



Karlsruher Institut für Technologie

Fakultät für Maschinenbau

Institut für Automation und angewandte Informatik

Optimierung von Deep-Learning-Methoden zur Auswertung biologischer 3D-Bildstapel

Masterarbeit in Elektrotechnik und Informationstechnik

eingereicht von

David Exler

Erstprüfer: Prof. Dr. Gerardo Hernandez-Sosa

Zweitprüfer: Prof. Dr. Markus Reischl

2025

Inhaltsverzeichnis

1 Einleitung	1
2 Theorie	3
2.1 Überblick	3
2.2 Methoden	3
2.2.1 Benchmark	3
2.2.2 Segmentierung	4
2.2.3 Klassifikation	5
2.3 Literaturrecherche	10
2.3.1 Benchmark	10
2.3.2 Segmentierungsmodelle	11
2.3.3 Klassifikator	12
2.4 Offene Probleme	14
2.5 Zielsetzung	15
3 Neues Konzept	17
3.1 Überblick	17
3.2 Daten	18
3.3 Neues Kriterium: Injektive Panoptische Qualität	20
3.4 Klassifikatormethoden	23
3.4.1 Übersicht	23
3.4.2 Encoder	24
3.4.3 Vorverarbeitung	24
3.4.4 Vortraining	26
3.4.5 Klassifikations-Kopf	28
3.5 Segmentierung	30
3.5.1 Modelle	30
3.5.2 Nachverarbeitungsmethoden	30
4 Implementierung	31
4.1 Überblick	31
4.2 Segmentierungsmodelle	31
4.3 Klassifikatoren	32
4.4 3D-Zelldaten-Pipeline	34
4.4.1 Übersicht	34

4.4.2	Segmentierung	35
4.4.3	Labeling-App	36
4.4.4	Methodenvergleich	38
4.4.5	Training	39
4.4.6	Visualisierung	40
4.5	Umsetzung	41
5	Ergebnisse	43
5.1	Überblick	43
5.2	Hardware	43
5.3	Segmentierung	43
5.4	Klassifikation	47
5.4.1	Überblick	47
5.4.2	Encoder	49
5.4.3	Vortraining	50
5.4.4	Klassifikations-Kopf	53
5.4.5	Vorverarbeitung	55
5.4.6	Wichtigkeit der Marker	57
6	Diskussion	59
6.1	Überblick	59
6.2	Segmentierung	59
6.3	Klassifikation	61
7	Zusammenfassung und Ausblick	69
A	Anhang	73
A.1	SWINV2 Architektur	73
A.2	Injektive Panoptische Qualität (IPQ)-Ergebnisse	74
A.3	Einzelne Ergebnisse aller Klassifikator Kombinationen	80
	Quellenverzeichnis	83

Einleitung 1

Die vorliegende Arbeit behandelt die automatisierte Optimierung von Deep-Learning-Methoden zur Segmentierung und Klassifikation dreidimensionaler Bildstapel. Zur Optimierung der Segmentierung wird ein neues Qualitätskriterium für Segmentierungsmodelle eingeführt und zur Optimierung der Klassifikation werden diverse neue Methoden sowie ein Framework zum automatisierten Vergleich der Methoden vorgestellt. Die eingeführten Methoden werden an einem Datensatz aus dreidimensionalen Mikroskopaufnahmen von Myotubenkulturen angewandt. Myotuben sind mehrkernige Muskelzellfäden [1, 2]. Sie repräsentieren ein intermediäres Stadium der Muskelentwicklung, in dem sich die grundlegende Organisation der Muskelfaser bildet [3]. Menschliche Myotube-Kulturen [4] können für diverse Forschungszwecke als Modellsysteme eingesetzt werden. Sie werden beispielsweise verwendet, um Muskelkrankheiten zu modellieren [5], Antworten auf neue Medikamente vorherzusagen [6] und synthetische Muskeln [7] sowie Muskelregeneration [8] zu erforschen. In den meisten Fällen werden die Myotuben, ihre Zellkerne (Nuclei) und die zugehörigen, umliegenden Strukturen eingefärbt und unter dem Mikroskop analysiert [9, 10, 11]. Die Bilddaten entstehen unter unterschiedlichen Herstellungsbedingungen, Färbungen und Aufnahmegeräten, was eine generalisierte Automatisierung erschwert [12, 13, 14, 15].

Im Zuge der vorliegenden Arbeit werden nach dem Protokoll von Couturier et al. [16] hergestellte in-vitro-Kulturen von Myotuben analysiert und die Ergebnisse automatisiert ausgewertet. Aus den Daten werden interpretierbare Merkmale wie die Zellkernanzahl, die Verteilung der Zellkernklassen und die Verteilung der Volumina der Instanzen extrahiert. Diese Merkmale ermöglichen die Überwachung der Entwicklung der Myotuben ohne manuellen Aufwand. Alle hierzu angewandten Methoden sind auf Basis quantitativer Vergleiche aktueller Forschung gewählt und bestmöglich an die Anforderungen angepasst.

Um die Analyse biologischer 3D-Daten effizient und in großem Umfang durchführen zu können, sind Bildverarbeitungsprogramme erforderlich [17], da die manuelle Auswertung anspruchsvoll und zeitintensiv ist [5]. Myotuben ordnen sich in Forschungsumgebungen zu chaotischen Netzen mit Verschränkungen und Überkreuzungen an [18]. Aktuell sind Bildverarbeitungsmethoden nicht in der Lage, zuverlässig einzelne Myotuben in einem dreidimensionalen Bild zu erkennen und von ihrem Anfang bis zum Ende zu verfolgen [19]. Besonders die Bündel, die in der Entwicklung der Myotuben häufig entstehen, verhindern die getrennte Segmentierung der einzelnen Myotuben [7]. Außerdem stellen die dreidimensionalen Daten eine große Herausforderung für Hard- und Software dar [20, 21, 22, 23]. Besonders herausfordernd sind dabei die stark erhöhten Speicher- und GPU-

Anforderungen sowie die geringere Zahl an etablierten Methoden und Datensätzen [24, 25, 26]. Der Mehrwert einer dritten räumlichen Dimension kann die Ergebnisse allerdings wesentlich verbessern, was die Verarbeitung von 3D-Daten zu einem zentralen Forschungsaspekt macht [27, 28].

Das übergeordnete Ziel der vorliegenden Arbeit ist es, die Segmentierung und Klassifikation der Nuclei in den Myotubenkulturen zu optimieren. Hierzu müssen ein optimales Modell aus etablierten 3D-Segmentierungsmodellen für Nuclei gewählt, Annotationen für die Klassifikationen durch Expert*Innen erfasst und Methoden zur Klassifikation von 3D-Bildstapeln optimiert werden. Um diese Anforderungen zu erfüllen, liefert die Arbeit folgende Neuheitswerte:

- Ein neues Qualitätskriterium zur Bewertung von Segmentierungsmodellen hinsichtlich der Eignung der entstehenden Segmentierungsmasken zur Extraktion interpretierbarer Merkmale.
- Eine Anwendung, die zeiteffizient Expertenwissen zu den Klassen von Nuclei in dreidimensionalen Daten erfasst.
- Eine neue Methode, um Klassifikatoren durch semi-supervised-Vortraining stärker zu generalisieren.
- Zwei neue Klassifikations-Kopf-Architekturen für dreidimensionale Daten, die einen Klassifikator an verschiedene räumliche Verteilungen von Informationen in dreidimensionalen biologischen Daten anpassen.
- Zwei Vorverarbeitungsmethoden für dreidimensionale Nuclei, die grundlegende Aussagen über die Lokalisierung der zur Klassifikation relevanten Informationen ermöglichen.
- Ein automatisiertes Framework, das Anwender*Innen ohne Vorkenntnisse die Segmentierung und das zeiteffiziente Annotieren von Nuclei in neuen Datensätzen ermöglicht und einen Klassifikator auf Basis umfassender Vergleiche der eingeführten Methoden optimiert.
- Ein optimierter Klassifikator für die vorliegende Myotubenkultur.

Die nachfolgende Ausarbeitung ist wie folgt strukturiert. In Kapitel 2 werden die Grundlagen, die relevante Literatur und der Stand der Technik beschrieben. Außerdem sind dort offene Probleme sowie die Ansätze, die die vorliegende Arbeit verfolgt, um sie zu lösen, dargestellt. Die Methodik (Kapitel 3) beschreibt das neue, praktische Konzept, das die Arbeit einführt, und Kapitel 4 (Implementierung) behandelt die praktische Umsetzung dieser Methoden. In Kapitel 5 (Ergebnisse) werden die Ergebnisse der durchgeföhrten Experimente ungewertet dargestellt; im Kapitel 6 (Diskussion) werden sie anschließend ausgewertet. Mit dem Kapitel 7 (Zusammenfassung und Ausblick) schließt die Ausarbeitung ab, stellt kurz die gesamte Arbeit dar und liefert einen Ausblick auf zukünftige Ziele. Der Code dieser Arbeit ist verfügbar unter: github.com/DavidExler/Masterarbeit.

Theorie 2

2.1 Überblick

Im nachfolgenden Kapitel wird der theoretische Hintergrund der vorliegenden Thesis behandelt. Hierzu werden sowohl für die Arbeit relevante Methoden als auch die Literatur verwandter Projekte und Studien zum aktuellen Stand der Technik beleuchtet. Da die Arbeit im Kern die panoptische Segmentierung von Zelldaten behandelt, werden hier Methoden zur Instanzsegmentierung und Klassifikation vorgestellt. Die Methoden zur Instanzsegmentierung sind auf Künstliche Intelligenz (**KI**)-Methoden konzentriert; von klassischen Ansätzen wird abgesehen. Für die Klassifikation wird von grundlegenden Erklärungen der Deep-Learning-Methoden abgesehen. Stattdessen werden einzelne moderne Ansätze sowie relevante Methoden der Visualisierung oder Optimierung von Klassifikatoren angeführt. Um die Methoden in einem vergleichbaren und reproduzierbaren Aufbau zu demonstrieren, werden zudem Benchmark-Methoden beleuchtet. Dem theoretischen Hintergrund wird der Neuheitswert der Arbeit gegenübergestellt, um den Beitrag der vorliegenden Arbeit zur Forschung zu verdeutlichen.

2.2 Methoden

2.2.1 Benchmark

Um die Leistungsfähigkeit der im Rahmen der vorliegenden Thesis eingeführten Methoden zu prüfen, sind umfangreiche Datensätze erforderlich. Für jede isolierbare Aufgabe muss ein Datensatz gewählt werden, der Annotationen entsprechend dieser Aufgabe enthält. Die Herausforderung bei der Auswahl eines Datensatzes besteht darin, dass die darin enthaltenen Daten (Quelldaten) den Daten *ähnlich* sein müssen, für die die Anwendung entworfen wird (Zieldaten). So wird sichergestellt, dass sich die auf den Quelldaten gemessene Qualität der Anwendung sinnvoll auf die Zieldaten übertragen lässt [29]. Der Begriff *Ähnlichkeit* ist mehrdeutig, und das Definieren von Merkmalen in Daten, die *Ähnlichkeit* messen, ist anspruchsvoll. Metriken, die als Maß für *Ähnlichkeit* herangezogen werden, müssen maßgeschneidert zur Anwendung passen und sind bereits breit erforscht [30, 31, 32]. Daten können mithilfe passender Metriken *Ähnlichkeits*-Gruppen, sogenannten *Domänen*, zugeordnet werden [33]. Beispiele für Domänenunterschiede in biomedizinischen Bildaufnahmen sind verschiedene Farbmarker, Aufnahmegeräte oder Zoomstufen. In Abb. 2.1 sind Zellkerne verschiedener Domänen abgebildet.

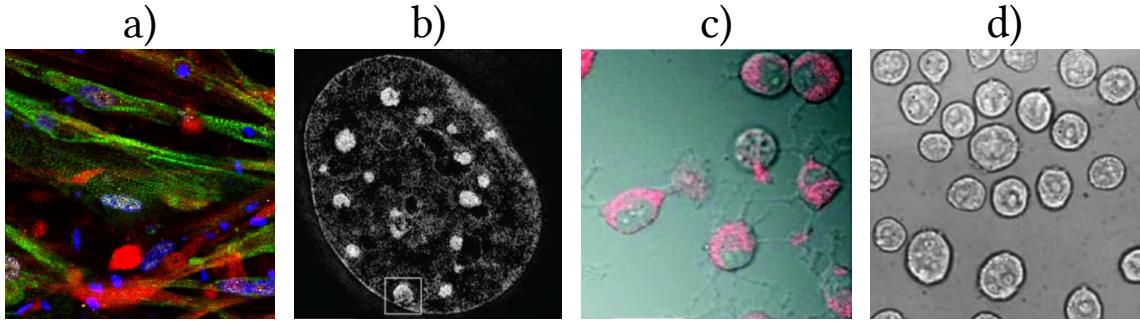


Abb. 2.1 | Vier Aufnahmen von Zellen. Die Bilddomänen lassen sich durch die verschiedenen Marker und Aufnahmetechniken klar unterscheiden. a) 3D-Bildstapel aus den Daten der vorliegenden Arbeit mit verschiedenen Fluoreszenzmarkern, aufgenommen mit einem Leica TCS SP8 Konfokalmikroskop. b) 3D-SIM-Superauflösungsmikroskopie mit DAPI-Färbung [34]. c) Konfokale Fluoreszenzmikroskopie mit antikörperbasierten Farbstoffen [35]. d) Klassische Hellfeldmikroskopie [36].

Intuitiv gehören die sichtbaren Objekte zur Klasse *Zellkern*, aber der Stil unterscheidet sich stark. Sie unterscheiden sich also in ihrer jeweiligen Domäne. In der Bildverarbeitung ist es essenziell, die Domäne der Quelldaten im Hinblick auf die Aufgabe der Applikation zu berücksichtigen [37, 38]. Hierzu kann ein Datensatz mit passender Domäne gewählt werden oder eine Domänenanpassung durchgeführt werden [39, 40, 41].

Ghosh et al. definieren Domänenanpassung: „Gegeben Quell- und Ziel-Domänen D_s und D_t sowie die Aufgaben τ_s und τ_t , zielen Domänenadaptations-basierte Verfahren darauf ab, ein Modell mit Parametern θ zu erlernen, das für die Zielaufgabe geeignet ist, wenn $D_s \neq D_t$ und $\tau_s = \tau_t$.“ [42].

2.2.2 Segmentierung

Segmentierung ist die Aufgabe, Pixel mit semantischen Annotationen zu klassifizieren (semantische Segmentierung [43]), einzelne Objekte voneinander abzugrenzen (Instanzsegmentierung [44]) oder beides zu kombinieren (panoptische Segmentierung [45]) [46]. In Abb. 2.2 sind beispielhaft Annotationen der verschiedenen Segmentierungsarten zu sehen [45]. Links zu sehen ist ein exemplarisches Originalbild. Daneben sind von links nach rechts Segmentierungsmasken für semantische, Instanz- und panoptische Segmentierung zu sehen. In der semantischen sowie in der panoptischen Maske sind die Farben der Objekte mit einer interpretierbaren Objektklasse verknüpft.

Für die verschiedenen Segmentierungsarten werden Architekturen an die jeweilige Aufgabe angepasst [47, 48]. Modelle sind dabei in der Regel nach dem Vorbild des *U-Net* [49] aus einem Merkmalsextraktor (Encoder) und einem Vorhersagenetz (Decoder) aufgebaut [50]. Der Encoder nutzt zur Merkmalsextraktion beispielsweise Bildfaltungen mit Kernen, deren optimale Gewichte anhand annotierter Daten gelernt werden [51]. Dabei verringert der Encoder iterativ die Größe der Eingabe jeder Schicht des Netzes in X-, Y- und im dreidimensionalen Fall in Z-Richtung, erhöht dabei aber die Informationstiefe pro verbleibendem Pixel, bis ein hochdimensionaler Merkmalsvektor übrig bleibt. Der Decoder

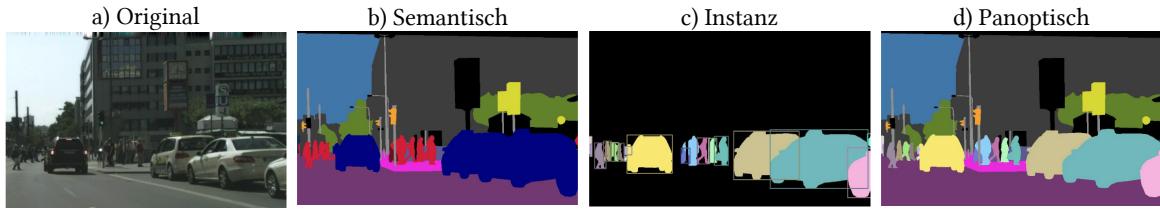


Abb. 2.2 | Die verschiedenen Arten der Segmentierung. Links ist das Originalbild zu sehen, rechts sind die zugeordneten pixelweise Annotationen farblich eingezeichnet. Gleiche Farben bedeuten gleiche Annotationen. In der semantischen Maske sind gleiche Annotationen mehrerer Objekte derselben semantischen Klasse zu finden. In der Instanz-Maske ist jedem Objekt eine individuelle Annotation zugeordnet, unabhängig von dessen Klasse. In der panoptischen Maske sind auch einzelne Objekte getrennt, den Annotationen verschiedener Klassen werden allerdings noch semantische Klassen zugeordnet [45].

hebt, meist durch transponierte Bildfaltungen [52, 53], die räumliche Auflösung schrittweise wieder an, indem er die Bilddimensionen vergrößert und die Merkmalskanäle gleichzeitig reduziert [54]. Über sogenannte Skip connections [49] werden dabei Merkmale aus den entsprechenden Encoder-Schichten mit den Decoder-Stufen verknüpft, sodass sowohl globale Kontextinformationen als auch feine Strukturen für die Segmentierung erhalten bleiben [55, 56].

Zur Nachbearbeitung von Segmentierungsmasken kann das Watershed-Verfahren eingesetzt werden [57]. Dabei wird zunächst aus der Segmentierungsmaske ein Gradientenbild erzeugt und anschließend eine Überflutungssimulation durchgeführt, bei der regionale Minima als Startpunkte dienen. Der Algorithmus eignet sich insbesondere zur Trennung überlagerter Instanzen.

Biologische und medizinische Bilddaten sind oft dreidimensionale Volumina. Dreidimensionale Daten erhöhen sowohl den Rechenaufwand für die Segmentierung als auch die Komplexität von Segmentierungsmodellen und stellen damit eine besondere Herausforderung dar [58, 59, 60]. Methoden für zweidimensionale Segmentierung lassen sich anpassen, um direkt mit dreidimensionalen Daten zu operieren [61]. Auch explizite 3D-Segmentierungsmethoden werden erforscht [62, 63]. Da die Auflösung in Z-Richtung oft geringer ist als die räumliche Auflösung, werden häufig 2.5D-Methoden verwendet, die die Beziehungen zwischen Volumenschichten gesondert modellieren [64]. Auch 2D-Methoden werden für 3D-Segmentierung eingesetzt, indem einzelne 2D-Schichten des Volumens segmentiert und anschließend durch Nachverarbeitung zusammengefügt werden [65]. Die besten Ergebnisse liefern in der Regel 3D-Methoden [66].

2.2.3 Klassifikation

Klassifikation beschreibt das Zuordnen einer Kategorie oder Klasse, zu der eine gegebene Stichprobe gehört [67]. Hierzu werden die Merkmale des Objekts, das in der Stichprobe präsentiert wird, durch Beobachtung oder Messung erfasst [68]. Nach wiederholter Extraktion der Merkmale von Objekten verschiedener Klassen werden Muster in den Merkmalen gesucht, um Regeln für die Zuweisung von Objekten zu Klassen auf Basis dieser

Muster festzulegen [69, 70]. Sowohl die Algorithmen zur Merkmalsextraktion als auch zum Ableiten der Muster und Regeln können mit unterschiedlich hohem Rechenaufwand, Abstraktionsgrad und Maß an Generalisierbarkeit implementiert werden [71]. Zur Merkmalsextraktion werden klassisch beschreibende Eigenschaften des Objekts berechnet und miteinander kombiniert [72]. Als Eigenschaften eignen sich beispielsweise die Verteilung der Farbkanäle, eine Charakterisierung der Textur oder die Fläche des Objekts [73, 74]. Eine weitere verbreitete Eigenschaft ist eine Kombination von Parametern der Fourier-Entwicklung einer Kontur, die aus der diskreten komplexen Zahlenfolge

$$c[n] = x[n] + i \cdot y[n], \quad n = 0, \dots, N - 1, \quad (2.1)$$

mithilfe der diskreten Fourier-Transformation

$$F[k] = \sum_{n=0}^{N-1} c[n] \cdot e^{-2\pi i \frac{kn}{N}}, \quad k = 0, \dots, N - 1, \quad (2.2)$$

gebildet werden [75, 76]. Hierbei sind $x[n]$ und $y[n]$ die Koordinaten des n -ten von N equidistanten Stützpunkten entlang der Kontur des Objekts, $c[n]$ ihre komplexe Darstellung und $F[k]$ die Fourier-Transformation der komplexen Darstellungen. Die Ergebnisse der Fourier-Transformationen mehrerer Stützpunkte werden anschließend als Eigenschaften verwendet. Merkmalsvektoren werden häufig abstrahiert und in ihrer Dimensionalität reduziert, beispielsweise durch eine Principal Component Analysis [77, 78]. Sind keine Annotationen verfügbar, werden diese Metriken zum Clustering verwendet [69, 79]. Wenn nur wenige Annotationen vorhanden sind, können semi-supervised-Verfahren angewandt werden, die insbesondere die Ähnlichkeit zwischen Stichproben ohne Annotationen herausarbeiten [80, 81]. Eine prominente Methode des semi-supervised-Lernens ist das Label-Spreading, das mithilfe einer Kernfunktion [82, 83] die Dimensionen von Merkmalsvektoren ändert und in einen alternativen Merkmalsraum transformiert [84]. Verschiedene Kernfunktionen wie die Radiale Basis Funktion [85]

$$\phi(x, y) = \exp(-\gamma \|x - y\|^2), \quad \gamma > 0 \quad (2.3)$$

werden für das Label-Spreading eingesetzt [86]. Hierbei sind $x, y \in \mathbb{R}^d$ die Koordinaten der Stichprobe im Merkmalsraum, γ ein Parameter, der die Breite der Radialbasisfunktion steuert, und $\phi(x, y)$ der Wert der Radialen-Basis-Funktion. Die meist genutzten Methoden der Klassifikation sind logikbasierte Ansätze wie Entscheidungsbäume, Perzeption-basierte Ansätze wie neuronale Netze, statistische Ansätze wie Bayes'sche Netzwerke oder Nächster-Nachbar-Verfahren und Support-Vector-Maschinen [87]. Diese Methoden basieren direkt auf Ähnlichkeiten zwischen den Merkmalen unbekannter Stichproben und Stichproben mit bekannter Klasse [88]. Moderne Anwendungen nutzen zur Merkmalsextraktion verschiedene Deep-Learning-basierte Methoden [89]. Vor allem Convolutional Neural Networks (CNNs) [90] und Vision Transformers (ViTs) [91] können aus Bildern aussagekräftige, abstrakte Merkmale extrahieren [92]. Ein Netz, das zur Merkmalsextraktion eingesetzt wird, wird als **Encoder** bezeichnet.

Als **Klassifikations-Kopf** wird der zusammenfassende Teil des Klassifikators bezeichnet; er gibt einen Zuversichtlichkeitswert für jede Klasse aus. Der State-of-the-Art für den Klassifikations-Kopf ist ein neuronales Netz, das auf Basis der abstrakten Merkmale des Encoders eine Zuversichtlichkeit für jede Klasse ausgibt [93]. Hierzu lernt in der Regel ein Multi-Layer-Perzeptron auf Basis von Trainingsdaten mit zugehöriger Annotation den Zusammenhang zwischen den Merkmalen und der assoziierten Klasse [94].

Für das Training von Klassifikatoren sind eine Loss-Funktion und häufig Vorverarbeitungsmethoden erforderlich. Der Cross-Entropy Loss [95]

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{n=1}^N -\log p(\tilde{y} = \tilde{y}_n \mid x_n, \theta) \quad (2.4)$$

ist die etablierte Loss-Funktion für das Training von Klassifikatoren [96, 97]. Hierbei ist der Loss L von der Annotation \tilde{y} und der Vorhersage \tilde{y}_n abhängig, die von den Eingangsdaten x_n und den Modellparametern θ bestimmt werden. Ein Problem der Cross-Entropy-Loss-Minimierung ist ihre Anfälligkeit gegenüber Rauschen in den Annotationen. Viele verschiedene Ansätze in der Forschung gehen dieses Problem an [98, 99, 100]. Eine häufig genutzte Methode ist die Minimierung des Generalized Cross Entropy Loss [101]

$$\arg \min_{\theta, w \in [0,1]^n} \sum_{i=1}^n w_i \mathcal{L}_q(f(x_i; \theta), y_i) - \mathcal{L}_q(k) \sum_{i=1}^n w_i, \quad (2.5)$$

wobei \mathcal{L}_q die generalisierte Form des Cross-Entropy-Losses ist, die durch den Parameter q reguliert wird. Dieser kontrolliert den Einfluss fehlerhafter oder unsicherer Trainingsbeispiele. Die Gewichte $w_i \in [0, 1]$ dienen der Gewichtung einzelner, besonders unsicherer Trainingsinstanzen. Das Modell $f(x_i; \theta)$ gibt die Vorhersage für die Eingabe x_i basierend auf den Modellparametern θ aus, während y_i die entsprechende Zielannotation ist.

Bilineare Interpolation ist ein Verfahren zur Bildvorverarbeitung, das die Bilddimension erhöht, indem Werte für neue Pixel zwischen bestehenden Pixeln geschätzt werden [102]. Zur Schätzung des neuen Wertes wird dabei ein gewichtetes Mittel aus den vier benachbarten Pixeln genommen:

$$\hat{f}(x, y) = (1-p)(1-q)f_{i,j} + p(1-q)f_{i+1,j} + (1-p)qf_{i,j+1} + pqf_{i+1,j+1}, \quad (2.6)$$

wobei $\hat{f}(x, y)$ der neue Wert, $p, q \in [0, 1]$ die relativen Abstände zu den Nachbarpixeln, i, j die Indizes der Nachbarpixel und f die Intensitäten der Nachbarpixel sind.

Normierungsmethoden werden während des Trainings eingesetzt, um die Daten zu regulieren und Signale weder unverhältnismäßig groß noch verschwindend klein werden zu lassen. Batch normalization normalisiert die Eingänge einer Schicht über ein mini batch, indem für jedes abstrakte Merkmal der Schicht, das bei der Dimensionsreduktion des Bildes entsteht, eine Transformation durchgeführt wird [103]. Für die Transformation wird

der Mittelwert des mini batches vom Wert abgezogen und durch die Standardabweichung geteilt. Layer normalization verfolgt einen ähnlichen Ansatz, berechnet den Mittelwert und die Varianz aber pro Schicht von allen Neuronen [104]. Beide stabilisieren das Training tiefer neuronaler Netze und lassen die Normalisierung als Bestandteil der Modellarchitektur lernen, statt sie nur als Vorverarbeitungsschritt durchzuführen [105, 106].

Als Vergleichsmetrik der Klassifikation wird für gewöhnlich die Genauigkeit des getesteten Netzes auf den annotierten Daten verwendet:

$$\text{Genauigkeit} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{\hat{y}_i = y_i\}, \quad (2.7)$$

wobei N die Anzahl der Vorhersagen, \hat{y}_i die Vorhersage des Klassifikators und y_i die Annotation sind.

Um das Verhalten von Klassifikatoren zu visualisieren, gibt es verschiedene Methoden. Eine dieser Methoden ist Grad-CAM [107]. Grad-CAM steht für Gradient-weighted Class Activation Mapping und ist eine Methode, um räumlich aufgelöste Relevanzkarten aus tiefen neuronalen Netzen zu erzeugen. Dabei werden die Gradienten einer bestimmten Klasse bezüglich der Aktivierungen einer Schicht des Modells verwendet, um Regionen des Eingabebildes zu finden, die besonders stark zur Vorhersage beitragen. Die resultierende „Wichtigkeits“-Karte wird auf die Eingabe zurückprojiziert und wird typischerweise als farbige Heatmap dargestellt. Abb. 2.3 zeigt zwei Beispiele einer solchen Heatmap [107].



Abb. 2.3 | Grad-CAM Beispiel. Links ist das Originalbild zu sehen. Daneben sind zwei Heatmaps platziert, die mit Grad-CAM die Wichtigkeit räumlicher Regionen des Bilds für die Klassen „Katze“ und „Hund“ visualisieren.

Mithilfe der extrahierten Gradientenkarten lassen sich neben der Wichtigkeit bestimmter räumlicher Regionen der Eingangsdaten auch die relative Wichtigkeit der Eingangskanäle bestimmen. Durch die Integration des Gradientenfelds werden stabilere und interpretierbare Aussagen möglich. Die Integration gleicht lokale Schwankungen der Gradienten aus. Mithilfe der Kosinusähnlichkeit, die den Winkel zwischen zwei Vektoren im Merkmalsraum beschreibt und damit die Übereinstimmung ihrer Richtungen quantifiziert, werden die integrierte und die nicht integrierte Karte auf Konsistenz geprüft. Eine hohe Kosinusähnlichkeit weist darauf hin, dass beide Karten auf ähnliche Eingabemuster reagieren

und das Modell intern konsistente Repräsentationen der relevanten Merkmale lernt.

Eine weitere Möglichkeit, die Effizienz eines Klassifikators zu visualisieren, ist ein Scatterplot, der den Merkmalsraum, in den der Klassifikator die Eingangsdaten abbildet, niederdimensional darstellt. Im Scatterplot werden Stichproben aus verschiedenen Klassen eingetragen. Die Clusterbildung im Scatterplot visualisiert, wie gut die Klassen im Merkmalsraum getrennt sind, und liefert eine Schätzung der Güte des Klassifikators auf den vorliegenden Daten. Um den Merkmalsraum in eine visualisierbar niedrige Dimension zu projizieren, wird häufig die t-Distributed Stochastic Neighbor Embedding (**t-SNE**) eingesetzt [108, 109]. Die **t-SNE** ist eine nichtlineare Methode zur Dimensionsreduktion, speziell zur Visualisierung hochdimensionaler Daten in zwei oder drei Dimensionen. Die Methode modelliert die Abstände zwischen den Datenrepräsentationen im hochdimensionalen Raum als Wahrscheinlichkeitsverteilungen und definiert eine Transformation in eine niederdimensionale Repräsentation mit den entsprechenden Wahrscheinlichkeitsverteilungen. Mit einem Optimierungsalgorithmus minimiert die **t-SNE** die Kullback-Leibler-Divergenz der beiden Wahrscheinlichkeitsverteilungen, um die Entfernung der Datenpunkte in beiden Räumen möglichst ähnlich zu machen. Abb. 2.4 zeigt exemplarisch zwei solcher Scatterplots für den MNIST-Datensatz [110]. Die Scatterplots visualisieren den Merkmalsraum des gleichen Klassifikators nach einer Epoche des Trainings (links) und nach 25 Epochen des Trainings (rechts). Links sind die Klassen-Cluster schlechter getrennt als rechts.

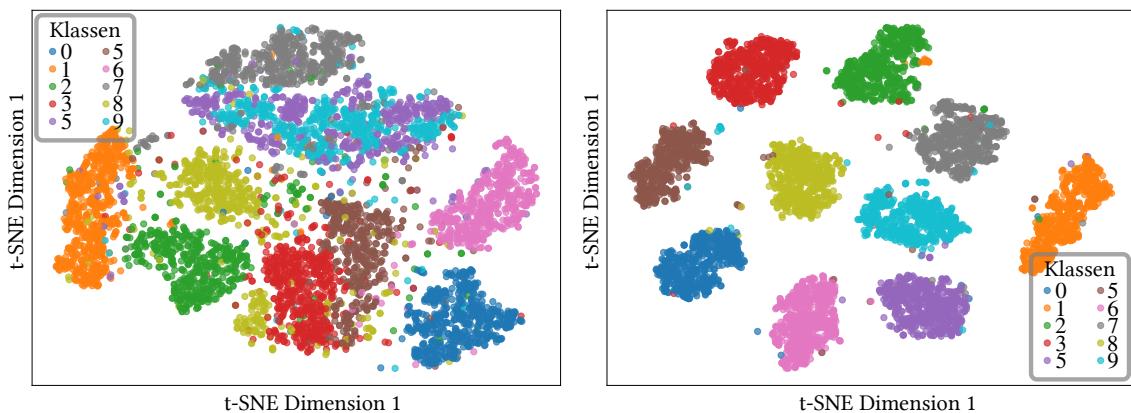


Abb. 2.4 | Zwei Scatterplots. Beide Scatterplots visualisieren den Merkmalsraum eines Klassifikators, links nach der ersten Epoche des Trainings, rechts nach 25 Epochen. Mithilfe der **t-SNE** wird die hochdimensionale interne Repräsentation der Merkmale in eine zweidimensionale Ansicht projiziert. Einige Stichproben des MNIST-Datensatzes sind in dieser zweidimensionalen Ansicht platziert und ihre Klassen sind farblich markiert.

Dreidimensionale Daten stellen eine besondere Herausforderung für die Klassifikation dar [111]. Vortrainierte Methoden der 2D-Klassifikation wie **CNN**-Encodern können auf einzelne Schichten eines 3D-Volumens angewandt werden, und die extrahierten Merkmale können anschließend aneinandergereiht werden, aber explizite 3D-Methoden sind oft besser [112]. Die Methoden können auch an dreidimensionale Daten angepasst werden,

indem ihre 2D-Operationen auf drei Dimensionen ausgeweitet werden, wodurch der Rechenaufwand steigt [113]. Angepasste Methoden wie 3D-CNNs sind nicht in der Lage, Beziehungen zu erfassen, wenn die relevanten Bildregionen räumlich weit voneinander entfernt sind und schräg in den drei Dimensionen verteilt sind [114]. Um dieses Problem zu lösen, schlägt die Literatur self-attention-Mechanismen über die Z-Dimension vor, die auch nicht-lokale Informationen erfassen [115, 116].

2.3 Literaturrecherche

2.3.1 Benchmark

Da das Annotieren von Zelldaten für die Segmentierung mit erheblichem manuellen Aufwand verbunden ist und zusätzlich Expertenwissen voraussetzt, sind Datensätze hierfür selten. Einige prominente Datensätze mit Annotationen für eine Instanzsegmentierung, deren Domänen zu den Zieldaten der Anwendung dieser Arbeit *ähnlich* sind, sind:

- LiveCell [117], ein manuell annotierter und von Expert*Innen validierter Datensatz aus 5.239 2D-Bildern. Die Daten sind mit Phasenkontrastmikroskopie gesammelt und enthalten 1.686.352 individuelle Zellen von acht verschiedenen Zelltypen.
- YeaZ [118], ein zweiteiliger Datensatz von Hefe-Zellen aus 87 Phasenkontrast-Bildern mit insgesamt 10.422 Zellen und 614 Hellfeld-Bildern mit insgesamt 3.841 Zellen in 6 Beleuchtungsstufen aufgenommen. Die Annotationen sind semi-maniell erstellt, da die Phasenkontrast-Bilder manuell und die Lichtfeld-Bilder aus den Phasenkontrast-Segmentierungsmasken annotiert wurden.
- DeepBas [119, 120], ein Datensatz von *B. subtilis strain SH130* Bakterien. Er besteht aus Weitfeldaufnahmen (Fluoreszenz), aufgenommen mit einem inversen Mikroskop, bestehend aus sieben manuell annotierten Bildern mit je 46 bis 335 Zellen.
- die Cell Tracking Challenge [121], eine Sammlung aus 13 Datensätzen verschiedener Mikroskopiemodalitäten, die sich zur Messung der Segmentierungs- und Verfolgungsfähigkeiten für verschiedene Zelltypen eignen.
- MoNuSeg [122], eine Zusammenstellung manuell annotierter Gewebeschnitte aus sieben verschiedenen Organen. Über 21.000 Zellen sind pro Bild in den 30 Bildern mit verschiedenen Färbungen und Aufnahmetechniken verteilt.
- TissueNet [123] ist ein umfassender Datensatz mit über 1.000.000 Zellen aus verschiedenen Gewebearten und unterschiedlichen Aufnahmetechniken.
- S_BIAD1518 [124, 125], ein Datensatz, der neben manuell annotierten Bildern von acht verschiedenen Zellarten synthetisch erzeugte Daten enthält. Mithilfe von SpCycleGAN [126] wurden dazu auf Basis simulierter Annotationen Bilder generiert, die anstreben, die Merkmale der realen Bilder zu reproduzieren. Es handelt sich um 3D-Multispektraldaten, aufgenommen mittels Fluoreszenzbildgebung.

Aufgrund des geringen Volumens an frei zugänglichen Daten sind diese Sammlungen auch für das Training von Segmentierungsnetzen begehrte. Neben annotierten Datensätzen bietet die Literatur auch Methoden zum eigenständigen Erzeugen domänenpezifischer Datensätze [127, 128, 129, 130, 131]. Beispielsweise können 3D-Trainingsdaten mit realistischer Zellform und -ausrichtung sowie umgebenden Markern synthetisch erzeugt und durch ein Generative Adversarial Network an eine gewünschte Bilddomäne angepasst werden [132].

2.3.2 Segmentierungsmodelle

Foundation-Modelle sind für viele moderne KI-Anwendungen unerlässlich [133]. Sie werden zunächst für allgemeine Aufgaben vortrainiert und anschließend auf spezifische Anwendungen angepasst (fine-tuning), meist unter Einfrieren von Teilen der Gewichte [134]. Auch Segmentierungsmodelle profitieren stark von umfangreichem Vortraining [135]. In der aktuellen Forschung werden verschiedene Foundation-Modelle zur Segmentierung angewandt [136, 137, 138, 139]. Ein prominentes Exemplar ist das Segment Anything Model (**SAM**) von Meta AI [140] (siehe Abb. 2.5). Es besteht aus einem Bild-Encoder, einem Prompt-Encoder und einem Masken-Decoder.

