

# **Synthese von 3D-Mikroskopdaten für die Deep-Learning-gestützte Analyse**

Zur Erlangung des akademischen Grades eines

**DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)**

von der KIT-Fakultät für Maschinenbau des  
Karlsruher Instituts für Technologie (KIT)  
angenommene

**DISSERTATION**

von

**M.Sc. Roman Bruch**

Tag der mündlichen Prüfung:

11.07.2025

Hauptreferent:

Prof. Dr.-Ing. Markus Reischl

Korreferentin:

Prof. Dr. rer. nat. Britta Nestler

Korreferent:

Prof. Dr. rer. nat. Rüdiger Rudolf



# Abstract

In materials science and biology, high-dimensional image data offers considerable potential for the development of new materials and active substances. However, in order to utilise the potential of this complex data, precise and automated analysis methods are required. Deep learning methods have proven to be efficient for analysing complex data. However, their training requires large amounts of annotated data, the manual creation of which is almost impossible to realise with high-dimensional data. One solution to this challenge is the synthesis of training data. A central contribution of this work is the development of SimOptiGAN, a method for generating synthetic 3D training data. A major innovation of this method is the targeted integration of rare elements. In addition, for the first time, biophysical simulations are integrated into the synthesis process to improve the arrangement of objects. Comprehensive comparisons prove that the method developed in this work significantly increases the segmentation performance on real data compared to existing approaches. Another major contribution of this study is the development of the IMS framework, which overcomes the limitation of the number of simultaneously detectable markers in fluorescence microscopy. Experimental results demonstrate that the IMS framework achieves a strong correlation between synthetic and real image signals, thereby enabling improvements in classification performance. Thanks to their modular structure, the methods developed in this thesis can be flexibly adapted to different application areas of image analysis. The data synthesis methodology developed in this thesis therefore makes an important contribution to the automated analysis of high-dimensional data, and thus enables a more efficient utilisation of such complex data sets.

# Kurzfassung

In der Werkstoffkunde und Biologie bieten hochdimensionale Bilddaten ein erhebliches Potenzial für die Entwicklung neuartiger Materialien und Wirkstoffe. Um das Potenzial dieser komplexen Daten zu nutzen, bedarf es jedoch präziser und automatisierter Analysemethoden. Deep-Learning-Methoden haben sich für die Analyse komplexer Daten als leistungsfähig erwiesen. Ihr Training setzt jedoch große Mengen an annotierten Daten voraus, deren manuelle Erstellung bei hochdimensionalen Daten kaum realisierbar ist. Eine Lösung für diese Herausforderung stellt die Synthese von Trainingsdaten dar. Ein zentraler Beitrag der vorliegenden Arbeit ist die Entwicklung von SimOptiGAN, einer Methode zur Generierung synthetischer 3D-Trainingsdaten. Eine wesentliche Neuerung der Methode besteht in der gezielten Einbindung seltener Elemente. Zudem werden erstmals biophysikalische Simulationen in den Syntheseprozess integriert, um die Anordnung von Objekten zu verbessern. Umfassende Vergleiche zeigen, dass die entwickelte Methode die Segmentierungsleistung auf realen Daten im Vergleich zu bestehenden Ansätzen signifikant verbessert. Ein weiterer wesentlicher Beitrag ist die Entwicklung des Frameworks IMS, das die Limitierung der Anzahl gleichzeitig erfassbarer Marker in der Fluoreszenzmikroskopie überwindet. In den durchgeführten Experimenten zeigt IMS eine hohe Korrelation zwischen synthetischen und realen Signalen und ermöglicht zudem eine Verbesserung der Klassifikationsleistung. Dank der modularen Struktur lassen sich die entwickelten Methoden flexibel an unterschiedliche Anwendungsbereiche der Bildanalyse anpassen. Die in dieser Arbeit entwickelte Methodik zur Datensynthese leistet somit einen wichtigen Beitrag zur automatisierten Analyse hochdimensionaler Daten und ermöglicht dadurch eine effizientere Nutzung solcher komplexen Datensätze.



# Vorwort

Die vorliegende Arbeit entstand in meiner Zeit als Doktorand an dem Institut für Molekular- und Zellbiologie an der Hochschule Mannheim und dem Institut für Automation und angewandte Informatik am Karlsruher Institut für Technologie. In dieser Zeit habe ich viel Motivation und Unterstützung erhalten, wofür ich mich im Folgenden bedanken will.

