technischen Weiterentwicklung des Geräts stellt auch die KI-basierte Verbesserung der Bildqualität eine interessante Forschungsfrage dar. Da pro Aufnahmesequenz des Home-OCT mehrere Messungen durchgeführt werden, wäre es interessant zu untersuchen, inwiefern eine Auswahl der qualitativ besten Scans sinnvoll miteinander fusioniert werden kann, um den SNR zu verbessern und so die Erkennung von IRF und SRF zu ermöglichen.

Unabhängig von diesen noch offenen Forschungsfragen wurden im Rahmen dieser Arbeit bereits wichtige Erfolge bei der automatischen Analyse von anatomischen und pathologischen Strukturen in OCT-Bilddaten erreicht. Die erzielten Ergebnisse demonstrieren das gewinnbringende Potenzial FCN-basierter Segmentierungsmethoden, die sich aufgrund ihrer Flexibilität nicht allein auf die hier betrachteten Anwendungsbereiche der OCT beschränken. Ihre einfache Erweiterbarkeit ermöglicht es außerdem, ihre Robustheit weiter zu erhöhen, sodass anatomische Strukturzusammenhänge auch unter pathologischen Bedingungen stärker berücksichtigt werden können. Aufgrund ihrer Stärken stellen Deep-Learning-Methoden heute den aktuellen Stand der Technik dar, deren Vorhersagen mit denen von klinischen Experten vergleichbar sind [De Fauw u. a., 2018]. Um die klinische Akzeptanz von Deep-Learning-Methoden zu fördern, bedarf es außer der Aufschlüsselung ihres Black-Box-Charakters auch der Steigerung ihrer Zuverlässigkeit, deren Relevanz bei der Bewertung von Segmentierungsergebnissen verdeutlicht worden ist. Zusammengefasst stellen die in dieser Arbeit entwickelten Segmentierungsverfahren einen wichtigen Beitrag zur automatischen Analyse von OCT-Bilddaten dar, deren Einsatz die medizinische Versorgung in vielfältiger Weise unterstützen kann.

Anhang A

Ergänzende Ergebnisse zu Kapitel 6

A.1 Qualitative Ergebnisbeispiele von PED-Segmentierungen und ihrer Unsicherheitsschätzung

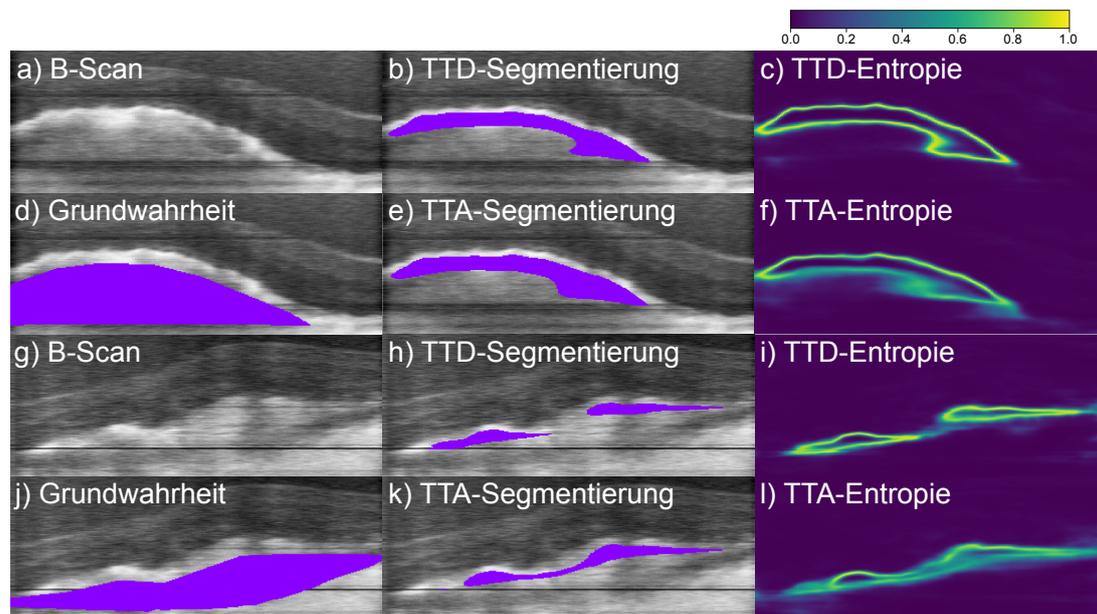


Abb. A.1: Zwei Segmentierungsbeispiele sehr großer PEDs mit ihren jeweiligen epistemischen (TTD) und aleatorischen (TTA) Unsicherheiten. Angesichts ihrer Unterrepräsentation im Bilddatensatz können sehr große PED-Strukturen weder ausreichend gut segmentiert noch ihre Unsicherheiten zuverlässig geschätzt werden.

Anhang B

Eigene Publikationen

- [Ehrhardt u. a., 2014] Ehrhardt, J., Kepp, T., Schmidt-Richberg, A. und Handels, H. „Joint multi-object registration and segmentation of left and right cardiac ventricles in 4D cine MRI“. In: *SPIE Medical Imaging 2014: Image Processing*. Bd. 9034. 2014, S. 90340M.
- [Kepp u. a., 2015] Kepp, T., Ehrhardt, J. und Handels, H. „Evaluation verschiedener Ansätze zur 4D-4D-Registrierung kardiologischer MR-Bilddaten“. In: *Bildverarbeitung für die Medizin 2015*. Springer Vieweg, 2015, S. 95–100.
- [Opfer u. a., 2016] Opfer, R., Suppa, P., Kepp, T., Spies, L., Schippling, S., Huppertz, H.-J., Initiative, A. D. N. u. a. „Atlas based brain volumetry: how to distinguish regional volume changes due to biological or physiological effects from inherent noise of the methodology“. *Magnetic Resonance Imaging* 34 (4), 2016, S. 455–461.
- [Suppa u. a., 2016] Suppa, P., Hampel, H., Kepp, T., Lange, C., Spies, L., Fiebach, J. B., Dubois, B., Buchert, R., Initiative, A. D. N. u. a. „Performance of hippocampus volumetry with FSL-FIRST for prediction of alzheimer’s disease dementia in at risk subjects with amnesic mild cognitive impairment“. *Journal of Alzheimer’s Disease* 51 (3), 2016, S. 867–873.
- [Egger u. a., 2017] Egger, C., Opfer, R., Wang, C., Kepp, T., Sormani, M. P., Spies, L., Barnett, M. und Schippling, S. „MRI FLAIR lesion segmentation in multiple sclerosis: does automated segmentation hold up with manual annotation?“ *NeuroImage: Clinical* 13, 2017, S. 264–270.
- [Kepp u. a., 2017] Kepp, T., Koinzer, S., Handels, H. und Brinkmann, R. „Registrierung von nicht sichtbaren Laserbehandlungsarealen der Retina in Live-Aufnahmen des Fundus“. In: *Bildverarbeitung für die Medizin 2017*. Springer Vieweg, 2017, S. 331–336.
- [Kepp u. a., 2018] Kepp, T., Droigk, C., Casper, M., Evers, M., Salma, N., Manstein, D. und Handels, H. „Segmentation of subcutaneous fat within mouse skin in 3D OCT image data using random forests“. In: *SPIE Medical Imaging 2018: Image Processing*. Bd. 10574. 2018, S. 1057426-1–1057426-8.