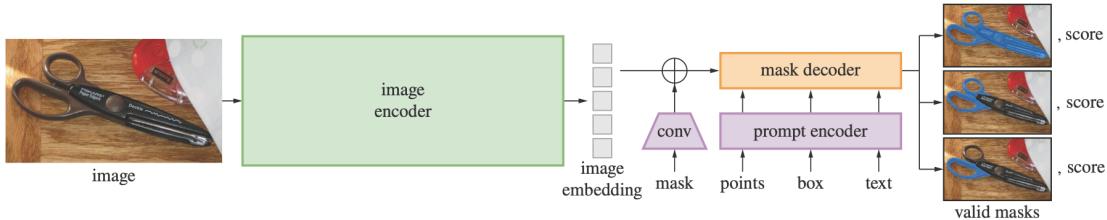


Abb. 2.5 | Architektur des **SAM**-Modells. Eingabebilder werden durch einen Bildencoder in Repräsentationen umgewandelt. Zusätzliche optionale Hinweise zum zu segmentierenden Objekt werden durch Bildfaltungen oder einen Prompt-Encoder repräsentiert. Anschließend prädiziert der Decoder mehrere mögliche Masken und die zugehörigen Zuversichtlichkeiten [140].

Als Bild-Encoder dient ein Vision Transformer [141] mit Vortraining als Masked Autoencoder [142] und zusätzlichem Training für höhere Bildauflösung. Der Prompt-Encoder ist mehrstufig. Ein angelernter Positional Encoder generiert Repräsentationen aus Positions-Nutzereingaben wie Punkten und Boxen. Für textuelle Prompts wird der Encoder des CLIP-Modells [143] verwendet. Außerdem werden Bildfaltungen als Encoder auf Maskeneingaben angewandt. Mithilfe dieser Encoder wird dem Modell eine Repräsentation des zu segmentierenden Bildes sowie optional Hinweise auf das erwünschte Ergebnis bereitgestellt, die bereits semantische Informationen und abstrakte Bildmerkmale enthalten. Aus diesen Repräsentationen generiert der Decoder anschließend mehrere mögliche Masken mit zugehörigen Zuversichtlichkeiten, aus denen ein finales Segmentierungsergebnis ausgewählt wird.

Das **SAM**-Modell wurde bereits für viele explizite Mikroskopie-Zelldaten-Anwendungen

angepasst [144, 145, 146]. Auch für bestehende biologische Segmentierungsanwendungen, wie etwa Cellpose[147], wurde **SAM** auf Zelldaten angepasst [148]. Dieser *Fine-tune* nennt sich CellposeSAM. Er kombiniert den Bild-Encoder von **SAM** mit dem *Flow*-Segmentierungsansatz von Cellpose. Dabei generiert der Bild-Encoder direkt Vektoren, die Zwischenrepräsentationen, sogenannte *Flows*, darstellen. Diese *Flows* werden pixelweise zu einem Gradientenfeld überführt. Mithilfe der Gradienten werden Objektinstanzen vorhergesagt. Abb. 2.6 zeigt diesen Ablauf als Diagramm.

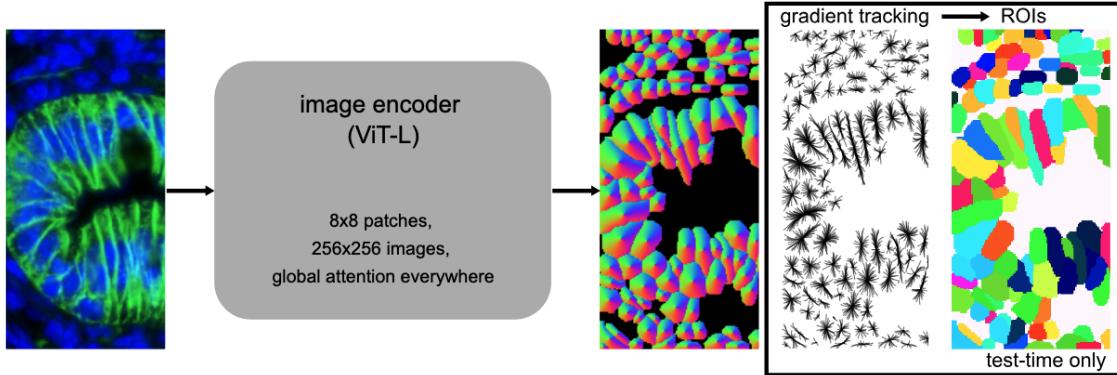


Abb. 2.6 | Ablauf des CellposeSAM-Modells. Eingabebilder werden durch einen Bild-Encoder (ViT-L) direkt zu sogenannten *Flows* umgeformt, einer Repräsentation von vorhergesagten Objektmerkmalen, deren Werte von der relativen Position innerhalb des detektierten Objekts abhängen. Die Gradienten der *Flows* werden verfolgt (gradient tracking) und aus dem daraus entstehenden Gradientenfeld werden Segmentierungsinstanzen (ROIs) vorhergesagt [148].

Deepcell [149, 150, 151] bietet weitere Zellsegmentierungsmodelle. Das Deepcell-Caliban-Modell [152] nutzt als Encoder eine EfficientNetV2L-Architektur [153], an deren Ausgabeschichten eine Pyramidenstruktur zur Merkmalsfusion angeschlossen ist. Eine Besonderheit des Netzes ist, dass Eingabebildern zusätzlich Koordinatenkarten hinzugefügt werden. Als Decoder dienen drei Segmentierungsköpfe, die verschiedene Transformationen der annotierten Trainingsmasken vorhersagen.

In der Literatur ist außerdem das nnU-Net [154] sehr verbreitet, ein Segmentierungsframework, das sich automatisch an neue biomedizinische Aufgaben anpasst. Es konfiguriert Vorverarbeitung, Netzwerkarchitektur, Training und Nachbearbeitung dynamisch auf Basis der Eigenschaften des jeweiligen Datensatzes. Die Leistungsfähigkeit des Ansatzes ergibt sich nicht aus einer neuen Architektur oder einer neuen Lernmethode, sondern aus der konsequenten Automatisierung und Systematisierung von Entwurfsentscheidungen.

2.3.3 Klassifikator

Für Klassifikatoren werden in der Regel nur Encoder vorgenommen. Der Klassifikations-Kopf muss an die Klassen des vorliegenden Problems angepasst werden [92, 134, 155]. State-of-the-Art für Bild-Encoder sind **CNNs** oder **ViTs**, die auf dem ImageNet-Datensatz [156] vorgenommen werden [157, 158]. Klassifikatoren profitieren stark von ImageNet-Vortraining

[159, 160].

ResNet ist ein Residual Neural Network, ein CNN mit sogenannten residual connections [161]. Diese residual connections verbinden den Ein- und Ausgang modularer Faltungsschichten und verbessern die Leistungsfähigkeit tiefer neuronaler Netze [161, 162, 163] (siehe Abb. 2.7 a). In ihrem Paper stellen die Autoren fünf unterschiedlich tiefe Varianten der **ResNet**-Architektur vor. Jede Variante enthält fünf Blöcke mit residual connections. Die Blöcke bestehen aus Faltungen mit verschiedenen Kernelgrößen und Strides, batch normalization [164] und der ReLU [165] Aktivierungsfunktion.

EfficientNet V2 ist ein Nachfolger der EfficientNet-Modellfamilie [153, 166]. Die Architektur basiert auf modularen Blöcken von Bildfaltungsoperatoren mit besonders kleinen Faltungskernen und Squeeze-and-Excitations, genannt MBConv [166, 167] und Fused-MBConv [168] (siehe Abb. 2.7 b).

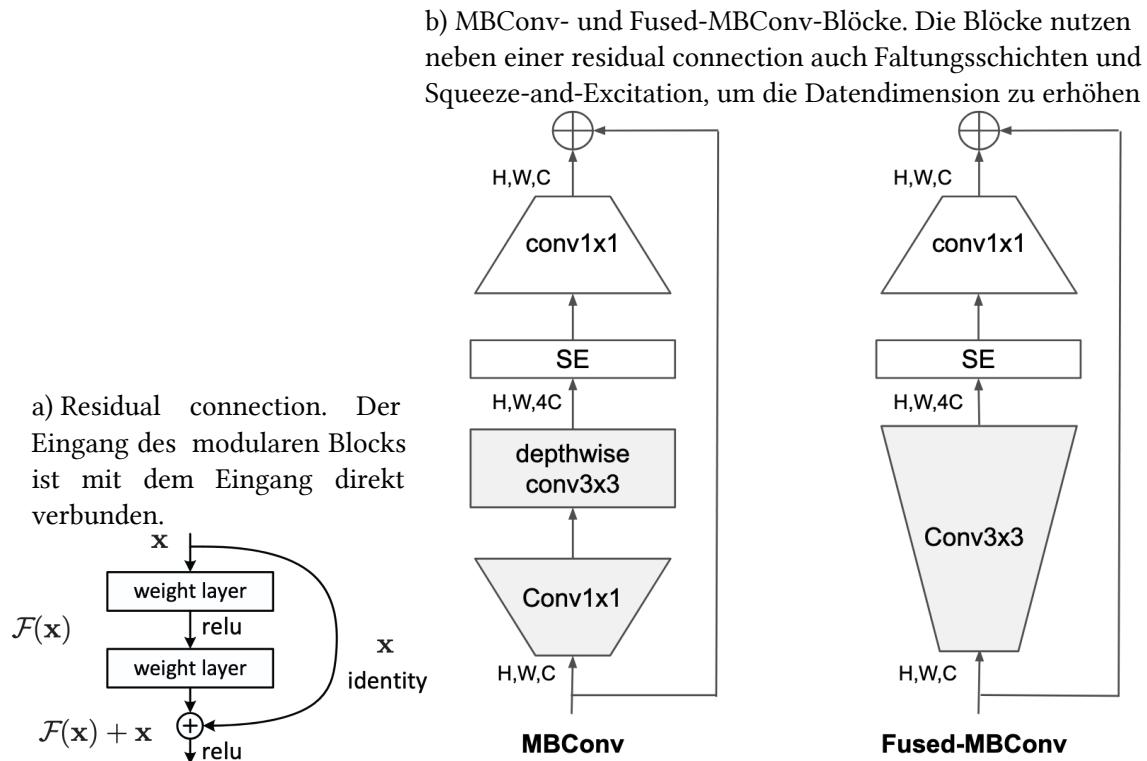


Abb. 2.7 | Diagramme a) der Residual Connections [161] und b) der MBConv-Blöcke und der Fused-MBConv-Blöcke [153].

ConvNeXt [169] ist eine CNN-Modellfamilie mit besonders großen Faltungskernen. Die **ConvNeXt**-Architektur umfasst fünf modulare Blöcke mit Faltungen und residual connections, wie die ResNet-Architektur [161]. Allerdings verändert dabei **ConvNeXt** einige Details der ResNet-Architektur, wie beispielsweise die GeLU-Aktivierungsfunktion [170] und layer normalization [104].

Swin Transformer [24] ist eine beliebte Modellfamilie der ViTs. Ihr Nachfolger, die **Swin Transformer V2**-Familie [171], vergrößert die Modelle weiter. Die Architektur kombiniert Bildausschnitte mit einem Positionsbias. Hierzu werden ein Bildfenster z und dessen relative Koordinaten im Bild, Δx und Δy , in einem attention-Mechanismus zusammengeführt. Die Positionen werden in einem MLP-Netz verarbeitet, während das Bildfenster mit drei verschiedenen Gewichtsmatrizen multipliziert wird. Mithilfe einer Kosinusähnlichkeitsfunktion, der Softmax-Funktion [172], der elementweisen Multiplikation sowie der Addition werden diese Ergebnisse in einen Merkmalsraum überführt. Zwei layer normalizations [104], ein weiteres MLP-Netz und residual connections vervollständigen anschließend den modularen **Swin Transformer V2**-Block. Dieser Aufbau ist in Abb. A.1 im Anhang dargestellt. Der Abschnitt A.1 des Anhangs zeigt die Architektur als Diagramm.

2.4 Offene Probleme

Einzelne Myotuben lassen sich nicht mithilfe eines Segmentierungsmodells aus der Literatur instanzsegmentieren. Selbst für Expert*Innen sind in dichten Strukturen Myotuben-Instanzen nicht immer eindeutig trennbar. Nicht viele Segmentierungsmodelle für Nuclei sind erhältlich, insbesondere für dreidimensionale Daten. Die verfügbaren Modelle verhalten sich je nach Datensatz unterschiedlich, und ihre Eignung für bestimmte Aufgaben muss in jeder Anwendung individuell geprüft werden. Die Daten der vorliegenden Arbeit wurden gemäß dem Protokoll in Couturier et al. [16] erstellt und umfassen Myotubenkulturen sowie deren Nuclei und sind mit insgesamt fünf verschiedenen Fluoreszenzmarkern versehen. Für diese spezifischen Aufnahmebedingungen und Marker der vorliegenden Daten gibt es keinen angepassten Klassifikator. Der Erfolg eines Übertrags verschiedener vortrainierter Encoder und etablierter Methoden auf die vorliegenden Daten ist unvorhersehbar, da sich die gelernten Merkmalsräume möglicherweise nicht für die Klassifikation der neuartigen dreidimensionalen Daten eignen. Eine weitere Fragestellung ist deshalb, wie Klassifikationsmethoden mit der Kombination der Marker umgehen. Für jeden Datensatz mit neuen Zellkernklassen und Aufnahmebedingungen muss nicht nur ein neues Modell trainiert, sondern auch Methoden- und Hyperparameteroptimierung durchgeführt werden, um optimale Klassifikatorleistung zu erzielen. Des Weiteren ist der Umgang mit dreidimensionalen Daten, insbesondere in Umgebungen mit geringer Rechenleistung, ein offenes Problem. Diverse Lösungen existieren, um dreidimensionale Daten mit Expertenwissen zu versehen. Diesen Lösungen fehlt bislang ein Arbeitsablauf, der Daten unmittelbar segmentiert und vorbereitet, um relevante Bildausschnitte direkt aus dem Datensatz zu extrahieren, sodass Expert*Innen ausschließlich annotieren müssen.

2.5 Zielsetzung

Im Zuge der vorliegenden Arbeit soll die automatische Extraktion interpretierbarer Eigenarten der Myotubenkulturen ermöglicht werden. Zu diesem Ziel definiert die vorliegende Arbeit folgende Zwischenziele:

- Es soll ein Segmentierungsmodell gefunden werden, das die Eigenschaften wie Zellkernvolumina und die lokale Nucleidichte möglichst unverändert aus den vorliegenden, dreidimensionalen Daten extrahieren kann. Dazu wird ein neues Bewertungskriterium für die Instanzsegmentierung eingeführt und auf einige etablierte Modelle angewandt.
- Das Segmentierungsmodell soll dann genutzt werden, um einen Ablauf zu schaffen, in dem Expert*Innen die Klassen der Nuclei besonders zeiteffizient annotieren können, um einen Klassifikator zu trainieren. Durch das Anreichern dreidimensionaler Daten mit den Segmentierungsmasken und das automatische Fokussieren einzelner Nuclei sollen sowohl die Rechenzeit optimiert werden als auch der Aufwand, einzelne Nuclei entlang drei Dimensionen zu suchen, eliminiert werden. Hierzu wird eine neue Anwendung entwickelt, die 3D-Zelldaten liest und anschließend eine Oberfläche zur Annotierung bereitstellt.
- Die entstehenden Annotationen sollen direkt in einen Trainingsablauf für Klassifikatoren integriert werden. Hierbei müssen Klassifikatoren mit dreidimensionalen Daten variabler Tiefe umgehen können. Ein ausführlicher Methodenvergleich verschiedener Encoder, Klassifikations-Köpfe, Vorverarbeitungsmethoden sowie Vortrainingsmethoden soll für Nutzer*Innen ohne Programmierkenntnisse ermöglicht werden. Dazu wird eine neue Anwendung entwickelt, die die erstellten Annotationen und Segmentierungsmasken nutzt, um einen Klassifikator zu trainieren. Nutzer*Innen können in einer grafischen Oberfläche verschiedene Methoden zum Vergleich auswählen, und in einem automatisierten Prozess werden Klassifikatoren aller Kombinationen trainiert und verglichen.
- Außerdem sollen die ausgelesenen Eigenschaften der eingegebenen Zelldaten leicht zugänglich sein. In einer weiteren neuen Anwendung werden automatisch die Voraussagen des Klassifikators mit dem besten Ergebnis im Methodenvergleich genutzt, um Nutzer*Innen Graphen mit den Eigenschaften der Zellkultur darzubieten.
- Zuletzt sollen alle neu entwickelten Module zu einer Gesamtanwendung zusammengefasst und getestet werden. In einer Parameterbestimmung wird das Optimum für die vorliegenden Aufnahmen explizit mithilfe der neuen Anwendung bestimmt.

Neues Konzept 3

3.1 Überblick

Das nachfolgende Kapitel beschreibt und diskutiert im Detail das angewandte Konzept der vorliegenden Thesis. Es behandelt die selbstentwickelten Beiträge zu den Methoden sowie die vorliegenden dreidimensionalen Bildstapel der Myotubenkulturen. Auf die in Abschnitt 3.2 beschriebenen Daten werden die eingeführten Methoden angewandt, um deren Effektivität zu demonstrieren. In Abschnitt 3.3 wird ein neues Bewertungskriterium für 3D-Instanzsegmentierungsmodelle eingeführt. Es bewertet Modelle im Hinblick auf ihre Eignung, interpretierbare Merkmale von Nuclei unverändert zu extrahieren. Abschnitt 3.4 führt Methoden zur Klassifikation von 3D-Daten ein. Diese Methoden passen den Klassifikator an verschiedene Eigenschaften eines Datensatzes an.

Die gesamte Methodik wird in einer modularen Anwendung umgesetzt, die 3D-Daten als Eingabe annimmt und interpretierbare Eigenschaften, wie die Verteilung der Klassen und das Volumen der anwesenden Nuclei, ausgibt. Diese Anwendung wird hier als 3D-Zelldaten-Pipeline bezeichnet. Abb. 3.1 bietet einen Überblick über die Methodik. Die Anwendung kann regulär angewandt (Inferenz) oder optimiert werden (Optimierung). Zur Optimierung werden die 3D-Daten einem Ablauf für den Vergleich der Segmentierungsmodelle oder der Klassifikatormethoden zur Verfügung gestellt. Zuerst wird das neu eingeführte Bewertungskriterium für Segmentierungsmodelle Injektive Panoptische Qualität ([IPQ](#)) für verschiedene Modelle berechnet (siehe Abschnitt 3.3). Dieses Bewertungskriterium quantifiziert die Eignung der Segmentierungsmodelle zur Extraktion der gewünschten Zellkerneigenschaften. Mithilfe der [IPQ](#)-Werte wird das beste Segmentierungsmodell für die Anwendung gewählt. Die Architektur und Parameter des optimalen Modells werden in den Inferenzablauf eingesetzt.

Um die Klassifikatormethoden zu optimieren, werden die 3D-Daten und die extrahierten Segmente in die neu entwickelte Labeling-App eingegeben. Die Labeling-App ermöglicht die zeiteffiziente Annotation von Nuclei, indem relevante Bildausschnitte automatisch anhand der Segmente extrahiert werden. Mit den erstellten Annotationen und den 3D-Daten werden iterativ verschiedene Klassifikatoren trainiert. Diese Klassifikatoren ergeben sich aus Kombinationen der verfügbaren Klassifikatormethoden. Die Klassifikatormethoden werden in Abschnitt 3.4 beschrieben und umfassen 1. Encoder-Architekturen, 2. Klassifikations-Kopf-Architekturen, 3. Vorverarbeitungsmethoden und 4. Vortrainingsmethoden. Unter den Methoden sind sowohl etablierte als auch neu entwickelte Ansätze. Anhand der Genauigkeit der trainierten Klassifikatoren auf einem separaten Validierungs-

anteil des Datensatzes werden die Methoden verglichen und eine optimale Konfiguration ausgegeben. Die Architektur und Parameter dieser Konfiguration werden dann in den Inferenzablauf der Anwendung eingesetzt.

Am Ende des Ablaufs werden die klassifizierten Segmente verwendet, um verschiedene Grafiken zu erzeugen, die interpretierbare Eigenschaften visualisieren.

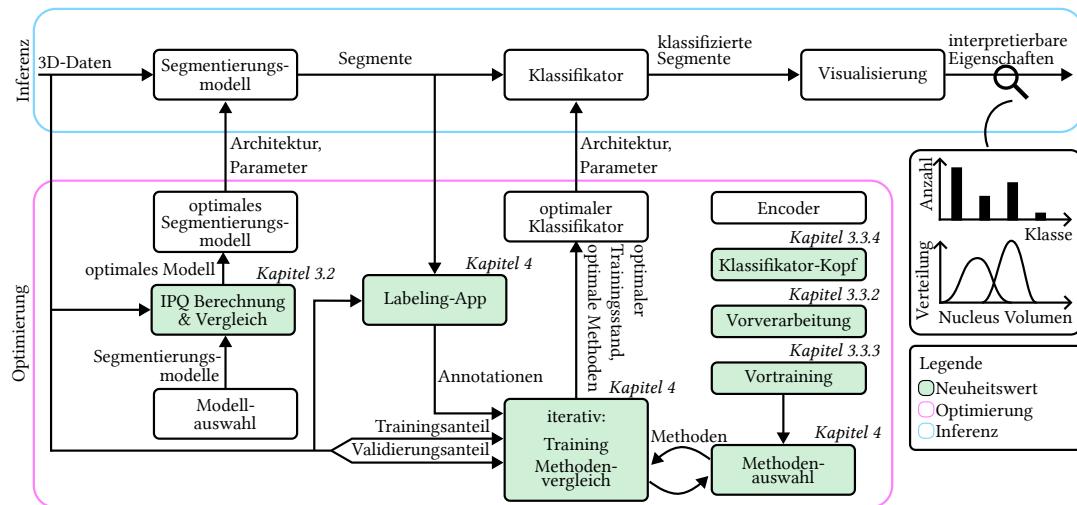


Abb. 3.1 | Methodik der vorliegenden Arbeit. Die Anwendung kann regulär angewandt (Inferenz) oder optimiert werden (Optimierung). Die Optimierung ist zweiteilig. Zur Optimierung der Segmentierungsmodele wird das neu entwickelte Injektive Panoptische Qualität (IPQ)-Bewertungskriterium eingesetzt. Für den Klassifikator werden Kombinationen verschiedener Encoder, Klassifikations-Köpfe, Vorverarbeitungsmethoden und Vortrainingsmethoden iterativ trainiert und verglichen.

3.2 Daten

Die zugrunde liegenden Bilddaten der vorliegenden Arbeit stammen aus in-vitro-Kulturen von Myotuben, die aus humanen induzierten pluripotenten Stammzellen (hiPSC) differenziert wurden. Diese Zellen wurden nach dem in Couturier et al. [16] beschriebenen Protokoll hergestellt und anschließend mittels Immunfluoreszenzfärbung markiert. Von den Zellen wurden 3D-Bildstapel mit voxelbasierten Intensitätswerten, die jeweils einem Fluoreszenzkanal zugeordnet sind, mit einem Konfokalmikroskop aufgenommen. Die Kulturen enthalten Myotuben und deren Nuclei in fünf Fluoreszenzkanälen, die entsprechend mit fünf unterschiedlichen Fluoreszenzmarkern angefärbt wurden:

- DAPI (Nuclei),
- α -Actinin (sarcomerisches Strukturprotein),
- Dystrophin (Membran-assoziiertes Muskelprotein),

- Synaptophysin (präsynaptisches Vesikelprotein),
- α -Bungarotoxin (Bindung an nikotinische Acetylcholinrezeptoren, nAChR) und
- S100 β (Schwannzellmarker).

Jeder dieser Marker färbt einen anderen Zellbestandteil an, sodass unter dem Mikroskop sichtbar ist, wo sich Kerne, Muskelfasern und synaptische Strukturen befinden. Die Bildaufnahme erfolgte mit einem Leica TCS SP8 Konfokalmikroskop, unter Verwendung von 405-nm-, 488-nm-, 561-nm- und 633-nm-Lasern und einem 20x-Objektiv. Je Kanal ergibt sich eine räumliche Auflösung von 1024 x 1024 Pixeln mit 568 nm pro Pixel. Die Anzahl der Z-Schichten beträgt 24 bis 64. Die Proben wurden in Matrigel eingebettet und unter physiologisch relevanten Kulturbedingungen (37 °C, 5% CO₂) im maturierenden Medium kultiviert, das u. a. Wachstums- und Differenzierungsfaktoren wie BDNF, GDNF, IGF-1, CHIR99021, DMH1, SB431542, Retinsäure und Purmorphamin enthielt. Das resultierende Material stellt ein menschliches in-vitro-Modell der neuromuskulären Verbindung dar. Abb. 3.2 zeigt exemplarisch je eine 2D-Schicht von einem der Marker-Kanäle.

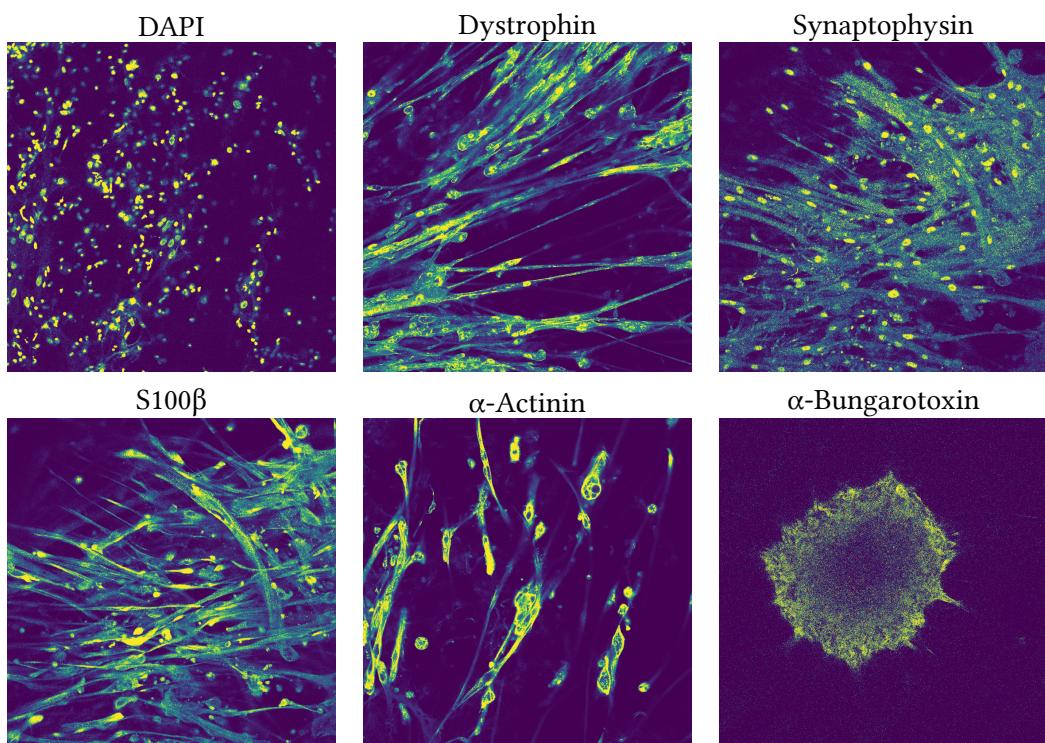


Abb. 3.2 | Beispiele von 2D-Schnitten der verfügbaren Färbungen der vorliegenden 3D-Bildstapel.

Die eingeführten Methoden werden als Experiment auf diese Daten angewandt, sie sind aber explizit entworfen, um Anpassungen an neue Datensätze nahtlos zu ermöglichen. Alle Experimente auf diesen Daten dienen der vorliegenden Arbeit als Fallstudie, um die Effektivität der vorgestellten Methoden zu demonstrieren.

3.3 Neues Kriterium: Injektive Panoptische Qualität

Zur Wahl des Segmentierungsmodells wird das neue Bewertungskriterium Injektive Panoptische Qualität (IPQ) eingeführt. Es besteht aus drei Faktoren, die jeweils eine Fehlerart bei Instanzsegmentierungsmasken prüfen. Als Datensatz wird aus den in Abschnitt

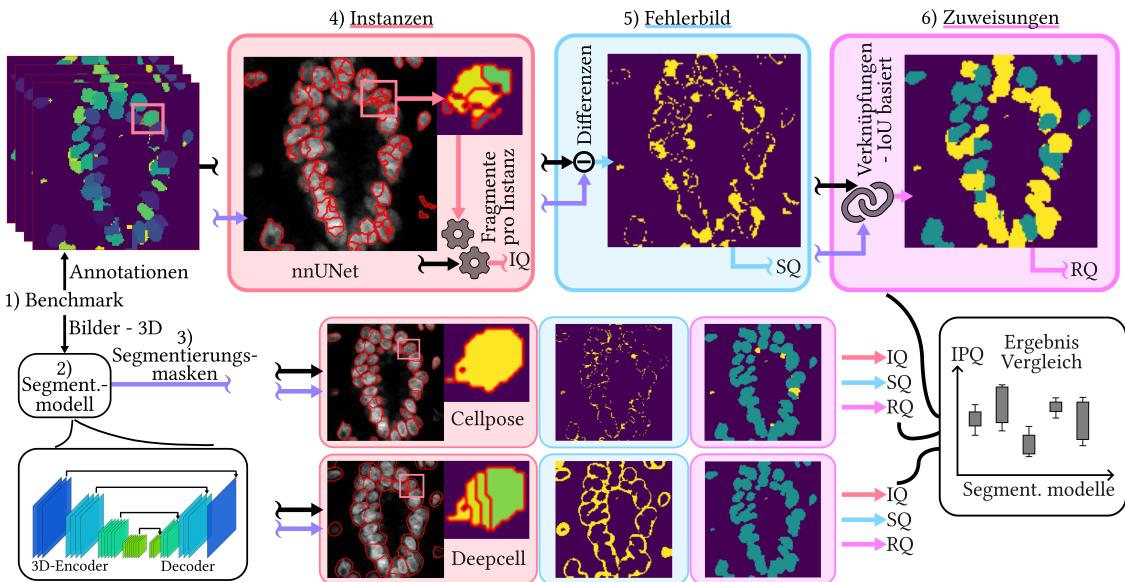


Abb. 3.3 | IPQ Visualisierung - Das Ablaufdiagramm stellt den Prozess dar, durch den das Segmentierungsmodell gewählt wird, das für die Anwendung der vorliegenden Arbeit eingesetzt wird. Ein peer-reviewed Benchmark-Datensatz aus dreidimensionalen Bildern (1) von diversen Zellkulturen mit den dazugehörigen Annotationen wird links eingegeben. Das zu bewertende Segmentierungsmodell (2) führt für die Bilder des Benchmarks eine Inferenz durch, um Segmentierungsmasken (3) bereitzustellen. Die entstehenden Masken werden anschließend zur Berechnung der neu eingeführten IPQ (siehe Abschnitt 3.3) eingesetzt. Der Ablauf der Bewertung ist in drei Phasen gegliedert. Aus jeder Maske werden zuerst die einzelnen Fragmente extrahiert, die sich mit einer einzelnen Instanz der Annotation überlagern (4). Außerdem wird ein Fehlerbild durch Berechnung der Intersection over Union (IoU) hergestellt, hier als logische XOR-Fläche der Maske und der Annotation als Platzhalter dargestellt (5). Zuletzt wird ein Zuweisungsbild erstellt, das die True Positives (TPs), False Positives (FPs) und False Negatives (FNs) festhält (6). Aus diesen drei Teilen wird jeweils ein Faktor zur Berechnung der IPQ gebildet. Durch Vergleiche der Ergebnisse verschiedener Modelle lässt sich dann das optimale Segmentierungsmodell für die Anwendung wählen.

2.3.1 vorgestellten Benchmarks der S_BIAD1518-Datensatz [124, 125] verwendet, da dieser nicht in den Trainingsdaten eines zu testenden Segmentierungsmodells vorkommt. Der Datensatz verfügt über manuelle Annotationen für Instanzsegmentierungsmasken und ist durch synthetisch generierte Daten und Masken erweitert. Im Gegensatz zu selbstentwickelten synthetischen Daten weicht die Bilddomäne dieses Benchmarks zwar stärker von der Domäne der Zieldaten ab, aber dafür sind die Daten an eine Veröffentlichung mit standardisiertem Peer-Review-Prozess gebunden.

Die Aufgabe des Bewertungskriteriums ist es, zu quantifizieren, wie gut sich eine vorliegende Instanzsegmentierung eignet, um reale Eigenschaften einer Aufnahme, wie die Zellkernanzahl, die Größe der Zellkerne und die lokale Zellkerndichte auszuwerten. Das neu

entwickelte Kriterium ist eine Erweiterung der Panoptic Quality (PQ) [45]. Durch die PQ wird die IoU für individuelle Instanzen bewertet und es werden False Positive (FP) sowie False Negative (FN) Detektionen bestraft. Zusätzlich werden durch einen neuen Faktor Verletzungen der injektiven Abbildung von segmentierten Nuclei auf die Instanzen der Annotation negativ bewertet werden, da die genaue Anzahl der Nuclei und die räumliche Dichte der Nuclei eine bedeutungsvolle Metrik für die Auswertung sind. Zur Berechnung wird im ersten Schritt der folgende Brute-Force-Algorithmus angewandt, der die Zuordnung von Segmentierungsinstanzen zu Annotationsinstanzen durchführt.

Algorithm 1 Brute-Force-Annotation-Zuordnung für jede Segmentierungsinstanz

Eingabe: $maske_{Vorhersage}$, $maske_{Annotation}$

Ausgabe: $annotation_{opt}$, IoU_{opt}

Für $id_{Instanz}$ in $|maske_{Vorhersage}|$ true:

$Instanz \leftarrow maske_{prediction}[id_{Instanz}]$

Für $annotation$ in $maske_{Annotation}$ true:

$IoU \leftarrow IoU(annotation, Instanz)$

Wenn $IoU > IoU_{opt}[id_{Instanz}]$ dann:

$IoU_{opt}[id_{Instanz}] \leftarrow iou$

$annotation_{opt}[id_{Instanz}] \leftarrow annotation_id$

EndeWenn

EndeFür

EndeFür

Rückgabe $annotation_{opt}$, IoU_{opt}

Die nachfolgende Formel zeigt das IPQ-Bewertungskriterium unterteilt in die 3 Faktoren Segmentation-Quality (SQ), Recognition-Quality (RQ) und die neu eingeführte Injective-Quality (IQ):

$$IPQ = \underbrace{\frac{k_1 \times \sum_{(p,g) \in TP} \text{IoU}\left(\bigcup_{p_i \in p} p_i, g\right)}{|TP|}}_{\text{Segmentation-Quality (SQ)}} \times \underbrace{\frac{k_2 \times |TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{Recognition-Quality (RQ)}} \times \underbrace{\frac{k_3 \times |GT|}{\sum_{p \in P} (\max(1, n_p - 1))}}_{\text{Injective-Quality (IQ) (neu)}}, \quad (3.1)$$

wobei:

- k_1, k_2, k_3 Optionale Vorfaktoren zur Gewichtung der drei Teile der Metrik sind,
- TP die Menge aller (p, g) -Tupel ist, wobei g eine Annotationsinstanz und p der Vektor aller zugehörigen Segmentierungsinstanzen ist,

- $|TP| \in \mathbb{Z}$ die Anzahl der korrekt erkannten Instanzen bezeichnet, also Annotationsinstanzen mit $\text{IoU} > 0,5$,
- $\text{IoU}(\bigcup_{p_i \in p} p_i, g) \in [0, 1]$ die IoU zwischen allen Segmentierungsinstanzen p_i in der TP -Instanz p und der zugehörigen Annotationsinstanz g beschreibt,
- $|FP| \in \mathbb{Z}$ die Anzahl der falsch-positiven Segmentierungen ist, d. h. vorhergesagte Instanzen ohne Annotationsentsprechung,
- $|FN| \in \mathbb{Z}$ die Anzahl der nicht erkannten Annotationsinstanzen ist, also Annotationsinstanzen ohne zugehörige Vorhersage,
- $|GT| \in \mathbb{Z}$ die Anzahl der Annotationsinstanzen ist,
- P die Menge aller Segmentierungsinstanzen ist, ungeachtet der Annotationszuordnung,
- $p \subseteq P$ ein Vektor aller Segmentierungsinstanzen, die der gleichen Annotationsinstanz zugeordnet sind, ist,
- n_p die Dimension des Vektors p ist,
- $SQ \in [0, 1]$ ein Faktor ist, der die Qualität der Segmentierung anhand der IoU von der segmentierten und der erwarteten Instanz vergleicht,
- $RQ \in [0, 1]$ ein Faktor ist, der bewertet, wie vollständig und fehlerfrei das Segmentierungsmodell die vorhandenen Nuclei findet und, ob es dabei zu Halluzinationen kam,
- $IQ \in [0, 1]$ ein Faktor ist, der das Unterteilen von Nuclei durch das Segmentierungsmodell zu bestrafen. Wird ein Nucleus durch mehrere Instanzen der Segmentierungsmaske dargestellt, wird n_p größer als eins und der Faktor sinkt,
- $IPQ \in [0, 1]$ ein Maß für die panoptische Segmentierungsqualität mit der Voraussetzung von injektiver Abbildung der Segmentierungsmasken-Instanzen auf die Annotationsinstanzen darstellt, wobei höhere Werte bessere Übereinstimmung bedeuten,

Der SQ -Faktor bestraft Fehler, die das Volumen von Nuclei verändern, mit dem RQ -Faktor werden Veränderungen der Zellkernanzahl bestraft und der IQ -Faktor bestraft Veränderungen der lokalen Zellkerndicht.