Zuerst danke ich meinem Doktorvater Herr Prof. Markus Reischl, der durch sein Engagement und seine Unterstützung maßgeblich zum Gelingen dieser Dissertation beigetragen hat. Er hat nicht nur ideale Rahmenbedingungen für meine Promotion geschaffen, sondern mich auch durch seine ausgezeichnete fachliche und menschliche Betreuung stets begleitet. Besonders schätze ich sein stetig offenes Ohr für meine Anliegen und die pragmatischen Tipps, die mir insbesondere in herausfordernden Situationen geholfen haben. Die anregenden Diskussionen mit ihm waren eine wertvolle Inspirationsquelle und haben wesentlich dazu beigetragen, meine Arbeit voranzutreiben. Auch danke ich ihm für den sportlichen Ehrgeiz und die bewältigten sportlichen Herausforderungen.

Zudem danke ich Frau Prof. Britta Nestler für die Übernahme des Korreferats und das damit verbundene Interesse an dieser Arbeit. Ebenso schätze ich ihr wertvolles Feedback, das zum Erfolg dieser Arbeit beigetragen hat.

Auch danke ich meinem Betreuer Prof. Rüdiger Rudolf, der mir die Möglichkeit eröffnet hat, eine Promotion in diesem spannenden Forschungsbereich durchzuführen. Seine stetige Unterstützung sowie die hervorragende Betreuung haben mich während der gesamten Arbeit begleitet und entscheidend geprägt. Besonders dankbar bin ich für die gewinnbringenden und ideengebenden Diskussionen, die meine Forschung stets bereichert haben. Seine ansteckende Begeisterung für

die Biologie und die Erforschung neuer Technologien war inspirierend und sehr motivierend. Ebenso schätze ich die Freiheiten, die er mir bei der Wahl meines Forschungsthemas eingeräumt hat.

Mein herzlicher Dank gilt auch meinen geschätzten Kolleginnen und Kollegen aus Mannheim und Karlsruhe, die mich während meiner Promotion sowohl fachlich als auch persönlich unterstützt haben. Besonders dankbar bin ich den Kolleginnen und Kollegen der Arbeitsgruppe aus Mannheim für die Erklärungen komplexer biologischer Prozesse und die Hilfsbereitschaft und Unterstützung bei der Durchführung von Experimenten, die für den Fortschritt meiner Arbeit von großem Wert waren. Im Speziellen danke ich Mario Vitacolonna, Florian Keller, Elina Nürnberg, Nathalie Couturier, Tatjana Straka, Sarah Hörner und Tiziana Cesetti. In diesem Zuge möchte ich auch dem Projektpartner Prof. Simeon Sauer für die gewinnbringende Zusammenarbeit danken.

Den Kolleginnen und Kollegen aus Karlsruhe bin ich für die fachlichen Diskussionen und wertvollen Ratschläge dankbar. Die gemeinsamen Erlebnisse auf Konferenzen, Retreats und privaten Events bleiben mir in bester Erinnerung. Nicht zuletzt danke ich für den sportlichen Ausgleich, der mir geholfen hat, neue Energie zu schöpfen. Im Speziellen danke ich Marcel Schilling, Luca Rettenberger, Moritz Böhlend, Oliver Neumann, André Orth, Hawo Höfer, Lorenz Wühl, Friedrich Münke, Jan Schützke und Tim Scherr.

Auch danke ich meiner Freundin Eileen, welche mich in der letzten Phase begleitet, unterstützt und motiviert hat. Zuletzt möchte ich meinen Eltern Ulla und Detlef sowie meinem Bruder Carsten für die unermessliche Unterstützung, das Vertrauen, die stetige Motivation und die sehr geschätzten Ratschläge auf meinem Lebensweg danken.

Bruchsal, im März 2025

Roman Bruch

# Inhaltsverzeichnis

<b>Abstract</b>	<b>i</b>
<b>Kurzfassung</b>	<b>ii</b>
<b>Vorwort</b>	<b>iii</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Motivation	1
1.2 Erhebung von hochdimensionalen Bilddaten	2
1.2.1 Übersicht	2
1.2.2 Konfokale Mikroskopie	4
1.2.3 Mikroskopische biologische Daten	7
1.3 Datenauswertung	8
1.3.1 Übersicht	8
1.3.2 Segmentierung	10
1.3.3 Reduktion des Annotationsaufwands	12
1.4 Synthese von experimentellen Daten	14
1.4.1 Übersicht	14
1.4.2 Simulationsbasierte Methoden	17
1.4.3 Datengetriebene Methoden	19
1.4.4 Hybride Methoden	22
1.4.5 Synthese von Bildmodalitäten	23
1.5 Offene Probleme	25
1.6 Zielsetzung und Struktur der Arbeit	27
<b>2 Neue Methodik zur Datensynthese</b>	<b>29</b>
2.1 Übersicht	29