- [Salma u. a., 2018] Salma, N., Evers, M., Casper, M. J., Droigk, C., Kepp, T., Handels, H. und Manstein, D. „Mouse model of cold-induced localized fat loss (selective cryolipolysis)“. In: *Lasers in Surgery and Medicine*. Bd. 50. 2018, S19–S20.
- [Kepp u. a., 2019a] Kepp, T., Droigk, C., Casper, M., Evers, M., Hüttmann, G., Salma, N., Manstein, D., Heinrich, M. P. und Handels, H. „Segmentation of mouse skin layers in optical coherence tomography image data using deep convolutional neural networks“. *Biomedical Optics Express* 10 (7), 2019, S. 3484–3496.
- [Kepp u. a., 2019b] Kepp, T., Ehrhardt, J., Heinrich, M. P., Hüttmann, G. und Handels, H. „Topology-Preserving Shape-Based Regression Of Retinal Layers In OCT Image Data Using Convolutional Neural Networks“. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, 2019, S. 1437–1440.
- [Uzunova u. a., 2019] Uzunova, H., Ehrhardt, J., Kepp, T. und Handels, H. „Interpretable explanations of black box classifiers applied on medical images by meaningful perturbations using variational autoencoders“. In: *SPIE Medical Imaging 2019: Image Processing*. Bd. 10949. 2019, S. 264–271.
- [Kepp u. a., 2020] Kepp, T., Sudkamp, H., von der Burchard, C., Schenke, H., Koch, P., Hüttmann, G., Roider, J., Heinrich, M. P. und Handels, H. „Segmentation of retinal low-cost optical coherence tomography images using deep learning“. In: *SPIE Medical Imaging 2020: Computer-Aided Diagnosis*. Bd. 11314. 2020, 113141O.
- [Hüttmann u. a., 2021] Hüttmann, G., Wunderlich, M., Koch, P., Sudkamp, H., Theisen-Kunde, D., Kepp, T., Handels, H., von der Burchard, C. und Roider, J. „Improved SELFF-OCT using motion correction and artifact detection“. *Investigative Ophthalmology & Visual Science* 62 (8), 2021, S. 2527–2527.
- [Sudkamp u. a., 2021] Sudkamp, H. M., von der Burchard, C., Moltmann, M., Koch, P., Endt, M. vom, Müntz, M., Ehlken, C., Kepp, T., Theisen-Kunde, D., Birngruber, R. u. a. „Self-examination low-cost full-field OCT (SELFF-OCT) allows treatment decision for age related macular degeneration (AMD) and diabetic macular edema (DME) with high sensitivity and specificity: clinical pilot study“. In: *Ophthalmic Technologies XXXI*. Bd. 11623. 2021, 116230Z.
- [Andresen u. a., 2022a] Andresen, J., Kepp, T., Ehrhardt, J., von der Burchard, C., Roider, J. und Handels, H. „Deep learning-based simultaneous registration and unsupervised non-correspondence segmentation of medical images with pathologies“. *International Journal of Computer Assisted Radiology and Surgery*, 2022, S. 1–12.
- [Andresen u. a., 2022b] Andresen, J., Kepp, T., Wang-Evers, M., Ehrhardt, J., Manstein, D. und Handels, H. „Unsupervised Segmentation of Wounds in Optical Coherence Tomography Images Using Invariant Information Clustering“. In: *Bildverarbeitung für die Medizin 2022*. Springer Vieweg, 2022. Akzeptiert.

[Kepp u. a., 2022] Kepp, T., Andresen, J., Sudkamp, H., von der Burchard, C., Roider, J., Hüttmann, G., Ehrhardt, J. und Handels, H. „Epistemic and Aleatoric Uncertainty Estimation for PED Segmentation in Home OCT Images“. In: *Bildverarbeitung für die Medizin 2022*. Springer Vieweg, 2022. Akzeptiert.

Literaturverzeichnis

- [Adhi u. a., 2013] Adhi, M. und Duker, J. S. „Optical coherence tomography—current and future applications“. *Current Opinion in Ophthalmology* 24 (3), 2013, S. 213.
- [Altaf u. a., 2019] Altaf, F., Islam, S. M. S., Akhtar, N. und Janjua, N. K. „Going Deep in Medical Image Analysis: Concepts, Methods, Challenges, and Future Directions“. *IEEE Access* 7, 2019, S. 99540–99572.
- [Antony u. a., 2012] Antony, B. J., Abràmoff, M. D., Sonka, M., Kwon, Y. H. und Garvin, M. K. „Incorporation of texture-based features in optimal graph-theoretic approach with application to the 3D segmentation of intraretinal surfaces in SD-OCT volumes“. In: *SPIE Medical Imaging 2012: Image Processing*. Bd. 8314. 2012, 83141G.
- [Antony u. a., 2013] Antony, B. J., Abràmoff, M. D., Harper, M. M., Jeong, W., Sohn, E. H., Kwon, Y. H., Kardon, R. und Garvin, M. K. „A combined machine-learning and graph-based framework for the segmentation of retinal surfaces in SD-OCT volumes“. *Biomedical Optics Express* 4 (12), 2013, S. 2712–2728.
- [Apostolopoulos u. a., 2017] Apostolopoulos, S., De Zanet, S., Ciller, C., Wolf, S. und Sznitman, R. „Pathological OCT retinal layer segmentation using branch residual u-shape networks“. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2017, S. 294–301.
- [Asgari u. a., 2019] Asgari, R., Orlando, J. I., Waldstein, S., Schlanitz, F., Baratsits, M., Schmidt-Erfurth, U. und Bogunović, H. „Multiclass segmentation as multi-task learning for drusen segmentation in retinal optical coherence tomography“. In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. 2019, S. 192–200.
- [Banfield u. a., 2006] Banfield, R. E., Hall, L. O., Bowyer, K. W. und Kegelmeyer, W. P. „A comparison of decision tree ensemble creation techniques“. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (1), 2006, S. 173–180.
- [BenTaieb u. a., 2016] BenTaieb, A. und Hamarneh, G. „Topology aware fully convolutional networks for histology gland segmentation“. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2016, S. 460–468.

- [Bhagat u. a., 2009] Bhagat, N., Grigorian, R. A., Tutela, A. und Zarbin, M. A. „Diabetic macular edema: pathogenesis and treatment“. *Survey of Ophthalmology* 54 (1), 2009, S. 1–32.
- [Bille, 2019] Bille, J. F. „High Resolution Imaging in Microscopy and Ophthalmology: New Frontiers in Biomedical Optics“. *High Resolution Imaging in Microscopy and Ophthalmology*, 2019.
- [Bogunović u. a., 2019] Bogunović, H., Venhuizen, F., Klimscha, S., Apostolopoulos, S., Bab-Hadiashar, A., Bagci, U., Beg, M. F., Bekalo, L., Chen, Q., Ciller, C. u. a. „RETOUCH: The Retinal OCT Fluid Detection and Segmentation Benchmark and Challenge“. *IEEE Transactions on Medical Imaging* 38 (8), 2019, S. 1858–1874.
- [Bouma u. a., 2000] Bouma, B. E., Tearney, G. J., Compton, C. C. und Nishioka, N. S. „High-resolution imaging of the human esophagus and stomach in vivo using optical coherence tomography“. *Gastrointestinal Endoscopy* 51 (4), 2000, S. 467–474.
- [Boykov u. a., 2001] Boykov, Y. Y. und Jolly, M.-P. „Interactive graph cuts for optimal boundary & region segmentation of objects in ND images“. In: *Proceedings eighth IEEE International Conference on Computer Vision (ICCV)*. Bd. 1. 2001, S. 105–112.
- [Boykov u. a., 2004] Boykov, Y. und Kolmogorov, V. „An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision“. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (9), 2004, S. 1124–1137.
- [Boykov u. a., 2006a] Boykov, Y. und Funka-Lea, G. „Graph Cuts and Efficient N-D Image Segmentation“. *International Journal of Computer Vision* 70, 2006, S. 109–131.
- [Boykov u. a., 2006b] Boykov, Y. und Veksler, O. „Graph cuts in vision and graphics: Theories and applications“. In: *Handbook of Mathematical Models in Computer Vision*. Springer, 2006, S. 79–96.
- [Brandl u. a., 2016] Brandl, C., Stark, K., Wintergerst, M., Heinemann, M., Heid, I. und Finger, R. „Epidemiologie der altersbedingten Makuladegeneration“. *Der Ophthalmologe* 113 (9), 2016, S. 735–745.
- [Breiman, 2001] Breiman, L. „Random forests“. *Machine Learning* 45 (1), 2001, S. 5–32.
- [Breiman, 1996] Breiman, L. „Bagging predictors“. *Machine Learning* 24, 1996, S. 123–140.
- [Calonder u. a., 2012] Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C. und Fua, P. „BRIEF: Computing a local binary descriptor very fast“. *IEEE*