3.4 Klassifikatormethoden

3.4.1 Übersicht

Jedem instanzsegmentierten Nucleus wird eine Klasse zugewiesen, um die Ausgabe zur panoptischen Segmentierungsmaske zu erweitern. Erst die panoptische Segmentierungsmaße ermöglicht das automatische Extrahieren interpretierbarer Eigenschaften aus den Daten. Durch diese panoptische Maske und die extrahierten Eigenschaften der Myotubenkultur wird ein klarer Überblick über den aktuellen Stand der Zellkultur geboten.

Für die Klassenzuweisung ist ein Klassifikator notwendig, der einen Bildausschnitt mit einer Nucleus-Instanz als Eingabe annimmt und eine Klasse als Ausgabe ausgibt. Um diesen Klassifikator optimal zu entwerfen, wird ein umfangreicher Benchmark aus den Zieldaten erstellt, mit dem die vorgestellten Methoden verglichen werden. Benchmarks aus der Literatur umfassen weder dieselben Klassen noch dieselben Objektmerkmale. Deshalb wird ein eigener, kein etablierter Benchmark verwendet. Für das Training wird einheitlich der Adam-Algorithmus [173] mit einer Lernrate von 0,0001 eingesetzt. Außerdem wird der Cross-Entropy-Loss [95] verwendet.

Aus den annotierten Bilddaten werden für jede Anwendung ein Test- und ein Trainingsanteil im Verhältnis eins zu neun extrahiert. Alle betrachteten Variationen des Klassifikators werden ausschließlich mit den Trainingsdaten trainiert und ihre Leistung ausschließlich anhand der Testdaten getestet. Beide Anteile des Datensatzes werden durch Augmentierung erweitert und anschließend in Batches zusammengefasst. Zur Datenaugmentierung werden die folgenden Methoden eingesetzt:

- **Rotation:** Mit einer Wahrscheinlichkeit von 50% werden die Eingabedaten um 90° in der XY-Ebene rotiert.
- **Spiegelung:** Ebenfalls mit einer Wahrscheinlichkeit von 50% erfolgt eine Spiegelung entlang der Z-Achse.
- **Gaußsches Rauschen:** Mit einer Wahrscheinlichkeit von 20% wird Rauschen mit einem Mittelwert von 0 und einer Standardabweichung von 0,01 hinzugefügt.

Prominente Encoder aus der Literatur werden vergleichend eingesetzt. Darüber hinaus werden hier verschiedene Methoden der Vorverarbeitung, des Vortraining und der Klassifikations-Kopf-Architektur eingeführt.

Im Folgenden sind diese Methoden einzeln beschrieben. Da jede mögliche Kombination mit jedem Netz zu trainieren einen unausführbar hohen Rechenaufwand bedeutet, wird eine Vorauswahl von Kombinationen getroffen.

3.4.2 Encoder

Insgesamt sechs verschiedene Bild-Encoder werden für den Methodenvergleich eingesetzt. Tab. 3.1 zeigt diese sechs Encoder.

Tab. 3.1 | Vergleich der sechs vortrainierten Modelle, deren Encoder für den Methodenvergleich eingesetzt werden, hinsichtlich ihrer Genauigkeit auf dem ImageNet-Datensatz [156] und ihrer Parameter. Angegeben sind sowohl die Top-1-Genauigkeit (Acc@1) als auch die Top-5-Genauigkeit (Acc@5), also ob die korrekte Klasse die zuversichtlichste Vorhersage, bzw. unter den fünf zuversichtlichen Vorhersagen ist.

Name	Acc@1 (ImageNet)	Acc@5 (ImageNet)	Parameter (Mio)
ResNet18	69.76%	89.08%	11.7
ResNet101	77.37%	93.55%	44.5
Swin V2	84.11%	96.87%	87.9
ConvNeXt	84.41%	96.98%	197.8
EfficientNet V2	85.81%	97.79%	118.5
CellposeSAM	-	-	305

In der Tabelle sind die Namen, die Anzahl der Parameter und, falls vorhanden, die Top-1-Genauigkeit (Acc@1) und die Top-5-Genauigkeit (Acc@5) auf dem ImageNet-Datensatz angegeben. Bis auf das Segmentierungsmodell CellposeSAM handelt es sich um Encoder, die aus Klassifikatoren stammen, die auf dem ImageNet-Datensatz [156] vorgenommen sind. Swin V2 und der CellposeSAM-Encoder sind ViT-Architekturen, während die anderen Encoder CNNs sind.

3.4.3 Vorverarbeitung

Da sich in vielen Bildausschnitten zwei oder mehr Nuclei überschneiden, werden zwei Vorverarbeitungsmethoden eingeführt, die dem Klassifikator signalisieren, welcher der sichtbaren Nuclei klassifiziert werden soll. Diese Methoden unterscheiden sich darin, wie die Segmentierungsmaske des Nucleus dem Klassifikator zugänglich gemacht wird. Die erste Methode, hier Masken-Methode genannt, ersetzt den Nucleus-Kanal durch die Segmentierungsmaske des gesuchten Nucleus. Das Ziel ist, das Risiko zu minimieren, dass umliegende Nuclei das Klassifikationsergebnis verfälschen. Die Klassifikationsentscheidung wird mit der Masken-Methode von der Geometrie des Nucleus abhängig gemacht. Mit der Methode geht ein Verlust der Oberflächenmerkmale einher. Außerdem hängt das Klassifikationsergebnis bei dieser Vorverarbeitungsart stärker von der Qualität der Segmentation ab. Ein Erfolg der Methode weist auf einen hohen Informationsgehalt in der Geometrie hin.

Die zweite Methode wird hier Distanz-Methode genannt. Mit der Distanz-Methode wird der Nucleus-Kanal mit einer Entfernungsmaske gewichtet. Hierzu wird pixelweise der originale Nucleus-Kanal mit einer Transformation des Abstands aller Pixel außerhalb der

Segmentierungsmaske wie folgt multipliziert:

$$I'(x) = I(x) \cdot \exp\left(-\frac{1}{\sigma} \min_{y \in \neg M} \|x - y\|_2\right), \quad (3.2)$$

wobei:

- $I(x) \in [0, 1]$ der Intensitätswert des Nucleus-Kanal an der Position x ist,
- $I'(x) \in [0, 1]$ der Intensitätswert des neuen, transformierten Nucleus-Kanal an der Position x ist,
- $x \in \Omega \subset \mathbb{N}^3$ die Position eines Voxels im diskreten Bildraum ist,
- $M \subseteq \Omega$ die Segmentierungsmaske und $\neg M = \Omega \setminus M$ deren Komplement im Bildraum sind,
- und $\sigma \in \mathbb{R}^+$ ein Parameter zur Steuerung des exponentiellen Abfalls ist.

Die Verwendung der Distanz-Methode hat zum Ziel, die Oberflächenmerkmale des Nucleus zu erhalten. Außerdem wird mit der Vorverarbeitungsmethode der Einfluss eventuell fehlerhafter Segmentierungsmasken durch die kontinuierliche Abstandstransformation minimiert. Allerdings ist hierdurch auch das Risiko einer Einflussnahme auf das Klassifikationsergebnis durch umliegende Nuclei nicht vollständig eliminiert, sondern nur vermindert. In Abb. 3.4 sind die Nucleus-Kanäle der verschiedenen Methoden und die Entfernungsmaske dargestellt.

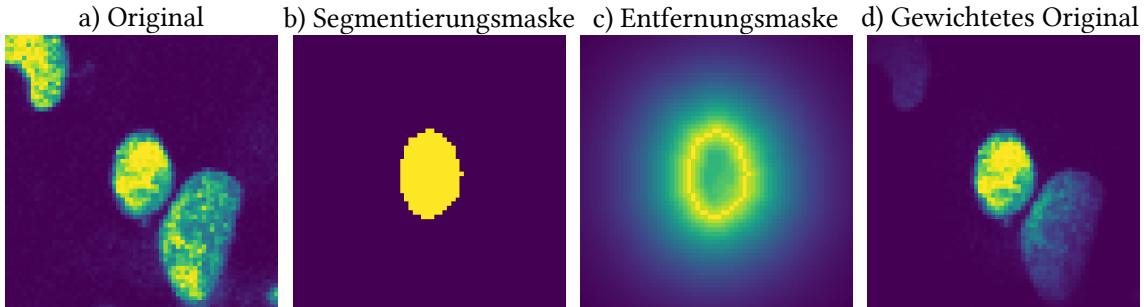


Abb. 3.4 | 2D-Schnitte aus ausgeschnittenen 3D-Bildbereichen zur Darstellung der beiden Vorverarbeitungsmethoden. a) Ausgeschnittener Bereich des originalen Nucleus-Kanals mit mehreren, intuitiv trennbaren Nuclei. b) Segmentierungsmaske des Nucleus. Die Masken-Methode ersetzt den originalen Nucleus-Kanal durch diese Maske. Die binäre Maske zeigt keine Oberflächenmerkmale des Nucleus, lediglich geometrische Merkmale. c) Entfernungsmaske des Nucleus. Diese wird pixelweise mit dem Nucleus-Kanal für die Distanz-Methode multipliziert. d) Mit der Entfernungsmaske gewichtetes Original. Diese Darstellung wird mit der Distanz-Methode an den Klassifikator übergeben. Zu sehen ist, dass der nahegelegene ungewünschte Nucleus noch stellenweise mit hoher Intensität vorhanden ist.

Je nach Modellarchitektur müssen die Bilddaten noch skaliert werden, bevor sie den vortrainierten Modellen übergeben werden können, da die Klassifikatoren Eingaben konstanter Größe benötigen. Dazu wird einfache bilineare Interpolation verwendet (siehe Abschnitt 2.2.3). Um Einfluss auf die Ergebnisse durch eine ungleichmäßige Verteilung der

Annotationen auf die vorhandenen Klassen zu vermeiden, wird vor dem Training die Anzahl der Stichproben pro Klasse extrahiert. Mithilfe der Anzahlverteilung werden die Gradienten dann stärker gewichtet, die zu unterrepräsentierten Klassen gehören. Weil hier mit rechenzeitintensiven dreidimensionalen Daten umgegangen werden muss, ist auch ein dynamisches Speichermanagement Teil der Vorverarbeitungsmethoden. Die Anwendung aller beschriebenen Vorverarbeitungsmethoden wird in einem neu entwickelten 'Retreiver' zusammengefasst. Dieser Retreiver extrahiert die aktuell gewünschten Bildausschnitte, wendet die Vorverarbeitungsmethoden an, zählt die Stichproben pro Klasse und verschiebt nur die notwendigen Daten auf die GPU.

3.4.4 Vortraining

Aus der Literatur sind verschiedene Methoden des Vortrainings bekannt. Hier werden:

- kein Vortraining,
- semi-supervised, und
- fully-supervised Vortraining

betrachtet. Die Abb. 3.5 zeigt die hier umgesetzten Methoden.

Kein Vortraining Ohne Vortraining startet der Encoder, der den Großteil der Gewichte umfasst, mit zufällig initialisierten Gewichtswerten. Da diese zufälligen Werte keine sinnvollen Merkmale extrahieren, wird ein besonders langes Training mit den Zieldaten durchgeführt. Jedes Modell wird jeweils für 75 Epochen trainiert.

Semi supervised Die semi-supervised-Annotationen werden mithilfe eines few-shot-gestützten Cluster-Algorithmus erstellt, der hier als Pseudo-Labler bezeichnet wird. Ein*e Expert*In erstellt hierzu Annotationen von wenigen Nuclei. Danach werden aus den restlichen Segmentierungsmasken einige Merkmale extrahiert und zu einem Vektor zusammengefasst. Zuerst werden das Volumen, die Oberfläche und die Achsenlängen jedes Nucleus direkt bestimmt. Außerdem wird die Exzentrizität aus dem Verhältnis der längsten und der kürzesten Achse berechnet. Für die Kompaktheit wird das Volumen der Maske durch die kleinste mögliche Begrenzungsbox geteilt. Darüber hinaus wird aus der Z-Schicht, in der die Segmentierungsmaske am größten ist, die 2D-Kontur erfasst. Aus dieser Kontur wird eine komplexe Zahlenfolge berechnet und der Absolutwert der ersten zehn Fourier-Koeffizienten als einzelne weitere Merkmale dem Merkmalsvektor hinzugefügt (siehe Abschnitt 2.2.3).

Der Pseudo-Labler normalisiert die Werte der Merkmalsvektoren zu einem Mittelwert von Null und einer Varianz von Eins und wendet eine Principal Component Analysis [77] an, um redundante Informationen zu entfernen. Das Ziel dabei ist, einen Mittelweg zwischen Informationserhalt und Gefahr des Overfitting sowie zwischen Rechenaufwand zu finden. Mithilfe des Label-Spreading-Algorithmus [84], mit einer Radialen Basis Funktion [85]

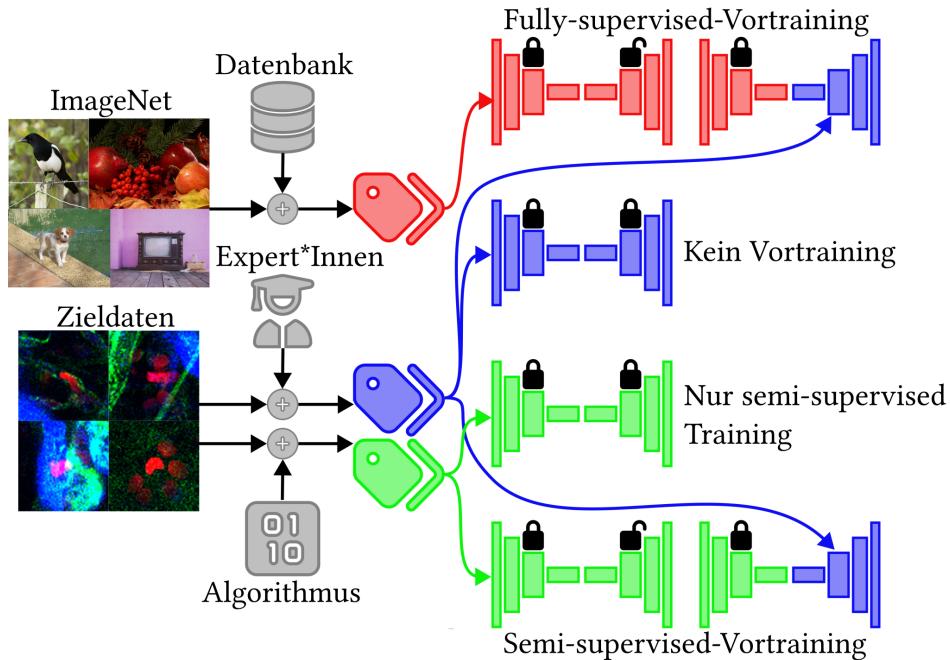


Abb. 3.5 | Übersicht über die Vortrainingsmethoden. Links zu sehen sind die beiden verfügbaren Bildermengen, ImageNet und die Zieldaten. Rechts von diesen Bildermengen werden den Bildern Annotationen hinzugefügt, entweder durch die ImageNet-Datenbank, Expert*Innen oder einen Algorithmus. Jeder Encoder (linke Seite eines Netzwerks) und jeder Klassifikations-Kopf (rechte Seite eines Netzwerks) werden mit einer dieser drei Annotationsmengen trainiert. Die Farben der Annotation und des Netzwerks zeigen dabei die Zuordnung. Mit offenen oder geschlossenen Schaltern über den Encodern und Klassifikations-Köpfen ist dargestellt, ob die Gewichte eingefroren werden. Vier verschiedene Versionen jedes Klassifikators werden hier trainiert. Das erste Netzwerk wird auf den ImageNet-Daten vorgenommen. Anschließend wird mit Expert*Innen-Annotationen der Klassifikations-Kopf neu trainiert. Das Zweite erhält kein Vortraining; es ist komplett mit den Zieldaten trainiert. Für das dritte Netzwerk werden lediglich die Annotationen des semi-supervised Algorithmus eingesetzt. Die letzte Variante wird semi-supervised vorgenommen und anschließend mit den Zieldaten fine-tuned.

als Kernelfunktion, werden die manuellen Annotationen über die Struktur der Daten auf alle Stichproben ausgebrettet. Durch den Pseudo-Labeller entstehen Annotationen für die Daten ohne manuelle Annotationen. Diese neuen Annotationen werden eingesetzt, um in 25 Epochen sowohl die Encoder, als auch die Klassifikations-Köpfe zu trainieren, mit dem Ziel unter geringem Aufwand für die Expert*Innen umfangreiche Klassifikatoren zu trainieren. Optional werden die Gewichte des Encoders hiernach, bis auf die letzten beiden Schichten, eingefroren und nur der Klassifikations-Kopf wird in weiteren 35 Epochen mit dem Trainingsdatensatz der Zieldaten trainiert. Das Ziel dieses Vorgehens ist es, eine stärkere Generalisierung zu erreichen, indem Overfitting bei der Merkmalsextraktion vermieden wird. Da der Encoder mit anderen Daten vorgenommen wird, ist zu erwarten, dass er eine sinnvolle Merkmalsextraktion lernt, ohne auf die expliziten Merkmale der individuellen Stichproben im Trainingsdatensatz angepasst zu sein. Dadurch sind die Beziehungen zwischen Merkmalen und Klassen, die der Klassifikations-Kopf lernt, nicht nur auf die Merkmale des Trainingsdatensatzes beschränkt.

Fully-supervised Das fully-supervised Vortraining bezieht sich hier auf die Initialisierung eines Encoders mit den Gewichten einer entsprechenden Veröffentlichung. Diese Gewichte entstehen durch Vortraining auf dem ImageNet-Datensatz oder stammen aus dem **SAM**-Encoder. Bis auf die letzten 20 bis 30 Prozent der Schichten werden alle Encoder-Gewichte während des Trainings eingefroren. In 50 Epochen werden dann die verbleibenden Encoder-Schichten und der Klassifikations-Kopf trainiert. Die Merkmalsextraktion wird aus einem Datensatz einer anderen Domäne gelernt, was Overfitting verhindert. Nur der Klassifikations-Kopf wird auf Zieldaten trainiert, wobei aufgrund der diversifizierten Merkmale des Encoders eine hohe Generalisierbarkeit angestrebt wird.

3.4.5 Klassifikations-Kopf

An die Encoder werden verschiedene Klassifikations-Köpfe angehängt. Hier werden zwei neue Klassifikations-Kopf-Architekturen nach dem Vorbild vergleichbarer Anwendungen aus der Literatur eingeführt (siehe Abschnitt 2.2.3). Die etablierten Architekturen der Literatur beruhen auf 3D-CNNs und self-attention-Mechanismen über die Z-Dimension.

Abb. 3.6 zeigt die beiden Architekturen systematisch. Der erste Klassifikator wird hier *Volumen-Klassifikator* genannt. Er interpretiert die Merkmale, die der Encoder generiert, als Volumen und generiert daraus mithilfe von 3D-Faltungen und Pooling eine neue Repräsentation. Diese Repräsentation wird dann durch Linear Layers und Leaky ReLu als Aktivierungsfunktion zu vier Ausgabe-Klassen umgeformt. Die Idee des *Volumen-Klassifikators* ist es, die Merkmale, die der Encoder generiert, möglichst vollständig zu erfassen und alle räumlichen Beziehungen, auch in Z-Richtung, festzustellen. Hierbei ist das Ziel, dass durch die 3D-Faltungen eine domänenspezifische Interpretation der Merkmale gelernt wird, sodass die neuen Merkmale nach dem anschließenden Pooling aussagekräftig und niederdimensional sind.

Außerdem wird hier der *Schichten-Klassifikator* eingeführt. Der *Schichten-Klassifikator* betrachtet die einzelnen Schichten des Bilds anhand der individuellen Merkmalsschichten, die der Encoder ausgibt, mithilfe eines Self-Attention-Mechanismus über die Z-Dimension. Dazu werden die räumlichen X- und Y-Dimensionen durch einen spatial average zusammengefasst. Durch eine multihead attention mit vier attention-Köpfen mit embedding dimension 256 wird eine Repräsentation aus den individuellen Schichten der Merkmale erstellt. Lineare Layers und Leaky ReLu als Aktivierungsfunktion formen anschließend die Repräsentation zu den vier Klassen um. Für den *Schichten-Klassifikator* ist das Ziel, dass durch die Vereinfachung der Daten aussagekräftige, schichtenweise Merkmale entstehen und dass diese räumlich invariant sind, da der betrachtete Nucleus in den Bildfenstern zentriert ist.

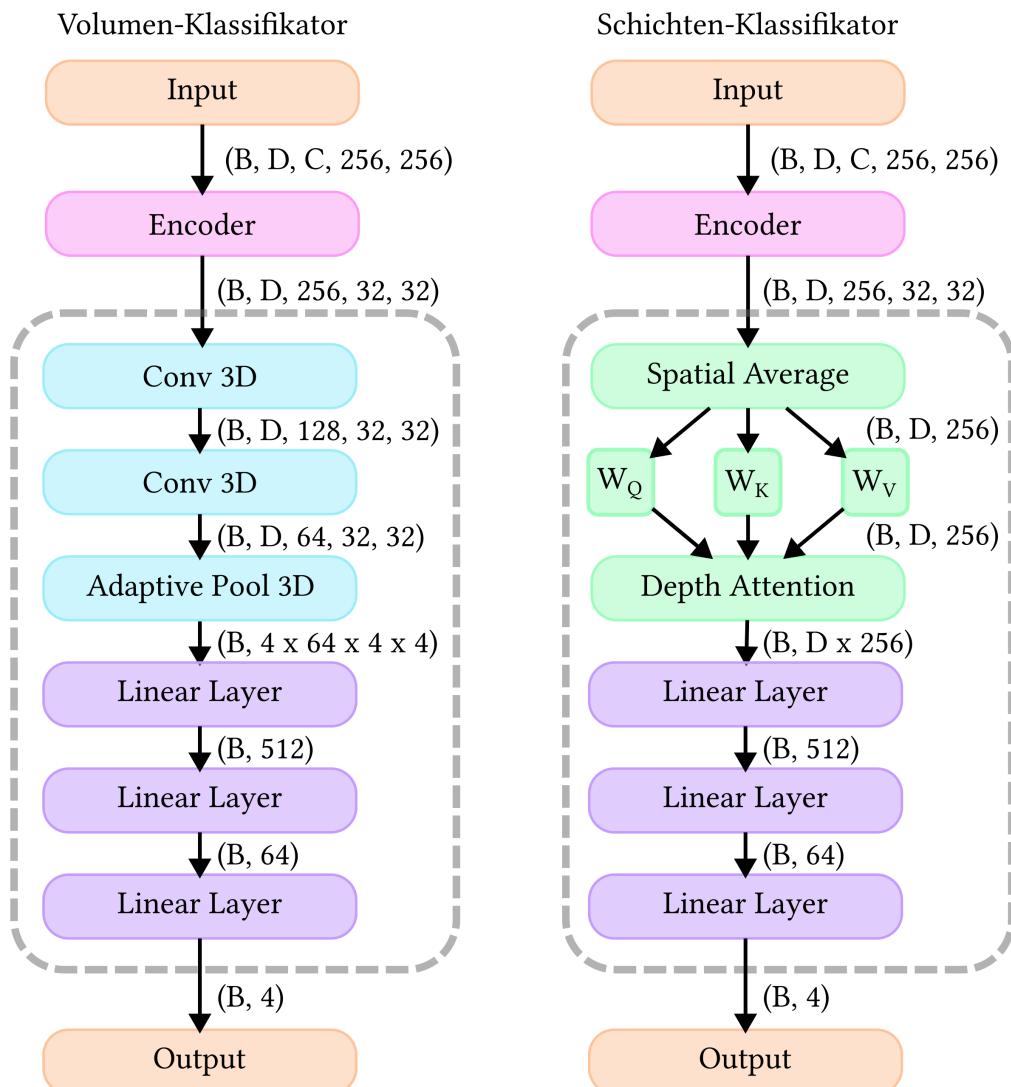


Abb. 3.6 | Architektur der beiden Klassifikatoren. Der Encoder wird jeweils modular ausgetauscht. Links zu sehen ist die Architektur des *Volumen-Klassifikators*, der die Merkmale, die der Encoder generiert, als Volumen interpretiert und mithilfe von 3D-Faltungen und Pooling daraus eine Repräsentation erzeugt. Diese Repräsentation wird anschließend durch Linear Layers in vier Ausgabe-Klassen umgeformt. Rechts ist der *Schichten-Klassifikator* zu sehen. Die räumlichen X- und Y-Dimensionen werden im *Schichten-Klassifikator* durch einen spatial average zusammengefasst. Durch eine Multihead Attention mit vier Attention-Köpfen und Embedding-Dimension 256 wird dann eine Repräsentation aus den individuellen Schichten der Merkmale erstellt. Lineare Layers formen anschließend die Repräsentation zu den vier Klassen um.

3.5 Segmentierung

3.5.1 Modelle

Für die Instanzsegmentierung der Nuclei werden die folgenden drei Modelle eingesetzt:

- Ein Modell des nnU-Net Framework, das selbstkonfigurierte Modelle basierend auf der U-Net-Architektur erstellt [154].
- Das DeepCell-Caliban-Modell, das Bildfaltungen und speziell entwickelten Nachverarbeitungsstrategien vereint.
- Cellpose-SAM, das die Architekturen der Cellpose-Modelle mit der Architektur und den Gewichten des Foundation-Model [SAM](#) vereint.

3.5.2 Nachverarbeitungsmethoden

Zur Nachbearbeitung der Instanzsegmentierungsmasken wird der Instanztrenner eingeführt. Diese Methode dient dazu, separate Instanzen mit denselben numerischen Annotation (IDs) zu trennen und mit einzigartigen Zahlenwerten zu versehen. Pro Instanz werden hierzu ein zufälliges Pixel betrachtet und davon ausgehend alle Pixel mit direkter Verbindung über die Segmentierungsmaske gesucht. Diese verbundenen Pixel werden als neue Instanz mit einzigartiger ID abgelegt, bis keine Pixel ohne Verbindung übrig bleiben. Eine weitere Methode der Nachbearbeitung, die auf Segmentierungsmasken eingesetzt werden kann, ist der Watershed-Algorithmus (siehe Abschnitt 2.2.2). Der Algorithmus trennt überlagerte Instanzen, die das Modell als einzelne Instanzen segmentiert hat.

Implementierung 4

4.1 Überblick

Im nachfolgenden Kapitel wird die Implementierung aller relevanten Methoden mit Fokus auf die praktische Anwendung in Form von Nutzerschnittstellen oder als Entwicklerskript erläutert. Eine neu entwickelte Anwendung, die 3D-Zelldaten-Pipeline, vereint verschiedene Software-Module, die jeweils ein Zwischenziel der vorliegenden Arbeit umsetzen. Die Anwendung wird über eine grafische Nutzeroberfläche bedient, die auch ohne Programmierkenntnisse genutzt werden kann. Die beschriebenen Funktionen sind im bereitgestellten Repository zu finden.

4.2 Segmentierungsmodelle

Abb. 4.1 zeigt den Signalfluss der Anwendung der Segmentierungsmodelle sowie ihrer Bewertung. Die Segmentierungsmodelle werden zur Evaluation in einfachen Entwicklerskripten über die Kommandozeile oder in Jupyter Notebooks ausgeführt. Die Eingabedaten werden in einem Python-Entwicklerskript vorab normalisiert. Für das nnU-Net-Modell (siehe Abschnitt 3.5.1) wird auf die Eingabedaten zunächst eine Konvertierung angewandt, die die Daten durch Padding in die geforderte Größe bringt und zu .nii.gz-Dateien umformt. Eine Parameterbestimmung für die Anzahl der Wiederholungen, die Normalisierungsart und die Datensatz-Eigenschaften wird anhand der Parameter-Vorlage der Autoren durchgeführt. Das Modell wird direkt anhand der Architektur und der Gewichte der Veröffentlichung initialisiert und über die Kommandozeile auf alle Daten angewandt. Die entstehenden Instanzen werden in einen Instanztrenner gegeben (siehe Abschnitt 3.5.2), um den Instanzen einzigartige IDs zuzuweisen.

Für das Deepcell-Modell (siehe Abschnitt 3.5.1) wird direkt das bereitgestellte Jupyter Notebook der Deepcell-Veröffentlichung auf die Benchmarkdaten angewandt [152]. Die Ergebnisse werden daraufhin mit dem Watershed-Algorithmus der OpenCV-Bibliothek nachbearbeitet (siehe Abschnitt 3.5.2).

Das CellposeSAM-Modell (siehe Abschnitt 3.5.1) wird als Entwicklerskript weitgehend mit der Architektur, den Gewichten und den vorgeschlagenen Parametern der Autoren angewandt. In einer Parameterbestimmung wird der durchschnittliche Durchmesser der Nuclei an die eingegebenen Daten angepasst. Auf die Ergebnisse wird keine Nachbearbeitung angewandt.

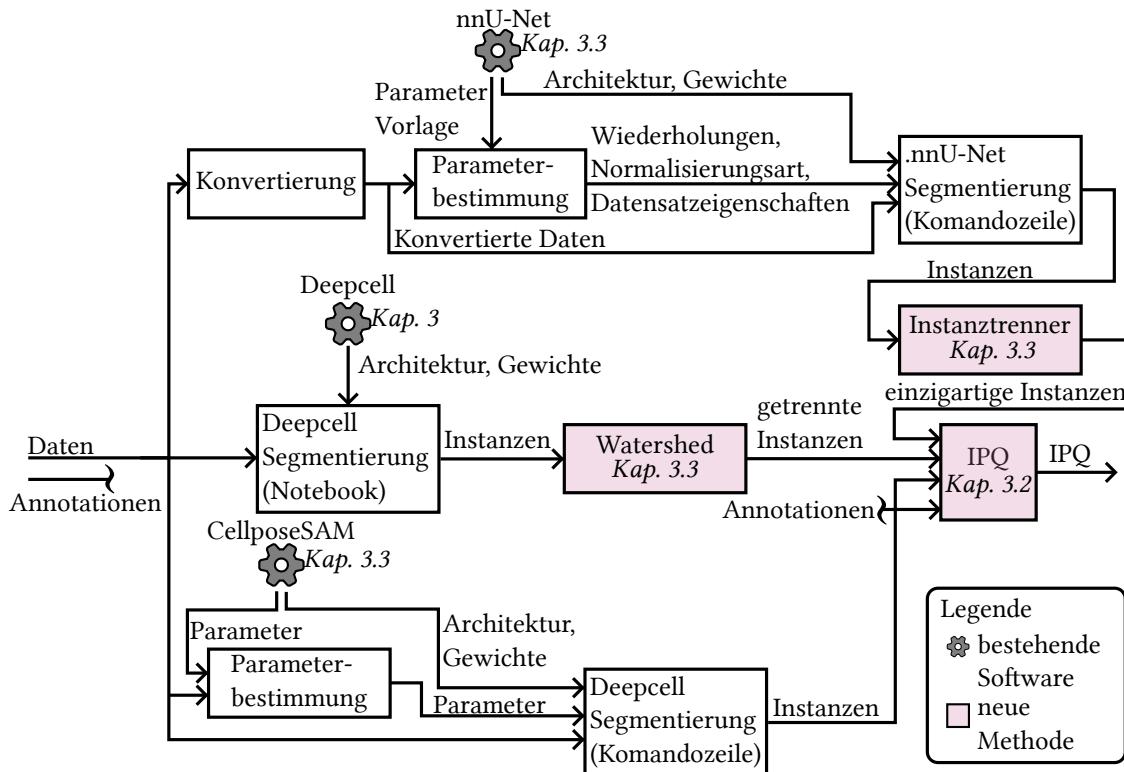


Abb. 4.1 | Signalflussdiagramm der praktischen Anwendung der Segmentierungsmodelle.

Die Instanzen aller drei Modelle sowie die Annotationen werden in einem Jupyter Notebook zur Berechnung der IPQ eingegeben (siehe Abschnitt 3.3). Mithilfe einer neu entwickelten Funktion werden hier die Ergebnisse mit der IPQ-Metrik bewertet.

4.3 Klassifikatoren

Für die Klassifikatoren werden vier Software-Module entwickelt, wie in Abb. 4.2 dargestellt. Das erste Modul nimmt als Eingabe eine Methodenauswahl (siehe Abschnitt 3.4), die verwendet werden soll, und gibt einen Klassifikator zurück. Mithilfe der Pytorch-Bibliothek werden Encoder und, je nach Vortrainingsmethode, die zugehörigen Gewichte geladen. Außerdem werden neu entwickelte Klassen von PyTorch-Modellen für das Erstellen des Klassifikations-Kopfs aufgerufen (siehe Abschnitt 3.4.5).

In das zweite Modul wird ebenfalls eine Methodenauswahl eingegeben und abhängig von der Vortrainings- und Vorverarbeitungsmethode ein entsprechender Datensatz zurückgegeben (siehe Abschnitt 3.4.4 und Abschnitt 3.4.3). Dieser Datensatz wird als Pytorch Data-Loader erstellt und mit einer Batch-Größe und Augmentierungen der Monai-Bibliothek versehen. In diesen Dataloader wird eine neu entwickelte Retreiver-Instanz (siehe Abschnitt 3.4.3) eingebettet. Der Retreiver wendet die entsprechenden Vorverarbeitungsmethoden

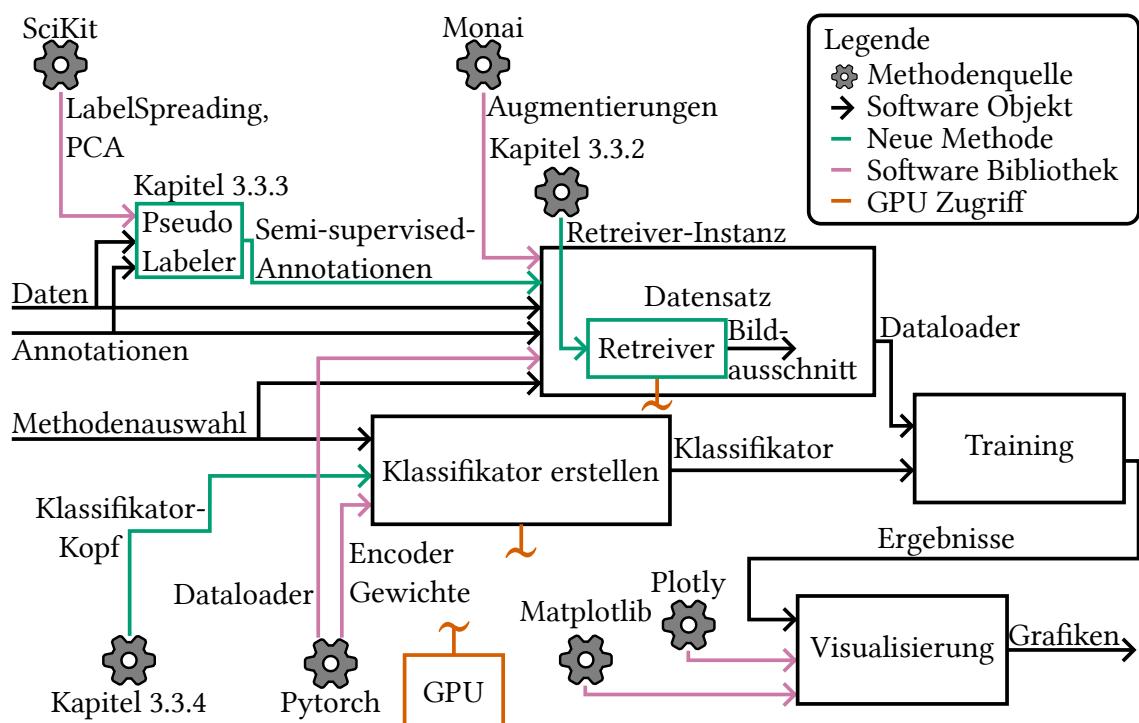


Abb. 4.2 | Signalflussdiagramm der praktischen Umsetzung der Klassifikatoren. Zu sehen sind die Module, aus denen sich die Anwendung der Klassifikatoren zusammensetzt. Links sind als Eingabedaten mit zugehörigen Annotationen sowie eine Kombination aus Methoden und rechts Grafiken, die die Ergebnisse visualisieren, als Ausgabe zu sehen. Verschiedene Python-Bibliotheken werden genutzt (Rosa), aber auch neue Methoden werden eingesetzt (Grün). Der Klassifikator und der Retriever, der dynamisch Bildausschnitte lädt, greifen auf die GPU zu (Orange).

auf die Bilddaten an und ermöglicht es, die großen, dreidimensionalen Daten nicht alle auf einmal als Variablen im Datensatz unterzubringen, sondern nur die Indizes von Bildern und Instanzen. Diese Indizes werden dann genutzt, um dynamisch nur die gewünschten Bildausschnitte auf die GPU zu laden. In einem Pseudo-Labler-Modul werden zusätzlich Semi-supervised-Annotationen mithilfe der PCA und der Label-Spreading-Funktion der SciKit-Bibliothek erstellt (siehe Abschnitt 3.4.4).

Das dritte Modul erhält einen Klassifikator und einen Datensatz und führt einen Trainingsdurchlauf durch. Es speichert die Ergebnisse und die Gewichte der neu trainierten Klassifikatoren in automatisch benannten Dateien.

Im vierten Modul werden mithilfe der plotly-Bibliothek Grafiken aus den eingegebenen Ergebnissen erstellt und, mithilfe der Matplotlib-Bibliothek, separat als .png gespeichert.

4.4 3D-Zelldaten-Pipeline

4.4.1 Übersicht

In Abb. 4.3 ist eine Übersicht der Software-Architektur der 3D-Zelldaten-Pipeline dargestellt. Die Anwendung vereint die in der vorliegenden Arbeit eingeführten Methoden (siehe Kapitel 3).