2.2	Synthese von Trainingsdaten . . . . .	33
2.2.1	Übersicht . . . . .	33
2.2.2	SimOptiGAN – Eine neue Methode zur Symbiose von Simulation und Deep Learning . . . . .	36
2.2.3	Einbringung biophysikalischer Modellierung in den Syntheseprozess . . . . .	45
2.2.4	Bewertung von synthetischen Trainingsdaten . . . . .	62
2.3	Synthese von Bildmodalitäten . . . . .	67
2.3.1	Übersicht . . . . .	67
2.3.2	Neues Framework zur Synthese von Bildmodalitäten . . . . .	69
2.3.3	Bewertung von synthetischen Markersignalen . . . . .	72
2.3.4	Zellzyklus-Klassifikation . . . . .	78
2.3.5	Voraussetzungen für eine erfolgreiche Synthese von Bildmodalitäten . . . . .	82
<b>3</b>	<b>Implementierung . . . . .</b>	<b>84</b>
3.1	Übersicht . . . . .	84
3.2	Implementierung der entwickelten Methoden . . . . .	85
3.2.1	Synthese von Trainingsdaten . . . . .	85
3.2.2	Synthese von Bildmodalitäten . . . . .	86
3.2.3	Spherolyzer - Eine neue Software für die Analyse von Drug-Screenings . . . . .	87
3.3	Grafische Benutzeroberflächen . . . . .	89
3.3.1	Spherolyzer – Zugehörige Benutzerschnittstelle . . . . .	89
3.3.2	3D-ClassAnnotator – Grafische Oberfläche zur Klassenannotation in 3D-Daten . . . . .	90
<b>4</b>	<b>Datensätze . . . . .</b>	<b>92</b>
4.1	Übersicht . . . . .	92
4.2	Datensatz zur Untersuchung des Einflusses seltener Elemente . . . . .	93
4.3	Datensätze zum Methodenvergleich der Trainingsdatensynthese . . . . .	97
4.3.1	Übersicht . . . . .	97
4.3.2	Reale Bilder zum Training der GAN-Modelle . . . . .	98
4.3.3	Biophysikalisch simulierte Zellgrenzen . . . . .	100

4.3.4	Händische Ground-Truth zur Bewertung der Segmentierungsleistung . . . . .	100
4.4	Datensatz zur Untersuchung der Synthese von Bildmodalitäten . . .	102
4.5	Datensatz zur Zellzyklus-Klassifikation anhand von synthetischen Markern . . . . .	104
<b>5</b>	<b>Anwendung . . . . .</b>	<b>107</b>
5.1	Übersicht . . . . .	107
5.2	Synthese von Trainingsdaten . . . . .	108
5.2.1	Übersicht . . . . .	108
5.2.2	SimOptiGAN reduziert die Domänenlücke durch hybriden Ansatz . . . . .	108
5.2.3	Verbesserte Segmentierung durch Einbringung seltener Elemente . . . . .	110
5.2.4	Eingeführte Methoden zeigen Vorteile in Aufwand und Segmentierungsgüte . . . . .	113
5.2.5	Membransynthese . . . . .	129
5.2.6	SimOptiGAN ermöglicht Datensynthese auch in anderen Forschungsbereichen . . . . .	130
5.2.7	Zusammenfassung . . . . .	132
5.3	Synthese von Bildmodalitäten . . . . .	136
5.3.1	Übersicht . . . . .	136
5.3.2	Methodenvergleich . . . . .	137
5.3.3	Zellzyklus-Klassifikation mit synthetischem Marker . . . . .	143
5.3.4	Bleed-Through-Test . . . . .	147
5.3.5	Machbarkeitsanalyse für weitere Marker . . . . .	151
5.3.6	Zusammenfassung . . . . .	157
<b>6</b>	<b>Zusammenfassung und Ausblick . . . . .</b>	<b>160</b>
<b>A</b>	<b>Anhang . . . . .</b>	<b>165</b>
A.1	Einführung generativer Netzwerke . . . . .	165
A.1.1	Generative Adversarial Networks . . . . .	165
A.1.2	Denoising-Diffusion-Probabilistic-Models (DDPM) . . . . .	173
A.2	Einführung in die Zellzyklusstadien . . . . .	175