-
- Transactions on Pattern Analysis and Machine Intelligence* 34 (7), 2012, S. 1281–1298.
- [Calderon-Delgado u. a., 2018] Calderon-Delgado, M., Tjiu, J.-W., Lin, M.-Y. und Huang, S.-L. „High resolution human skin image segmentation by means of fully convolutional neural networks“. In: *2018 International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD)*. 2018, S. 31–32.
- [Cao u. a., 2020] Cao, J., Liu, X., Zhang, Y. und Wang, M. „A Multi-task Framework for Topology-guaranteed Retinal Layer Segmentation in OCT Images“. In: *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2020, S. 3142–3147.
- [Chen u. a., 2012] Chen, X., Niemeijer, M., Zhang, L., Lee, K., Abràmoff, M. D. und Sonka, M. „Three-dimensional segmentation of fluid-associated abnormalities in retinal OCT: probability constrained graph-search-graph-cut“. *IEEE Transactions on Medical Imaging* 31 (8), 2012, S. 1521–1531.
- [Chen u. a., 2013] Chen, Q., Leng, T., Zheng, L., Kutzscher, L., Ma, J., Sisternes, L. de und Rubin, D. L. „Automated drusen segmentation and quantification in SD-OCT images“. *Medical Image Analysis* 17 (8), 2013, S. 1058–1072.
- [Chiu u. a., 2010] Chiu, S. J., Li, X. T., Nicholas, P., Toth, C. A., Izatt, J. A. und Farsiu, S. „Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation“. *Optics Express* 18 (18), 2010, S. 19413–19428.
- [Chiu u. a., 2012] Chiu, S. J., Izatt, J. A., O’Connell, R. V., Winter, K. P., Toth, C. A. und Farsiu, S. „Validated automatic segmentation of AMD pathology including drusen and geographic atrophy in SD-OCT images“. *Investigative Ophthalmology & Visual Science* 53 (1), 2012, S. 53–61.
- [Chiu u. a., 2015] Chiu, S. J., Allingham, M. J., Mettu, P. S., Cousins, S. W., Izatt, J. A. und Farsiu, S. „Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema“. *Biomedical Optics Express* 6 (4), 2015, S. 1172–1194.
- [Chinn u. a., 1997] Chinn, S. R., Swanson, E. A. und Fujimoto, J. G. „Optical coherence tomography using a frequency-tunable optical source.“ *Optics Letters* 22 5, 1997, S. 340–2.
- [Chopra u. a., 2021] Chopra, R., Wagner, S. K. und Keane, P. A. „Optical coherence tomography in the 2020s—outside the eye clinic“. *Eye* 35 (1), 2021, S. 236–243.
- [Criminisi u. a., 2011] Criminisi, A., Shotton, J., Konukoglu, E. u. a. „Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning“. *Foundations and Trends® in Computer Graphics and Vision* 7 (2–3), 2011, S. 81–227.

- [Dabov u. a., 2009] Dabov, K., Foi, A., Katkovnik, V. und Egiazarian, K. „BM3D image denoising with shape-adaptive principal component analysis“. In: *Signal Processing with Adaptive Sparse Structured Representations (SPARS'09)*. 2009.
- [de Boer u. a., 2003] de Boer, J. F., Cense, B., Park, B. H., Pierce, M. C., Tearney, G. J. und Bouma, B. E. „Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography.“ *Optics Letters* 28 21, 2003, S. 2067–9.
- [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. u. a. „Clinically applicable deep learning for diagnosis and referral in retinal disease“. *Nature Medicine* 24 (9), 2018, S. 1342.
- [DeBuc, 2011] DeBuc, D. C. „A review of algorithms for segmentation of retinal image data using optical coherence tomography“. *Image Segmentation* 1, 2011, S. 15–54.
- [Devalla u. a., 2018] Devalla, S. K., Renukanand, P. K., Sreedhar, B. K., Subramanian, G., Zhang, L., Perera, S., Mari, J.-M., Chin, K. S., Tun, T. A., Strouthidis, N. G. u. a. „DRUNET: a dilated-residual U-Net deep learning network to segment optic nerve head tissues in optical coherence tomography images“. *Biomedical Optics Express* 9 (7), 2018, S. 3244–3265.
- [DOG, 2020] Deutsche Ophthalmologische Gesellschaft (DOG), Retinologische Gesellschaft e. V. (RG) und Berufsverband der Augenärzte Deutschlands e. V. (BVA). „Stellungnahme der DOG, der RG und des BVA zur Anti-VEGF-Therapie bei der neovaskulären altersabhängigen Makuladegeneration“. *Der Ophthalmologe* 117 (8), 2020, S. 746–754.
- [Dollár, 2022] Dollár, P. *Piotr's Computer Vision Matlab Toolbox (PMT)*. <https://github.com/pdollar/toolbox>. (Besucht am 01.03.2022).
- [Dössel u. a., 2014] Dössel, O. und Buzug, T. M. *Medizinische Bildgebung*. Walter de Gruyter GmbH & Co. KG, 2014.
- [Drexler u. a., 2008] Drexler, W. und Fujimoto, J. G. *Optical Coherence Tomography: Technology and Applications*. Springer Science & Business Media, 2008.
- [Dufour u. a., 2012] Dufour, P. A., Ceklic, L., Abdillahi, H., Schroder, S., De Dzanet, S., Wolf-Schnurrbusch, U. und Kowal, J. „Graph-based multi-surface segmentation of OCT data using trained hard and soft constraints“. *IEEE Transactions on Medical Imaging* 32 (3), 2012, S. 531–543.
- [Edmonds u. a., 1972] Edmonds, J. und Karp, R. M. „Theoretical improvements in algorithmic efficiency for network flow problems“. *Journal of the ACM (JACM)* 19 (2), 1972, S. 248–264.