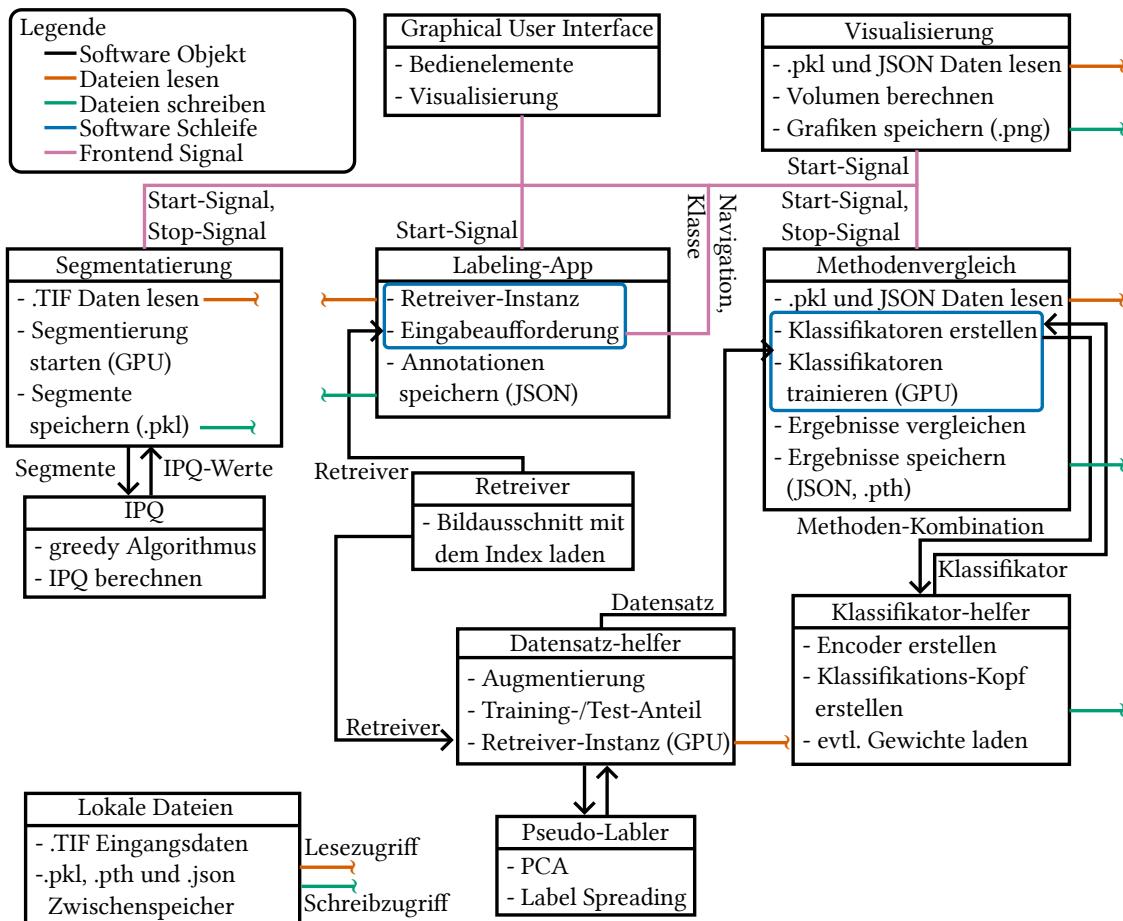


Abb. 4.3 | Übersicht der Architektur der neu entwickelten 3D-Zelldaten-Pipeline. Als Rechtecke zu sehen sind die Module, die die 3D-Zelldaten-Pipeline bilden. Schwarze Pfeile stellen die Weitergabe von Software Objekten zwischen den Modulen dar. In Orange und Grün sind Zugriffe auf lokale Dateien dargestellt, Orange ist Dateien lesen und Grün Dateien schreiben. Software-Schleifen sind als blaue Rechtecke dargestellt. Rosa Verbindungen stellen die Signale von und an das Graphical User Interface (GUI) dar.

Oben zentral zu sehen ist das Graphical User Interface (GUI), die grafische Schnittstelle mit Bedienelementen und Visualisierungen für die Interaktion mit Nutzer*Innen. Vom GUI gehen Start- und Stop-Signale an die wichtigsten Module (Segmentierung, Labeling-App, Visualisierung und Methodenvergleich) weiter. Diese vier Module werden direkt von Nutzer*Innen bedient. Das Segmentierung-Modul greift auf lokale Dateien zu, um Bilder

einzulesen, und startet die Segmentierung für diese Daten mit verschiedenen Modellen (siehe Abschnitt 4.2). Die gefundenen Segmente werden dann an das IPQ-Modul gegeben, das die IPQ-Metrik für die gefundenen Segmente zurückgibt, insofern Annotationen zu den Daten existieren (siehe Abschnitt 3.3).

Das Labeling-App-Modul enthält eine Schleife aus einer Retriever-Instanz und einer Eingabeaufforderung (siehe Abschnitt 3.4.3). Diese Retriever-Instanz greift auf lokale Dateien zu, um je nach Nutzereingabe einen Bildausschnitt zu laden. Den Bildausschnitten weisen Nutzer*Innen Annotationen zu, die anschließend lokal gespeichert werden.

Im Methodenvergleich werden die Annotationen und Segmentierungsmasken eingelesen. In einer Schleife werden mithilfe eines Klassifikator-Helpers Klassifikatoren aus verschiedenen Methoden-Kombinationen erstellt und trainiert. Der Datensatz enthält eine Retriever-Instanz (siehe Abschnitt 3.4.3) und wird von einem Datensatz-Helper-Modul erstellt. Ein Pseudo-Labler-Modul erstellt Semi-supervised-Annotationen für den Datensatz, die je nach Vortrainingsmethode in der Schleife verwendet werden können (siehe Abschnitt 3.4.4). Die Ergebnisse der verschiedenen Klassifikatoren werden anschließend verglichen und gespeichert. Das Visualisierung-Modul liest die Ergebnisse des Klassifikators und stellt diese in Grafiken dar.

Abb. 4.4 zeigt das GUI der 3D-Zelldaten-Pipeline. In den fünf Tabs (Segmentierung, Labeling-App, Methodenvergleich, Training und Visualisierung) werden die Aufgaben der Anwendung erledigt. Als Frontend der App dient eine Dash-Anwendung, die eine einfache HTML-Seite bereitstellt. Das Backend ist in Python entwickelt, und die gesammelten Daten werden als JSON gespeichert. Zwischenergebnisse wie trainierte Modelle und verarbeitete Daten werden als nicht interpretierbare Python-spezifische Datentypen abgelegt. Hierzu ist ein Docker mit Zugriff auf das lokale Dateiensystem versehen und über Kubernetes betrieben.

4.4.2 Segmentierung

Abb. 4.4 zeigt den Tab der 3D-Zelldaten-Pipeline, in dem die Segmentierung der Zellkerne erfolgt. Im Eingabefeld 'Eingabe Ordner' wird ein Ordner, relativ zu dem Speicherort der Anwendung, angegeben, aus dem TIF-Bilder gelesen werden. Unter 'Ausgabe Ordner' wird angegeben, wo die Instanzsegmentierungsmasken gespeichert werden. Mit dem Knopf 'Starte Segmentierung' wird die Segmentierung gestartet und mit dem 'Abbrechen'-Knopf wieder abgebrochen. Das Textfeld 'Segmentation running...' blinkt im Takt von einer Sekunde, solange die Segmentierung läuft.

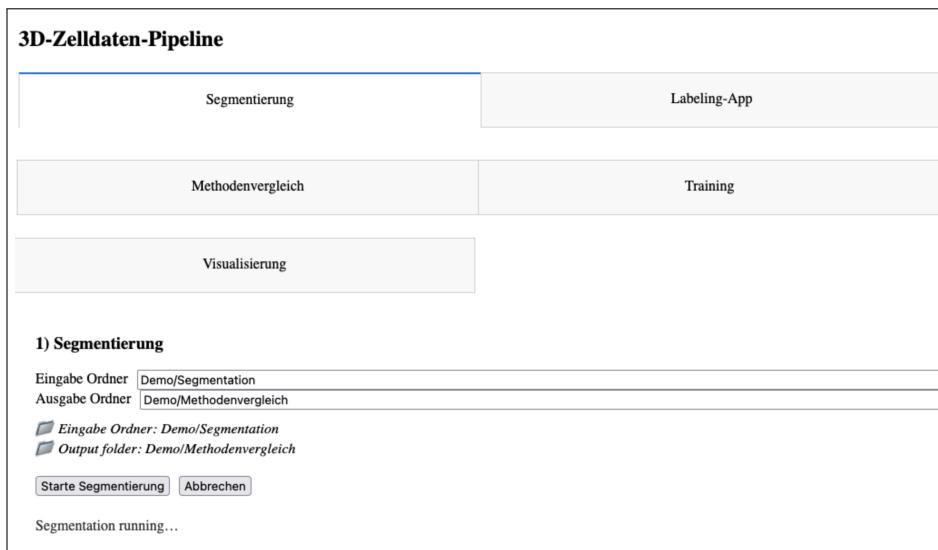


Abb. 4.4 | Ausschnitt der 3D-Zelldaten-Pipeline. Oben sind fünf Tabs zu sehen, aus denen die auszuführende Aufgabe gewählt werden kann. Diese Tabs sind zur besseren Darstellung untereinander statt nebeneinander angeordnet. Darunter ist ein Ausschnitt des ausgewählten 'Segmentierung'-Tabs zu sehen.

4.4.3 Labeling-App

Um den Klassifikator zu trainieren, werden einige Zieldaten Annotationen mit der neu entwickelten Labeling-App hinzugefügt. Die Anforderungen an die Labeling-App sind ein nutzerfreundlicher, zeiteffizienter Ablauf und Zugänglichkeit ohne Programmiererfahrung. Funktional muss die Labeling-App imstande sein, einzelne, zu klassifizierenden Nuclei zu visualisieren und Eingabemöglichkeiten zu bieten, mit denen die Expert*Innen die Klasse des Nucleus eintragen können. Diese eingetragenen Annotationen müssen sinnvoll gespeichert werden. Abb. 4.5 zeigt das neu entwickelte GUI der Labeling-App.

Mit den Bedienelementen werden die Anforderungen an die Funktionalität erfüllt. Zentral sind zwei Fenster zu sehen, die je einen 2D-Schnitt des ausgewählten Nucleus anzeigen. Diese Schnitte werden von einem neu entwickelten Retriever dynamisch geladen (siehe Abschnitt 3.4.4). Das linke Bild zeigt den Nucleus stark vergrößert, das rechte Bild zeigt die Umgebung des Nucleus. Auf den Bildern ist jeweils ein Kasten gezeichnet, der den ausgewählten Nucleus umrandet, um Nuclei, die dicht aneinander liegen, zu unterscheiden. Mit dem Schieberegler unter den Fenstern wird ausgewählt, welche der Schichten, in denen der Nucleus vorhanden ist, gezeigt werden soll. Über dem Fenster sind der Index des aktuell dargestellten Nucleus und der des Bilds angegeben. Links neben dem Fenster befinden sich Bedienelemente mit den folgenden Funktionen:

- *Previous picture*: Vorheriges Bild auswählen,
- *Next picture*: Nächstes Bild auswählen,
- *Previous nucleus*: Vorherigen Nucleus auswählen,

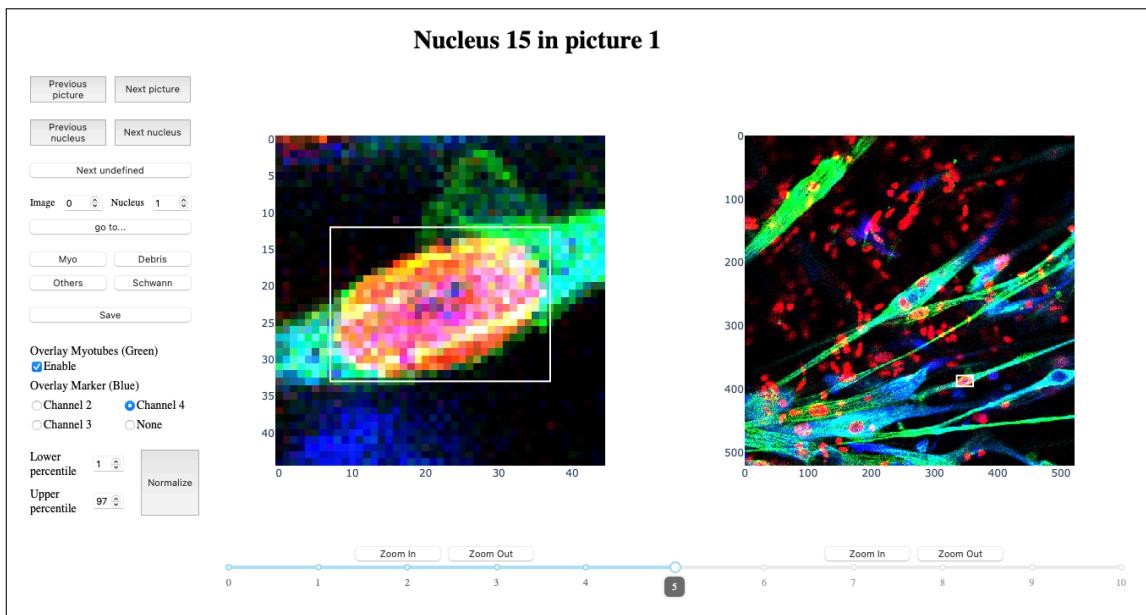


Abb. 4.5 | GUI der neu entwickelten Labeling-App. In der Mitte wird der Nucleus angezeigt, der annotiert werden soll, und links sind Bedienelemente zu sehen. Unter dem dargestellten Nucleus befindet sich ein Schieberegler, der die Navigation entlang der Z-Achse ermöglicht. Über die Bedienelemente wird zwischen Bildern und Zellen umgeschaltet, die Klasse des Nucleus bestimmt, der gezeigte Ausschnitt wird mit prozentualen Schwellenwerten normalisiert und die Fenstergröße geändert.

- *Next nucleus*: Nächsten Nucleus auswählen,
- *Next undefined*: Nächsten Nucleus ohne eingetragene Annotation auswählen,
- *Image*: Eingabefeld für das auszuwählende Bild,
- *Nucleus*: Eingabefeld für den auszuwählenden Nucleus,
- *go to...*: Bild und Nucleus auswählen, wie in den Eingabefeldern *Image* und *Nucleus* definiert,
- *Myo*: 'Myotuben-Zellkern'-Klasse als Annotation des ausgewählten Nucleus definieren,
- *Debris*: 'Debris'-Klasse als Annotation des ausgewählten Nucleus definieren,
- *Other*: 'Andere'-Klasse als Annotation des ausgewählten Nucleus definieren,
- *Schwann*: 'Schwannzellen-Zellkern'-Klasse als Annotation des ausgewählten Nucleus definieren,
- *Save*: Manuell die festgelegten Annotationen speichern. Die Labeling-App speichert außerdem eigenständig periodisch,

- *Overlay Myotubes (Green)*: Mit einem Haken bei 'Enable' wird der Marker, der die Myotuben einfärbt, in Grün eingeblendet,
- *Overlay Marker (Blue)*: Einer oder keiner der restlichen vorhandenen Marker wird in Blau eingeblendet,
- *Lower Percentile*: Unteren prozentualen Schwellwert wählen, der bei der nächsten Normalisierung angewandt werden soll,
- *Upper Percentile*: Oberen prozentualen Schwellwert wählen, der bei der nächsten Normalisierung angewandt werden soll,
- *Normalize*: Normalisierung lokal auf den Ausschnitt des aktuell ausgewählten Nucleus anwenden. Intensitätswerte, die im beziehungsweise über dem Perzentil der eingetragenen Schwellwerte liegen, werden hierbei zusammengefasst,
- *Zoom In*: Anzuzeigenden Ausschnitt verkleinern und
- *Zoom Out*: Anzuzeigenden Ausschnitt vergrößern.

Aufgrund der Anforderung einer nutzerfreundlichen, zeiteffizienten Bedienung sind alle Berechnungen, die während der Nutzung der Labeling-App ausgeführt werden, darauf ausgelegt, die Rechenzeit zu minimieren. Hierzu werden alle Bilder und Masken bei der Initialisierung der Labeling-App in den Cache geladen. Des Weiteren ist eine Python-Klasse angelegt, die zusätzlich das aktuell ausgewählte Bild und den ausgewählten Nucleus speichert. Erst wenn eine Änderung der Auswahl vorgenommen wird, wird ein neuer Nucleus oder ein neues gesamtes Bild geladen – und selbst dann lediglich aus dem Cache.

4.4.4 Methodenvergleich

Der Tab 'Methodenvergleich' ist in Abb. 4.6 dargestellt. Die Eingabefelder 'Eingabe Ordner' und 'Ausgabe Ordner' bestimmen, aus welchem und in welchen Ordner, relativ zum Speicherort der Anwendung, die Daten gelesen oder geschrieben werden. Im Eingabe Ordner müssen dreidimensionale Bilder, Segmentierungsmasken und Annotationen vorhanden sein. Die vorangegangenen Tabs speichern die Daten direkt im erwarteten Format ab. Unter der Anzeige der gewählten Ordner sind einige Checkbox-Felder vorhanden. Diese sind in die vier Gruppen Encoder, Klassifikations-Kopf, Vorverarbeitung (Nucleus-Kanal) und Vortraining unterteilt. Alle hier gewählten Methoden werden nachfolgend verwendet und in jeder Kombination trainiert. Die verfügbaren Methoden sind in Abschnitt 3.4 beschrieben. Mit dem 'Start'-Knopf wird der Methodenvergleich gestartet und mit dem 'Abbrechen'-Knopf wieder abgebrochen. Nach Abschluss des Vergleichs werden die Voraussagen des Modells mit der höchsten Genauigkeit auf dem Test-Anteil der Eingangsdaten für alle Nuclei im Datensatz abgelegt, auch für die Daten ohne Annotationen. Außerdem wird die Kombination der Methoden gespeichert, mit der die höchste Genauigkeit erzielt wurde.

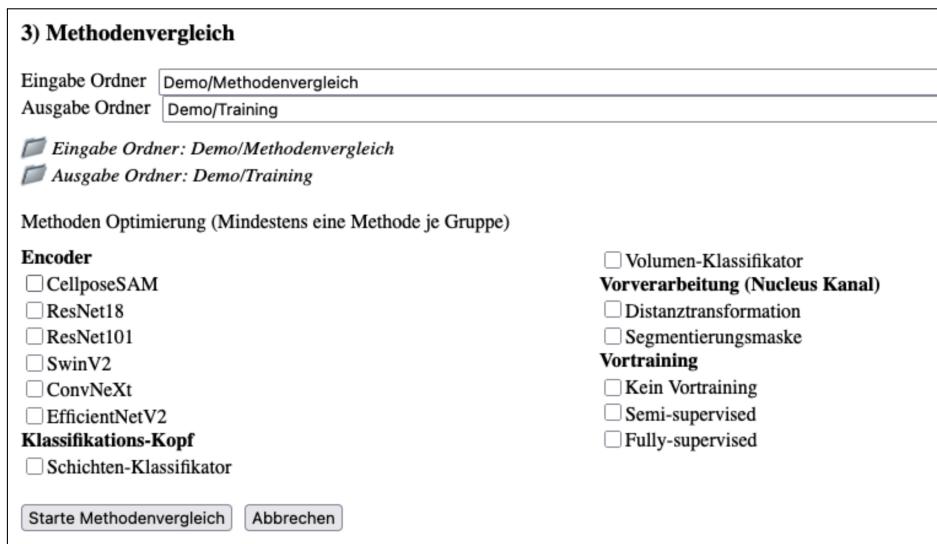


Abb. 4.6 | Ausschnitt des Tabs der 3D-Zelldaten-Pipeline, der den Vergleich der Klassifikatormethoden ermöglicht. Methoden der Kategorien Encoder, Klassifikations-Kopf, Vorverarbeitung und Vortraining können ausgewählt werden, um einen Vergleich aller Kombinationen der Methoden zu starten.

4.4.5 Training

Im 'Training'-Tab (Abb. 4.7) wird der Klassifikator mit der zuvor ermittelten optimalen Kombination von Methoden in mehr Epochen erneut trainiert. Dabei können auch mehr Annotationen eingesetzt werden, die während des Methodenvergleichs nachgeliefert wurden. Das Training über diesen Tab wird außerdem in mehr Epochen durchgeführt als das Training des 'Methodenvergleich'-Tabs. Dieser Tab ist optional, das Training aus dem Methodenvergleich direkt genutzt werden. In den Eingabefeldern 'Eingabe Ordner' und 'Ausgabe Ordner' wird angegeben, welche Ordner verwendet werden sollen. Im Eingabe Ordner muss eine Datei vorhanden sein, die die beste Kombination der Methoden auszeichnet. Nachdem das Training mit dem 'Starte Training'-Knopf gestartet und vollendet wurde, werden die Vorhersagen für alle Nuclei der Eingabedaten im Ausgabe Ordner abgelegt. Sind hier keine Annotationen verfügbar, wird eine einfache Inferenz mit dem trainierten Modell im Eingabe Ordner durchgeführt und die Vorhersagen gespeichert.

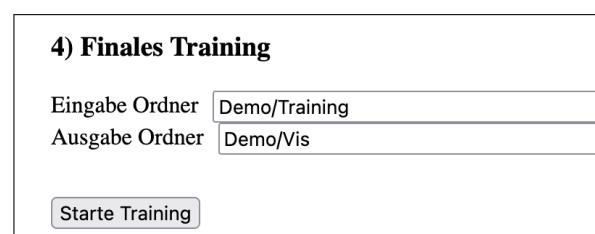


Abb. 4.7 | Ausschnitt des Tabs der 3D-Zelldaten-Pipeline, der das finale Training des Klassifikators durchführt. Hier wird ein Ordner gewählt, aus dem die beste Kombination von Methoden gelesen werden soll, und anschließend wird ein längeres Training gestartet.

4.4.6 Visualisierung

Im letzten Tab werden die Ergebnisse der Anwendung dargestellt (siehe Abb. 4.8). Mithilfe der beiden Eingabefelder wird festgelegt, aus welchen Ordnerne die Ergebnisse stammen sollen und wohin die erstellten Grafiken gespeichert werden. Mit dem 'Starte Visualisierung'-Knopf wird der Algorithmus gestartet, der die interpretierbaren Eigenschaften aus den Ergebnissen ausliest. Daraufhin werden unten auf der Seite drei Grafiken dargestellt. Links wird beispielhaft eine Instanzsegmentierungsmaße einer Schicht eines der 3D-Bilder gezeigt. In der Mitte wird ein Balkendiagramm dargestellt, das die Anzahl der Nuclei pro Klasse zeigt. Das rechte Balkendiagramm zeigt die Verteilung der Nucleusvolumina in Pixeln hoch drei. Um dieses Volumen zu interpretieren, muss die Umrechnung von Pixeln in Micrometer je nach Auflösung des Aufnahmegeräts und der Zoom-Stufe manuell durchgeführt werden.

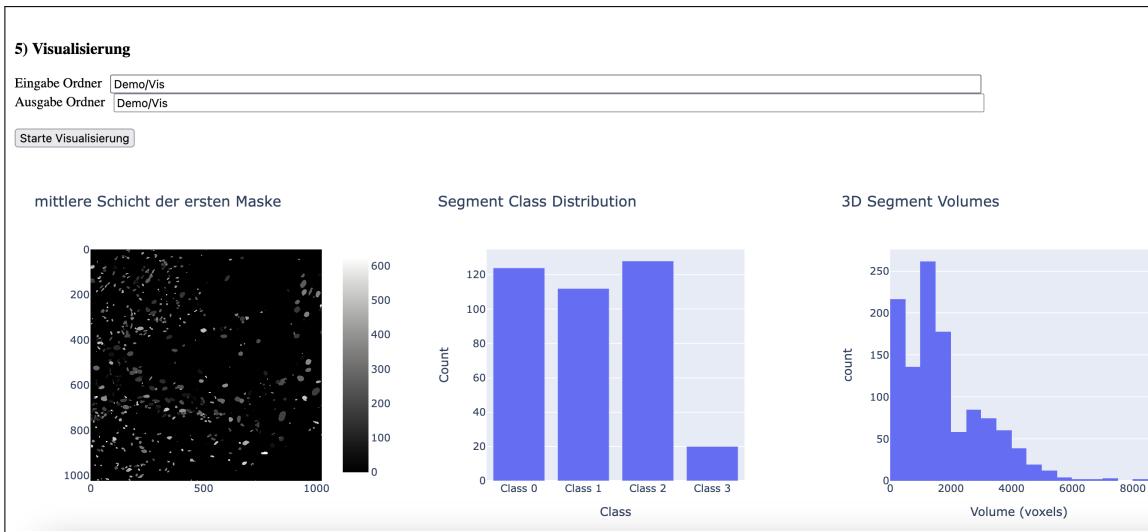


Abb. 4.8 | 'Visualisierung'-Tab der 3D-Zelldaten-Pipeline. Hier werden die Ergebnisse der Anwendung in interpretierbaren Grafiken dargestellt.

4.5 Umsetzung

Die Tabelle 4.1 stellt den Umfang des Codes der einzelnen Module dar.

Tab. 4.1 | Übersicht der Implementierungskomponenten mit Codeumfang und genutzten Bibliotheken

Modul	Beschreibung	Zeilen Code	Verwendete Bibliotheken
Segmentierung			
Vorverarbeitung	Vorverarbeitung der 3D-Bildstapel	40	tqdm, NumPy
Deepcell	Deepcell Segmentierung	125	SciKit-Image, NumPy, de- epcell, deep- cell_toolbox
nnU-Net	nnU-Net Segmentierung	107	nnunetv2
CellposeSAM	CellposeSAM Segmentierung	45	NumPy, cellpo- se
Nachbearbeitung und Bewertung	Nachbearbeitung der Segmentie- rungsmasken, greedy-Algorithmus und IPQ-Berechnung	600	NumPy, Open- CV, SciKit- Image, SciPy
Klassifikation			
Retriever	Dynamisches Laden von Bildaus- schnitten auf die GPU	100	dash, plotly
Pseudo-Labler	Semi-supervised-Annotationen er- stellen	215	SciKit-Image, NumPy
Visualisierungen	Kontur-Bilder, Grad-CAM, Scatter- plots, etc	650	SciKit-Image, NumPy, Mat- plotlib, SciPy, Torch
Klassifikator erstellen	Encoder- und Klassifikations-Kopf- Code	300	Torch, cellpose
Datensatz	Vorverarbeitungsmethoden, Daten- satz erstellen	200	Torch, NumPy
3D-Zelldaten- Pipeline			
Frontend	GUI der 3D-Zelldaten-Pipeline, Webserver, Threading	800	dash, plotly
Webserver backend	Unterstützende Funktionen für das Frontend-serving	100	-

Modul	Beschreibung	Zeilen Code	Verwendete Bibliotheken
Training	Klassifikator-Training, Management von Modellcheckpoints und aktiven Trainings-Threads	380	Monai, tqdm, Pytorch, SciKit-Learn
Methodenauswahl	Auswahl der Methoden	60	-
Inferenz	Inferenz mit einem trainierten Klassifikator	160	Torch, SciKit-Learn
Labeling-App	Backend für die Labeling-App	360	NumPy
Visualisierung	Frontend Grafiken erstellen und speichern	190	plotly, pandas, NumPy, Matplotlib
Gesamt		4325	

Ergebnisse 5

5.1 Überblick

Das nachfolgende Kapitel beschreibt die Ergebnisse der durchgeführten Experimente. Die Experimente werden in mehreren Durchläufen der neu entwickelten 3D-Zelldaten-Pipeline zur Auswertung von 3D-Zelldaten durchgeführt. Auf Grundlage der Messergebnisse der 3D-Zelldaten-Pipeline wird eine statistische Auswertung durchgeführt. Zuerst werden die Bewertungen der Instanzsegmentierungsmasken anhand der [IPQ](#)-Metrik und anschließend die der Klassifikatoren anhand ihrer Genauigkeit betrachtet.

Die [IPQ](#)-Ergebnisse werden in die drei Faktoren: Segmentation-Quality ([SQ](#)), Recognition-Quality ([RQ](#)) und die neu eingeführte Injective-Quality ([IQ](#)) gegliedert (siehe Abschnitt [3.3](#)). Daraus wird ersichtlich, dass jedes der Segmentierungsmodelle in einem der Faktoren dominiert, insgesamt aber CellposeSAM den anderen Modellen signifikant überlegen ist. Auch die Ergebnisse der Klassifikation werden unterteilt, um den Einfluss einzelner Methoden auf die Genauigkeit des Klassifikators zu identifizieren. Vergleiche der Ergebnisse verschiedener Methoden zeigen, dass die neu eingeführten Methoden wie der Pseudo-Labler (siehe Abschnitt [3.4.4](#)) je nach Anforderung dem Stand der Technik entsprechen. Außerdem zeigen die Ergebnisse, dass die Anwendung der 3D-Zelldaten-Pipeline optimale Methoden für die Segmentierung und Klassifikation von Zelldaten effizient ermittelt und gegenüber nicht-optimalen Methoden einen signifikanten Qualitätsunterschied aufweist. Mithilfe ausführlicher Analysen lassen sich des Weiteren grundlegende Erkenntnisse über das Verhalten von Klassifikatoren im Umgang mit biologischen 3D-Bildstapeln gewinnen.

5.2 Hardware

Für die Anwendung wird eine NVIDIA GeForce RTX 3090 Ti mit 24 GB VRAM verwendet. Der verwendete Server verfügt über eine 12th-Gen-Intel(R) Core(TM) i9-12900KF-CPU mit 16 Kernen und 64 GB RAM.

5.3 Segmentierung

Für die Wahl eines Segmentierungsmodells wird in Abschnitt [3.3](#) das Bewertungskriterium [IPQ](#) eingeführt. Außerdem wird in Abschnitt [2.3.1](#) der annotierte S_BIAD1518-Datensatz vorgestellt. Die [IPQ](#) wird auf dem Datensatz mit den Masken der drei vortrainierten Seg-

mentierungsmodelle und zur Validierung anhand der Annotationen durchgeführt. Alle Annotationen erreichen einen IPQ-Wert von genau Eins. Abb. 5.1 zeigt exemplarisch einen 2D-Schnitt eines 3D-Bildes mit roten Linien als Konturen der Annotation bzw. der vorhergesagten Segmentierungsmaske je Modell. Die prädizierten Instanzsegmentierungsmasken für die dreidimensionalen Benchmarkdaten der drei Segmentierungsmodelle unterscheiden sich optisch deutlich (siehe Abb. 5.1). Masken des nnUNet-Modells sehen kantiger aus als die Masken anderer Modelle und weisen oft eine raue Kontur mit hervorstehenden Extremitäten oder Einkerbungen auf. Deepcell-Masken sehen glatt und ausgebreitet aus. Durch die Watershed-Nachverarbeitung sind teilweise sichtbare Artefakte entstanden, an Stellen, an denen das Trennen der Instanzen für Watershed nicht möglich ist. Die CellposeSAM-Masken sehen intuitiv am besten aus, sie haben oft nahezu eliptische Konturen und trennen Instanzen so wie es für das menschliche Auge sinnvoll erscheint. Das spiegelt sich auch in deutlichen Unterschieden in der IPQ wider.

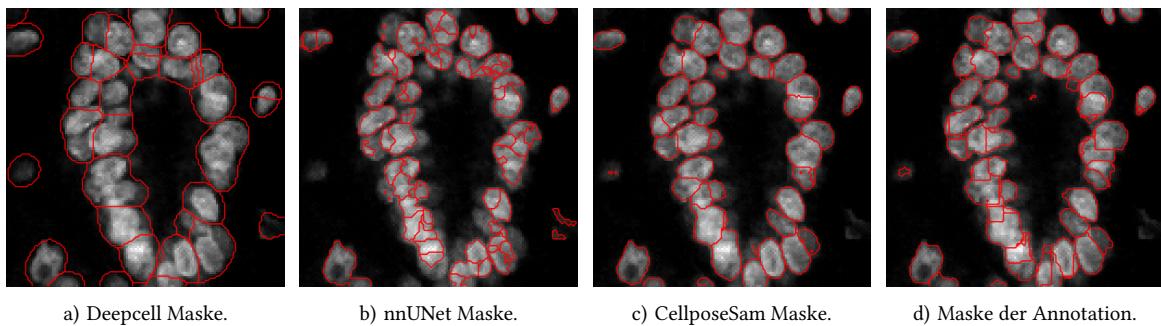


Abb. 5.1 | Darstellung der Segmentierungsmasken der verschiedenen Segmentierungsmodelle sowie der Annotation als Konturen auf einem zweidimensionalen Durchschnitt einer dreidimensionalen Stichprobe des S_BIAD1518-Datensatzes.

Die Ergebnisse jedes Segmentierungsmodells sind einzeln und für jedes Bild im Anhang A.2 angehängt. Eine Zusammenfassung der IPQ-Ergebnisse ist in den Boxplots in Abb. 5.2 gegeben. Das zentrale Ergebnis ist, dass CellposeSAM die besten IPQ-Werte liefert. Mit einem Mittelwert von 0,64 liegt die IPQ von CellposeSAM über dem Mittelwert bei nnUNet (0,04) und Deepcell (0,02), wie höchst signifikante einseitige t-Tests zeigen. Dennoch zeigt sich, dass CellposeSAM lediglich in der Kategorie SQ den höchsten Mittelwert aufweist. Die RQ der nnUNet-Masken ist höher als die der CellposeSAM-Masken. Ebenso ist die IQ der Deepcell-Masken höher als die IQ der CellposeSAM-Masken. Beide Beobachtungen sind durch höchst signifikante t-Tests gestützt.

Obwohl nnUNet und Deepcell jeweils eine Metrik dominieren, wird ihr IPQ-Wert durch die beiden niedrigsten Faktoren stark heruntergezogen, während CellposeSAM in jeder Metrik gut, wenn auch nicht am besten, abschneidet. Außerdem zeigen die Boxplots viele Ausreißer in den Daten, was den Unterschieden zwischen den Bildkategorien, die im Datensatz enthalten sind, geschuldet sein kann.

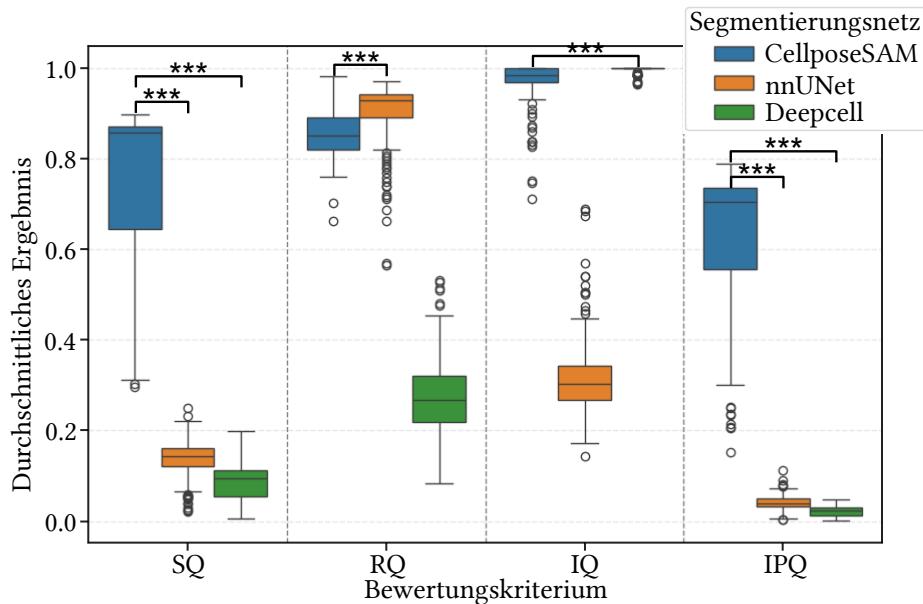


Abb. 5.2 | Ergebnisse der IPQ-Berechnungen mit den Faktoren $k_i = 1$ (siehe Formel (3.3)). Die X-Achse unterteilt die Daten in die Kriterien Segmentation-Quality (SQ), Recognition-Quality (RQ), Injectve-Quality (IQ) und Injektive Panoptische Qualität (IPQ), wie in der Formel (3.3) beschrieben. Für jede Metrik sind drei farbige Boxplots zu sehen, jeweils einer pro Segmentierungsmodell. Die Sterne zeigen die Signifikanz relevanter t-Tests an.

In Abb. 5.3 sind Beispiele für die einzelnen Fehlerarten zu sehen. Das IQ-Beispiel zeigt die Verletzung injektiver Abbildung der Nuclei auf die Annotation anhand einer nnUNet-Maske. Zentral ist hier ein nicht elliptischer Nucleus zu sehen.

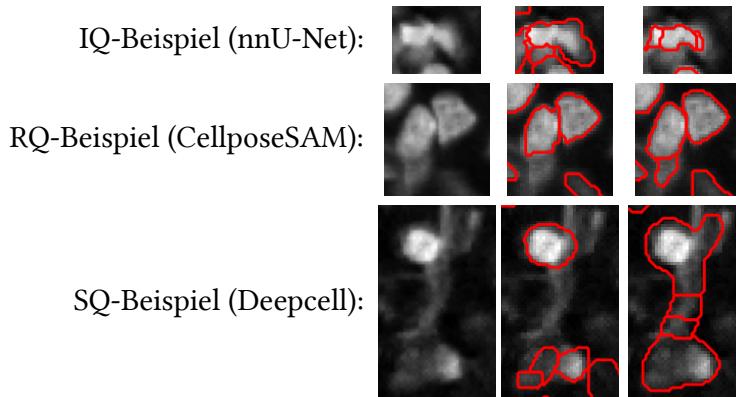


Abb. 5.3 | Exemplische Darstellung einzelner Ausprägungen verschiedener Fehlerarten. Hierzu sind jeweils, in dieser Reihenfolge von links nach rechts, ein Ausschnitt eines 2D-Durchschnitts einer Stichprobe, die Annotation als Kontur und die vorhergesagte Maske eines Modells zu sehen. Die erste Zeile zeigt ein Beispiel für einen schlechten IQ-Wert anhand einer nnUNet-Maske. In der zweiten Zeile ist ein RQ-Fehler anhand einer CellposeSAM-Maske zu sehen und unten ist ein Beispiel für einen schlechten SQ-Wert anhand einer Deepcell-Maske zu sehen.