A.3 Beschreibung des Modelltrainings und der Anwendung . . . . .	175
A.3.1 StarDist-Modelle . . . . .	175
A.3.2 SimOptiGAN . . . . .	176
A.3.3 SimOptiGAN+ . . . . .	177
A.3.4 Mem2NucGAN-P . . . . .	177
A.3.5 Mem2NucGAN-U . . . . .	178
A.4 Ablation Study bezüglich des zusätzlichen Segmentierungs-Diskriminators von DualGenGAN . . . . .	179
A.5 Statistische Analyse der Segmentierung . . . . .	180
A.6 Grafikspeicher bei Laufzeitexperiment zur Synthese von Bildmodalitäten . . . . .	180
A.7 Klassifikationsnetzwerk Zell-Zyklus-Klassifikation . . . . .	181
<b>Abkürzungen und Symbole . . . . .</b>	<b>184</b>
<b>Abbildungsverzeichnis . . . . .</b>	<b>193</b>
<b>Tabellenverzeichnis . . . . .</b>	<b>197</b>
<b>Eigene Veröffentlichungen . . . . .</b>	<b>198</b>
<b>Literaturverzeichnis . . . . .</b>	<b>201</b>

# 1 Einleitung

## 1.1 Motivation

In einer Ära zunehmender Digitalisierung und Automatisierung gewinnt die automatische Auswertung von Daten im Bereich des Maschinenbaus zunehmend an Bedeutung. Die stetig wachsende Datenmenge, die durch fortschrittliche Sensortechnologien und automatisierte Fertigungsprozesse generiert wird, bietet ein enormes Potenzial für die Optimierung von Produktionsprozessen, die Entwicklung von optimierten Materialien und Wirkstoffen und die Verbesserung und Sicherung der Qualität von Produkten. Unter dem Stichwort Industrie 4.0 werden etwa digitale Zwillinge eingesetzt, welche beispielsweise ein digitales Abbild reeller Maschinen darstellen [1]. Hierbei können Messdaten von einer realen Maschine und deren Produkten erhoben und mittels Datenverarbeitung Prognosen bezüglich optimaler Betriebsbereiche oder Wartungsintervalle getroffen werden. Im Bereich des autonomen Fahrens werden Sensordaten genutzt und kombiniert, um daraus Entscheidungen abzuleiten [2]. Im Bereich der Materialwissenschaften ermöglicht eine automatisierte Auswertung von Informationen, welche durch Experimente, Simulationen und Messungen gewonnen werden, maßgeschneiderte Werkstoffe mit spezifischen Eigenschaften zu entwickeln [3, 4]. Durch automatische Extraktionen von relevanten Merkmalen können komplexe Datensätze effizient analysiert und so neue Einblicke gewonnen werden.

Auch in der biomedizinischen Forschung spielt die automatische Auswertung von Daten eine entscheidende Rolle. Die Analyse von bildgebenden Verfahren wie MRT, CT und Mikroskopie erfordert oft die Identifizierung und Quantifizierung von Strukturen, Mustern und Anomalien, für deren Erkennung ein hohes Maß an

Erfahrung und Fachwissen benötigt wird. Durch den Einsatz von Deep Learning und anderen fortgeschrittenen Analysemethoden können Forscher automatisch relevante Informationen extrahieren, Krankheiten diagnostizieren und Therapien und Wirkstoffe optimieren [5, 6, 7].

Klassischerweise lässt sich der Ablauf einer automatisierten Datenauswertung in die folgenden Abschnitte unterteilen. Der erste entscheidende Schritt liegt in der Aufnahme oder Erhebung der Messdaten. Hierbei ist es von zentraler Bedeutung, die gewünschten Eigenschaften optimal zu messen, um etwa ein gutes Signal-Rausch-Verhältnis zu erzielen. Bei Mikroskopen ist unter anderem die Lichtintensität und die Auflösung unter Berücksichtigung der Nyquist-Rate einzustellen und auf eine angemessene Qualität der Proben zu achten. Nach der Erhebung der Daten folgt die Datenaufbereitung, bei der verschiedene Messdaten zusammengeführt oder in ein für die Weiterverarbeitung geeignetes Format umgewandelt werden. Beispielsweise sind die Aufnahmen von mehreren Proben in eine geordnete Struktur zu bringen. Anschließend kann eine Vorverarbeitung der Daten erfolgen, bei der etwa die Aufnahmen gefiltert oder dekonvolviert werden, um Rauschen und Lichtstreuung zu reduzieren. Nach dieser Vorverarbeitung findet die eigentliche Datenanalyse statt, bei der etwa die vorhandenen Strukturen erkannt und klassifiziert werden. Im Zuge der Daten-Nachbearbeitung können die Ergebnisse visualisiert oder statistische Tests durchgeführt werden.

## 1.2 Erhebung von hochdimensionalen Bilddaten

### 1.2.1 Übersicht

Zweidimensionale Bilddaten sind aus dem täglichen Leben allgemein bekannt, beispielsweise durch die Fotografie. 3D-Bilddaten sind dagegen im alltäglichen Leben selten anzutreffen. Bei 3D-Aufnahmen fügt sich die zusätzliche Dimension der Tiefe hinzu, was eine umfassendere Erfassung von Objekten ermöglicht, aber



auch eine Herausforderung für die Auswertung darstellt. Es gibt zwei gängige Darstellungsformen für 3D-Bilddaten: Rastergrafiken und Punktwolken.