-
- [Fang u. a., 2017] Fang, L., Cunefare, D., Wang, C., Guymer, R. H., Li, S. und Farsiu, S. „Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search“. *Biomedical Optics Express* 8 (5), 2017, S. 2732–2744.
- [Fercher u. a., 1995] Fercher, A. F., Hitzenberger, C. K., Kamp, G. und El-Zaiat, S. Y. „Measurement of intraocular distances by backscattering spectral interferometry“. *Optics Communications* 117 (1-2), 1995, S. 43–48.
- [Fercher, 2010] Fercher, A. F. „Optical coherence tomography - development, principles, applications.“ *Zeitschrift für medizinische Physik* 20 4, 2010, S. 251–76.
- [Ford u. a., 1962] Ford, D. R. und Fulkerson, D. R. *Flows in Networks*. Princeton University Press, 1962.
- [Fujimoto u. a., 2016] Fujimoto, J. und Huang, D. „Foreword: 25 Years of Optical Coherence Tomography.“ *Investigative Ophthalmology & Visual Science* 57 9, 2016, OCTi–OCTii.
- [Gal u. a., 2015] Gal, Y. und Ghahramani, Z. „Bayesian convolutional neural networks with Bernoulli approximate variational inference“. *arXiv preprint arXiv:1506.02158*, 2015.
- [Gal u. a., 2016] Gal, Y. und Ghahramani, Z. „Dropout as a bayesian approximation: Representing model uncertainty in deep learning (ICML)“. In: *International Conference on Machine Learning*. 2016, S. 1050–1059.
- [Gambichler u. a., 2005] Gambichler, T., Moussa, G., Sand, M., Sand, D., Altmeyer, P. und Hoffmann, K. „Applications of optical coherence tomography in dermatology“. *Journal of Dermatological Science* 40 (2), 2005, S. 85–94.
- [Gambichler u. a., 2007] Gambichler, T., Regeniter, P., Bechara, F. G., Orlikov, A., Vasa, R., Moussa, G., Stücker, M., Altmeyer, P. und Hoffmann, K. „Characterization of benign and malignant melanocytic skin lesions using optical coherence tomography in vivo“. *Journal of the American Academy of Dermatology* 57 (4), 2007, S. 629–637.
- [Ganaye u. a., 2019] Ganaye, P.-A., Sdika, M., Triggs, B. und Benoit-Cattin, H. „Removing segmentation inconsistencies with semi-supervised non-adjacency constraint“. *Medical Image Analysis* 58, 2019, S. 101551.
- [Garvin u. a., 2008] Garvin, M. K., Abramoff, M. D., Kardon, R., Russell, S. R., Wu, X. und Sonka, M. „Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-D graph search“. *IEEE Transactions on Medical Imaging* 27 (10), 2008, S. 1495–1505.
- [Garvin u. a., 2009] Garvin, M. K., Abramoff, M. D., Wu, X., Russell, S. R., Burns, T. L. und Sonka, M. „Automated 3-D intraretinal layer segmentation of macular

- spectral-domain optical coherence tomography images“. *IEEE Transactions on Medical Imaging* 28 (9), 2009, S. 1436–1447.
- [Glorot u. a., 2010] Glorot, X. und Bengio, Y. „Understanding the difficulty of training deep feedforward neural networks“. In: *Proceedings of the thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2010, S. 249–256.
- [Glocker u. a., 2012] Glocker, B., Pauly, O., Konukoglu, E. und Criminisi, A. „Joint classification-regression forests for spatially structured multi-object segmentation“. In: *European Conference on Computer Vision (ECCV)*. 2012, S. 870–881.
- [Goldberg u. a., 1988] Goldberg, A. V. und Tarjan, R. E. „A new approach to the maximum-flow problem“. *Journal of the ACM (JACM)* 35 (4), 1988, S. 921–940.
- [Graves, 2011] Graves, A. „Practical variational inference for neural networks“. *Advances in Neural Information Processing Systems* 24, 2011.
- [Greig u. a., 1989] Greig, D. M., Porteous, B. T. und Seheult, A. H. „Exact maximum a posteriori estimation for binary images“. *Journal of the Royal Statistical Society: Series B (Methodological)* 51 (2), 1989, S. 271–279.
- [Haeker u. a., 2007] Haeker, M., Sonka, M., Kardou, R., Shah, V. A., Wu, X. und Abramoff, M. D. „Automated segmentation of intraretinal layers from macular optical coherence tomography images“. In: *SPIE Medical Imaging 2007: Image Processing*. Bd. 6512. 2007, S. 651214.
- [Hamwood u. a., 2018] Hamwood, J., Alonso-Caneiro, D., Read, S. A., Vincent, S. J. und Collins, M. J. „Effect of patch size and network architecture on a convolutional neural network approach for automatic segmentation of OCT retinal layers“. *Biomedical Optics Express* 9 (7), 2018, S. 3049–3066.
- [Handels, 2009] Handels, H. *Medizinische Bildverarbeitung: Bildanalyse, Mustererkennung und Visualisierung für die computergestützte ärztliche Diagnostik und Therapie*. Springer-Verlag, 2009.
- [Hattenbach u. a., 2016] Hattenbach, L. O., Framme, C., Junker, B., Pielen, A., Agostini, H. T. und Maier, M. „Intraoperative Echtzeit-OCT in der Makulachirurgie“. *Der Ophthalmologe* 113, 2016, S. 656–662.
- [He u. a., 2016] He, K., Zhang, X., Ren, S. und Sun, J. „Deep Residual Learning for Image Recognition“. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, S. 770–778.
- [He u. a., 2019] He, Y., Carass, A., Liu, Y., Jedyak, B. M., Solomon, S. D., Saidha, S., Calabresi, P. A. und Prince, J. L. „Deep learning based topology guaranteed surface and MME segmentation of multiple sclerosis subjects from retinal OCT“. *Biomedical Optics Express* 10 (10), 2019, S. 5042–5058.

-
- [He u. a., 2021] He, Y., Carass, A., Liu, Y., Jedynek, B. M., Solomon, S. D., Saidha, S., Calabresi, P. A. und Prince, J. L. „Structured layer surface segmentation for retina OCT using fully convolutional regression networks“. *Medical Image Analysis* 68, 2021, S. 101856.
- [Heinrich u. a., 2018] Heinrich, L., Funke, J., Pape, C., Nunez-Iglesias, J. und Saalfeld, S. „Synaptic cleft segmentation in non-isotropic volume electron microscopy of the complete drosophila brain“. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2018, S. 317–325.
- [Heibel u. a., 2020] Heibel, H. D., Hooey, L. und Cockerell, C. J. „A Review of Non-invasive Techniques for Skin Cancer Detection in Dermatology“. *American Journal of Clinical Dermatology*, 2020, S. 1–12.
- [Ho, 1995] Ho, T. K. „Random decision forests“. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition (ICDAR)*. Bd. 1. 1995, S. 278–282.
- [Hori u. a., 2006] Hori, Y., Yasuno, Y., Sakai, S., Matsumoto, M., Sugawara, T., Madjarova, V. D., Yamanari, M., Makita, S., Yasui, T., Araki, T. u. a. „Automatic characterization and segmentation of human skin using three-dimensional optical coherence tomography“. *Optics Express* 14(5), 2006, S. 1862–1877.
- [Hu u. a., 2021] Hu, K., Liu, D., Chen, Z., Li, X., Zhang, Y. und Gao, X. „Embedded Residual Recurrent Network and Graph Search for the Segmentation of Retinal Layer Boundaries in Optical Coherence Tomography“. *IEEE Transactions on Instrumentation and Measurement*, 2021.
- [Huang u. a., 2017] Huang, G., Liu, Z., Van Der Maaten, L. und Weinberger, K. Q. „Densely Connected Convolutional Networks“. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, S. 4700–4708.
- [Huang u. a., 1991] Huang, D., Swanson, E. A., Lin, C. P., Schuman, J. S., Stinson, W. G., Chang, W., Hee, M. R., Flotte, T., Gregory, K., Puliafito, C. A. u. a. „Optical coherence tomography“. *Science* 254(5035), 1991, S. 1178–1181.
- [Ioffe u. a., 2015] Ioffe, S. und Szegedy, C. „Batch normalization: Accelerating deep network training by reducing internal covariate shift“. *arXiv preprint arXiv:1502.03167*, 2015.
- [ISAPS, 2020] ISAPS. *ISAPS International Survey on Aesthetic/Cosmetic Procedures*. 2020. URL: https://www.isaps.org/wp-content/uploads/2022/01/ISAPS-Global-Survey_2020.pdf.
- [Isensee u. a., 2021] Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J. und Maier-Hein, K. H. „nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation“. *Nature Methods* 18(2), 2021, S. 203–211.