Dieser Nucleus wird vom nnUNet-Modell in mehrere Segmente unterteilt, wodurch die räumliche Konzentration der Nuclei überschätzt wird. Nicht-eliptische Nuclei führen häufig zu dieser Art von Fehlern.

Mit dem Fehler gehen **RQ**-Fehler einher, weil zwei der Segmente im Vergleich zur Annotation zu klein sind, um als **TP** erkannt zu werden. In der zweiten Zeile ist anhand einer CellposeSAM-Maske ein **RQ**-Beispiel dargestellt. Das Segmentierungsmodell hat hier einen Nucleus halluziniert. Der Nucleus unten links, direkt unter dem Linken der beiden großen Nuclei, ist in der Annotation nicht zu finden. Dieser Fehler geht nicht mit einem schlechteren **SQ**- oder **IQ**-Wert einher. Durch den Fehler wird die Anzahl der Nuclei überschätzt.

Zuletzt ist ein **SQ**-Beispiel dargestellt. Neben und unter dem sichtbaren Nucleus oben sind vermutlich Schatten von Nuclei aus anderen Z-Ebenen zu sehen, die vom Deepcell-Modell als Nuclei erfasst wurden. Dadurch ist das Segment, das den sichtbaren Nucleus abbildet, deutlich zu groß, was die **IoU**-Werte beeinträchtigt. Mit diesem Fehler geht auch ein schlechterer **RQ**-Wert einher, da der Schatten zu mehreren Halluzinationen geführt hat. Da diese Halluzinationen nicht mit den Annotationssegmenten überschneiden, ist der **IQ**-Wert hier nicht betroffen. Der Fehler führt zu einer Fehleinschätzung des Nucleivolumens.

Abb. 5.4 zeigt ein Beispiel für eine Instanzsegmentierungsmaske der Zieldaten mit dem CellposeSAM-Modell.

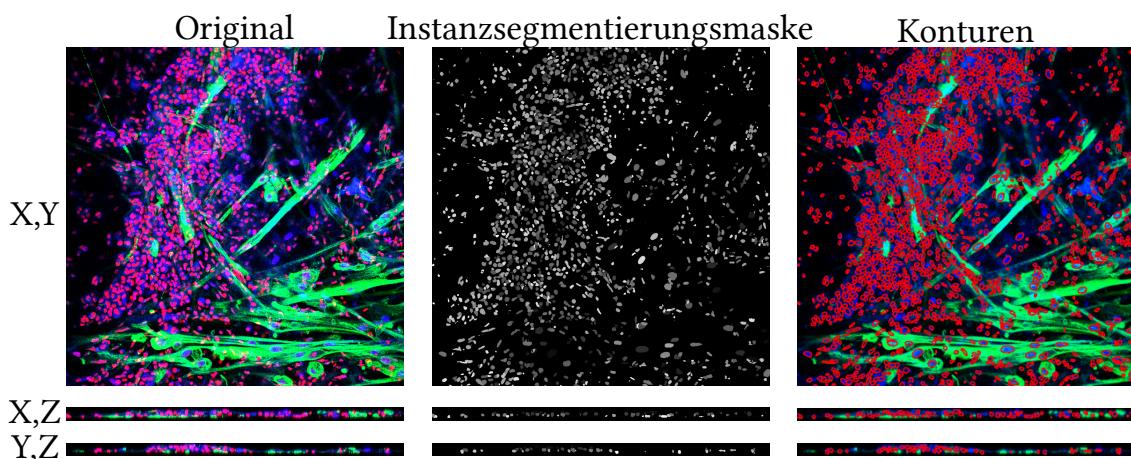


Abb. 5.4 | Exemplarische Darstellung einer Zieldaten-Instanzsegmentierungsmaske des CellposeSAM-Modells. Links ist eine 2D-Schicht des Original-Bildes zu sehen, daneben die Instanzsegmentierungsmaske und rechts eine Überlagerung von den Konturen der Instanzen über die Marker-Kanäle des Originals. Zu jeder der drei Ansichten sind auch 2D-Schnitte der X-Z-Ebene und der Y-Z-Ebene gegeben.

Links sind exemplarisch 2D-Schichten eines 3D-Eingabevolumens zu sehen. Die Schichten sind Schnitte durch die drei Koordinaten-Ebenen des Volumens. Daneben ist die Instanzsegmentierungsmaske als Maskenbild und als Konturen auf den Marker-Kanälen des Eingabebilds dargestellt. Zu sehen ist, dass die CellposeSAM-Masken auch intuitiv sehr gut zur Eingabe passen.

5.4 Klassifikation

5.4.1 Überblick

Die in Abschnitt 3.4 eingeführten Methoden zur Klassifikation werden anhand eines separaten Anteils des manuell annotierten Datensatzes durch die neu entwickelte 3D-Zelldaten-Pipeline getestet. Der manuell annotierte Datensatz enthält 125 Stichproben von Myotuben-Zellkernen, 95 Stichproben der Klasse Debris, 148 Stichproben der Klasse „Andere“ und nur 16 Stichproben der Klasse Schwannzellen-Nucleus. Die Klassen Eins bis Vier sind hier der Reihe nach Myotuben-Kerne, Debris, „Andere“ und Schwannzellen-Kerne. Der Abschnitt A.3 des Anhangs zeigt alle Ergebnisse der verschiedenen Methodenkominationen tabellarisch. Um die Genauigkeit der Klassifikatoren anhand des Test-Anteils möglichst repräsentativ zu bestimmen, werden manuell 50% statt 20% der Schwannzellen-Nuclei in den Test-Anteil der Daten eingesetzt. Die Genauigkeit, also der prozentuale Anteil richtiger Vorhersagen, auf dem Test-Anteil des Zieldatensatzes wird als Kriterium verwendet.

Da die Modelle während des Trainings rauschbehaftete Verläufe der Genauigkeit aufweisen und es zu Overfitting kommen kann, wird Early-Stopping implementiert, wobei die mittlere Genauigkeit des Klassifikators auf dem Test-Anteil der manuell annotierten Daten über fünf Epochen hinweg gemessen wird. Dazu wird die geglättete Genauigkeit berücksichtigt. Wenn diese viermal in Folge sinkt oder gleich bleibt, wird das Training unterbrochen. Abb. 5.5 zeigt exemplarisch zwei Trainingsverläufe, die dieses Phänomen belegen.

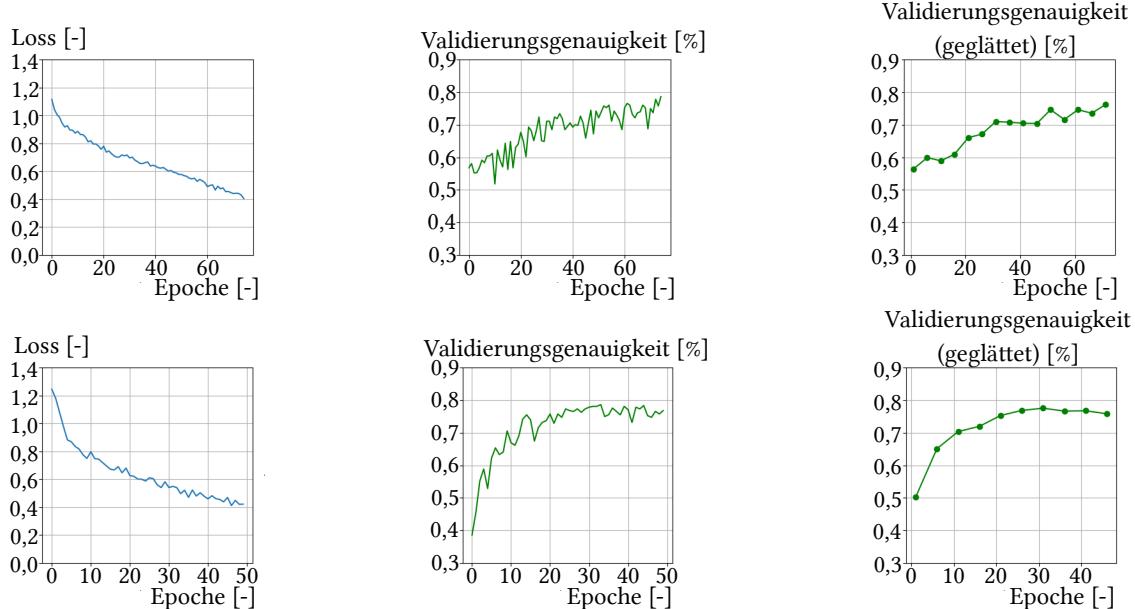


Abb. 5.5 | Die Diagramme zeigen den Trainingsverlauf eines Klassifikators. Links ist der Verlauf des Loss zu sehen. In der Mitte und rechts sind die Verläufe der Validierungsgenauigkeit zu sehen. Rechts ist die Kurve zur Anschaulichkeit durch Mittelung von je fünf Werten geglättet. Der obere Verlauf zeigt einen Klassifikator ohne Overfitting, im unteren Verlauf ist Overfitting zu erkennen.

Im oberen Verlauf ist zu sehen, dass die Genauigkeit des Klassifikators tendenziell steigt. Die geglättete Genauigkeit zeigt hier zwar rauschbehaftetes Verhalten mit einigen lokalen Minima, aber nach ein bis zwei Berechnungsschritten steigt der Wert wieder. Der Verlauf ist zwar nicht monoton steigend, aber es liegt nahe, dass bei Fortsetzung des Trainings weitere Verbesserungen zu erwarten sind. Im unteren Verlauf ist Overfitting zu erkennen, da die geglättete Genauigkeit über vier Berechnungsschritte hinweg gesunken oder gleich geblieben ist. Der Trainings-Loss nimmt in den gleichen Epochen weiter ab. Das Modell passt sich also weiter an die Trainingsdaten an, ohne dass es besser darin wird, ungesenehe Nuclei zu erkennen. Die Genauigkeit auf dem Test-Anteil des Datensatzes sinkt etwa ab Epoche 30 und erreicht ihr Maximum in Epoche 32. Dementsprechend wird die Genauigkeit des Klassifikators in Epoche 32 für den Vergleich herangezogen. Mit der Prädiktion erweitern die Klassifikatoren die Instanzsegmentierungsmasken zu panoptischen Segmentierungsmasken. Abb. 5.6 zeigt exemplarisch eine panoptische Maske.

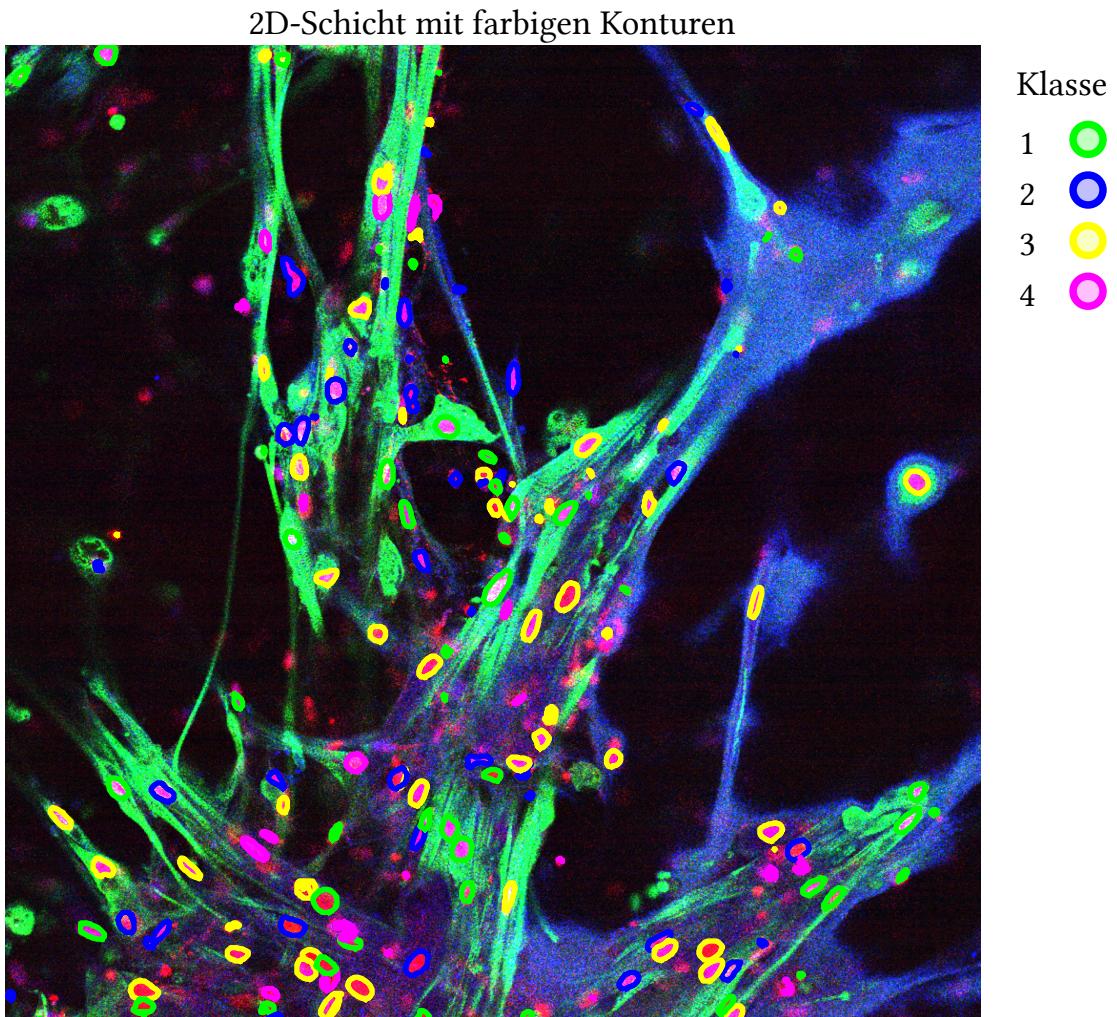


Abb. 5.6 | Panoptische Segmentierungsmaske, erstellt durch die Kombination der Instanzsegmentierungs-maske des CellposeSAM-Segmentierungsmodells und der prädizierten Klassen eines Klassifikators.

Die Klassen eins bis vier sind der Reihe nach Myotuben-Kerne, Debris, „Andere“ und Schwannzellen-Kerne. Die Konturfarben der Nuclei kennzeichnen ihre Klassen.

Ganz rechts in der Mitte der 2D-Schicht ist ein Beispiel für eine fehlklassifizierung zu sehen. Die Kontur ist gelb, was bedeutet, die Vorhersage ist die Klasse „Andere“. Tatsächlich spricht aber die grüne Umgebung des Nucleus dafür, dass es sich um einen Myotuben-Zellkern handelt, der untypischerweise nicht elliptisch, sondern annähernd kreisförmig ist, weil die Muskelfaser orthogonal zur Bildebene verläuft.

5.4.2 Encoder

In Abb. 5.7 ist die Genauigkeit der Klassifikatoren pro Encoder gegeben. Zu sehen ist sowohl das Maximum, als auch der Durchschnitt der Genauigkeitswerte der verschiedenen Methodenkombinationen jedes Klassifikators. Die Klassifikatoren sind nach dem aufsteigenden Maximalwert sortiert.

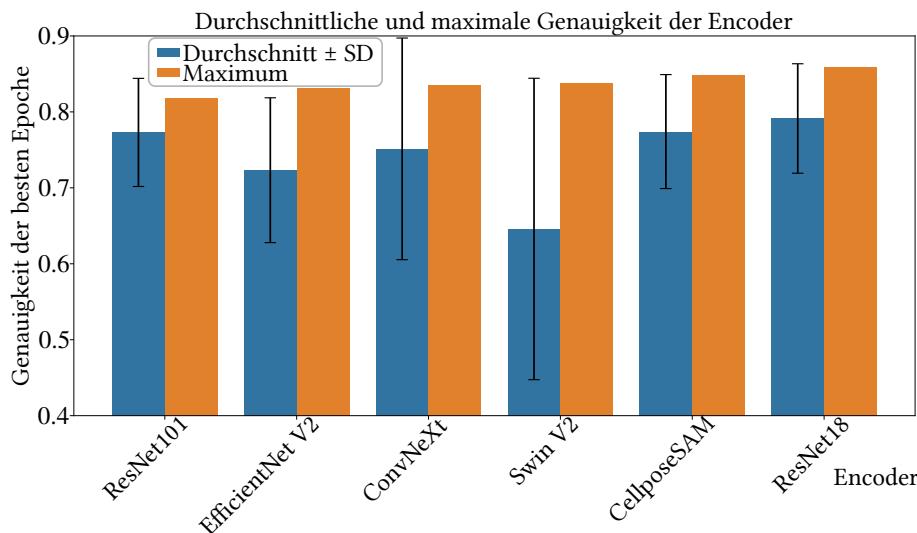


Abb. 5.7 | Das Balkendiagramm zeigt an der Y-Achse die Genauigkeit der Klassifikatoren bei der Verwendung der verschiedenen Encoder. Auf der X-Achse sind die Encoder in Gruppen angeordnet. Jede Gruppe enthält einen Maximalwert (Orange) und einen Durchschnittswert (Blau), da jeder Encoder mit verschiedenen Kombinationen von Methoden getestet wird. Die Encoder sind nach aufsteigendem Maximalwert von links nach rechts sortiert.

Zu sehen ist, dass der beste Wert vom ResNet18-Encoder stammt. Mit 85,9% Genauigkeit auf dem Test-Anteil des Datensatzes ist diese Kombination das gefundene Optimum. Auch im Mittelwert haben die Klassifikatoren mit dem ResNet18-Encoder mit 79,1% die höchste Genauigkeit. CellposeSAM und ResNet101 haben als Encoder die zweit- und dritthöchsten durchschnittlichen Genauigkeiten mit 77,4% und 77,3%, jeweils. Während CellposeSAM allerdings die zweithöchste maximale Genauigkeit aufweist, hat ResNet101 die geringste. Besonders auffällig ist Swin V2. Mit 64,6% hat es die niedrigste durchschnittliche Genauigkeit, aber der Maximalwert von 83,9% ist der dritthöchste. Besonders häufig werden

mit beiden Klassifikations-Köpfen die Klassen Schwannzellen-Kern mit Myotuben-Kernen und die Klasse „Andere“ mit Debris verwechselt. Abb. 5.8 zeigt einige Beispiele dieser Verwechslung.

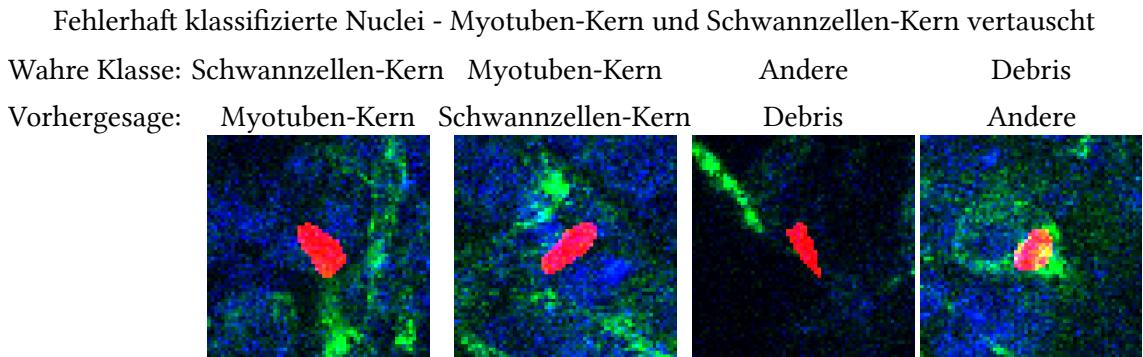


Abb. 5.8 | Beispiele für die Verwechslung von Schwannzellen-Kernen und Myotuben-Kernen sowie „Andere“ und Debris

5.4.3 Vortraining

Abb. 5.9 zeigt eine Übersicht über die Effektivität der verwendeten Vortrainingsmethoden.

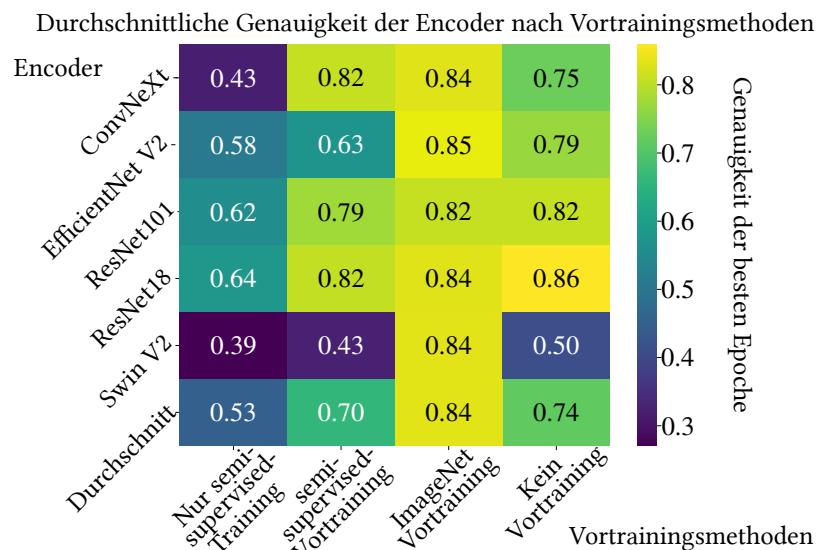


Abb. 5.9 | Die Heatmap zeigt die Genauigkeiten der Klassifikatoren bei der Verwendung der verschiedenen Vortrainingsmethoden. Auf der Y-Achse sind die verschiedenen Encoder aufgeführt. Die X-Achse zeigt die getesteten Methoden des Vortrainings an. Die Farbe der Felder und der Wert darin zeigen die Durchschnittsgenauigkeiten aller Klassifikatoren, die die jeweilige Vortrainingsmethode nutzen. Die letzte Zeile zeigt den Durchschnitt der Werte aller Encoder, abhängig von der Vortrainingsmethode.

Um die Ergebnisse möglichst unabhängig von den anderen Klassifikatormethoden zu betrachten, werden hier nur die Klassifikatoren mit der besten Vorverarbeitungsmethode für

den Encoder und dem Volumen-Klassifikator als Klassifikations-Kopf betrachtet. Auf der vertikalen Achse sind die verschiedenen Encoder sowie der Durchschnitt aller Encoder zu sehen. Auf der horizontalen Achse sind die vier Vortrainingsmethoden aufgeführt. Links ist nur semi-supervised-Training zu sehen. Daneben steht das semi-supervised-Vortraining, also das Vortraining mit den Pseudo-Labler-Annotationen, Einfrieren der Encoder-Gewichte und Fortsetzen des Trainings mit vollständig annotierten Daten zu sehen. Rechts ist das fully-supervised ImageNet-Vortraining mit eingefrorenen Encoder-Gewichten und „kein Vortraining“ aufgetragen (siehe Abschnitt 3.4.4).

Zu sehen ist, dass das Vortraining mit semi-supervised-Daten durchschnittlich deutlich schlechtere Ergebnisse liefert als das fully-supervised ImageNet-Vortraining und etwas schlechtere Ergebnisse als kein Vortraining. Im Durchschnitt führt das Fortführen des Trainings mit den annotierten Daten nach dem semi-supervised-Vortraining zu einer Steigerung der Genauigkeit um 13 Prozentpunkte. Für den ConvNeXt-Encoder ist der Vorteil des fortgesetzten Trainings besonders hoch mit 39 Prozentpunkten. Der ConvNeXt-Encoder ist außerdem mit dem semi-supervised-Vortraining um sieben Prozentpunkte genauer als ohne Vortraining. Auch die beiden ResNet-Modelle erzielen mit dem semi-supervised-Vortraining ähnlich gute Ergebnisse wie mit anderen Vortrainingsmethoden. Im Vergleich zum ImageNet-Vortraining, das deutlich mehr Rechenaufwand erfordert, sind die Klassifikatoren mit ResNet18 und ResNet101 nur um zwei bzw. drei Prozentpunkte schlechter, wenn sie die Annotationen des Pseudo-Lablers als Vortraining nutzen. Beide haben außerdem eine höhere Genauigkeit als der Durchschnitt ohne Vortraining. Sie sind beide auch ohne Fortsetzung des Trainings nach dem semi-supervised-Vortraining bereits 64% und 62% genau.

Besonders auffällig ist auch hier das Swin V2-Modell, das ohne fully-supervised-Vortraining nur eine durchschnittliche Genauigkeit von 44,0% erreicht. Vortraining auf dem ImageNet-Datensatz führt im Vergleich dazu bei diesem Encoder zu einer Steigerung der Genauigkeit um 40 Prozentpunkte. Wie in Abschnitt 5.4.2 beschrieben, hat die Genauigkeit des Swin V2-Encoders einen hohen Maximalwert, aber einen geringen Durchschnittswert. Der Vergleich der Vortrainingsmethoden zeigt, dass das an der schlechten Leistung des Swin V2-Encoders ohne ImageNet-Vortraining liegt.

Abb. 5.10 zeigt zwei Scatterplots. Beide visualisieren Projektionen des Merkmalsraums des Swin V2-Encoders in zwei Dimensionen mittels t-SNE. Die linke Abbildung zeigt den Merkmalsraum des Encoders ohne Vortraining und die rechte den Merkmalsraum nach ImageNet-Vortraining. Zu sehen ist, dass der Swin V2-Encoder keinen sinnvollen Merkmalsraum erlernt. Das Vortraining auf dem ImageNet-Datensatz ist umfangreicher und diverser als auf den Zieldaten und führt dadurch zu einem generalisierten Merkmalsraum. ImageNet-Vortraining liefert im Durchschnitt die besten Ergebnisse mit einer Genauigkeit von 84%. Außerdem ist die Varianz entlang der Encoder mit dem ImageNet-Vortraining ($5,6 \cdot 10^{-5}$) deutlich geringer als ohne Vortraining (0,016). Dennoch erreicht ResNet18 ohne Vortraining die beste Genauigkeit mit 86%.

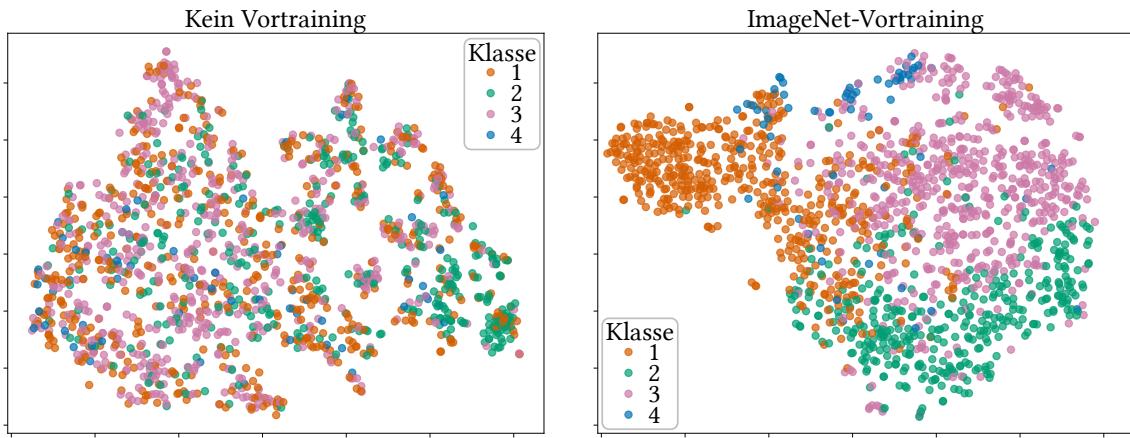


Abb. 5.10 | 2D-Projektionen des Merkmalraums zweier Klassifikatoren mittels t-SNE. Der linke Scatterplot zeigt den Merkmalsraum eines Swin V2-Encoders ohne Vortraining, der rechte den eines ImageNet-vortrainierten Swin V2-Encoders.

Des Weiteren wird die durchschnittliche Genauigkeit pro Klasse betrachtet. Abb. 5.11 zeigt die durchschnittliche Genauigkeit pro Klasse und Sterne als Hinweise auf signifikante Ergebnisse von paarweisen t-Tests. In jedem Balken des Diagramms ist je ein Beispiel der Klassen ohne Vorverarbeitung als 2D-Schnitt gegeben. In Rot sind die Nucleus-Kanäle dargestellt, Grün und Blau sind Marker-Kanäle. Die Nuclei der Klasse Myotuben-Kern Schwannzellen-Kern sind optisch sehr ähnlich, beide sind lang und dünn. Die Unterscheidung ist nur durch die Betrachtung der umliegenden Strukturen möglich.

Wie eingangs beschrieben, sind im Datensatz nur 16 Beispiele der Klasse Schwannzellen-Nucleus enthalten, davon acht im Trainings- und acht im Test-Anteil. Zu sehen ist, dass insbesondere die Schwannzellen-Kerne schlecht erkannt werden. Wie in Abschnitt 3.4.3 beschrieben, wird dieses Problem der ungleichen Verteilung der Klassen mithilfe des Retrievers durch eine Gewichtung der Gradienten unterrepräsentierter Klassen angegangen. In der Abbildung ist jedoch zu sehen, dass dieser Ansatz das Problem nicht vollständig löst. Die unterrepräsentierte Klasse der Schwannzellen-Nuclei wird dennoch schlechter erkannt als alle anderen Klassen. Beinahe alle Methoden ergeben isoliert betrachtet eine ähnliche Verteilung der Genauigkeit pro Klassen, nur das Vortraining hat einen Einfluss. Vortraining mit semi-supervised-Annotationen und anschließendes Fortsetzen des Trainings mit manuell annotierten Daten führen zu Klassifikatoren, die teilweise wesentlich ausgewogener die Klassen erkennen. Mit dem Volumen-Klassifikator und der Masken-Methode zur Vorverarbeitung erreichen der ResNet101-Encoder 83,2%, der ResNet18-Encoder 73,5% und der ConvNeXt-Encoder 56,3% Genauigkeit für die Klasse der Schwannzellen-Nuclei. Die durchschnittliche Genauigkeit für die Schwannzellen-Klasse mit diesen Methodenkombinationen beträgt 71%. Das ist im Vergleich zu den 21,7% durchschnittlicher Genauigkeit für die Schwannzellen-Klasse ein starker Genauigkeitszuwachs. Die durchschnittliche Genauigkeit ist nach dem semi-supervised-Vortraining also geringer, aber dafür wird die unterrepräsentierte Klasse der Schwannzellen-Kerne nicht, wie bei anderen Vortrainingsmethoden, schlechter erkannt.

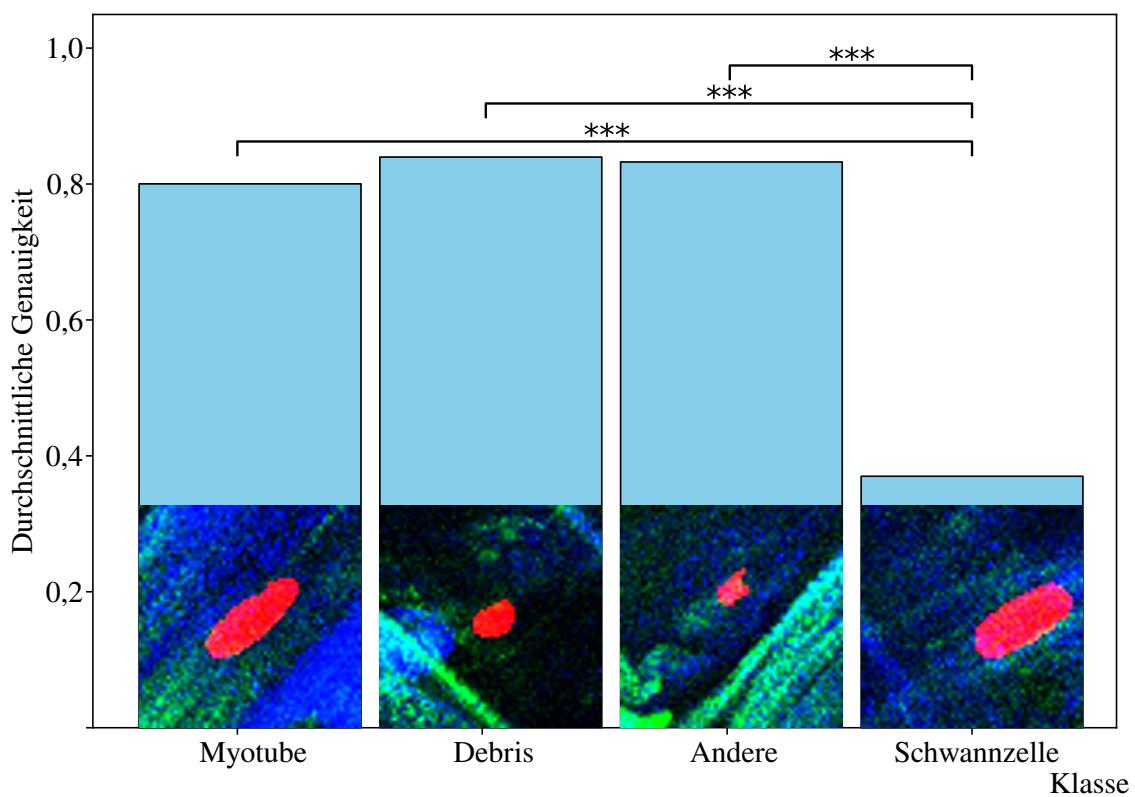


Abb. 5.11 | Durchschnittliche Genauigkeit pro Klasse aller Klassifikatoren. Die Klassen sind Myotuben-Kerne, Debris, „Andere“ und Schwannzellen-Kerne. Unten in den Balken ist je ein Beispiel für die entsprechende Klasse als 2D-Schnitt gegeben.

5.4.4 Klassifikations-Kopf

Abb. 5.12 zeigt eine Gegenüberstellung der Genauigkeit beider Klassifikations-Köpfe in einem Balkendiagramm mit Fehlerbalken. Dabei werden nur die Klassifikatoren berücksichtigt, bei denen sich die Methodenkombination ausschließlich im verwendeten Klassifikations-Kopf unterscheidet. Das ist notwendig, da nicht alle möglichen Kombinationen trainiert wurden und ein Vergleich mit unvollständigen Kombinationen keine aussagekräftigen Ergebnisse liefert. Die blauen Balken des Balkendiagramms zeigen jeweils die Ergebnisse der Architekturen, die den Schichten-Klassifikator nutzen, die orangefarbenen Balken die Ergebnisse des Volumen-Klassifikators. Auf der X-Achse sind als Gruppen zuerst die Encoder und zuletzt der Durchschnitt aufgetragen, und auf der Y-Achse die durchschnittliche Genauigkeit der Modelle mit dem jeweiligen Klassifikations-Kopf. An den Daten lässt sich erkennen, dass der Einfluss des Klassifikations-Kopfes nicht homogen ist. Mit dem ConvNeXt-Encoder ist der Schichten-Klassifikator performanter als der Volumen-Klassifikator, während es bei allen anderen Encodern umgekehrt ist. Für den CellposeSAM Encoder ist der Genauigkeitsunterschied 12,6 Prozentpunkte groß, für EfficientNetV2 einen halben Prozentpunkt.

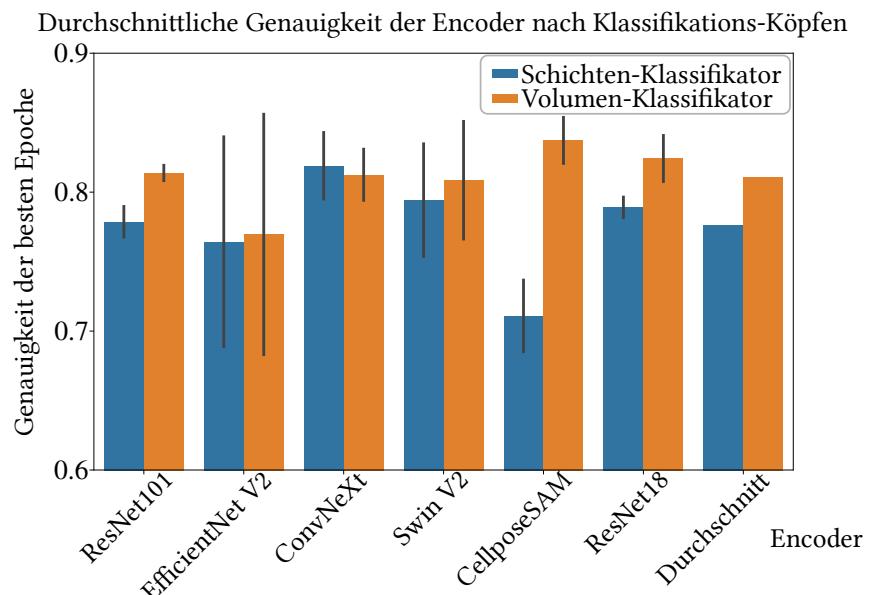


Abb. 5.12 | Genauigkeitswert der Klassifikatoren unter Verwendung eines bestimmten Klassifikations-Kopfes. Auf der Y-Achse ist der Durchschnitt der Genauigkeiten der besten Trainingsepochen aller Klassifikatoren, die diesen Klassifikations-Kopf nutzen, aufgetragen. Die X-Achse zeigt Gruppen von Encodern, jeweils besetzt mit einem Wert für den Schichten- und den Volumen-Klassifikator. Rechts zu sehen sind Balken, die den Durchschnitt der Werte aller Encoder, abhängig von dem Klassifikations-Kopf, zeigen.

Die Daten zeigen also, dass die Intensität des Unterschieds in der Genauigkeit der beiden Klassifikations-Köpfe stark vom Encoder abhängt. Die Balken rechts, die je den Durchschnitt aller Modelle mit einem Klassifikations-Kopf darstellen, zeigen signifikant, dass der Volumen-Klassifikator bessere Ergebnisse liefert.

Mit dem Schichten-Klassifikator werden häufig Nuclei fehlerhaft klassifiziert, die nicht in der mittleren Schicht des Bildausschnitts vorkommen. Die Nuclei sind oft in den obersten Schichten der Eingabedaten zu finden. Wenn sie nicht mindestens zwölf Ebenen von der obersten Z-Ebene entfernt sind, schneidet der Retriever Bildausschnitte aus, in denen die Nuclei in Z-Richtung nicht zentriert sind. Abb. 5.13 zeigt fünf Ebenen eines Nucleus, für die das zutrifft.