Rastergrafiken stellen eine strukturierte Anordnung von Datenpunkten in einem Raster oder Gitter dar. Im Fall von 3D-Rastergrafiken werden die Elemente nicht mehr als Pixel, sondern als Voxel bezeichnet. Jedes Voxel enthält Informationen über die Intensität. 3D-Rastergrafiken eignen sich für die Darstellung von Intensitäten innerhalb von Volumina. Bekannte Beispiele für 3D-Rastergrafiken sind CT- oder MRT-Aufnahmen. Punktwolken hingegen sind unstrukturierte Darstellungen, bei denen Punkte flexibel im Raum angeordnet sind. Jeder Punkt besitzt eine genaue Positionsinformation und kann zusätzlich Informationen wie die Intensität enthalten [8]. Punktwolken eignen sich für die Darstellung von dreidimensionalen Oberflächen, wie sie beispielsweise von Tiefenkameras oder Lidar-Sensoren erfasst werden. Generell unterscheiden sich die Methoden zur Auswertung der beiden Darstellungsformen. Weiterhin wurden auch Methoden entwickelt, welche die Darstellungsform der Punktwolken in Rastergrafiken überführen [9, 10, 11, 12]. Da bei Punktwolken nur die Oberflächen erfasst werden, erfolgt die Umwandlung unter Verlust der Tiefendimension zu 2D-Bilddaten. In dieser Arbeit wird sich auf die Darstellungsform der Rastergrafik konzentriert.

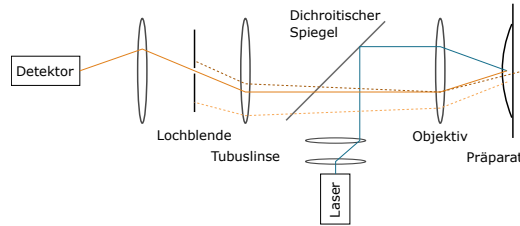
Verfahren wie MRT, CT, Ultraschall und Mikroskopie ermöglichen die Erfassung dreidimensionaler Daten. CT-Aufnahmen werden beispielsweise in der Werkstoffforschung und -prüfung für zerstörungsfreie Untersuchungen von Proben eingesetzt [13], während sie in der Medizin die Visualisierung des Inneren des menschlichen Körpers ermöglichen [14]. Mikroskopaufnahmen in der Biologie ermöglichen die Untersuchung von dreidimensionalen Zellstrukturen [15]. Mit der Aufnahme von unterschiedlichen Kanälen oder einer zusätzlichen zeitlichen Komponente ergeben sich sogar 4D- oder 5D-Daten. Hochdimensionale Daten ermöglichen vielfältige Anwendungsmöglichkeiten, stellen jedoch aufgrund ihrer hohen Datenkomplexität und -größe spezielle Anforderungen an Speicherung und Auswertung dar.

## 1.2.2 Konfokale Mikroskopie

Es stehen zahlreiche Mikroskopietechniken zur Verfügung, die zur Aufnahme von Proben eingesetzt werden können. Eine umfassende Übersicht über diese Techniken findet sich in [16]. In dieser Arbeit liegt der Fokus auf der Fluoreszenzmikroskopie unter Verwendung von Konfokalmikroskopen. Die konfokale Mikroskopie ist eine leistungsfähige Bildgebungstechnik, die insbesondere in den Bereichen Biologie und Materialwissenschaft Anwendung findet. In Kombination mit fluoreszierenden Stoffen ermöglicht sie die hochauflösende Aufnahme von großen 3D-Strukturen in toto (im Ganzen) mit einem guten Kontrast.

Die Grundlage der Fluoreszenzmikroskopie stellen Fluorophore dar. Das sind Moleküle, welche die Eigenschaft besitzen, Photonen einer bestimmten Energie zu absorbieren und Photonen mit geringerer Energie zu emittieren [17]. Das emittierte Signal weist aufgrund seiner geringeren Energie immer eine größere Wellenlänge auf als das anregende Signal. Diese Eigenschaft ermöglicht es, im Vergleich zu Durchlichtmikroskopen einen hohen Kontrast zu erzielen. Bei der Filterung des anregenden Signals hebt sich das emittierte Licht stark von einem ansonsten dunklen Hintergrund ab. Dafür ist allerdings eine klare Trennung des anregenden Signals von dem emittierten Signal notwendig. Unter regulären Bedingungen ergibt sich selbst bei einer Anregung mit einer genau abgegrenzten Wellenlänge ein kontinuierliches Spektrum [16], welches sich mit der anregenden Quelle überlagern kann. Weiterhin ist die Intensität des emittierten Lichts deutlich geringer als die des anregenden Lichts [15]. Zur Trennung der Signale werden unter anderem dichroitische Spiegel eingesetzt, welche Licht anhand ihrer Wellenlänge entweder reflektieren oder durchlassen [15]. In der Anwendung sind diese Spiegel so ausgelegt, dass sie das anregende Licht reflektieren und das emittierte Licht mit der größeren Wellenlänge durchlassen. Zusätzlich werden noch Anregungs- und Emissionsfilter eingesetzt, um die Signaltrennung weiter zu verbessern.