- [Jähne, 2013] Jähne, B. *Digitale Bildverarbeitung*. Springer-Verlag, 2013.
- [Jang u. a., 2001] Jang, I.-K., Tearney, G. und Bouma, B. „Visualization of tissue prolapse between coronary stent struts by optical coherence tomography: comparison with intravascular ultrasound“. *Circulation* 104 (22), 2001, S. 2754–2754.
- [Josse u. a., 2011] Josse, G., George, J. und Black, D. „Automatic measurement of epidermal thickness from optical coherence tomography images using a new algorithm“. *Skin Research and Technology* 17 (3), 2011, S. 314–319.
- [Joy u. a., 2021] Joy, T. T., Sedai, S. und Garnavi, R. „Analyzing epistemic and aleatoric uncertainty for drusen segmentation in optical coherence tomography images“. In: *AIII Workshop on Trustworthy AI for Healthcare*. 2021.
- [Kafieh u. a., 2013] Kafieh, R., Rabbani, H. und Kermani, S. „A review of algorithms for segmentation of optical coherence tomography from retina“. *Journal of Medical Signals and Sensors* 3 (1), 2013, S. 45.
- [Karri u. a., 2016] Karri, S., Chakraborti, D. und Chatterjee, J. „Learning layer-specific edges for segmenting retinal layers with large deformations“. *Biomedical Optics Express* 7 (7), 2016, S. 2888–2901.
- [Kendall u. a., 2017a] Kendall, A., Badrinarayanan, V. und Cipolla, R. „Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding“. In: *Proceedings of the British Machine Vision Conference (BMVC)*. BMVA Press, 2017, S. 571–5712.
- [Kendall u. a., 2017b] Kendall, A. und Gal, Y. „What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?“ In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*. NIPS’17. Curran Associates Inc., 2017, S. 5580–5590.
- [Kepp u. a., 2018] Kepp, T., Droigk, C., Casper, M., Evers, M., Salma, N., Manstein, D. und Handels, H. „Segmentation of subcutaneous fat within mouse skin in 3D OCT image data using random forests“. In: *SPIE Medical Imaging 2018: Image Processing*. Bd. 10574. 2018, S. 1057426-1–1057426-8.
- [Kepp u. a., 2019a] Kepp, T., Droigk, C., Casper, M., Evers, M., Hüttmann, G., Salma, N., Manstein, D., Heinrich, M. P. und Handels, H. „Segmentation of mouse skin layers in optical coherence tomography image data using deep convolutional neural networks“. *Biomedical Optics Express* 10 (7), 2019, S. 3484–3496.
- [Kepp u. a., 2019b] Kepp, T., Ehrhardt, J., Heinrich, M. P., Hüttmann, G. und Handels, H. „Topology-Preserving Shape-Based Regression Of Retinal Layers In OCT Image Data Using Convolutional Neural Networks“. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, 2019, S. 1437–1440.
- [Kepp u. a., 2020] Kepp, T., Sudkamp, H., von der Burchard, C., Schenke, H., Koch, P., Hüttmann, G., Roider, J., Heinrich, M. P. und Handels, H. „Segmentation

-
- of retinal low-cost optical coherence tomography images using deep learning“. In: *SPIE Medical Imaging 2020: Computer-Aided Diagnosis*. Bd. 11314. 2020, 113141O.
- [Kepp u. a., 2022] Kepp, T., Andresen, J., Sudkamp, H., von der Burchard, C., Roeder, J., Hüttmann, G., Ehrhardt, J. und Handels, H. „Epistemic and Aleatoric Uncertainty Estimation for PED Segmentation in Home OCT Images“. In: *Bildverarbeitung für die Medizin 2022*. Springer Vieweg, 2022. Akzeptiert.
- [Kervadec u. a., 2019] Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J. und Ayed, I. B. „Boundary loss for highly unbalanced segmentation“. In: *International Conference on Medical Imaging with Deep Learning (MIDL)*. 2019, S. 285–296.
- [Khan u. a., 2020] Khan, A. I. und Al-Habsi, S. „Machine learning in computer vision“. *Procedia Computer Science* 167, 2020, S. 1444–1451.
- [Kingma u. a., 2014] Kingma, D. P. und Ba, J. „Adam: A method for stochastic optimization“. *arXiv preprint arXiv:1412.6980*, 2014.
- [Klein u. a., 2017] Klein, K. B., Bachelor, E. P., Becker, E. V. und Bowes, L. E. „Multiple same day cryolipolysis treatments for the reduction of subcutaneous fat are safe and do not affect serum lipid levels or liver function tests“. *Lasers in Surgery and Medicine* 49 (7), 2017, S. 640–644.
- [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*. 2012, S. 1097–1105.
- [Lakshminarayanan u. a., 2017] Lakshminarayanan, B., Pritzel, A. und Blundell, C. „Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles“. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*. NIPS’17. Curran Associates Inc., 2017, S. 6405–6416.
- [Lang u. a., 2013a] Lang, A., Carass, A., Hauser, M., Sotirchos, E. S., Calabresi, P. A., Ying, H. S. und Prince, J. L. „Retinal layer segmentation of macular OCT images using boundary classification“. *Biomedical Optics Express* 4 (7), 2013, S. 1133–1152.
- [Lang u. a., 2013b] Lang, A., Carass, A., Sotirchos, E., Calabresi, P. und Prince, J. L. „Segmentation of retinal OCT images using a random forest classifier“. In: *SPIE Medical Imaging 2013: Image Processing*. Bd. 8669. 2013, 86690R.
- [LeCun u. a., 2015] LeCun, Y., Bengio, Y. und Hinton, G. „Deep learning“. *Nature* 521 (7553), 2015, S. 436–444.

- [Leitgeb u. a., 2003] Leitgeb, R., Hitzenberger, C. und Fercher, A. F. „Performance of fourier domain vs. time domain optical coherence tomography“. *Optics Express* 11 (8), 2003, S. 889–894.
- [Leitgeb u. a., 2004] Leitgeb, R. A., Drexler, W., Unterhuber, A., Hermann, B., Bajraszewski, T., Le, T. M., Stingl, A. und Fercher, A. F. „Ultrahigh resolution Fourier domain optical coherence tomography.“ *Optics Express* 12 10, 2004, S. 2156–65.
- [Li u. a., 2006a] Li, K., Wu, X., Chen, D. Z. und Sonka, M. „Optimal surface segmentation in volumetric images—a graph-theoretic approach“. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (1), 2006, S. 119–134.
- [Li u. a., 2006b] Li, Y., Shekhar, R. und Huang, D. „Corneal pachymetry mapping with high-speed optical coherence tomography“. *Ophthalmology* 113 (5), 2006, S. 792–799.
- [Li u. a., 2015] Li, A., Cheng, J., Yow, A. P., Wall, C., Wong, D. W. K., Tey, H. L. und Liu, J. „Epidermal segmentation in high-definition optical coherence tomography“. In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. 2015, S. 3045–3048.
- [Lim u. a., 2012] Lim, J. H., Wickremasinghe, S. S., Xie, J., Chauhan, D. S., Baird, P. N., Robman, L. D., Hageman, G. und Guymer, R. H. „Delay to treatment and visual outcomes in patients treated with anti-vascular endothelial growth factor for age-related macular degeneration“. *American Journal of Ophthalmology* 153 (4), 2012, S. 678–686.
- [Liu u. a., 2018] Liu, X., Cao, J., Fu, T., Pan, Z., Hu, W., Zhang, K. und Liu, J. „Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning“. *IEEE Access* 7, 2018, S. 3046–3061.
- [Liu u. a., 2020] Liu, X., Cao, J., Wang, S., Zhang, Y. und Wang, M. „Confidence-guided Topology-preserving Layer Segmentation for Optical Coherence Tomography Images with Focus-column Module“. *IEEE Transactions on Instrumentation and Measurement*, 2020.
- [Long u. a., 2015] Long, J., Shelhamer, E. und Darrell, T. „Fully convolutional networks for semantic segmentation“. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, S. 3431–3440.
- [Loo u. a., 2018] Loo, J., Fang, L., Cunefare, D., Jaffe, G. J. und Farsiu, S. „Deep longitudinal transfer learning-based automatic segmentation of photoreceptor ellipsoid zone defects on optical coherence tomography images of macular telangiectasia type 2“. *Biomedical Optics Express* 9 (6), 2018, S. 2681–2698.
- [Lu u. a., 2019] Lu, D., Heisler, M., Lee, S., Ding, G. W., Navajas, E., Sarunic, M. V. und Beg, M. F. „Deep-learning based multiclass retinal fluid segmentation and