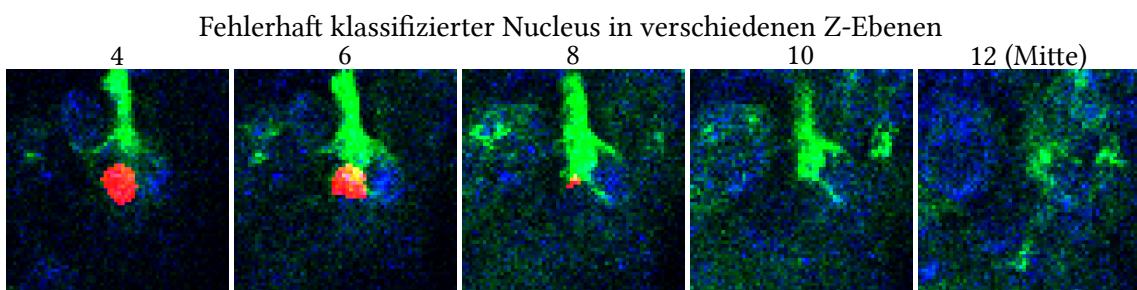


Abb. 5.13 | Fünf 2D-Schichten eines fehlklassifizierten Nucleus. Die Z-Ebene ist über den Abbildungen angegeben. Zu sehen ist, dass der Nucleus nicht zentriert ist und nicht in der mittleren Schicht präsent ist.

Der Nucleus ist mit der Masken-Methode vorverarbeitet und in Rot zu sehen. Er erscheint nicht in den Schichten zehn und zwölf. Schicht acht zeigt gerade die untere Kante des Nucleus.

5.4.5 Vorverarbeitung

In Abb. 5.14 sind die durchschnittlichen Genauigkeiten der Modelle gezeigt, die jeweils eine der Vorverarbeitungsmethoden nutzen.

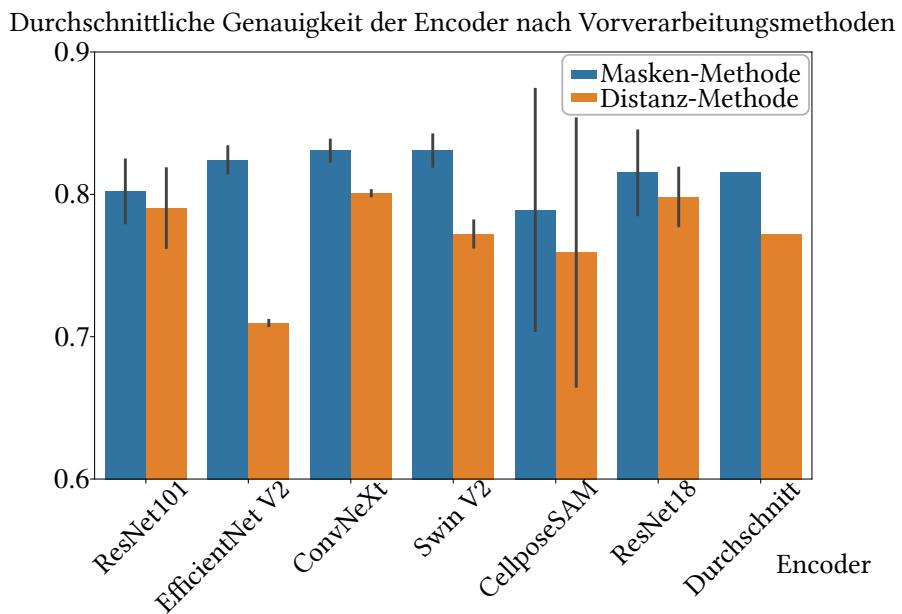


Abb. 5.14 | Genauigkeitswert der Klassifikatoren bei Verwendung einer bestimmten Vorverarbeitungsart. Auf der Y-Achse ist der Durchschnitt der Genauigkeiten der besten Trainingsepochen aller Klassifikatoren, die diese Vorverarbeitungsart nutzen, aufgetragen. Die X-Achse zeigt Gruppen von Encodern, jeweils besetzt mit einem Wert für die Masken-Methode, die den Nucleus-Kanal des Bildes durch die Segmentierungsmaske ersetzt, und die Distanz-Methode, die eine Distanztransformation auf den Nucleus-Kanal anwendet. Rechts zu sehen sind Balken, die den Durchschnitt der Werte aller Encoder, abhängig von der Vorverarbeitungsmethode, zeigen.

Wie für den Vergleich Klassifikations-Kopf-Architekturen sind auch hier die Klassifikatoren ausgeschlossen, deren Methodenkombination mit nur einer der beiden Vorverarbeitungsmethoden trainiert wurde. Auf der Y-Achse ist die Genauigkeit aufgetragen, auf der X-Achse sind die Encoder und der Durchschnitt aller Encoder als Gruppen. Die Ergebnisse zeigen deutlich, dass die Masken-Methode, also das Ersetzen des Nucleus-Kanals durch die Segmentierungsmaske, zu einer höheren Genauigkeit führt. Mit der Distanz-Methode, also dem Anwenden einer Distanztransformation der Segmentierungsmaske auf den Nucleus-Kanal, beträgt die Genauigkeit durchschnittlich 77,2%, mit der Masken-Methode sind es 81,5%. Die Differenz ist hoch signifikant. Zu sehen ist aber, dass der Einfluss der Vorverarbeitung unterschiedlich stark ist, je nach Encoder. Während die Genauigkeit für Modelle

mit dem ResNet101-Encoder mit der Masken-Methode um 0,8 Prozentpunkte gegenüber der Distanz-Methode steigt, macht die Vorverarbeitung beim EfficientNet V2-Encoder 11,5 Prozentpunkte aus.

Mit der Distanz-Methode treten häufig Fehler auf, wenn Nuclei dicht aneinander liegen. Abb. 5.15 zeigt Nuclei mit der Distanz-Methode und der Masken-Methode, für die mit der Distanz-Methode eine schlechtere Genauigkeit erzielt wird als mit der Masken-Methode.

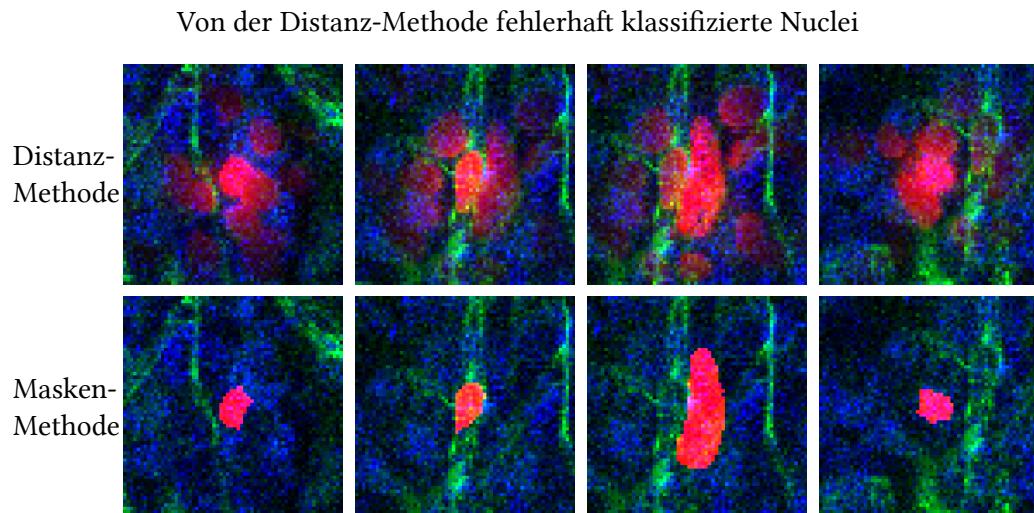


Abb. 5.15 | Beispiele für Nuclei, die häufig mit der Masken-Methode korrekt und mit der Distanz-Methode fehlerhaft klassifiziert werden. In der oberen Zeile sind die Nuclei zu sehen, die mit der Distanz-Methode vorverarbeitet sind. Darunter ist jeweils der gleiche Nucleus mit der Masken-Methode zu sehen.

Die Nuclei in den Beispielen sind alle in Clustern angeordnet und berühren umliegende Nuclei. Mit der Masken-Methode sind die umliegenden Nuclei ausgeblendet, mit der Distanz-Methode gehen sie in den betrachteten Nucleus über. Intuitiv lassen sich die Nuclei auch mit der Distanz-Methode trennen, wie die Ergebnisse zeigen, ist das den Klassifikatoren aber nicht immer möglich.

Mithilfe von Grad-CAM [107] wird außerdem das Verhalten der Klassifikatoren exemplarisch visualisiert. Abb. 5.16 zeigt den Nucleus-Kanal und einen Marker-Kanal aus zwei 2D-Schichten. Die beiden 2D-Schichten sind aus unterschiedlichen 3D-Stapeln, links einer, der mit der Masken-Methode, und rechts einer, der mit der Distanz-Methode vorverarbeitet wurde. Zu sehen sind jeweils Heatmaps für zwei Klassifikatoren. Beide Heatmaps zeigen gradientenbasiert die Aktivierungen in einer Schicht des Merkmalsraums auf die Eingangsdaten zurückprojiziert. Die Heatmaps sind jeweils zur besseren Interpretierbarkeit einer 2D-Schicht der Segmentierungsmaske des Nucleus und eines Marker-Kanals überlagert. Dadurch wird räumlich aufgezeigt, welche Regionen der Eingangsdaten zur Entscheidung für eine bestimmte Klasse besonders beigetragen haben. Die beiden Nucleus-Kanal-Ansichten zeigen, dass Regionen direkt um den Nucleus herum für die Klassifikationsentscheidung wichtiger sind als die Oberfläche des Nucleus. Selbst unter Verwendung der

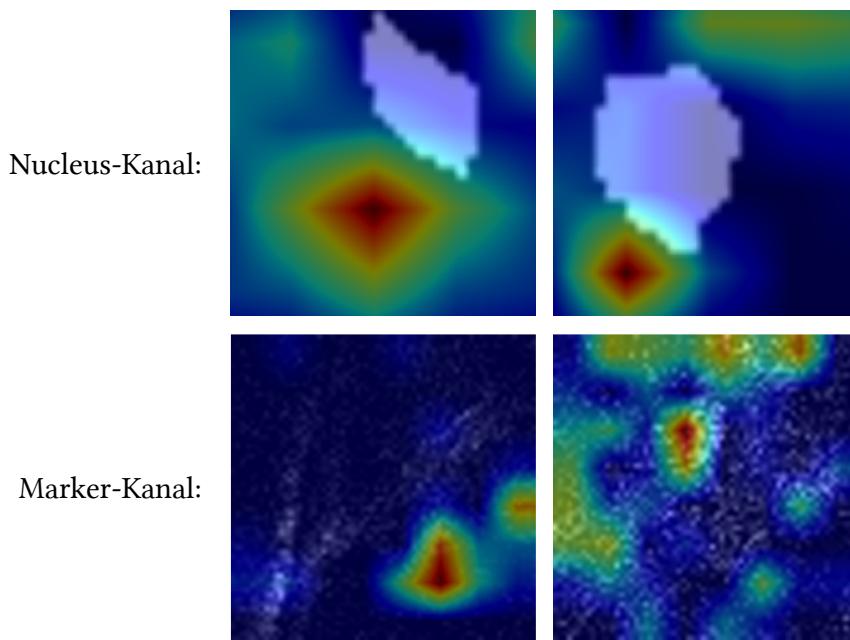


Abb. 5.16 | Grad-CAM-Visualisierung für zwei exemplarische 2D-Schichten, jeweils mit einem Nucleus-Kanal und einem Marker-Kanal. Links ist ein Beispiel einer Stichprobe gegeben, rechts mit der Distanz-Methode. Die Nucleus-Kanäle zeigen die Segmentierungsmasken der Nuclei in Weiß, überlagert mit den gradientenbasierten Heatmaps, die die Wichtigkeit räumlicher Regionen für die Klassifikation darstellen. In beiden Nucleus-Kanälen ist die Wichtigkeit der Nucleus-Oberfläche nur sehr gering. In den Marker-Kanälen ist zu sehen, dass Regionen mit erkennbaren Strukturen besonders wichtig sind.

Distanz-Methode, die zum Ziel hat, Oberflächen-Merkmale zu erhalten und dafür klare Kanten und eine eindeutige Geometrie der Segmentierungsmaske augibt, wird die Oberfläche kaum betrachtet. Des Weiteren sind in den Heatmaps der Marker-Kanäle zu sehen, dass auch abseits des Nucleus-Kanals eine intuitiv sinnvolle Lokalisierung der Aufmerksamkeit des Klassifikators erfolgt. Die Maxima der Heatmaps sind auf sichtbare Strukturen im Marker-Kanal platziert.

5.4.6 Wichtigkeit der Marker

Eine weitere angestellte Untersuchung ist die Quantifizierung der Wichtigkeiten der verschiedenen Marker-Kanäle. Auch für Expert*Innen ist es schwer, eindeutige Beziehungen zwischen den Strukturen, die durch bestimmte Marker gefärbt werden, und den verschiedenen Klassen der Nuclei herzustellen. Zur Abschätzung der globalen Markerrelevanz werden die Gradienten der Modellvorhersage bis auf die Eingangsebene zurückprojiziert und die betragsmäßigen Gradienten über die räumlichen Dimensionen gemittelt. Die Wichtigkeitswerte, die sich ergeben, sind nahezu gleich verteilt.

Abb. 5.17 zeigt exemplarisch Gradientenfelder einer 2D-Schicht aus einer Stichprobe der dreidimensionalen Zieldaten sowie deren Kosinusähnlichkeit. In den rohen Gradienten sind die sensitivsten Bereiche des Modells direkt aus den Ableitungen der Vorhersage be-

züglich der Eingabe ablesbar. Diese Sensitivitäten zeigen starkes Rauschen, weil die Effekte punktweiser Variationen für jeden Pixel der Eingabedaten dargestellt werden. In der mittleren Darstellung sind die integrierten Gradienten dargestellt. Durch die Integration sind Rauscheffekte deutlich reduziert, und eine stabilere, interpretierbarere Verteilung der relevanten Regionen im Gradientenfeld ist sichtbar. Zur Beurteilung der Konsistenz des Klassifikators werden die Ähnlichkeiten zwischen den Gradientenfeldern quantitativ verglichen. Die Felder haben eine Kosinusähnlichkeit von 0,913 und damit eine hohe Übereinstimmung in der Richtung der Wichtigkeitsverteilungen. Der Klassifikator reagiert also auf ähnliche Muster in beiden Feldern. Die Region mit besonders dichten Maxima im Gradientenfeld ist durch rote Pfeile gekennzeichnet. Die Kosinusähnlichkeit ist in der rechten Abbildung pixelweise dem Eingabebild überlagert. Der hohe Wert spricht für eine stabile und konsistente Aufmerksamkeitslokalisierung des Klassifikators, was wiederum auf eine robuste interne Repräsentation der relevanten Merkmale hindeutet.

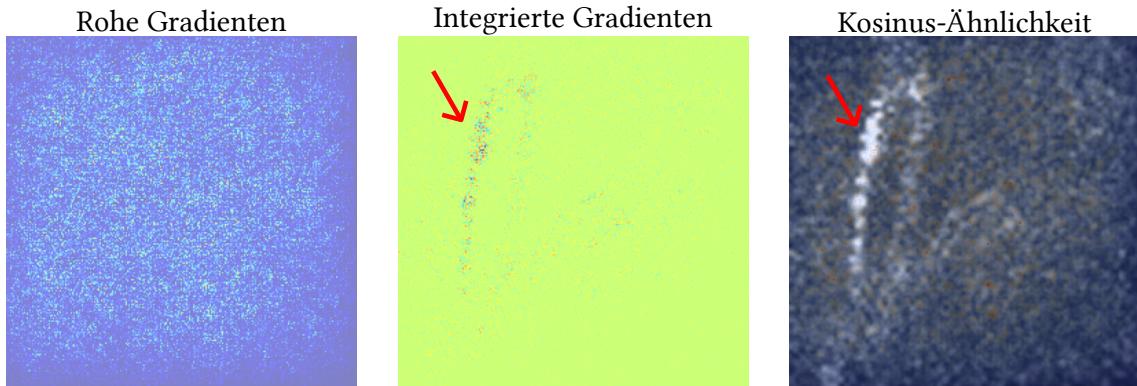


Abb. 5.17 | Gradientenfelder einer 2D-Schicht einer Stichprobe der dreidimensionalen Zieldaten und deren Kosinusähnlichkeit. Die Kosinusähnlichkeit ist der 2D-Schicht überlagert. Das rechte Gradientenfeld ist integriert, um hochfrequente lokale Unterschiede zu entfernen. Die Untersuchung der Kosinusähnlichkeit soll zeigen, ob der Klassifikator konsistent auf bestimmte Strukturen der Eingabe reagiert. Mit den roten Pfeilen ist die Region mit besonders dichten lokalen Maxima der integrierten Maxima gekennzeichnet. Diese Region stellt eine Struktur dar, die für den Klassifikator konsistent zur Entscheidung beiträgt.

Diskussion 6

6.1 Überblick

Die nachfolgenden Abschnitte diskutieren die Ergebnisse der vorliegenden Arbeit. Dabei wird sowohl auf die quantitative Auswertung der durchgeführten Experimente eingegangen als auch auf die Bewertung der angewandten Methoden. Es wird gezeigt, dass die neu entwickelte IPQ-Metrik zur Bewertung von Instanzsegmentierungsmodellen, im Bezug auf ihre Eignung, interpretierbare Merkmale zu extrahieren, geeignet ist. Außerdem wird gezeigt, dass die neu entwickelte Methode des Vortraining einen positiven Einfluss auf die Klassifikationsgenauigkeit haben kann und, dass die 3D-Zelldaten-Pipeline effizient optimale Deep-Learning-Methoden für neue Bilddomänen identifiziert.

6.2 Segmentierung

Durch einen Vergleich der Masken mit den Annotationen in Abb. 5.2 und den zugehörigen Ergebnissen der einzelnen Bewertungskriterien sind die Schwächen und Stärken der individuellen Modelle ersichtlich. Die nnUNet-Masken sind sichtbar kleiner als die Nucleus-Instanzen, was eine schlechte Segmentation-Quality (SQ) bewirkt. Oft zerteilen mehrere nnuNet-Masken eine Nucleus-Instanz, was von der Injectve-Quality (IQ) bestraft wird. Wie auch die gute Recognition-Quality (RQ) zeigt, findet dafür nnUNet sehr zuverlässig die vorhandenen Nuclei mit mindestens einer Maske. Mit dem nnUNet-Modell sind aufgrund der kleinen Segmente, die das Modell vorhersagt, Überschneidungen zwischen Annotations- und Segmentierungsmaske gering. Ein Weg, die Leistung des Modells in Bezug auf die IPQ-Metrik zu verbessern, ist deshalb vermutlich, das Modell, beispielsweise durch fine-tuning, auf größere Segmente anzupassen. Eingangsdaten, beispielsweise durch einen Mittelwert-Filter, kleiner zu dimensionieren und mehrere Bilder aneinander gereiht einzugeben kann auch zu Verbesserungen führen, dabei besteht allerdings das Risiko, dass der Informationsverlust durch den Filter das Ergebnis negativ beeinflusst. Die Instanzen können in der Nachbearbeitung auch durch Dilatation vergrößert werden. Das erhöht vermutlich die SQ, ändert aber nichts an der schwachen IQ.

Deepcell (siehe Abb. 5.1a) überschätzt die Nuclei, wodurch die SQ stark abnimmt. Das Ergebnis sind Masken, die intuitiv zu groß sind und oft mehr als einen Nucleus enthalten. Das bedeutet auch, dass einige Nuclei nicht von einer eigenen Maske identifiziert werden, wodurch sie als FN-Instanzen kategorisiert und eine schlechte RQ bedingt. Durch diese Übersegmentierung wird vermieden, dass Instanzen der Annotation durch die Deepcell-

Masken geteilt werden, was zu einer guten **IQ** führt. Außerdem ist durch die großen Segmente die **SQ** besonders schlecht, weil die Größe der Segmente den Nenner der **IoU** erhöht. Die Deepcell-Masken können hinsichtlich der **IPQ** von weiterer Nachverarbeitung profitieren. Mithilfe einer Erosion können die Masken kleiner gemacht werden, was auch zu einer besseren Leistung der Watershed-Nachverarbeitung führen kann. Umgekehrt kann nach der Erosion die Maske auch Kerben an der Kontur einzelner Instanzen aufweisen, die zur Spaltung der Instanz durch die Watershed-Nachverarbeitung führen. Auch für das Deepcell-Modell ist fine-tuning zur Anpassung an die Größe der Nuclei vermutlich ein Weg, die Qualität zu verbessern.

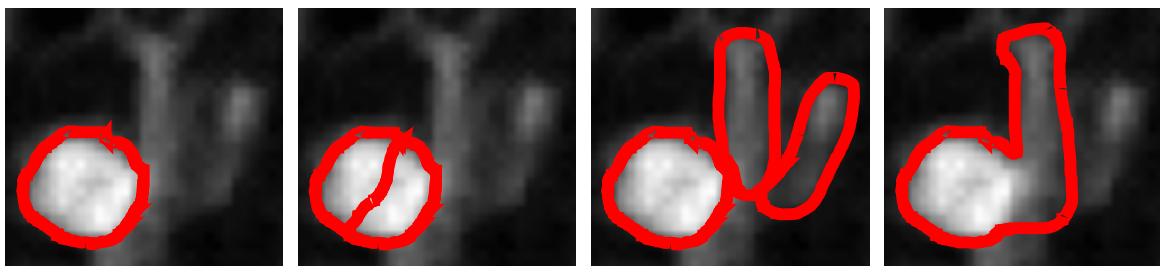
Sowohl für das nnUNet- als auch für das Deepcell-Modell gibt es zwar Vorschläge für Methoden zur Verbesserung der **IPQ** auf dem Benchmark-Datensatz, aber die Übertragbarkeit dieser Methoden auf die Zieldaten ist fraglich. Die Bilddomäne sowie Eigenschaften wie Größe, Exzentrizität und Rundheit der Zieldaten unterscheiden sich von denen des Benchmarks. Da keine Annotationen für die Instanzsegmentierung der Zieldaten verfügbar sind, wird von der Anwendung der Methoden dieser Arbeit abgesehen, um kein Overfitting auf den Benchmark-Datensatz zu riskieren.

Anhand der Interpretation der Werte und der exemplarischen Fehler in Abb. 5.3 wird die Effektivität der neu entwickelten **IPQ**-Metrik ersichtlich. Verschiedene Fehlerarten, die zu unterschiedlichen Verfälschungen der interpretierbaren Merkmale der Daten führen, werden gezielt von der Metrik erfasst. Die adressierten Fehlerarten umfassen:

- Linien auf der Oberfläche eines Nucleus, die als Kontur des Nucleus interpretiert werden. Der Nucleus wird geteilt und zwei getrennte Instanzen werden prädiziert,
- Schatten von Strukturen in anderen Z-Ebenen des 3D-Volumens oder sonstige Artefakte im Nucleus-Kanal, die einem Nucleus überlagert sind, werden als Teil des Nucleus prädiziert und
- Artefakte im Nucleus-Kanal, die als eigenständiger Nucleus prädiziert werden.

In der Praxis gehen die Fehler oft mit Fehlern einer anderen Art einher, was darauf zurückzuführen ist, dass die Modelle nicht demselben Denkverhalten folgen wie menschliche Beobachter*Innen. Abb. 6.1 zeigt idealisiert die Fehlerarten, die mit der **IPQ**-Metrik erkannt werden. Alle Fehler führen für den betrachteten Nucleus zu einem Faktor von 0,5 in ihrer Kategorie und zu einem optimalen Wert von Eins in den anderen Kategorien. Hier wird deutlich, dass die Metrik geeignet ist um:

- Fehleinschätzungen der Konzentration von Nuclei durch den neu eingeführten **IQ**-Wert zu bestrafen,
- Fehleinschätzungen des Volumens der Nuclei durch einen geringen **SQ**-Wert zu bestrafen und
- Fehleinschätzungen der Anzahl der Nuclei durch einen geringen **RQ**-Wert zu bestrafen.



a) Annotationsmaske als Kontur.
b) Idealer IQ-Fehler. Ein optimal segmentierter Nucleus wird in zwei Segmente gespalten.
c) Idealer RQ-Fehler. Es werden zusätzliche Nuclei halluziniert.
d) Idealer SQ-Fehler. Der Nucleus wird in doppelter Größe vorhergesagt.

Abb. 6.1 | Darstellung der Segmentierungsmasken verschiedener idealer Fehler und der Annotation als Kontur.

Von der Metrik wird allerdings nicht erfasst, inwieweit die Geometrie mit der Geometrie der Annotation übereinstimmt. Die **SQ** erfasst zwar, ob die Fläche der Segmentierungsmaske sich mit der Annotation deckt, aber Veränderungen der geometrischen Eigenschaften werden nicht gezielt bestraft. Änderungen der Geometrie sind gegenüber einfachen Änderungen der Fläche besonders zu bestrafen, da der Klassifikator anhand der Segmentierungsmaske im Anschluss die Klasse des Nucleus hervorsagt und bei der Klassifikation die Form des Nucleus besonders wichtig ist.

6.3 Klassifikation

Die Interpretation der Ergebnisse der Genauigkeit von Klassifikatoren mit verschiedenen Encodern, Klassifikations-Köpfen, Vorverarbeitungsmethoden und Vortrainingsmethoden zeigt, dass die 3D-Zelldaten-Pipeline effizient optimale Klassifikator-Methoden identifiziert. Wie die Ergebnisse zeigen, ermöglicht die Anwendung der 3D-Zelldaten-Pipeline das Training eines Klassifikators, der deutlich leistungsfähiger ist als ein Klassifikator mit zufällig gewählten Methoden – und das ganz ohne Programmierkenntnisse oder Vorerfahrung mit Deep-Learning-Methoden seitens der Anwender*Innen.

Die Ergebnisse zeigen, dass die beiden ResNet-Encoder im Durchschnitt besonders hohe Genauigkeiten erzielen. Da die ResNet-Encoder auch die Encoder mit der geringsten Parameteranzahl sind, liegt die Vermutung nahe, dass die gewählten Encoder zu groß sind und schlechte Ergebnisse einem Architektur-bedingten Overfitting geschuldet sind.

Abb. 6.2 zeigt die durchschnittliche Genauigkeit aller Klassifikatoren als Funktion der Parameterzahl des verwendeten Encoders. Große Modelle sind anfälliger für Overfitting. Intuitiv ist aber kein Zusammenhang zwischen der Parameteranzahl und der durchschnittlichen Genauigkeit zu erkennen und auch die Spearman-Korrelationsanalyse deutet nicht auf einen Zusammenhang hin. Das bedeutet, der Hauptgrund, warum Swin V2, EfficientNet V2 und ConvNeXt unter dem Durchschnitt abschneiden, ist nicht, dass der Umfang der Daten zu gering ist, um die Parameter ausreichend zu trainieren. Da ResNet18 der genaueste Encoder ist, liegt dennoch die Vermutung nahe, dass noch kleinere Encoder die

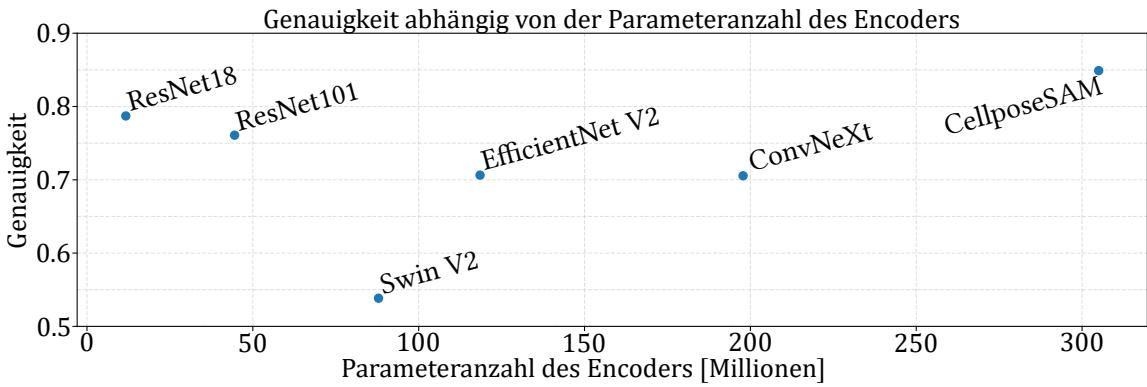


Abb. 6.2 | Durchschnittliche Genauigkeit aller Klassifikatoren abhängig von der Parameteranzahl ihres Encoders. Die X-Achse zeigt die Parameteranzahl, skaliert in Millionen, und die Y-Achse die durchschnittliche Genauigkeit.

Genauigkeit weiter erhöhen können.

Die Architektur und die Gewichte der Encoder bestimmen den Merkmalsraum, den die Encoder mit der internen Repräsentation der Eingangsdaten aufspannen. Dieser Merkmalsraum muss für eine präzise Klassifikation relevante Merkmale des klassifizierten Objekts abbilden. Der Swin V2-Encoder verfügt über eine Transformer-Architektur und schneidet im Durchschnitt besonders schlecht ab. Vermutlich sind die Ansprüche an den Trainingsumfang für Transformer-Architekturen besonders hoch, und in den relativ kurzen Trainingsabläufen wird kein ausreichender Merkmalsraum gelernt. Für die CNN-Architekturen ist ein kurzes Training weniger problematisch. Bei Betrachtung ausschließlich der CNN zeigt sich ein negativer Zusammenhang zwischen Parameteranzahl des Encoders und der durchschnittlichen Genauigkeit, allerdings ist die Stichprobenanzahl für eine Korrelationsanalyse sehr gering.

Der Volumen-Klassifikator ist signifikant besser als der Schichten-Klassifikator (siehe Abschnitt 5.4). Diese Erkenntnis legt nahe, dass die dreidimensionalen Faltungsschichten besser den Zusammenhang der Merkmale erfassen können als der Attention-Mechanismus des Schichten-Klassifikators. Das kann daran liegen, dass die dreidimensionale Faltung räumliche Mittlungen nur mit gelernten Gewichten durchführt, während die Spatial Average Operation die räumliche Beziehung der Merkmale großflächig vernachlässigt. Außerdem ist es möglich, dass die für die Klassifikation ausschlaggebenden Informationen entlang der Z-Dimension nicht ausreichend erhalten werden, was zu schlechter Leistung des Z-Richtung-fokussierten Schichten-Klassifikators führt. Stattdessen kann auch die Kombination aus räumlichen Merkmalen und Merkmalen entlang der Z-Achse wichtig sein, was die gute Leistung des Volumen-Klassifikators bedingen kann. Da der Schichten-Klassifikator mit dem ConvNeXt Encoder im Durchschnitt höhere Genauigkeiten erzielt, als mit dem Volumen-Klassifikator, ist die Abwägung zwischen den beiden Methoden dennoch für zukünftige Anwendungen sinnvoll. Bestimmte Encoder erfassen die relevanten Merkmale auf unterschiedliche Weise, und beide Klassifikations-Köpfe haben das Potential, Zusammenhänge aus bestimmten Merkmalsräumen besonders gut zu

erfassen. Besonders da die ImageNet-Encoder nicht auf dreidimensionale Eingabedaten vortrainiert werden, ist die Abwägung zwischen einem Fokus auf Merkmale entlang der Z-Achse und räumlichen Merkmalen essenziell. Auch für zukünftige Encoder besteht die Möglichkeit, dass sie Beziehungen von Merkmalen innerhalb einer 2D-Schicht der Eingabedaten so effizient und einheitlich zusammenfassen, dass der Attention-Mechanismus des Schichten-Klassifikators die Informationen besser als der Volumen-Klassifikator erhalten kann.

Auch für die Vorverarbeitung ist eine der Methoden im Durchschnitt deutlich überlegen. Die durchschnittliche Genauigkeit von Klassifikatoren mit der Masken-Methode ist signifikant höher als die Genauigkeit von Klassifikatoren mit der Distanz-Methode. Abschnitt 3.4 beschreibt die Vorteile und Risiken der Methoden. Insbesondere der Vorteil der Masken-Methode wird an den Ergebnissen sichtbar. Anhand der GradCAM-Heatmap in Abb. 5.16 ist zu erkennen, dass die Oberflächenmerkmale der Nuclei für die Klassifikation keine Rolle spielen. Ein Risiko der Masken-Methode ist ein Qualitätsverlust aufgrund der Eliminierung der Oberflächenmerkmale, aber selbst mit der Distanz-Methode werden diese Merkmale nicht betrachtet. Die Masken-Methode hebt die Geometrie der Nuclei besonders hervor, und sowohl die statistische Betrachtung der Ergebnisse als auch die Grad-CAM-Heatmaps legen nahe, dass diese Geometrie für die Klassifikation der Nuclei besonders wichtig ist. Mit der Distanz-Methode wird die Geometrie nicht besonders hervorgehoben. Unter Umständen wird die Geometrie sogar durch direkt umliegende Nuclei beeinträchtigt. Abb. 6.3 zeigt diesen Fall in zwei Beispielen. Links ist ein Beispiel für eine besonders starke Überschneidung gegeben. Die Geometrie des Nucleus ist hier ohne die angedeutete Kontur nicht ersichtlich.

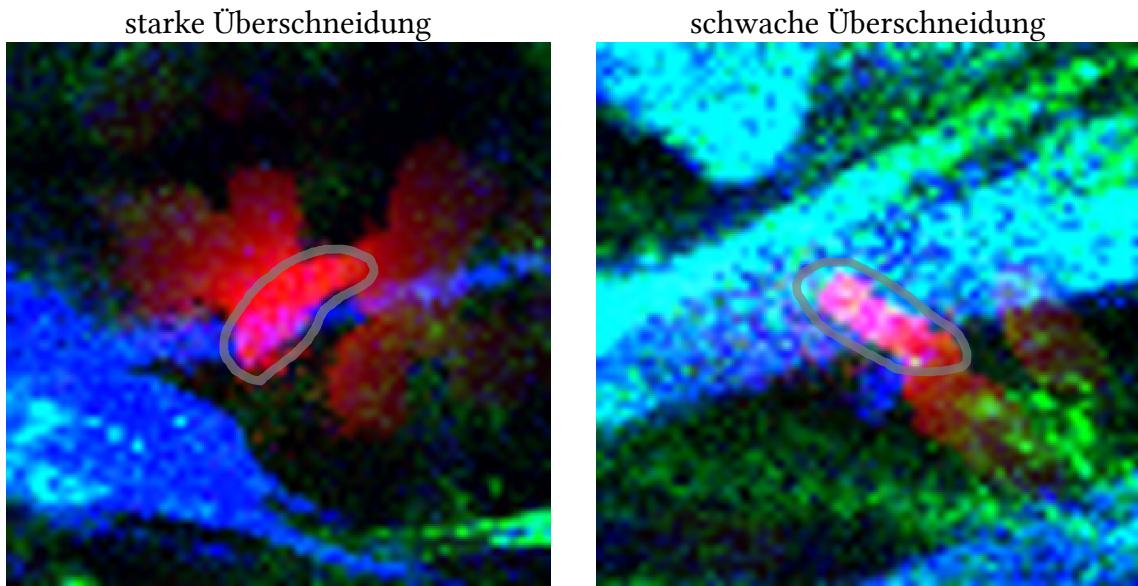


Abb. 6.3 | Beispiele für eine starke und eine schwache Überschneidung von Nuclei nach Anwendung der Distanz-Methode. In Grau ist je eine Umrandung der betrachteten Nuclei angedeutet.

Bei der Distanz-Methode wird die Segmentierungs-Maske nur zur Distanztransformation

verwendet. Die Information über die prädizierte Kontur geht dabei verloren, wenn umliegende Nuclei den betrachteten Nucleus direkt berühren. Die Intensität des berührenden Nucleus wird zwar gedämpft, aber die Kante des betrachteten Nucleus ist an der Berührungsstelle trotzdem verschwommen. Der berührende Nucleus wird in diesem Fall als Teil des betrachteten Objekts interpretiert. Der Einfluss der Vortrainingsmethoden ist besonders bedeutsam. Mit dem fully-supervised Vortraining ist die durchschnittliche Genauigkeit höher und konsistenter als mit anderen Vortrainingsmethoden. Das legt nahe, dass das rechenaufwendige Vortraining auf dem ImageNet-Datensatz mit den diversen Klassen und zahlreichen Stichproben für jede Klasse einen stark generalisierten Merkmalsraum erzeugt. Obwohl die Bilddomäne biologischer Zelldaten stark von der des ImageNet-Datensatzes abweicht, werden bedeutsame Bildmerkmale extrahiert. Auch bei der Adoption von 2D-Encodern auf 3D-Encoder bleibt der Merkmalsraum der fully-supervised-vortrainierten Encoder sinnvoll. Da die Zieldaten niemals vom Encoder gesehen werden, ist Overfitting weitgehend ausgeschlossen, was die geringe Varianz der Genauigkeiten innerhalb der Klassifikatoren mit fully-supervised-Vortraining erklärt.

Auch komplett ohne Vortraining wird ein sinnvoller Merkmalsraum gelernt, und die Klassifikatoren erreichen teilweise hohe Genauigkeitswerte. Der Swin V2-Encoder hat ohne Vortraining nur 50% Genauigkeit erreicht, was vermutlich darauf zurückzuführen ist, dass die aufwendige Transformer-Architektur nicht ausreichend mit dem geringen Trainingsumfang der Zieldaten lernen kann. Es wird keine sinnvolle Merkmalsextraktion gelernt, bevor der Encoder mit dem Overfitting beginnt. Das legt auch nahe, dass Encoder mit besonders hoher Parameteranzahl nicht ohne ImageNet-Vortraining auskommen. Die Beobachtung, dass der kleinste Encoder, der ResNet18-Encoder ohne Vortraining, das beste Ergebnis erzielt hat, unterstützt diese Aussage weiter. Abb. 6.4 zeigt die Genauigkeiten der verschiedenen Vortrainingsmethoden abhängig von der Parameteranzahl des verwendeten Encoders.