Abbildung 1.1 zeigt in vereinfachter Darstellung den optischen Aufbau eines konfokalen Fluoreszenzmikroskops, welches einen Laser als Lichtquelle nutzt. Lichtstrahlen des Lasers (blau), werden zuerst durch eine Linse aufgeweitet und



**Abbildung 1.1:** Vereinfachte Darstellung eines konfokalen Fluoreszenzmikroskops. Für eine erhöhte Übersichtlichkeit wurde jeweils nur eine begrenzende Seite des Lichtstrahls eingezeichnet. Abbildung in Anlehnung an [15]

so an den Durchmesser des rückseitigen Objektivs angepasst, wodurch eine optimale Fokussierung ermöglicht wird. Anschließend wird das Licht über einen dichroitischen Spiegel reflektiert, wonach es das Objektiv passiert und fokussiert auf die Probe trifft. Durch die Anregung fluoreszierender Stoffe wird Licht in einer anderen Wellenlänge emittiert (orange), welches das Objektiv passiert und auf den dichroitischen Spiegel zurückfällt. Da das emittierte Licht eine andere Wellenlänge als das anregende Licht aufweist, wird es nicht reflektiert, sondern kann den dichroitischen Spiegel passieren. Anhand einer weiteren Lochblende wird der Großteil der unerwünschten Strahlung, die nicht der Fokusebene entspringt, zurückgehalten (gestrichelte Linien). Zuletzt wird das Licht durch eine Linse auf den Detektor geleitet. Da durch die punktuelle Anregung und die Lochblende immer nur ein kleiner Punkt der Probe abgetastet wird, sind keine 2D-Bildsensoren notwendig. Als Detektoren haben sich Photoelektronenvervielfacher (engl. photomultiplier tube, PMT) etabliert. Um ein Bild der gesamten Probe zu erhalten, ist eine rasterförmige Abtastung der Probe notwendig, wozu in der Regel bewegliche Spiegel eingesetzt werden. Die einzeln abgetasteten Punkte werden softwareseitig zu einem vollständigen Bild zusammengesetzt.

Häufig werden mehrere Färbungen eingesetzt, um unterschiedliche Strukturen gleichzeitig sichtbar zu machen (Multiplexing). Dabei ist eine große Sorgfalt bei der Auswahl der verwendeten Färbungen bezüglich ihrer Emissionsspektren zu beachten. Idealerweise liegen diese so weit auseinander, dass es zu keiner Überlagerung kommt. In der Praxis ist dies aufgrund der Breite der Emissionsspektren gerade bei vielen gleichzeitigen Färbungen nicht immer möglich. Überlagern sich

Spektren, können auch mit perfekten Filtern die Signale nicht mehr vollständig voneinander getrennt werden und es ergibt sich ein unspezifisches Signal. Dabei wird auch von einem Bleed-Through oder Cross-Talk gesprochen, dem unerwünschten Durchscheinen einer Färbung in einen anderen Kanal. In Kombination mit einer Limitierung der verfügbaren Lichtquellen ergibt sich daraus eine Limitierung an gleichzeitig einsetzbaren Färbungen [15, 18].

Bei konfokalen Mikroskopen ist die Antwort des Systems unter idealen Bedingungen nahezu ortsunabhängig. Dabei handelt es sich um einen linearen Abbildungsprozess, bei dem das resultierende Signal als Summe einzelner Punkte beschrieben werden kann. Ein wichtiger Parameter der konfokalen Mikroskopie stellt die Punktspreizfunktion (PSF) dar. Diese beschreibt die Antwort eines Mikroskops auf einen idealisierten fluoreszierenden Punkt [19]. Mathematisch wird die Bildgebung über eine Faltung der Bildinhalte mit der PSF des Mikroskops beschrieben. Die PSF kann entweder berechnet oder experimentell bestimmt werden. In der Realität weisen alle Mikroskope leichte Abweichungen von den in den Modellen angenommenen Eigenschaften auf, wodurch es zu Differenzen zwischen der berechneten und gemessenen PSF kommt. Die Ortsunabhängigkeit des Systems gilt jedoch nur für idealisierte Proben ohne Medienübergänge. Durch den Übergang zwischen verschiedenen Medien innerhalb der Probe kommt es zu Lichtbrechungen und damit zu unerwünschten Streuungen des Lichtsignals. Das führt in tieferen Bildschichten zu einer veränderten PSF, was sich in einer abnehmenden Signalqualität und durch die geringere Signalintensität auch in einem geringeren Signal-Rausch-Verhältnis äußert.