-
- detection in optical coherence tomography images using a fully convolutional neural network“. *Medical Image Analysis* 54, 2019, S. 100–110.
- [Luo u. a., 2020] Luo, X., Chen, J., Song, T., Chen, Y., Wang, G. und Zhang, S. „Semi-supervised medical image segmentation through dual-task consistency“. *arXiv preprint arXiv:2009.04448*, 2020.
- [Ma u. a., 2020a] Ma, J., He, J. und Yang, X. „Learning geodesic active contours for embedding object global information in segmentation cnns“. *IEEE Transactions on Medical Imaging*, 2020.
- [Ma u. a., 2020b] Ma, J., Wei, Z., Zhang, Y., Wang, Y., Lv, R., Zhu, C., Gaoxiang, C., Liu, J., Peng, C., Wang, L. u. a. „How distance transform maps boost segmentation CNNs: an empirical study“. In: *Medical Imaging with Deep Learning (MIDL)*. 2020, S. 479–492.
- [Ma u. a., 2021] Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X. und Martel, A. L. „Loss odyssey in medical image segmentation“. *Medical Image Analysis*, 2021, S. 102035.
- [Maier u. a., 2015] 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, e0145118.
- [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. u. a. „ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI“. *Medical Image Analysis* 35, 2017, S. 250–269.
- [Manstein u. a., 2008] Manstein, D., Laubach, H., Watanabe, K., Farinelli, W., Zurakowski, D. und Anderson, R. R. „Selective cryolysis: A novel method of non-invasive fat removal“. *Lasers in Surgery and Medicine* 40 (9), 2008, S. 595–604.
- [Markomichelakis u. a., 2004] Markomichelakis, N. N., Halkiadakis, I., Pantelia, E., Peponis, V., Patelis, A., Theodossiadis, P. und Theodossiadis, G. „Patterns of macular edema in patients with uveitis: qualitative and quantitative assessment using optical coherence tomography“. *Ophthalmology* 111 (5), 2004, S. 946–953.
- [Maurer u. a., 2003] Maurer, C. R., Qi, R. und Raghavan, V. „A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions“. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (2), 2003, S. 265–270.
- [Medeiros u. a., 2004] Medeiros, F. A., Zangwill, L. M., Bowd, C. und Weinreb, R. N. „Comparison of the GDx VCC Scanning Laser Polarimeter, HRT II Confocal Scanning Laser Ophthalmoscope, and Stratus OCT Optical Coherence Tomograph for the Detection of Glaucoma“. *Archives of Ophthalmology* 122 (6), 2004, S. 827–837.

- [Mehrtash u. a., 2020] Mehrtash, A., Wells, W. M., Tempany, C. M., Abolmaesumi, P. und Kapur, T. „Confidence calibration and predictive uncertainty estimation for deep medical image segmentation“. *IEEE Transactions on Medical Imaging* 39 (12), 2020, S. 3868–3878.
- [Mehta u. a., 2020] Mehta, R., Filos, A., Gal, Y. und Arbel, T. „Uncertainty evaluation metric for brain tumour segmentation“. In: *Medical Imaging with Deep Learning (MIDL)*. 2020.
- [Montuoro u. a., 2017] Montuoro, A., Waldstein, S. M., Gerendas, B. S., Schmidt-Erfurth, U. und Bogunović, H. „Joint retinal layer and fluid segmentation in OCT scans of eyes with severe macular edema using unsupervised representation and auto-context“. *Biomedical Optics Express* 8 (3), 2017, S. 1874–1888.
- [Mrejen u. a., 2013] Mrejen, S. u. a. „Multimodal imaging of pigment epithelial detachment: a guide to evaluation“. *Retina* 33 (9), 2013, S. 1735–1762.
- [Navarro u. a., 2019] Navarro, F., Shit, S., Ezhov, I., Paetzold, J., Gafita, A., Peeken, J. C., Combs, S. E. und Menze, B. H. „Shape-aware complementary-task learning for multi-organ segmentation“. In: *International Workshop on Machine Learning in Medical Imaging*. 2019, S. 620–627.
- [Novosel u. a., 2017] Novosel, J., Vermeer, K. A., Jong, J. H. de, Wang, Z. und Vliet, L. J. van. „Joint segmentation of retinal layers and focal lesions in 3-D OCT data of topologically disrupted retinas“. *IEEE Transactions on Medical Imaging* 36 (6), 2017, S. 1276–1286.
- [Odena u. a., 2016] Odena, A., Dumoulin, V. und Olah, C. „Deconvolution and Checkerboard Artifacts“. *Distill*, 2016. URL: <http://distill.pub/2016/deconv-checkerboard>.
- [Orlando u. a., 2019] Orlando, J. I., Seeböck, P., Bogunović, H., Klimscha, S., Grechenig, C., Waldstein, S., Gerendas, B. S. und Schmidt-Erfurth, U. „U2-net: A bayesian u-net model with epistemic uncertainty feedback for photoreceptor layer segmentation in pathological oct scans“. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI)*. 2019, S. 1441–1445.
- [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: *NIPS 2017 Workshop on Autodiff*. 2017.
- [Pierre-Kahn u. a., 2005] Pierre-Kahn, V., Tadayoni, R., Haouchine, B., Massin, P. und Gaudric, A. „Comparison of optical coherence tomography models OCT1 and Stratus OCT for macular retinal thickness measurement“. *British Journal of Ophthalmology* 89, 2005, S. 1581–1585.
- [Qian, 1999] Qian, N. „On the momentum term in gradient descent learning algorithms“. *Neural Networks* 12 (1), 1999, S. 145–151.