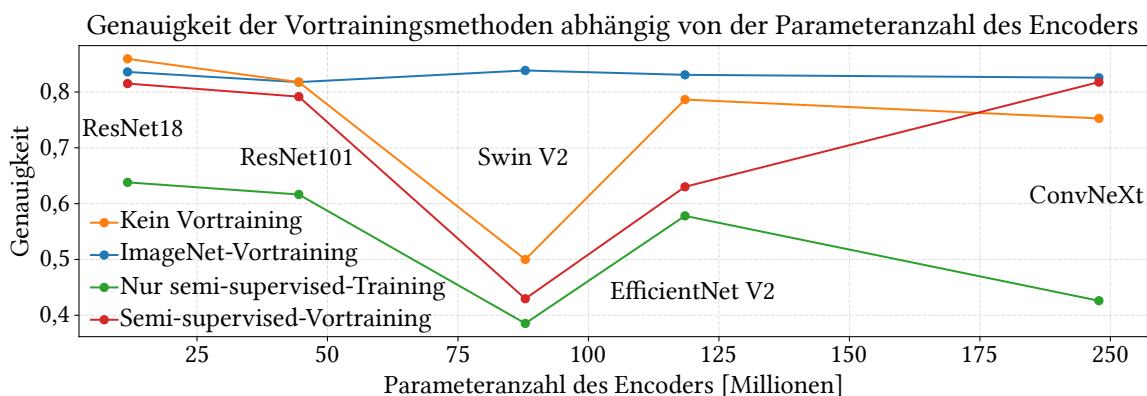


Abb. 6.4 | Genauigkeiten der verschiedenen Vortrainingsmethoden abhängig von der Parameteranzahl des verwendeten Encoders.

Die Abbildung zeigt, dass der Zusammenhang zwischen der Effizienz der Vortrainingsmethoden und der Parameteranzahl nicht trivial darzustellen ist. Mit den ResNet-Encodern

ist es möglich, das ImageNet-Vortaining auszulassen und die Merkmalsextraktion direkt aus den Zieldaten zu lernen. Sowohl ohne Vortraining, als auch mit dem semi-supervised-Vortraining ist die Genauigkeit der Klassifikatoren vergleichbar mit dem ImageNet-Vortraining. In Anbetracht des hohen Rechenaufwands, den das ImageNet-Vortraining erfordert, ist dieses Ergebnis besonders interessant. Beide Methoden erhöhen den Rechenaufwand beim fine-tuning eines Klassifikators auf eine neue Domäne, machen dafür aber das exzessive Vortraining überflüssig. Insgesamt ergibt sich hierdurch vermutlich eine geringere Generalisierbarkeit, aber für den angewandten Datensatz ist die Genauigkeit besser. Auch die beiden anderen CNN-Encoder können ohne ImageNet-Vortraining genutzt werden. Der EfficientNet V2-Encoder ist ohne Vortraining in der Lage, eine vergleichbare Genauigkeit zu erzielen wie mit dem fully-supervised-Vortraining. Dass das semi-supervised-Vortraining ein schlechtes Ergebnis liefert, ist demnach vermutlich auch dem kurzen Training geschuldet. Da der EfficientNet V2-Encoder ohne Vortraining einen sinnvollen Merkmalsraum direkt auf den Zieldaten lernen kann, liegt nahe, dass zu wenige Epochen mit den semi-supervised-Annotationen gelernt wurden, um diesen Merkmalsraum zu erfassen. Es ist auch möglich, dass der Klassifikator-Kopf zu kurz auf dem neu geschaffenen Merkmalsraum trainiert wurde, um die Klassifikationsentscheidung darin optimal zu lernen. Der ConvNeXt-Encoder hingegen hat in den gegebenen Epochen eine sinnvolle Repräsentation gelernt und ist durch das semi-supervised-Vortraining besser als ohne Vortraining. Vermutlich liegt das daran, dass der Encoder im umfangreichen semi-supervised-Vortraining eine besser generalisierte Repräsentation gelernt hat. Außerdem liegt es nahe, dass der ConvNeXt-Encoder ohne Vortraining den Merkmalsraum auf die Trainingdaten overfitted.

Die durchschnittliche Genauigkeit des Swin V2-Encoders ImageNet-Vortraining ist 44%, mit ImageNet-Vortraining ist sie bei 84 %. Das liegt vermutlich daran, dass der relativ große Encoder mit der komplexen ViT-Architektur aus den wenigen Stützvektoren der Repräsentation der Eingangsdaten keinen ausreichenden Merkmalsraum aufspannen kann. Außerdem ist es möglich, dass der Klassifikator-Kopf nur schwach generalisiert die Beziehung zwischen Merkmalen und Entscheidung lernen kann, weil der finale Merkmalsvektor der ViT-Architektur eine viel höhere Dimension hat.

Die Genauigkeit der meisten Encoder ist auch ohne ImageNet-Vortraining annehmbar, vor allem, da mithilfe der 3D-Zelldaten-Pipeline auch ein optimaler Encoder passend zur Vortrainingsmethode gewählt wird. Eine geringe Durchschnittsgenauigkeit ist nicht problematisch, wenn die maximale gefundene Genauigkeit hoch ist.

Trotz der hohen durchschnittlichen Genauigkeit sowohl des fully-supervised-Vortrainings als auch des Modells ohne Vortraining ist der resultierende Merkmalsraum nicht geeignet, um Myotuben-Kerne und Schwannzellen-Kerne zu unterscheiden. Die Unterscheidung dieser Klassen ist für die vorliegenden Daten wichtiger als die Unterscheidung der „Andere“-Klasse und der Debris-Klasse. Die Myotuben- und Schwannzellen-Klassen sind optisch sehr ähnlich und nur durch Expert*Innen unter Berücksichtigung der umliegenden Strukturen trennbar. Abb. 6.5 zeigt jeweils zwei Beispiele der beiden Klassen. Zu sehen sind die Nuclei in Rot und Myotuben in Grün, sowie der S100 β Marker. Die beiden Bei-

spiele aus derselben Reihe sind visuell kaum voneinander zu unterscheiden.

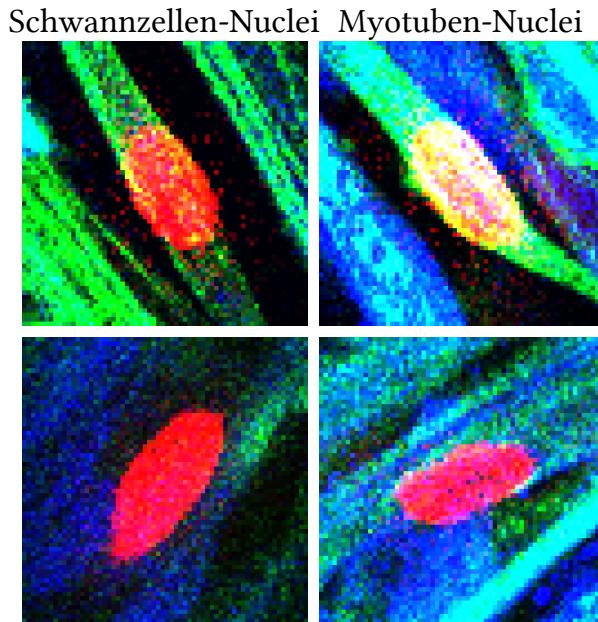


Abb. 6.5 | Beispiele von Nuclei der Schwannzellen-Klasse (links) und der Myotuben-Klasse (rechts). Die beiden Klassen werden von Klassifikatoren ohne semi-supervised-Vortraining schlecht voneinander getrennt. In den Reihen sind jeweils zwei Nuclei zu sehen, die visuell besonders ähnlich sind.

Die Schwannzellen-Klasse wird mit einer höchst signifikant geringeren Genauigkeit erkannt (siehe Abschnitt 5.4.6). Auch für die Klassifikatoren sind die beiden Klassen also schwer voneinander zu trennen. Wie in 5.4.3 beschrieben, sind nur Klassifikatoren, die mit der semi-supervised-Vortrainingsmethode trainiert wurden, in der Lage, die beiden Klassen zuverlässig zu trennen. Die neu entwickelte Vortrainingsmethode und der neu entwickelte Pseudo-Labler sind demnach wichtige Beiträge der vorliegenden Arbeit. Die Effizienz der Vortrainingsmethode zeigt insbesondere auch die Projektion der Klassenentscheidungen der Klassifikatoren auf einen niederdimensionalen, visualisierbaren Merkmalsraum.

In Abb. 6.6 sind zwei exemplarische Merkmalsräume als 2D-Projektionen zu sehen. Durch die Projektion gehen Distanzen zwischen den Stichproben entlang der vielen Dimensionen des Merkmalvektors verloren, aber mithilfe der t-SNE ist die Abbildung als Übersicht dennoch geeignet. Der linke Merkmalsraum ist aus einem ImageNet-Vortraining entstanden, der rechte aus einem semi-supervised-Vortraining. Zu sehen ist, dass im linken Scatterplot die Klassen Eins und Vier, also Myotuben- und Schwannzellen-Nuclei, eine besonders starke Überschneidung aufweisen. Für die Klassen „Andere“ und Debris sind dagegen separate Cluster erkennbar. Im rechten Scatterplot ist das Schwannzellen-Klassen-Cluster zwar immer noch nicht weit entfernt von anderen Clustern, aber wesentlich kompakter, als in der linken Abbildung.

Für die Unterscheidung der Myotuben- und Schwannzellen-Nuclei ist also das Vortraining

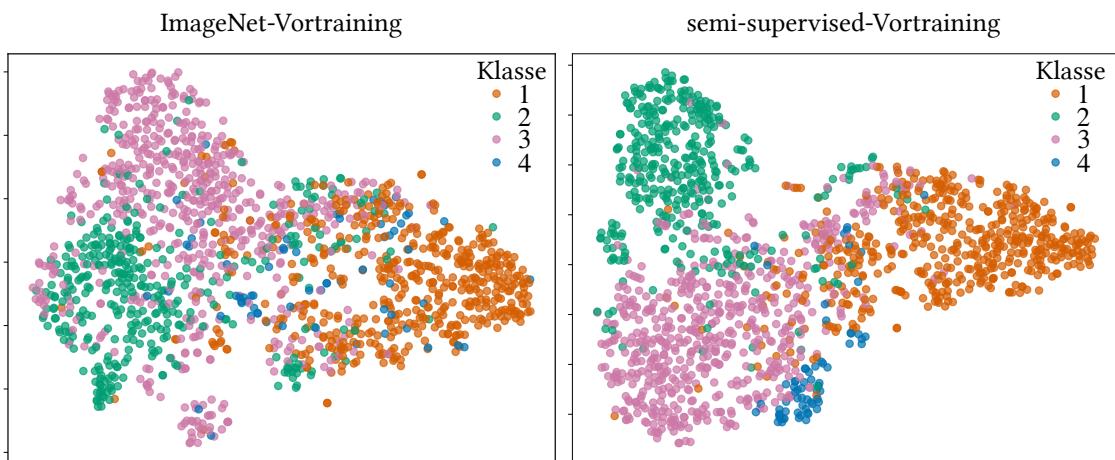


Abb. 6.6 | 2D-Projektionen des Merkmalraums zweier Klassifikatoren mittels t-SNE. Der linke Scatterplot zeigt den Merkmalsraum eines ImageNet-vortrainierten Klassifikators, der rechte den eines semi-supervised-vortrainierten Klassifikators. Die Werte der X- und Y-Achsen sind arbiträre Aktivierungsintensitäten ohne physische Interpretation.

mit semi-supervised-Annotationen besonders hilfreich. Vermutlich wird durch das umfangreiche Vortraining auf den Zieldaten dem Encoder ein so gut generalisierter Merkmalsraum antrainiert, dass der Klassifikations-Kopf die beiden Klassen trotz der Überschneidungen ihrer Merkmale trennen kann. In den Zieldaten hingegen sind diese Merkmale häufig vertreten und werden daher so hochauflösend extrahiert, dass es zur Trennung der Klassen ausreicht. Die Auflösung genau der Merkmale, die für das Trennen der beiden Klassen nötig sind, ist nach dem ImageNet-Datensatz-Vortraining vermutlich nicht sehr hoch.

Die durchschnittliche Genauigkeit der semi-supervised-vortrainierten Klassifikatoren ist zwar geringer als die der anderen Vortrainingsmethoden, dafür ist der entstandene Klassifikator aber besser generalisiert. Da im Test-Anteil der annotierten Daten nur acht der 81 Stichproben die Klasse Schwannzellen-Nucleus aufweisen, ist die Genauigkeit der semi-supervised-vortrainierten Klassifikatoren nicht zwingend repräsentativ für die tatsächliche Leistung. Für die Wahl einer Vortrainingsmethode müssen demnach die konkreten Ziele der Anwendung feststehen. Wenn einzelne Klassen besonders wichtig für die Interpretation der Ergebnisse sind, aber nicht häufig vorkommen, lohnt es sich, den semi-supervised-Ansatz zu testen, um eventuell durch die hohe Generalisierungsfähigkeit einen besseren Klassifikator zu erhalten.

Zusammenfassung und Ausblick 7

Die vorliegende Arbeit befasst sich mit der automatisierten Optimierung von Deep-Learning-Methoden. In der theoretischen Grundlage sind etablierte Deep-Learning-Methoden für die Instanzsegmentierung und Klassifikation gegeben. Vor allem Methoden, die für biologische 3D-Daten geeignet sind, werden beleuchtet. Zusätzlich sind Methoden zusammengefasst, die für die Anwendung, Auswertung und Visualisierung der vorgestellten Methoden wichtig sind. Zu den Methoden aus der Literatur wird außerdem eine Literaturrecherche angestellt, die explizit die aktuelle Forschung in dem Bereich darstellt. Hiernach werden die bisherigen Lücken in der Literatur dargestellt. Für dreidimensionale Bilddaten sind sowohl Methoden zur Segmentierung als auch zur Klassifikation verfügbar, jedoch kein umfassendes Framework, das Anwendung, Vergleich und Optimierung automatisiert. Dementsprechend wird das Ziel der vorliegenden Arbeit definiert. Durch die Arbeit soll der Aufwand ständig wiederkehrender Überlegungen und Vergleiche eliminiert werden, indem die Methodenauswahl und Evaluation von Deep-Learning-Methoden zur panoptischen Segmentierung von biologischen 3D-Daten automatisiert werden.

In der Methodik der Arbeit werden hierzu einige neu entwickelte Methoden eingeführt. Außerdem wird ein Datensatz dreidimensionaler Bildaufnahmen von Myotubenkulturen beschrieben, an dem die eingeführten Methoden demonstriert werden. Die Injektive Panoptische Qualität (**IPQ**) ist eine Metrik zur Bewertung von Instanzsegmentierungsmodellen hinsichtlich ihrer Eignung, interpretierbare Eigenschaften zu extrahieren. Durch die drei Faktoren der Metrik, werden Fehler in den Segmentierungsmasken bestraft die zu Änderungen der Nucleivolumina, Nucleianzahl und lokalen Dichte der Nuclei verursachen. Durch die Ergebnisse der durchgeföhrten Experimente wird deutlich, dass die **IPQ** die Leistung von Modellen effizient erfasst und interpretierbare Ergebnisse liefert. Des Weiteren werden verschiedene Encoder, Klassifikations-Köpfe, Vorverarbeitungsmethoden und Vortrainingsmethoden eingeföhrt. Diese Methoden sind in der 3D-Zelldaten-Pipeline implementiert, einer automatisierten Anwendung zum Vergleich der Leistungen der Methoden. Die Anwendung umfasst außerdem die neu eingeführte Labeling-App, eine Methode zur zeiteffizienten Annotation dreidimensionaler Bild-Datensätze. Mithilfe der 3D-Zelldaten-Pipeline werden Experimente an einem Datensatz von dreidimensionalen Myotubenkultur-Aufnahmen durchgeführt. Die Ergebnisse dieser Experimente belegen die Effizienz der 3D-Zelldaten-Pipeline. Außerdem zeigen sie, dass die Anwendung der semi-supervised-Vortrainingsmethode mit dem neu eingeführten Pseudo-Labler zwar zu einer geringeren Durchschnittsgenauigkeit, aber auch zu einer stärkeren Generalisierung des Klassifikators führt.

In den Ergebnissen sind des Weiteren grundlegende Erkenntnisse zu den Nuclei von Myotubenkulturen ersichtlich. Oberflächenmerkmale von Nuclei sind nicht hilfreich zur Unterscheidung von Nucleus-Klassen. Nur aus ihrer Geometrie wird eine konsistente Klassenentscheidung gelernt. Außerdem sind die Marker-Kanäle für die Klassenentscheidung genauso wichtig wie der Nucleus-Kanal.

Die Ergebnisse haben gezeigt, dass die Geometrie der Nuclei für die Klassifikation essenziell ist. Dementsprechend ist ein mögliches Anliegen kommender Forschung die Erweiterung der [IPQ](#)-Metrik, um Veränderungen der Geometrie durch das Segmentierungsmodell explizit zu bestrafen. Für diese Erweiterung kann ein neuer Faktor eingeführt werden, der die geometrischen Eigenschaften der Annotationen und der segmentierten Instanzen erfasst, beispielsweise als Parameter der Fourier-Entwicklung ihrer Konturen, und mit einem geeigneten Ähnlichkeitsmaß vergleicht. In Anbetracht der Ergebnisse der Klassifikator-Methoden kann zukünftige Forschung weitere, kleinere [CNN](#)-Encoder in Betracht ziehen, um die Hypothese zu prüfen, dass die Encoder zu groß sind um eine sinnvolle Repäsentation der räumlich relativ kleinen Nuclei zu finden. Mit neuen Datensätzen kann außerdem die Hypothese geprüft werden, dass die Oberflächenmerkmale von Nuclei für die Klassifikation unwichtig und die Geometrie ausschlaggebend sind, indem die neu eingeführten Vorverarbeitungsmethoden weiter verglichen werden. Besonders interessant ist für die kommende Forschung eine fortgeführte Analyse der vorgestellten Vortrainingsmethoden. Das semi-supervised-Vortraining führt zu einer geringeren Durchschnittsgenauigkeit, aber zu einer stärkeren Generalisierungsfähigkeit des Klassifikators. Kommende Forschung kann die Methode weiter optimieren, um die Durchschnittsgenauigkeit möglicherweise zu erhöhen und die Vorteile der Methode so besser anwendbar zu machen. Hierzu können weitere Eigenschaften der Nuclei erfasst und dem Merkmalsvektor hinzugefügt werden, auf den der Label-Spreading-Algorithmus angewandt wird, oder die PCA durch einen anderen Algorithmus ersetzt werden. Außerdem kann die anschließende Trainingsroutine mit mehr Epochen und geringerer Lernrate durchgeführt werden.

In Ausblick auf eine vollständige Instanzsegmentierung von Myotuben kann die panoptische Segmentierungsmaske der Nuclei verwendet werden, um Hinweise für ein Segmentierungsmodell wie [SAM](#) zu generieren. Die Myotuben-Zellkerne liegen innerhalb der Myotuben und sind mit ihrer Längsachse entlang der Hauptausrichtung der Myotuben orientiert. Anhand der Ausrichtungen und relativen Positionen der vorhergesagten Myotuben-Zellkerne lässt sich eine polynomische Kurve bestimmen, deren Verlauf weitgehend glatt ist. Diese Kurven können dem Segmentierungsmodell als Hinweise übergeben und so eventuell die Instanzsegmentierung von Myotuben ermöglichen. In einigen ersten Versuchen hiervon konnten zwar einige intuitiv sinnvolle Polynome extrahiert werden, aber auch einige fehlerhafte. Außerdem werden die Myotuben selbst mit den Polynomen als Hinweis durch das [SAM](#)-Modell nicht perfekt instanzsegmentiert.

KI Künstliche Intelligenz

GUI Graphical User Interface

SAM Segment Anything Model

t-SNE t-Distributed Stochastic Neighbor Embedding

IoU Intersection over Union

PQ Panoptic Quality

IPQ Injektive Panoptische Qualität

SQ Segmentation-Quality

RQ Recognition-Quality

IQ Injective-Quality

TP True Positive

FP False Positive

FN False Negative

CNN Convolutional Neural Network

ViT Vision Transformer

Anhang A

A.1 SWINV2 Architektur

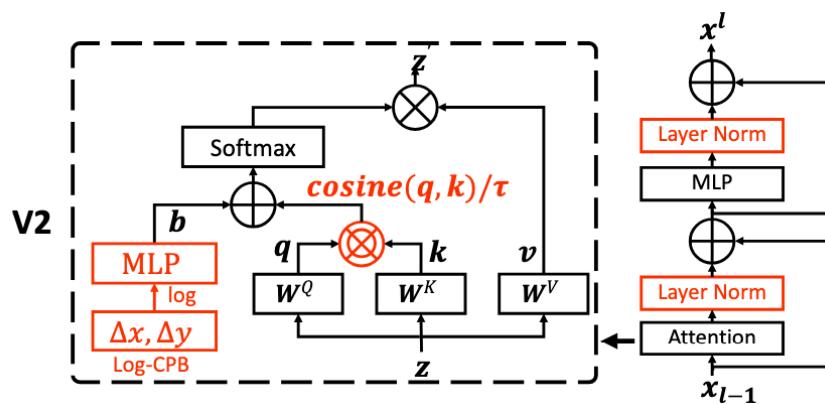


Abb. A.1 | Swin Transformer V2 Architektur [171]. Ein Bildfenster z und dessen relative Koordinaten im Bild Δx und Δy werden in einem Attention-Mechanismus zusammengeführt. Mithilfe einer Kosinus-Ähnlichkeitsfunktion, der Softmax-Funktion [172] und elementweiser Multiplikation sowie Addition werden diese Ergebnisse in einen Merkmalsraum überführt. Zwei layer normalization [104] Schichten, ein weiteres MLP-Netz und residual connections vervollständigen anschließend den modularen **Swin Transformer V2 Block**.

A.2 IPQ-Ergebnisse

Bild	CellposeSAM	nnUNet	Deepcell	Bild	CellposeSAM	nnUNet	Deepcell
1	0.315	0.082	0.010	2	0.349	0.066	0.015
3	0.326	0.139	0.008	4	0.405	0.138	0.017
5	0.313	0.082	0.006	6	0.323	0.139	0.008
7	0.349	0.066	0.015	8	0.406	0.139	0.018
9	0.656	0.249	0.013	10	0.449	0.052	0.026
11	0.500	0.056	0.022	12	0.500	0.056	0.025
13	0.452	0.052	0.026	14	0.645	0.039	0.063
15	0.529	0.056	0.112	16	0.645	0.039	0.063
17	0.529	0.056	0.112	18	0.670	0.148	0.058
19	0.304	0.048	0.015	20	0.297	0.197	0.013
21	0.464	0.058	0.077	22	0.604	0.086	0.021
23	0.516	0.114	0.029	24	0.585	0.151	0.057
25	0.544	0.115	0.029	26	0.507	0.097	0.017
27	0.576	0.172	0.030	28	0.502	0.118	0.030
29	0.697	0.022	0.040	30	0.627	0.025	0.014
31	0.646	0.023	0.031	32	0.613	0.029	0.030
33	0.586	0.143	0.022	34	0.874	0.149	0.141
35	0.862	0.159	0.054	36	0.868	0.180	0.080
37	0.879	0.148	0.079	38	0.854	0.131	0.154
39	0.852	0.167	0.121	40	0.874	0.210	0.044
41	0.846	0.222	0.095	42	0.881	0.143	0.096
43	0.876	0.160	0.109	44	0.880	0.139	0.067
45	0.860	0.149	0.104	46	0.838	0.121	0.099
47	0.872	0.146	0.113	48	0.852	0.138	0.080
49	0.863	0.174	0.098	50	0.867	0.162	0.094
51	0.854	0.142	0.091	52	0.872	0.199	0.118
53	0.871	0.134	0.121	54	0.873	0.154	0.076
55	0.877	0.157	0.114	56	0.880	0.135	0.126
57	0.857	0.148	0.075	58	0.861	0.183	0.099
59	0.865	0.151	0.085	60	0.850	0.184	0.085
61	0.857	0.170	0.154	62	0.881	0.160	0.097
63	0.851	0.113	0.102	64	0.861	0.173	0.170
65	0.885	0.119	0.105	66	0.875	0.185	0.144
67	0.863	0.169	0.084	68	0.853	0.171	0.129
69	0.850	0.154	0.158	70	0.867	0.146	0.110
71	0.859	0.139	0.132	72	0.856	0.133	0.153
73	0.898	0.114	0.094	74	0.879	0.123	0.098
75	0.860	0.233	0.117	76	0.879	0.162	0.138
77	0.855	0.133	0.099	78	0.842	0.185	0.137

Bild	CellposeSAM	nnUNet	Deepcell	Bild	CellposeSAM	nnUNet	Deepcell
79	0.876	0.167	0.137	80	0.872	0.154	0.117
81	0.880	0.161	0.088	82	0.879	0.160	0.142
83	0.879	0.124	0.106	84	0.869	0.141	0.096
85	0.856	0.158	0.100	86	0.837	0.163	0.062
87	0.863	0.149	0.108	88	0.867	0.187	0.119
89	0.874	0.135	0.086	90	0.845	0.141	0.099
91	0.860	0.181	0.199	92	0.868	0.168	0.111
93	0.828	0.139	0.131	94	0.858	0.146	0.084
95	0.886	0.144	0.093	96	0.863	0.115	0.102
97	0.862	0.139	0.112	98	0.863	0.123	0.133
99	0.856	0.168	0.130	100	0.875	0.129	0.111
101	0.870	0.150	0.077	102	0.863	0.142	0.092
103	0.843	0.145	0.125	104	0.856	0.161	0.197
105	0.879	0.143	0.091	106	0.873	0.135	0.142
107	0.870	0.148	0.105	108	0.845	0.143	0.055
109	0.869	0.177	0.097	110	0.855	0.171	0.112
111	0.894	0.143	0.118	112	0.865	0.130	0.051
113	0.863	0.175	0.112	114	0.884	0.154	0.116
115	0.848	0.137	0.099	116	0.874	0.154	0.104
117	0.857	0.171	0.061	118	0.876	0.137	0.091
119	0.871	0.143	0.050	120	0.872	0.151	0.111
121	0.876	0.128	0.138	122	0.873	0.180	0.104
123	0.857	0.170	0.108	124	0.870	0.185	0.131
125	0.861	0.183	0.085	126	0.577	0.050	0.078
127	0.560	0.107	0.037	128	0.572	0.090	0.037
129	0.585	0.113	0.059	130	0.591	0.088	0.013

Tab. A.1 | Einzelne SQ-Ergebnisse von jedem Segmentierungsnetz

Bild	CellposeSAM	nnUNet	Deepcell	Bild	CellposeSAM	nnUNet	Deepcell
1	0.794	0.741	0.271	2	0.958	0.872	0.176
3	0.808	0.722	0.212	4	0.946	0.845	0.225
5	0.797	0.739	0.275	6	0.814	0.716	0.222
7	0.961	0.868	0.179	8	0.950	0.839	0.225
9	0.844	0.664	0.347	10	0.942	0.565	0.509
11	0.862	0.760	0.430	12	0.862	0.769	0.382
13	0.937	0.570	0.514	14	0.902	0.892	0.306
15	0.953	0.921	0.347	16	0.902	0.892	0.306
17	0.953	0.921	0.347	18	0.663	0.688	0.453
19	0.702	0.751	0.308	20	0.773	0.712	0.184

Bild	CellposeSAM	nnUNet	Deepcell	Bild	CellposeSAM	nnUNet	Deepcell
21	0.773	0.825	0.144	22	0.956	0.930	0.167
23	0.947	0.813	0.341	24	0.922	0.789	0.452
25	0.945	0.879	0.346	26	0.935	0.821	0.348
27	0.964	0.824	0.329	28	0.904	0.783	0.418
29	0.937	0.948	0.323	30	0.947	0.870	0.346
31	0.930	0.918	0.381	32	0.938	0.903	0.321
33	0.980	0.804	0.532	34	0.847	0.937	0.261
35	0.872	0.909	0.300	36	0.833	0.927	0.218
37	0.813	0.943	0.295	38	0.862	0.957	0.289
39	0.816	0.951	0.229	40	0.820	0.940	0.196
41	0.819	0.914	0.259	42	0.826	0.951	0.308
43	0.820	0.953	0.330	44	0.794	0.881	0.168
45	0.823	0.944	0.206	46	0.867	0.890	0.158
47	0.862	0.930	0.283	48	0.843	0.940	0.364
49	0.870	0.909	0.149	50	0.813	0.906	0.263
51	0.761	0.945	0.294	52	0.813	0.901	0.323
53	0.826	0.947	0.227	54	0.877	0.935	0.178
55	0.877	0.940	0.275	56	0.847	0.961	0.168
57	0.864	0.935	0.265	58	0.833	0.950	0.224
59	0.813	0.943	0.250	60	0.829	0.908	0.267
61	0.872	0.941	0.283	62	0.826	0.960	0.226
63	0.816	0.943	0.082	64	0.857	0.950	0.289
65	0.800	0.910	0.263	66	0.840	0.937	0.330
67	0.855	0.952	0.317	68	0.816	0.942	0.176
69	0.872	0.945	0.200	70	0.794	0.933	0.235
71	0.864	0.943	0.294	72	0.836	0.933	0.217
73	0.840	0.973	0.272	74	0.800	0.947	0.286
75	0.862	0.931	0.182	76	0.893	0.943	0.220
77	0.833	0.928	0.238	78	0.911	0.935	0.222
79	0.862	0.919	0.174	80	0.847	0.908	0.229
81	0.794	0.945	0.224	82	0.893	0.963	0.323
83	0.870	0.952	0.206	84	0.791	0.908	0.240
85	0.887	0.935	0.272	86	0.847	0.936	0.218
87	0.909	0.937	0.162	88	0.909	0.927	0.348
89	0.806	0.933	0.289	90	0.773	0.920	0.168
91	0.806	0.924	0.152	92	0.840	0.933	0.289
93	0.852	0.927	0.271	94	0.813	0.917	0.118
95	0.820	0.962	0.296	96	0.833	0.966	0.268
97	0.823	0.950	0.351	98	0.895	0.945	0.326
99	0.872	0.939	0.217	100	0.870	0.938	0.377
101	0.813	0.937	0.240	102	0.829	0.934	0.245
103	0.926	0.940	0.240	104	0.850	0.927	0.220

Bild	CellposeSAM	nnUNet	Deepcell	Bild	CellposeSAM	nnUNet	Deepcell
105	0.877	0.942	0.204	106	0.877	0.935	0.213
107	0.820	0.943	0.215	108	0.812	0.952	0.320
109	0.781	0.929	0.255	110	0.872	0.948	0.275
111	0.847	0.949	0.261	112	0.826	0.939	0.278
113	0.862	0.951	0.214	114	0.820	0.941	0.365
115	0.864	0.933	0.220	116	0.840	0.922	0.317
117	0.885	0.941	0.216	118	0.847	0.908	0.305
119	0.862	0.927	0.226	120	0.847	0.913	0.237
121	0.800	0.909	0.235	122	0.820	0.935	0.274
123	0.864	0.937	0.267	124	0.862	0.938	0.364
125	0.840	0.893	0.280	126	0.984	0.876	0.267
127	0.922	0.797	0.305	128	0.958	0.808	0.476
129	0.934	0.780	0.528	130	0.922	0.832	0.481

Tab. A.2 | Einzelne RQ-Ergebnisse von jedem Segmentierungsnetz

Bild	CellposeSAM	nnUNet	Deepcell	Bild	CellposeSAM	nnUNet	Deepcell
1	0.826	0.429	1.000	2	0.748	0.389	1.000
3	0.894	0.506	1.000	4	0.861	0.458	1.000
5	0.832	0.439	1.000	6	0.900	0.520	1.000
7	0.753	0.395	1.000	8	0.869	0.474	1.000
9	0.992	0.673	1.000	10	0.991	0.689	1.000
11	0.957	0.401	1.000	12	0.954	0.385	1.000
13	0.991	0.685	1.000	14	0.966	0.320	0.991
15	0.840	0.242	0.986	16	0.966	0.320	0.991
17	0.840	0.242	0.986	18	0.986	0.405	1.000
19	0.712	0.334	0.987	20	0.932	0.570	1.000
21	0.838	0.171	0.985	22	0.875	0.233	1.000
23	0.950	0.505	1.000	24	0.944	0.400	0.987
25	0.933	0.447	1.000	26	0.909	0.427	1.000
27	0.959	0.541	1.000	28	0.940	0.501	1.000
29	0.957	0.143	1.000	30	0.954	0.307	1.000
31	0.954	0.201	1.000	32	0.935	0.277	1.000
33	0.935	0.416	1.000	34	1.000	0.261	1.000
35	0.985	0.324	1.000	36	1.000	0.326	0.968
37	1.000	0.301	1.000	38	1.000	0.261	0.985
39	0.986	0.321	1.000	40	1.000	0.365	1.000
41	0.972	0.405	0.986	42	1.000	0.280	1.000
43	1.000	0.308	1.000	44	1.000	0.283	1.000
45	0.986	0.279	1.000	46	0.968	0.238	1.000

Bild	CellposeSAM	nnUNet	Deepcell	Bild	CellposeSAM	nnUNet	Deepcell
47	1.000	0.281	1.000	48	0.985	0.261	1.000
49	1.000	0.307	1.000	50	1.000	0.308	1.000
51	0.985	0.275	1.000	52	1.000	0.368	0.984
53	1.000	0.251	1.000	54	1.000	0.281	1.000
55	1.000	0.302	0.986	56	1.000	0.253	1.000
57	0.984	0.272	1.000	58	1.000	0.339	1.000
59	1.000	0.316	1.000	60	0.986	0.363	1.000
61	0.985	0.309	1.000	62	1.000	0.330	1.000
63	0.983	0.211	1.000	64	0.986	0.312	0.971
65	1.000	0.235	1.000	66	1.000	0.343	1.000
67	1.000	0.324	1.000	68	0.984	0.311	1.000
69	0.984	0.286	1.000	70	1.000	0.267	1.000
71	0.985	0.265	0.985	72	0.983	0.244	1.000
73	1.000	0.213	0.985	74	1.000	0.240	1.000
75	1.000	0.420	1.000	76	1.000	0.291	0.984
77	0.986	0.264	1.000	78	0.985	0.340	1.000
79	1.000	0.304	1.000	80	1.000	0.305	1.000
81	1.000	0.311	1.000	82	1.000	0.302	0.986
83	1.000	0.234	1.000	84	0.984	0.256	1.000
85	0.986	0.323	0.986	86	0.970	0.322	1.000
87	1.000	0.278	0.984	88	1.000	0.335	1.000
89	1.000	0.271	1.000	90	0.983	0.267	1.000
91	0.984	0.331	1.000	92	1.000	0.278	1.000
93	0.970	0.276	0.985	94	1.000	0.275	1.000
95	1.000	0.269	0.986	96	1.000	0.228	1.000
97	0.985	0.262	1.000	98	0.983	0.239	1.000
99	0.984	0.293	1.000	100	1.000	0.258	0.985
101	1.000	0.284	1.000	102	0.984	0.258	1.000
103	0.985	0.281	1.000	104	0.983	0.278	1.000
105	1.000	0.276	1.000	106	1.000	0.245	0.966
107	1.000	0.275	1.000	108	0.971	0.282	1.000
109	1.000	0.325	0.985	110	0.985	0.328	0.985
111	1.000	0.249	1.000	112	1.000	0.263	1.000
113	1.000	0.320	1.000	114	1.000	0.300	0.985
115	0.985	0.256	0.970	116	1.000	0.303	0.986
117	1.000	0.306	1.000	118	1.000	0.288	0.986
119	1.000	0.278	1.000	120	1.000	0.296	1.000
121	1.000	0.256	1.000	122	1.000	0.316	1.000
123	0.986	0.342	1.000	124	1.000	0.325	1.000
125	1.000	0.349	1.000	126	0.979	0.287	1.000
127	0.923	0.361	1.000	128	0.975	0.465	1.000

Bild	CellposeSAM	nnUNet	Deepcell	Bild	CellposeSAM	nnUNet	Deepcell
129	0.969	0.540	1.000	130	0.938	0.442	1.000

Tab. A.3 | Einzelne IQ-Ergebnisse von jedem Segmentierungsnetz

Bild	CellposeSAM	nnUNet	Deepcell	Bild	CellposeSAM	nnUNet	Deepcell
1	0.207	0.026	0.003	2	0.250	0.022	0.003
3	0.235	0.051	0.002	4	0.330	0.053	0.004
5	0.207	0.027	0.002	6	0.237	0.052	0.002
7	0.253	0.023	0.003	8	0.335	0.055	0.004
9	0.549	0.112	0.005	10	0.419	0.020	0.013
11	0.412	0.017	0.009	12	0.411	0.016	0.010
13	0.419	0.020	0.013	14	0.563	0.011	0.019
15	0.424	0.013	0.038	16	0.563	0.011	0.019
17	0.424	0.013	0.038	18	0.438	0.041	0.026
19	0.152	0.012	0.005	20	0.214	0.080	0.002
21	0.300	0.008	0.011	22	0.505	0.019	0.003
23	0.464	0.047	0.010	24	0.509	0.048	0.025
25	0.480	0.045	0.010	26	0.431	0.034	0.006
27	0.532	0.077	0.010	28	0.426	0.046	0.012
29	0.625	0.003	0.013	30	0.566	0.007	0.005
31	0.573	0.004	0.012	32	0.537	0.007	0.010
33	0.537	0.048	0.012	34	0.741	0.036	0.037
35	0.741	0.047	0.016	36	0.724	0.054	0.017
37	0.714	0.042	0.023	38	0.736	0.033	0.044
39	0.685	0.051	0.028	40	0.717	0.072	0.009
41	0.673	0.082	0.024	42	0.728	0.038	0.030
43	0.718	0.047	0.036	44	0.699	0.035	0.011
45	0.697	0.039	0.021	46	0.703	0.026	0.016
47	0.752	0.038	0.032	48	0.708	0.034	0.029
49	0.751	0.049	0.015	50	0.705	0.045	0.025
51	0.640	0.037	0.027	52	0.709	0.066	0.038
53	0.720	0.032	0.027	54	0.765	0.040	0.014
55	0.769	0.045	0.031	56	0.746	0.033	0.021
57	0.729	0.038	0.020	58	0.717	0.059	0.022
59	0.703	0.045	0.021	60	0.695	0.061	0.023
61	0.736	0.049	0.043	62	0.728	0.051	0.022
63	0.683	0.022	0.008	64	0.728	0.051	0.048
65	0.708	0.026	0.027	66	0.736	0.059	0.048
67	0.737	0.052	0.027	68	0.684	0.050	0.023
69	0.729	0.042	0.032	70	0.688	0.036	0.026