Das Signal wird weiterhin von verschiedenen Rauschquellen beeinflusst. Ein großer Anteil des Rauschens wird durch Quanteneigenschaften des Lichts verursacht [19, 20]. Photonen werden nur mit einer bestimmten Wahrscheinlichkeit emittiert, was als Schrotrauschen bekannt ist und über die Poissonverteilung modelliert werden kann. Daneben gibt es weitere Rauschquellen wie den Dunkelstrom von Sensoren, bei dem auch ohne Lichteinfall ein Signal gemessen wird. Eine Übersicht von Rauschquellen und eine detaillierte Modellierung der Rauschanteile für Konfokalmikroskope findet sich in [20].

### 1.2.3 Mikroskopische biologische Daten

Zellkulturen spielen eine zentrale Rolle in der Wirkstoffforschung, da sie es ermöglichen, Wirkungsmechanismen zu untersuchen und die Toxizität sowie Spezifität von Substanzen in frühen Forschungsphasen zu überprüfen [21, 22, 23, 24]. Durch fluoreszierende Marker können spezifisch Zellbestandteile sichtbar gemacht und Zellstadien und Wirkungsmechanismen visualisiert werden. Beispiele sind Marker wie DAPI, Ki-67 oder Caspase-3, welche Zellkerne, Zellteilung und Zelltod sichtbar machen. Die Analyse von mikroskopischen Zellaufnahmen stellt eine besondere Herausforderung dar. Im Gegensatz zu natürlichen Aufnahmen von Straßenszenen oder CT- und MRT-Bildern, bei denen wenige größere Objekte vorliegen, handelt es sich bei zellbiologischen Aufnahmen um viele dicht beieinander liegende und vergleichsweise kleine Objekte. Um solche Aufnahmen auf Einzelzelebene zu analysieren, werden speziell auf diese Anwendungen zugeschnittene Algorithmen benötigt.

Zunehmend werden auch 3D-Zellkulturen wie Sphäroide und Organoide eingesetzt, die im Gegensatz zu herkömmlichen 2D-Kulturen eine größere Ähnlichkeit zu menschlichem Gewebe aufweisen [25, 26, 27, 28]. So bieten diese Kulturen größere Kontaktflächen zwischen den Zellen und Einflüsse wie Nähr- bzw. Wirkstoffgradienten kommen hinzu. Die Nutzung von 3D-Zellkulturen birgt allerdings auch Herausforderungen, wie die aufwendigere Kultivierung, Bildgebung und Auswertung. Die dreidimensionale Struktur führt bei Aufnahmen zu einer erhöhten Lichtbrechung, was die Erfassung und Quantifizierung von Signalen erschwert. Um die Lichtbrechung zu reduzieren und so auch tiefere Schichten aufnehmbar zu machen, werden Verfahren zur Gewebeklärung eingesetzt. Die Effektivität der Verfahren unterscheidet sich und ist unter anderem abhängig von der verwendeten Zelllinie [29]. Im Vergleich zu 2D-Kulturen werden aufgrund der starken Heterogenität der Daten oft komplexere Bildverarbeitungs- und Analysetechniken benötigt, um auch Strukturen in tiefen Bildebenen sicher zu erkennen. Charakteristisch für die Daten ist beispielsweise eine anisotrope Auflösung und eine ungleichmäßige und in der Tiefe abnehmende Helligkeit/Qualität, was zu einem

stark variierenden Signal-Rausch-Verhältnis führt. Zudem variiert die Morphologie der Zellen zwischen und innerhalb der Proben aufgrund von unterschiedlichen Zellarten und Zellstadien. Die Entwicklung von geeigneten Methoden zur Auswertung von 3D-Bildern ist daher von entscheidender Bedeutung, um das volle Potenzial dieser Zellkulturmodelle auszuschöpfen.