-
- [Ravishankar u. a., 2017] Ravishankar, H., Venkataramani, R., Thiruvankadam, S., Sudhakar, P. und Vaidya, V. „Learning and incorporating shape models for semantic segmentation“. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2017, S. 203–211.
- [Ronneberger u. a., 2015] Ronneberger, O., Fischer, P. und Brox, T. „U-net: Convolutional networks for biomedical image segmentation“. In: *International Conference on Medical Image Computing and Computer-assisted Intervention (MICCAI)*. 2015, S. 234–241.
- [Rosenfeld u. a., 2006] Rosenfeld, P. J., Brown, D. M., Heier, J. S., Boyer, D. S., Kaiser, P. K., Chung, C. Y. und Kim, R. Y. „Ranibizumab for neovascular age-related macular degeneration“. *New England Journal of Medicine* 355 (14), 2006, S. 1419–1431.
- [Roy u. a., 2017] Roy, A. G., Conjeti, S., Karri, S. P. K., Sheet, D., Katouzian, A., Wachinger, C. und Navab, N. „ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks“. *Biomedical Optics Express* 8 (8), 2017, S. 3627–3642.
- [Roy u. a., 2019] Roy, A. G., Conjeti, S., Navab, N., Wachinger, C., Initiative, A. D. N. u. a. „Bayesian QuickNAT: Model uncertainty in deep whole-brain segmentation for structure-wise quality control“. *NeuroImage* 195, 2019, S. 11–22.
- [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.
- [Sakata u. a., 2008] Sakata, L. M., Lavanya, R., Friedman, D. S., Aung, H. T., Gao, H., Kumar, R. S., Foster, P. J. und Aung, T. „Comparison of gonioscopy and anterior segment ocular coherence tomography in detecting angle closure in different quadrants of the anterior chamber angle“. *Ophthalmology* 115 (5), 2008, S. 769–774.
- [Salma u. a., 2018] Salma, N., Evers, M., Casper, M. J., Droigk, C., Kepp, T., Handels, H. und Manstein, D. „Mouse model of cold-induced localized fat loss (selective cryolipolysis)“. In: *Lasers in Surgery and Medicine*. Bd. 50. 2018, S19–S20.
- [Schmidt-Erfurth u. a., 2014] Schmidt-Erfurth, U., Chong, V., Loewenstein, A., Larsen, M., Souied, E., Schlingemann, R., Eldem, B., Monés, J., Richard, G. und Bandello, F. „Guidelines for the management of neovascular age-related macular degeneration by the European Society of Retina Specialists (EURETINA)“. *British Journal of Ophthalmology* 98 (9), 2014, S. 1144–1167.
- [Schippling u. a., 2015] Schippling, S., Balk, L., Costello, F., Albrecht, P., Balcer, L., Calabresi, P., Frederiksen, J., Frohman, E., Green, A., Klistorner, A. u. a. „Quality control for retinal OCT in multiple sclerosis: validation of the OSCAR-IB criteria“. *Multiple Sclerosis Journal* 21 (2), 2015, S. 163–170.

- [Schmidt-Erfurth u. a., 2016] Schmidt-Erfurth, U. und Waldstein, S. M. „A paradigm shift in imaging biomarkers in neovascular age-related macular degeneration“. *Progress in Retinal and Eye Research* 50, 2016, S. 1–24.
- [Sedai u. a., 2018] Sedai, S., Antony, B., Mahapatra, D. und Garnavi, R. „Joint segmentation and uncertainty visualization of retinal layers in optical coherence tomography images using Bayesian deep learning“. In: *Computational Pathology and Ophthalmic Medical Image Analysis*. Springer, 2018, S. 219–227.
- [Sedai u. a., 2019] Sedai, S., Antony, B., Rai, R., Jones, K., Ishikawa, H., Schuman, J., Gadi, W. und Garnavi, R. „Uncertainty guided semi-supervised segmentation of retinal layers in OCT images“. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2019, S. 282–290.
- [Seeböck u. a., 2019] Seeböck, P., Orlando, J. I., Schlegl, T., Waldstein, S. M., Bogunović, H., Klimescha, S., Langs, G. und Schmidt-Erfurth, U. „Exploiting epistemic uncertainty of anatomy segmentation for anomaly detection in retinal OCT“. *IEEE Transactions on Medical Imaging* 39 (1), 2019, S. 87–98.
- [Sheet u. a., 2013] Sheet, D., Chaudhary, A., Karri, S. P. K., Das, D., Katouzian, A., Banerjee, P., Navab, N., Chatterjee, J. und Ray, A. K. „In situ histology of mice skin through transfer learning of tissue energy interaction in optical coherence tomography“. *Journal of Biomedical Optics* 18 (9), 2013, S. 090503.
- [Sheet u. a., 2014] Sheet, D., Banerjee, S., Karri, S. P. K., Bag, S., Anura, A., Giri, A., Paul, R. R., Pal, M., Sarkar, B. C., Ghosh, R., Katouzian, A., Navab, N. und Ray, A. K. „Transfer learning of tissue photon interaction in optical coherence tomography towards in vivo histology of the oral mucosa“. *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, 2014, S. 1389–1392.
- [Sheet u. a., 2015] Sheet, D., Karri, S. P. K., Katouzian, A., Navab, N., Ray, A. K. und Chatterjee, J. „Deep learning of tissue specific speckle representations in optical coherence tomography and deeper exploration for in situ histology“. *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015, S. 777–780.
- [Simard u. a., 2003] Simard, P. Y., Steinkraus, D., Platt, J. C. u. a. „Best practices for convolutional neural networks applied to visual document analysis“. In: *ICDAR*. Bd. 3. 2003. 2003.
- [Simonyan u. a., 2015] Simonyan, K. und Zisserman, A. „Very Deep Convolutional Networks for Large-Scale Image Recognition“. In: *3rd International Conference on Learning Representations (ICLR)*. 2015.
- [Singer u. a., 2007] Singer, A. J., Wang, Z., McClain, S. A. und Pan, Y. „Optical coherence tomography: a noninvasive method to assess wound reepithelialization“. *Academic Emergency Medicine* 14 (5), 2007, S. 387–391.

-
- [Spital, 2018] Spital, G. „Therapie des diabetischen Makulaödems“. *Der Diabetologe* 14 (8), 2018, S. 577–589.
- [Srinivas u. a., 2004] Srinivas, S. M., de Boer, J. F., Park, B. H., Keikhanzadeh, K., Huang, H.-E. L., Zhang, J., Jung, W. G., Chen, Z. und Nelson, J. S. „Determination of burn depth by polarization-sensitive optical coherence tomography“. *Journal of Biomedical Optics* 9 (1), 2004, S. 207–213.
- [Srinivasan u. a., 2014] Srinivasan, P. P., Heflin, S. J., Izatt, J. A., Arshavsky, V. Y. und Farsiu, S. „Automatic segmentation of up to ten layer boundaries in SD-OCT images of the mouse retina with and without missing layers due to pathology“. *Biomedical Optics Express* 5 (2), 2014, S. 348–365.
- [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“. *The Journal of Machine Learning Research* 15 (1), 2014, S. 1929–1958.
- [Srivastava u. a., 2018] Srivastava, R., Yow, A. P., Cheng, J., Wong, D. W. und Tey, H. L. „Three-dimensional graph-based skin layer segmentation in optical coherence tomography images for roughness estimation“. *Biomedical Optics Express* 9 (8), 2018, S. 3590–3606.
- [Stanford Vision and Learning Lab, 2021a] Stanford Vision and Learning Lab. *CS231n: Convolutional Neural Networks for Visual Recognition*. 2021. URL: <https://cs231n.github.io/neural-networks-1/> (besucht am 01.03.2022).
- [Stanford Vision and Learning Lab, 2021b] Stanford Vision and Learning Lab. *CS231n: Convolutional Neural Networks for Visual Recognition*. 2021. URL: <https://cs231n.github.io/convolutional-networks/> (besucht am 01.03.2022).
- [Steinmetz u. a., 2021] Steinmetz, J. D., Bourne, R. R., Briant, P. S., Flaxman, S. R., Taylor, H. R., Jonas, J. B., Abdoli, A. A., Abrha, W. A., Abualhasan, A., Abu-Gharbieh, E. G. u. a. „Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study“. *The Lancet Global Health* 9 (2), 2021, e144–e160.
- [Sudkamp u. a., 2016] Sudkamp, H., Koch, P., Spahr, H., Hillmann, D., Franke, G., Müntz, M., Reinholz, F., Birngruber, R. und Hüttmann, G. „In-vivo retinal imaging with off-axis full-field time-domain optical coherence tomography“. *Optics Letters* 41 (21), 2016, S. 4987–4990.
- [Sudre u. a., 2017] Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S. und Cardoso, M. J. „Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations“. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, S. 240–248.