Bild	CellposeSAM	nnUNet	Deepcell	Bild	CellposeSAM	nnUNet	Deepcell
71	0.731	0.035	0.038	72	0.704	0.030	0.033
73	0.754	0.024	0.025	74	0.703	0.028	0.028
75	0.741	0.091	0.021	76	0.785	0.044	0.030
77	0.703	0.033	0.023	78	0.755	0.059	0.030
79	0.755	0.047	0.024	80	0.739	0.043	0.027
81	0.698	0.047	0.020	82	0.785	0.047	0.045
83	0.764	0.028	0.022	84	0.676	0.033	0.023
85	0.749	0.048	0.027	86	0.688	0.049	0.014
87	0.785	0.039	0.017	88	0.789	0.058	0.041
89	0.705	0.034	0.025	90	0.642	0.035	0.017
91	0.682	0.055	0.030	92	0.730	0.044	0.032
93	0.684	0.036	0.035	94	0.698	0.037	0.010
95	0.726	0.037	0.027	96	0.719	0.025	0.027
97	0.699	0.035	0.039	98	0.759	0.028	0.043
99	0.734	0.046	0.028	100	0.761	0.031	0.041
101	0.707	0.040	0.019	102	0.704	0.034	0.022
103	0.769	0.038	0.030	104	0.715	0.042	0.043
105	0.771	0.037	0.019	106	0.765	0.031	0.029
107	0.714	0.038	0.023	108	0.666	0.038	0.018
109	0.679	0.053	0.024	110	0.735	0.053	0.030
111	0.758	0.034	0.031	112	0.715	0.032	0.014
113	0.744	0.053	0.024	114	0.724	0.044	0.042
115	0.722	0.033	0.021	116	0.734	0.043	0.032
117	0.758	0.049	0.013	118	0.742	0.036	0.027
119	0.751	0.037	0.011	120	0.739	0.041	0.026
121	0.701	0.030	0.032	122	0.716	0.053	0.029
123	0.730	0.054	0.029	124	0.750	0.056	0.048
125	0.724	0.057	0.024	126	0.555	0.012	0.021
127	0.476	0.031	0.011	128	0.534	0.034	0.018
129	0.530	0.048	0.031	130	0.511	0.032	0.006

Tab. A.4 | Einzelne ipq-Ergebnisse von jedem Segmentierungsnetz

A.3 Einzelne Ergebnisse aller Klassifikator Kombinationen

Encoder	Klassifikations-Kopf	Vortraining	Mask Channel	Best accuracy
CellposeSAM	Schichten-Klassifikator	Fully-supervised	Masken-Methode	72,9%
CellposeSAM	Volumen-Klassifikator	Fully-supervised	Masken-Methode	84,9%
ResNet18	Schichten-Klassifikator	Fully-supervised	Masken-Methode	79,4%
ResNet18	Volumen-Klassifikator	Fully-supervised	Masken-Methode	83,6%
ResNet101	Schichten-Klassifikator	Fully-supervised	Masken-Methode	78,6%
ResNet101	Volumen-Klassifikator	Fully-supervised	Masken-Methode	81,8%
Efficientnet V2	Schichten-Klassifikator	Fully-supervised	Masken-Methode	81,8%
Efficientnet V2	Volumen-Klassifikator	Fully-supervised	Masken-Methode	83,1%
ConvNeXt	Schichten-Klassifikator	Fully-supervised	Masken-Methode	83,6%
ConvNeXt	Volumen-Klassifikator	Fully-supervised	Masken-Methode	82,6%
Swin V2	Schichten-Klassifikator	Fully-supervised	Masken-Methode	82,3%
Swin V2	Volumen-Klassifikator	Fully-supervised	Masken-Methode	83,9%
CellposeSAM	Schichten-Klassifikator	Fully-supervised	Distanz-Methode	69,3%
CellposeSAM	Volumen-Klassifikator	Fully-supervised	Distanz-Methode	82,6%
ResNet18	Schichten-Klassifikator	Fully-supervised	Distanz-Methode	78,4%
ResNet18	Volumen-Klassifikator	Fully-supervised	Distanz-Methode	81,2%
ResNet101	Schichten-Klassifikator	Fully-supervised	Distanz-Methode	77,1%
ResNet101	Volumen-Klassifikator	Fully-supervised	Distanz-Methode	81,0%
Efficientnet V2	Schichten-Klassifikator	Fully-supervised	Distanz-Methode	71,1%
Efficientnet V2	Volumen-Klassifikator	Fully-supervised	Distanz-Methode	70,8%
ConvNeXt	Schichten-Klassifikator	Fully-supervised	Distanz-Methode	80,2%
ConvNeXt	Volumen-Klassifikator	Fully-supervised	Distanz-Methode	79,9%
Swin V2	Schichten-Klassifikator	Fully-supervised	Distanz-Methode	76,6%
Swin V2	Volumen-Klassifikator	Fully-supervised	Distanz-Methode	77,9%
ResNet18	Volumen-Klassifikator	Kein Vortraining	Masken-Methode	85,9%
ResNet101	Volumen-Klassifikator	Kein Vortraining	Masken-Methode	81,8%
Efficientnet V2	Volumen-Klassifikator	Kein Vortraining	Masken-Methode	78,6%
ConvNeXt	Volumen-Klassifikator	Kein Vortraining	Masken-Methode	75,3%
Swin V2	Volumen-Klassifikator	Kein Vortraining	Masken-Methode	50,0%
ResNet18	Volumen-Klassifikator	Nur semi-supervised	Masken-Methode	63,8%
ResNet101	Volumen-Klassifikator	Nur semi-supervised	Masken-Methode	61,6%
Efficientnet V2	Volumen-Klassifikator	Nur semi-supervised	Masken-Methode	57,8%
ConvNeXt	Volumen-Klassifikator	Nur semi-supervised	Masken-Methode	42,6%
Swin V2	Volumen-Klassifikator	Nur semi-supervised	Masken-Methode	38,5%
ResNet18	Volumen-Klassifikator	Semi-supervised und Transfer	Masken-Methode	81,5%
ResNet101	Volumen-Klassifikator	Semi-supervised und Transfer	Masken-Methode	79,2%
Efficientnet V2	Volumen-Klassifikator	Semi-supervised und Transfer	Masken-Methode	63,0%
ConvNeXt	Volumen-Klassifikator	Semi-supervised und Transfer	Masken-Methode	81,8%
Swin V2	Volumen-Klassifikator	Semi-supervised und Transfer	Masken-Methode	43,0%

Tab. A.5 | Ergebnisse der einzelnen Methodenkombinationen der trainierten Klassifikatoren

Quellenverzeichnis

- [1] Warren H Lewis und Margaret R Lewis. „Behavior of cross striated muscle in tissue cultures“. In: *American Journal of Anatomy* 22.2 (1917), S. 169–194.
- [2] Emeka Enwere et al. „Role of the TWEAK-Fn14-cIAP1-NF-κB Signaling Axis in the Regulation of Myogenesis and Muscle Homeostasis“. In: *Frontiers in Immunology* 5 (Feb. 2014), S. 34.
- [3] Scott F Gilbert. *Developmental biology*. Englisch. 10. Aufl. Sunderland, Massachusetts: Sinauer Associates, 2014.
- [4] Irene A. Pogogeff und Margaret R. Murray. „Form and behavior of adult mammalian skeletal muscle in vitro“. en. In: *The Anatomical Record* 95.3 (1946), S. 321–335.
- [5] Antoine Weisrock et al. „MyoFInDer: An AI-Based Tool for Myotube Fusion Index Determination“. In: *Tissue Engineering Part A* 30.19-20 (Okt. 2024), S. 652–661.
- [6] Benjamin Lair et al. *MyoFuse: A fully AI-based workflow for automated quantification of skeletal muscle cell fusion in vitro*. en. Techn. Ber. Type: article. bioRxiv, Feb. 2025. Kap. New Results, S. 2025.02.17.638596.
- [7] Kyungchang Jeong et al. „SEPO-FI: Deep-learning based software to calculate fusion index of muscle cells“. In: *Computers in Biology and Medicine* 186 (März 2025), S. 109706.
- [8] Juergen Scharner und Peter S Zammit. „The muscle satellite cell at 50: the formative years“. In: *Skeletal Muscle* 1 (Aug. 2011), S. 28.
- [9] Pedro Veliça und Chris M. Bunce. „A quick, simple and unbiased method to quantify C2C12 myogenic differentiation“. eng. In: *Muscle & Nerve* 44.3 (Sep. 2011), S. 366–370.
- [10] Andy Nolan et al. „Fluorescent characterization of differentiated myotubes using flow cytometry“. eng. In: *Cytometry. Part A: The Journal of the International Society for Analytical Cytology* 105.5 (Mai 2024), S. 332–344.
- [11] Chibeza C. Agley et al. „An Image Analysis Method for the Precise Selection and Quantitation of Fluorescently Labeled Cellular Constituents“. In: *Journal of Histochemistry and Cytochemistry* 60.6 (Juni 2012), S. 428–438.
- [12] Min-Wen Jason Chua et al. „Assessment of different strategies for scalable production and proliferation of human myoblasts“. In: *Cell Proliferation* 52.3 (März 2019), e12602.

- [13] Aref Shahini et al. „Efficient and high yield isolation of myoblasts from skeletal muscle“. In: *Stem cell research* 30 (Juli 2018), S. 122–129.
- [14] Jessica Brunetti et al. „Nanopattern surface improves cultured human myotube maturation“. In: *Skeletal Muscle* 11.1 (Mai 2021), S. 12.
- [15] Gisela Nogales-Gadea et al. „Expression of Glycogen Phosphorylase Isoforms in Cultured Muscle from Patients with McArdle’s Disease Carrying the p.R771PfsX33 PYGM Mutation“. en. In: *PLOS ONE* 5.10 (Okt. 2010), e13164.
- [16] Nathalie Couturier et al. „Aberrant evoked calcium signaling and nAChR cluster morphology in a SOD1 D90A hiPSC-derived neuromuscular model“. In: *Frontiers in Cell and Developmental Biology* 12 (20. Juni 2024), S. 1429759.
- [17] Simon Noë et al. „The Myotube Analyzer: how to assess myogenic features in muscle stem cells“. en. In: *Skeletal Muscle* 12.1 (Juni 2022), S. 12.
- [18] Ahmed M. Abdelmoez et al. „Comparative profiling of skeletal muscle models reveals heterogeneity of transcriptome and metabolism“. In: *American Journal of Physiology - Cell Physiology* 318.3 (März 2020), S. C615–C626.
- [19] Haruki Inoue et al. „Automatic Quantitative Segmentation of Myotubes Reveals Single-cell Dynamics of S6 Kinase Activation“. eng. In: *Cell Structure and Function* 43.2 (Aug. 2018), S. 153–169.
- [20] Josh Moore et al. „OME-NGFF: a next-generation file format for expanding bioimaging data-access strategies“. en. In: *Nature Methods* 18.12 (Dez. 2021), S. 1496–1498.
- [21] Mingjie Pan et al. „DiffuseIR: diffusion models for isotropic reconstruction of 3D microscopic images“. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2023, S. 323–332.
- [22] Xuesong Li et al. „Three-dimensional structured illumination microscopy with enhanced axial resolution“. In: *Nature Biotechnology* 41.9 (2023), S. 1307–1319.
- [23] Neda Bagheri et al. „The new era of quantitative cell imaging—challenges and opportunities“. In: *Molecular cell* 82.2 (Jan. 2022), S. 241–247.
- [24] Ze Liu et al. „Swin transformer: Hierarchical vision transformer using shifted windows“. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, S. 10012–10022.
- [25] Jim James et al. „Segmentation of tomography datasets using 3D convolutional neural networks“. In: *Computational Materials Science* 216 (Jan. 2023), S. 111847.
- [26] Henry Chan et al. „Machine learning enabled autonomous microstructural characterization in 3D samples“. en. In: *npj Computational Materials* 6.1 (Jan. 2020), S. 1.
- [27] Yu Hirabayashi et al. „Deep learning for three-dimensional segmentation of electron microscopy images of complex ceramic materials“. en. In: *npj Computational Materials* 10.1 (März 2024), S. 46.

- [28] P. A. Midgley und M. Weyland. „3D electron microscopy in the physical sciences: the development of Z-contrast and EFTEM tomography“. In: *Ultramicroscopy*. Proceedings of the International Workshop on Strategies and Advances in Atomic Level Spectroscopy and Analysis 96.3 (Sep. 2003), S. 413–431.
- [29] Yaroslav Ganin und Victor Lempitsky. „Unsupervised domain adaptation by backpropagation“. In: *International conference on machine learning*. PMLR. 2015, S. 1180–1189.
- [30] Han Zhu et al. „Domain adaptation using class similarity for robust speech recognition“. In: *arXiv preprint arXiv:2011.02782* (2020).
- [31] Tahereh Koohi-Var und Morteza Zahedi. „Cross-domain graph based similarity measurement of workflows“. In: *Journal of Big Data* 5 (2018), S. 1–16.
- [32] Yuanzhe Cai et al. „Efficient algorithm for computing link-based similarity in real world networks“. In: *2009 Ninth IEEE International Conference on Data Mining*. IEEE. 2009, S. 734–739.
- [33] Wei Yuan, Jianfeng Gao und Hisami Suzuki. „An empirical study on language model adaptation using a metric of domain similarity“. In: *International Conference on Natural Language Processing*. Springer. 2005, S. 957–968.
- [34] Lothar Schermelleh et al. „Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy“. eng. In: *Science (New York, N.Y.)* 320.5881 (Juni 2008), S. 1332–1336.
- [35] Hong-Shang Peng und Daniel T. Chiu. „Soft fluorescent nanomaterials for biological and biomedical imaging“. en. In: *Chemical Society Reviews* 44.14 (2015), S. 4699–4722.
- [36] Rehan Ali et al. „Automatic segmentation of adherent biological cell boundaries and nuclei from brightfield microscopy images“. en. In: *Machine Vision and Applications* 23.4 (Juli 2012), S. 607–621.
- [37] Mei Wang und Weihong Deng. „Deep visual domain adaptation: A survey“. In: *Neurocomputing* 312 (2018), S. 135–153.
- [38] Xingchao Peng et al. „Visda: The visual domain adaptation challenge“. In: *arXiv preprint arXiv:1710.06924* (2017).
- [39] Tianyu Han, Lifeng Zhang und Shixiang Jia. „Bin similarity-based domain adaptation for fine-grained image classification“. In: *International Journal of Intelligent Systems* 37.3 (2022), S. 2319–2334.
- [40] Pedro O Pinheiro. „Unsupervised domain adaptation with similarity learning“. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, S. 8004–8013.
- [41] Yaroslav Ganin et al. „Domain-adversarial training of neural networks“. In: *Journal of machine learning research* 17.59 (2016), S. 1–35.

- [42] Soumyadeep Ghosh et al. „Domain Adaptation for Visual Understanding“. en. In: Hrsg. von Richa Singh et al. Cham: Springer International Publishing, 2020, S. 1–15.
- [43] Bharath Hariharan et al. „Simultaneous Detection and Segmentation“. en. In: Hrsg. von David Fleet et al. Bd. 8695. Cham: Springer International Publishing, 2014, S. 297–312.
- [44] John Winn und Jamie Shotton. „The layout consistent random field for recognizing and segmenting partially occluded objects“. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Bd. 1. IEEE, 2006, S. 37–44.
- [45] Alexander Kirillov et al. „Panoptic segmentation“. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, S. 9404–9413.
- [46] Shervin Minaee et al. „Image segmentation using deep learning: A survey“. In: *IEEE transactions on pattern analysis and machine intelligence* 44.7 (2021), S. 3523–3542.
- [47] Anurag Arnab und Philip HS Torr. „Pixelwise instance segmentation with a dynamically instantiated network“. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, S. 441–450.
- [48] Liang-Chieh Chen et al. „Masklab: Instance segmentation by refining object detection with semantic and direction features“. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, S. 4013–4022.
- [49] Olaf Ronneberger, Philipp Fischer und Thomas Brox. „U-Net: Convolutional Networks for Biomedical Image Segmentation“. en. In: Hrsg. von Nassir Navab et al. Bd. 9351. Cham: Springer International Publishing, 2015, S. 234–241.
- [50] Ross Girshick et al. „Rich feature hierarchies for accurate object detection and semantic segmentation“. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, S. 580–587.
- [51] Jonathan Long, Evan Shelhamer und Trevor Darrell. „Fully convolutional networks for semantic segmentation“. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, S. 3431–3440.
- [52] Matthew D. Zeiler et al. „Deconvolutional networks“. In: ISSN: 1063-6919. Juni 2010, S. 2528–2535.
- [53] Matthew D. Zeiler und Rob Fergus. „Visualizing and Understanding Convolutional Networks“. en. In: *Computer Vision – ECCV 2014*. Hrsg. von David Fleet et al. Cham: Springer International Publishing, 2014, S. 818–833.
- [54] Hyeonwoo Noh, Seunghoon Hong und Bohyung Han. „Learning Deconvolution Network for Semantic Segmentation“. In: ISSN: 2380-7504. Dez. 2015, S. 1520–1528.
- [55] Mohammadreza Mostajabi, Payman Yadollahpour und Gregory Shakhnarovich. „Feedforward semantic segmentation with zoom-out features“. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, S. 3376–3385.

- [56] Jiuxiang Gu et al. „Recent advances in convolutional neural networks“. In: *Pattern recognition* 77 (2018), S. 354–377.
- [57] Lamia Jaafar Belaid und Walid Mourou. „IMAGE SEGMENTATION: A WATERSHED TRANSFORMATION ALGORITHM“. en. In: *Image Analysis and Stereology* 28.2 (2009), S. 93–102.
- [58] Abdel Aziz Taha und Allan Hanbury. „Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool“. In: *BMC Medical Imaging* 15.1 (Aug. 2015), S. 29.
- [59] Yichi Zhang et al. „Bridging 2D and 3D segmentation networks for computation-efficient volumetric medical image segmentation: An empirical study of 2.5D solutions“. In: *Computerized Medical Imaging and Graphics* 99 (Juli 2022), S. 102088.
- [60] A. Avesta et al. „3D Capsule Networks for Brain Image Segmentation“. In: *American Journal of Neuroradiology* 44.5 (Apr. 2023), S. 562–568.
- [61] Jiancheng Yang et al. „Reinventing 2D Convolutions for 3D Images“. In: *IEEE Journal of Biomedical and Health Informatics* 25.8 (Aug. 2021), S. 3009–3018.
- [62] Anurag Arnab et al. „ViViT: A Video Vision Transformer“. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Okt. 2021, S. 6816–6826.
- [63] Yaoli Wang et al. „Vision Transformers for Image Classification: A Comparative Survey“. en. In: *Technologies* 13.1 (Jan. 2025), S. 32.
- [64] Alex Ling Yu Hung et al. „Csam: A 2.5 d cross-slice attention module for anisotropic volumetric medical image segmentation“. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2024, S. 5923–5932.
- [65] Hamed Karimi und Mohammad Hamghalam. „Segmentation of 3D MRI Using 2D Convolutional Neural Networks in Infants’ Brain“. en. In: *Multimedia Tools and Applications* 83.11 (März 2024), S. 33511–33526.
- [66] Arman Avesta et al. „Comparing 3D, 2.5D, and 2D Approaches to Brain Image Auto-Segmentation“. en. In: *Bioengineering* 10.2 (Feb. 2023), S. 181.
- [67] Keinosuke Fukunaga. „Statistical pattern recognition“. In: WORLD SCIENTIFIC, Aug. 1993, S. 33–60.
- [68] S.R. Kulkarni, G. Lugosi und S.S. Venkatesh. „Learning pattern classification-a survey“. In: *IEEE Transactions on Information Theory* 44.6 (Okt. 1998), S. 2178–2206.
- [69] A. K. Jain, M. N. Murty und P. J. Flynn. „Data clustering: a review“. In: *ACM Comput. Surv.* 31.3 (Sep. 1999), S. 264–323.
- [70] Shai Shalev-Shwartz und Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. en. Google-Books-ID: Hf6QAwAAQBAJ. Cambridge University Press, Mai 2014.
- [71] Marco Loog. „Supervised classification: Quite a brief overview“. In: *Machine Learning Techniques for Space Weather* (2018), S. 113–145.

- [72] Isabelle Guyon und André Elisseeff. „An Introduction to Feature Extraction“. en. In: Hrsg. von Isabelle Guyon et al. Berlin, Heidelberg: Springer, 2006, S. 1–25.
- [73] M. Kunaver und J.F. Tasic. „Image feature extraction - an overview“. In: Bd. 1. Nov. 2005, S. 183–186.
- [74] Wamidh K. Mutlag et al. „Feature Extraction Methods: A Review“. en. In: *Journal of Physics: Conference Series* 1591.1 (Juli 2020), S. 012028.
- [75] Charles T. Zahn und Ralph Z. Roskies. „Fourier Descriptors for Plane Closed Curves“. In: *IEEE Transactions on Computers* C-21.3 (März 1972), S. 269–281.
- [76] Frank P Kuhl und Charles R Giardina. „Elliptic Fourier features of a closed contour“. In: *Computer Graphics and Image Processing* 18.3 (März 1982), S. 236–258.
- [77] Karl Pearson. „LIII. On lines and planes of closest fit to systems of points in space“. In: *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2.11 (1901), S. 559–572.
- [78] H. Hotelling. „Analysis of a complex of statistical variables into principal components“. In: *Journal of Educational Psychology* 24.6 (1933), S. 417–441.
- [79] Rui Xu und D. Wunsch. „Survey of clustering algorithms“. In: *IEEE Transactions on Neural Networks* 16.3 (Mai 2005), S. 645–678.
- [80] Bernhard E. Boser, Isabelle M. Guyon und Vladimir N. Vapnik. „A training algorithm for optimal margin classifiers“. In: *Proceedings of the fifth annual workshop on Computational learning theory*. COLT '92. New York, NY, USA: Association for Computing Machinery, Juli 1992, S. 144–152.
- [81] David Yarowsky. „Unsupervised word sense disambiguation rivaling supervised methods“. In: *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*. ACL '95. USA: Association for Computational Linguistics, Juni 1995, S. 189–196.
- [82] Bernhard Schölkopf. „Support vector learning“. PhD Thesis. Oldenbourg München, Germany, 1997.
- [83] Alexander J. Smola und Bernhard Schölkopf. *Learning with kernels*. Bd. 4. Citeseer, 1998.
- [84] Dengyong Zhou et al. „Learning with Local and Global Consistency“. In: *Advances in Neural Information Processing Systems*. Bd. 16. MIT Press, 2003.
- [85] David Lowe und D. Broomhead. „Multivariable functional interpolation and adaptive networks“. In: *Complex systems* 2.3 (1988), S. 321–355.
- [86] Olivier Delalleau, Yoshua Bengio und Nicolas Le Roux. „Efficient non-parametric function induction in semi-supervised learning“. In: *International Workshop on Artificial Intelligence and Statistics*. PMLR, 2005, S. 96–103.
- [87] Sotiris B. Kotsiantis, Ioannis Zaharakis und P. Pintelas. „Supervised machine learning: A review of classification techniques“. In: *Emerging artificial intelligence applications in computer engineering* 160.1 (2007), S. 3–24.

- [88] G. Hughes. „On the mean accuracy of statistical pattern recognizers“. In: *IEEE Transactions on Information Theory* 14.1 (Jan. 1968), S. 55–63.
- [89] Asifullah Khan et al. „A survey of the vision transformers and their CNN-transformer based variants“. en. In: *Artificial Intelligence Review* 56.3 (Dez. 2023), S. 2917–2970.
- [90] Alex Krizhevsky, Ilya Sutskever und Geoffrey E Hinton. „ImageNet Classification with Deep Convolutional Neural Networks“. In: *Advances in Neural Information Processing Systems*. Bd. 25. Curran Associates, Inc., 2012.
- [91] Max Bain et al. „Frozen in Time: A Joint Video and Image Encoder for End-to-End Retrieval“. en. In: 2021, S. 1728–1738.
- [92] Jo Plested und Tom Gedeon. „Deep transfer learning for image classification: a survey“. In: *arXiv preprint arXiv:2205.09904* (Mai 2022). arXiv:2205.09904 [cs].
- [93] Alireza Ghods und Diane J Cook. „A Survey of Techniques All Classifiers Can Learn from Deep Networks: Models, Optimizations, and Regularization“. In: *arXiv preprint arXiv:1909.04791* (Sep. 2019). arXiv:1909.04791 [cs].
- [94] Jürgen Schmidhuber. „Deep learning in neural networks: An overview“. In: *Neural Networks* 61 (Jan. 2015), S. 85–117.
- [95] Sainbayar Sukhbaatar et al. „Training convolutional networks with noisy labels“. In: *arXiv preprint arXiv:1406.2080* (2014).
- [96] Elliott Gordon-Rodriguez et al. „Uses and Abuses of the Cross-Entropy Loss: Case Studies in Modern Deep Learning“. en. In: PMLR, Feb. 2020, S. 1–10.
- [97] Anqi Mao, Mehryar Mohri und Yutao Zhong. „Cross-Entropy Loss Functions: Theoretical Analysis and Applications“. en. In: PMLR, Juli 2023, S. 23803–23828.
- [98] Jacob Goldberger und Ehud Ben-Reuven. „Training deep neural-networks using a noise adaptation layer“. In: *International conference on learning representations*. 2017.
- [99] Bo Han et al. „Masking: A New Perspective of Noisy Supervision“. In: *Advances in Neural Information Processing Systems*. Bd. 31. Curran Associates, Inc., 2018.
- [100] Dan Hendrycks et al. „Using Trusted Data to Train Deep Networks on Labels Corrupted by Severe Noise“. In: *Advances in Neural Information Processing Systems*. Bd. 31. Curran Associates, Inc., 2018.
- [101] Zhilu Zhang und Mert Sabuncu. „Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels“. In: *Advances in Neural Information Processing Systems*. Bd. 31. Curran Associates, Inc., 2018.
- [102] P. R. Smith. „Bilinear interpolation of digital images“. In: *Ultramicroscopy* 6.2 (Jan. 1981), S. 201–204.
- [103] Sergey Ioffe und Christian Szegedy. „Batch normalization: Accelerating deep network training by reducing internal covariate shift“. In: *International conference on machine learning*. pmlr, 2015, S. 448–456.

- [104] Jimmy Lei Ba, Jamie Ryan Kiros und Geoffrey E Hinton. „Layer normalization“. In: *arXiv preprint arXiv:1607.06450* (2016).
- [105] Shibani Santurkar et al. „How Does Batch Normalization Help Optimization?“ In: *Advances in Neural Information Processing Systems*. Bd. 31. Curran Associates, Inc., 2018.
- [106] Jingjing Xu et al. „Understanding and Improving Layer Normalization“. In: *Advances in Neural Information Processing Systems*. Bd. 32. Curran Associates, Inc., 2019.
- [107] Ramprasaath R. Selvaraju et al. „Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization“. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Okt. 2017.
- [108] Laurens van der Maaten und Geoffrey Hinton. „Visualizing Data using t-SNE“. In: *Journal of Machine Learning Research* 9.86 (2008), S. 2579–2605.
- [109] T. Tony Cai und Rong Ma. „Theoretical Foundations of t-SNE for Visualizing High-Dimensional Clustered Data“. In: *Journal of Machine Learning Research* 23.301 (2022), S. 1–54.
- [110] Y. Lecun et al. „Gradient-based learning applied to document recognition“. In: *Proceedings of the IEEE* 86.11 (Nov. 1998), S. 2278–2324.
- [111] Du Tran et al. „A Closer Look at Spatiotemporal Convolutions for Action Recognition“. In: 2018, S. 6450–6459.
- [112] Christoph Feichtenhofer. „X3D: Expanding Architectures for Efficient Video Recognition“. In: 2020, S. 203–213.
- [113] Joao Carreira und Andrew Zisserman. „Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset“. In: 2017, S. 6299–6308.
- [114] Du Tran et al. „Learning Spatiotemporal Features With 3D Convolutional Networks“. In: 2015, S. 4489–4497.
- [115] Gedas Bertasius, Heng Wang und Lorenzo Torresani. „Is space-time attention all you need for video understanding?“ In: *Icml* 2.3 (2021).
- [116] Xiaolong Wang et al. „Non-Local Neural Networks“. In: 2018, S. 7794–7803.
- [117] Christoffer Edlund et al. „LIVECell—A large-scale dataset for label-free live cell segmentation“. In: *Nature methods* 18.9 (2021), S. 1038–1045.
- [118] Nicola Dietler et al. „A convolutional neural network segments yeast microscopy images with high accuracy“. In: *Nature communications* 11.1 (2020), S. 5723.
- [119] S. Holden und M. Conduit. *DeepBacs – Bacillus subtilis fluorescence segmentation dataset*. Data set. 2021.
- [120] Christoph Spahn et al. „DeepBacs: Bacterial image analysis using open-source deep learning approaches“. In: *bioRxiv* (2021).
- [121] Vladimír Ulman et al. „An objective comparison of cell-tracking algorithms“. In: *Nature methods* 14.12 (2017), S. 1141–1152.

- [122] Neeraj Kumar et al. „A dataset and a technique for generalized nuclear segmentation for computational pathology“. In: *IEEE transactions on medical imaging* 36.7 (2017), S. 1550–1560.
- [123] Noah F Greenwald et al. „Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning“. In: *Nature biotechnology* 40.4 (2022), S. 555–565.
- [124] Florian Kromp et al. „An annotated fluorescence image dataset for training nuclear segmentation methods“. In: *Nature Scientific Data* 7.262 (2020), S. 1–8.
- [125] Alain Chen et al. „3d ground truth annotations of nuclei in 3d microscopy volumes“. In: *bioRxiv* (2022), S. 2022–09.
- [126] Chichen Fu et al. „Three dimensional fluorescence microscopy image synthesis and segmentation“. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2018, S. 2221–2229.
- [127] Nazmiye Ceren Abay et al. „Privacy Preserving Synthetic Data Release Using Deep Learning“. en. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Hrsg. von Michele Berlingario et al. Bd. 11051. Cham: Springer International Publishing, 2019, S. 510–526.
- [128] Trivellore E. Raghunathan. „Synthetic Data“. en. In: *Annual Review of Statistics and Its Application* 8.1 (März 2021), S. 129–140.
- [129] Sergey I Nikolenko et al. *Synthetic data for deep learning*. Bd. 174. Springer, 2021.
- [130] Edward Choi et al. „Generating multi-label discrete patient records using generative adversarial networks“. In: *Machine learning for healthcare conference*. PMLR, 2017, S. 286–305.
- [131] Yingzhou Lu et al. „Machine learning for synthetic data generation: a review“. In: *arXiv preprint arXiv:2302.04062* (2023).
- [132] Roman Bruch et al. „Improving 3D deep learning segmentation with biophysically motivated cell synthesis“. In: *Communications Biology* 8.1 (2025), S. 43.
- [133] Rishi Bommasani et al. „On the opportunities and risks of foundation models“. In: *arXiv preprint arXiv:2108.07258* (2021).
- [134] Jason Yosinski et al. „How transferable are features in deep neural networks?“ In: *Advances in neural information processing systems* 27 (2014).
- [135] Jonas Dippel et al. „Transfer Learning for Segmentation Problems: Choose the Right Encoder and Skip the Decoder“. In: *arXiv preprint arXiv:2207.14508* (2022).
- [136] Huiyu Wang et al. „Max-deeplab: End-to-end panoptic segmentation with mask transformers“. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, S. 5463–5474.
- [137] Xueyan Zou et al. „Segment everything everywhere all at once“. In: *Advances in neural information processing systems* 36 (2023), S. 19769–19782.

- [138] Jitesh Jain et al. „Oneformer: One transformer to rule universal image segmentation“. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2023, S. 2989–2998.
- [139] Feng Li et al. „Mask dino: Towards a unified transformer-based framework for object detection and segmentation“. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2023, S. 3041–3050.
- [140] Alexander Kirillov et al. „Segment anything“. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2023, S. 4015–4026.
- [141] Alexey Dosovitskiy et al. „An image is worth 16x16 words: Transformers for image recognition at scale“. In: *arXiv preprint arXiv:2010.11929* (2020).
- [142] Kaiming He et al. „Masked autoencoders are scalable vision learners“. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, S. 16000–16009.
- [143] Alec Radford et al. „Learning Transferable Visual Models From Natural Language Supervision“. en. In: PMLR, Juli 2021, S. 8748–8763.
- [144] Anwai Archit et al. „Segment anything for microscopy“. In: *Nature Methods* (2025), S. 1–13.
- [145] Uriah Israel et al. „A foundation model for cell segmentation“. In: *arXiv preprint arXiv:2311.11004* (2023).
- [146] Alexandra D VandeLoo et al. „SAMCell: Generalized Label-Free Biological Cell Segmentation with Segment Anything“. In: *bioRxiv* (2025).
- [147] Carsen Stringer et al. „Cellpose: a generalist algorithm for cellular segmentation“. In: *Nature methods* 18.1 (2021), S. 100–106.
- [148] Marius Pachitariu, Michael Rariden und Carsen Stringer. „Cellpose-SAM: superhuman generalization for cellular segmentation“. In: *bioRxiv* (2025), S. 2025–04.
- [149] David A Van Valen et al. „Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments“. In: *PLoS computational biology* 12.11 (2016), e1005177.
- [150] Dylan Bannon et al. „DeepCell Kiosk: scaling deep learning–enabled cellular image analysis with Kubernetes“. In: *Nature methods* 18.1 (2021), S. 43–45.
- [151] Noah F Greenwald et al. „Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning“. In: *Nature biotechnology* 40.4 (2022), S. 555–565.
- [152] Erick Moen et al. „Accurate cell tracking and lineage construction in live-cell imaging experiments with deep learning“. In: *Biorxiv* (2019), S. 803205.
- [153] Mingxing Tan und Quoc Le. „Efficientnetv2: Smaller models and faster training“. In: *International conference on machine learning*. PMLR, 2021, S. 10096–10106.

- [154] Fabian Isensee et al. „nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation“. In: *Nature methods* 18.2 (2021), S. 203–211.
- [155] Jonas Dippel et al. „Transfer Learning for Segmentation Problems: Choose the Right Encoder and Skip the Decoder“. In: *arXiv preprint arXiv:2207.14508* (2022).
- [156] Olga Russakovsky et al. „ImageNet Large Scale Visual Recognition Challenge“. en. In: *International Journal of Computer Vision* 115.3 (Dez. 2015), S. 211–252.
- [157] Yang You et al. „ImageNet Training in Minutes“. en. In: *Proceedings of the 47th International Conference on Parallel Processing*. Eugene OR USA: ACM, Aug. 2018, S. 1–10.
- [158] Simon Kornblith, Jonathon Shlens und Quoc V. Le. „Do Better ImageNet Models Transfer Better?“ In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Juni 2019, S. 2661–2671.
- [159] Lucas Beyer et al. „Are we done with ImageNet?“ In: *arXiv preprint arXiv:2006.07159* (Juni 2020). arXiv:2006.07159 [cs].
- [160] Benjamin Recht et al. „Do imagenet classifiers generalize to imagenet?“ In: *International conference on machine learning*. PMLR, 2019, S. 5389–5400.
- [161] Kaiming He et al. „Deep residual learning for image recognition“. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, S. 770–778.
- [162] Kaiming He und Jian Sun. „Convolutional neural networks at constrained time cost“. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, S. 5353–5360.
- [163] Rupesh Kumar Srivastava, Klaus Greff und Jürgen Schmidhuber. „Highway networks“. In: *arXiv preprint arXiv:1505.00387* (2015).
- [164] Sergey Ioffe. „Batch renormalization: Towards reducing minibatch dependence in batch-normalized models“. In: *Advances in neural information processing systems* 30 (2017).
- [165] Vinod Nair und Geoffrey E Hinton. „Rectified linear units improve restricted boltzmann machines“. In: *Proceedings of the 27th international conference on machine learning (ICML-10)*. 2010, S. 807–814.
- [166] Mingxing Tan und Quoc Le. „Efficientnet: Rethinking model scaling for convolutional neural networks“. In: *International conference on machine learning*. PMLR, 2019, S. 6105–6114.
- [167] Mark Sandler et al. „Mobilenetv2: Inverted residuals and linear bottlenecks“. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, S. 4510–4520.
- [168] Suyog Gupta und Mingxing Tan. „EfficientNet-EdgeTPU: Creating accelerator-optimized neural networks with AutoML“. In: *Google AI Blog* 2.1 (2019).
- [169] Zhuang Liu et al. „A convnet for the 2020s“. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, S. 11976–11986.

- [170] Dan Hendrycks und Kevin Gimpel. „Gaussian error linear units (gelus)“. In: *arXiv preprint arXiv:1606.08415* (2016).
- [171] Ze Liu et al. „Swin transformer v2: Scaling up capacity and resolution“. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, S. 12009–12019.
- [172] John Bridle. „Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters“. In: *Advances in neural information processing systems* 2 (1989).
- [173] Diederik Kinga, Jimmy Ba Adam et al. „A method for stochastic optimization“. In: *International conference on learning representations (ICLR)*. Bd. 5. 6. California; 2015.

Erklärung

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Karlsruhe, den 5. November 2025

.....