- [Sudkamp u. a., 2018] Sudkamp, H., Hillmann, D., Koch, P., Endt, M. vom, Spahr, H., Müntz, M., Pfäffle, C., Birngruber, R. und Hüttmann, G. „Simple approach for aberration-corrected OCT imaging of the human retina“. *Optics Letters* 43 (17), 2018, S. 4224–4227.
- [Tajbakhsh u. a., 2020] Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z. und Ding, X. „Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation“. *Medical Image Analysis* 63, 2020, S. 101693.
- [Tennakoon u. a., 2017] Tennakoon, R., Gostar, A. K., Hoseinnejhad, R. und Bab-Hadiashar, A. „Retinal fluid segmentation and classification in OCT images using adversarial loss based CNN“. In: *Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH)*. 2017, S. 30–37.
- [Uzunova u. a., 2020] Uzunova, H., Ehrhardt, J. und Handels, H. „Generation of annotated brain tumor MRIs with tumor-induced tissue deformations for training and assessment of neural networks“. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2020, S. 501–511.
- [Vandenhende u. a., 2021] Vandenhende, S., Georgoulis, S., Van Gansbeke, W., Proesmans, M., Dai, D. und Van Gool, L. „Multi-task learning for dense prediction tasks: A survey“. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [Venhuizen u. a., 2017] Venhuizen, F. G., Ginneken, B. van, Liefers, B., Grinsven, M. J. van, Fauser, S., Hoyng, C., Theelen, T. und Sánchez, C. I. „Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks“. *Biomedical Optics Express* 8 (7), 2017, S. 3292–3316.
- [Vermeer u. a., 2011] Vermeer, K., Van der Schoot, J., Lemij, H. und de Boer, J. „Automated segmentation by pixel classification of retinal layers in ophthalmic OCT images“. *Biomedical Optics Express* 2 (6), 2011, S. 1743–1756.
- [von der Burchard u. a., 2017] von der Burchard, C. C., Tode, J., Ehlken, C. und Roeder, J. „2mm Central Macular Volume Scan Is Sufficient to Detect Exudative Age-related Macular Degeneration Activity in Optical Coherence Tomography“. *Investigative Ophthalmology & Visual Science* 58 (8), 2017, S. 374–374.
- [von der Burchard u. a., 2021] von der Burchard, C., Moltmann, M., Tode, J., Ehlken, C., Sudkamp, H., Theisen-Kunde, D., König, I., Hüttmann, G. und Roeder, J. „Self-examination low-cost full-field OCT (SELFF-OCT) for patients with various macular diseases“. *Graefes Archive for Clinical and Experimental Ophthalmology* 259 (6), 2021, S. 1503–1511.
- [Wang u. a., 2019] Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S. und Vercauteren, T. „Aleatoric uncertainty estimation with test-time augmentation for

-
- medical image segmentation with convolutional neural networks“. *Neurocomputing* 338, 2019, S. 34–45.
- [Weissman u. a., 2004] Weissman, J., Hancewicz, T. und Kaplan, P. „Optical coherence tomography of skin for measurement of epidermal thickness by shapelet-based image analysis“. *Optics Express* 12 (23), 2004, S. 5760–5769.
- [Wei u. a., 2020] Wei, H. und Peng, P. „The segmentation of retinal layer and fluid in SD-OCT images using mutex dice loss based fully convolutional networks“. *IEEE Access* 8, 2020, S. 60929–60939.
- [Welzel u. a., 2017] Welzel, J. und Schuh, S. „Noninvasive diagnosis in dermatology“. *JDDG: Journal der Deutschen Dermatologischen Gesellschaft* 15 (10), 2017, S. 999–1016.
- [Welzel u. a., 1997] Welzel, J., Lankenau, E., Birngruber, R. und Engelhardt, R. „Optical coherence tomography of the human skin“. *Journal of the American Academy of Dermatology* 37 (6), 1997, S. 958–963.
- [Welzel, 2001] Welzel, J. „Optical coherence tomography in dermatology: a review“. *Skin Research and Technology: Review article* 7 (1), 2001, S. 1–9.
- [Wikipedia, 2022] Wikipedia. *Convolutional Neural Network*. 2022. URL: https://de.wikipedia.org/wiki/Convolutional_Neural_Network (besucht am 01.03.2022).
- [Wintergerst u. a., 2017] Wintergerst, M. W., Schultz, T., Birtel, J., Schuster, A. K., Pfeiffer, N., Schmitz-Valckenberg, S., Holz, F. G. und Finger, R. P. „Algorithms for the automated analysis of age-related macular degeneration biomarkers on optical coherence tomography: a systematic review“. *Translational Vision Science & Technology* 6 (4), 2017, S. 10–10.
- [Xue u. a., 2020] Xue, Y., Tang, H., Qiao, Z., Gong, G., Yin, Y., Qian, Z., Huang, C., Fan, W. und Huang, X. „Shape-aware organ segmentation by predicting signed distance maps“. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Bd. 34. 07. 2020, S. 12565–12572.
- [Yadav u. a., 2014] Yadav, G., Maheshwari, S. und Agarwal, A. „Contrast limited adaptive histogram equalization based enhancement for real time video system“. In: *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. 2014, S. 2392–2397.
- [Yonekawa u. a., 2015] Yonekawa, Y., Miller, J. W. und Kim, I. K. „Age-related macular degeneration: advances in management and diagnosis“. *Journal of Clinical Medicine* 4 (2), 2015, S. 343–359.
- [Yushkevich u. a., 2006] Yushkevich, P. A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J. C. und Gerig, G. „User-Guided 3D Active Contour Segmen-

- tation of Anatomical Structures: Significantly Improved Efficiency and Reliability“. *Neuroimage* 31 (3), 2006, S. 1116–1128.
- [Zadeh u. a., 2017] Zadeh, S. G., Wintergerst, M. W., Wiens, V., Thiele, S., Holz, F. G., Finger, R. P. und Schultz, T. „CNNs enable accurate and fast segmentation of drusen in optical coherence tomography“. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, S. 65–73.
- [Zadeh u. a., 2019] Zadeh, S. G., Wintergerst, M. W. und Schultz, T. „Intelligent interaction and uncertainty visualization for efficient drusen and retinal layer segmentation in Optical Coherence Tomography“. *Computers & Graphics* 83, 2019, S. 51–61.
- [Zelickson u. a., 2009] Zelickson, B., Egbert, B. M., Preciado, J., Allison, J., Springer, K., Rhoades, R. W. und Manstein, D. „Cryolipolysis for noninvasive fat cell destruction: initial results from a pig model“. *Dermatologic Surgery* 35 (10), 2009, S. 1462–1470.
- [Zur u. a., 2018] Zur, D., Igllicki, M., Busch, C., Invernizzi, A., Mariussi, M., Loewenstein, A., Cebeci, Z., Chhablani, J. K., Chaikitmongkol, V., Couturier, A. u. a. „OCT biomarkers as functional outcome predictors in diabetic macular edema treated with dexamethasone implant“. *Ophthalmology* 125 (2), 2018, S. 267–275.