## 1.3 Datenauswertung

### 1.3.1 Übersicht

In der Vergangenheit wurden für die automatisierte Auswertung von Daten meist parametrisierbare Pipelines verwendet. Diese Pipelines wurden von Experten an die spezifischen Daten angepasst und konnten anschließend vom Benutzer durch verschiedene Parameter feinjustiert werden. Aufgrund der teilweise geringen Intuitivität mancher Parameter erfolgte deren Einstellung häufig nach dem Trial-and-Error-Prinzip, bei dem verschiedene Parameter getestet und die Ergebnisse dann visuell verglichen wurden. Diese Vorgehensweise führte jedoch zu einem hohen subjektiven Einfluss auf die Auswertung. In einigen Fällen wurden zur Überprüfung der Methode manuell annotierte Daten verwendet, die als Grundwahrheit (engl. ground truth) dienen. Mit solchen Daten konnte der Einfluss der Parameter überprüft und die Leistung der Pipeline objektiv bewertet werden. Bei einfachen klassischen Segmentierungsmethoden werden meist automatische Grenzwertverfahren wie das Otsu-Verfahren [30] eingesetzt, um eine erste Segmentierungsmaske zu erzeugen, die anschließend mit morphologischen Operationen verbessert wird [21, 31]. Fortgeschrittenere Pipelines kombinieren verschiedene Filter, um die Segmentierungsqualität zu verbessern, und nutzen spezielle Algorithmen zur Trennung von Objektinstanzen [24, 32, 33, 34, 35, 36, 37, 38, 39]. Tools wie XPIWIT [40] oder CellProfiler [35] ermöglichen es, auch ohne Programmierkenntnisse komplexe Pipelines zu erstellen und an neue Gegebenheiten anzupassen. Dennoch erfordert die Erstellung solcher Pipelines nach wie vor ein hohes Maß an Expertenwissen.

Durch den zunehmenden Erfolg von künstlicher Intelligenz hat sich die Herangehensweise zur Auswertung von Daten verändert. Gerade im medizinischen und biologischen Bereich mit heterogenen Bedingungen haben sich Deep-Learning-Verfahren als leistungsfähig erwiesen [41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51]. Wohingegen bei klassischen Methoden eine ausgewählte Pipeline parametrisiert wurde, wird nun eine Netzwerkarchitektur ausgewählt und das Netzwerk anhand einer großen Menge an händisch generierten Ground-Truth-Daten trainiert und validiert. Ein verbreitetes Problem, insbesondere bei großen leistungsfähigen Netzwerken, ist die benötigte Menge an Trainingsdaten, die für eine gute Leistung der Modelle erforderlich ist. Während die Annotation von 2D-Daten bereits zeitaufwendig und mühsam ist, stellt die Annotation von 3D-Daten – nicht nur wegen der größeren Datenmengen – eine nahezu unmögliche Aufgabe dar. Eingeschränkte 2D-Visualisierungen erschweren die Vorstellung der räumlichen Struktur und führen zu inkonsistenten Annotationen zwischen den Ebenen. Weiterhin führt die stärkere Signalstreuung in  $z$ -Richtung und das nicht isotrope Auflösungsvermögen zu unscharfen Objektgrenzen, woraus sich auch bei runden Objekten häufig zylinderförmige Annotationen ergeben. Zudem sorgt die Signalstreuung vor allem in tieferen Regionen für Probleme bei der Abtrennung einzelner Objekte. Bei komplexen 3D-Daten ist eine händische Annotation daher nicht praktikabel. Der hohe Aufwand für die händische Annotation wird häufig als Grund für die begrenzte Verbreitung von Deep-Learning-Verfahren zur 3D-Segmentierung angesehen [52, 53, 54, 55].

Gerade in diesen Fällen zeichnet sich eine weitere neue Vorgehensweise bei der Datenanalyse ab. Anstatt Trainingsdaten durch manuelle Annotation zu erzeugen, werden zunehmend synthetisch generierte Daten eingesetzt, bei denen die Ground-Truth schon bei der Generierung bekannt ist. Ziel ist es, die realen Daten möglichst originalgetreu abzubilden. Solche synthetischen Daten können einerseits klassisch mit Simulationen generiert werden, oder auch mit Deep-Learning-Ansätzen, welche häufig mit Methoden für unüberwachtes Lernen trainiert werden.

### 1.3.2 Segmentierung

Die Segmentierung von Bildern stellt einen wichtigen Schritt für die weitere Analyse von Bildinhalten dar [21, 31, 56, 57]. Im Unterschied zur Detektion, bei der nur die Koordinaten einer umliegenden Box (engl. bounding box) ausgegeben werden, besteht die Aufgabe bei der Segmentierung in der präzisen Lokalisierung von Objekten auf Pixelebene. Die Ergebnisse der Segmentierung werden als Segmentierungsmaske bezeichnet und in Bildform ausgegeben. Grundsätzlich wird zwischen der semantischen und der Instanz-Segmentierung unterschieden. Bei der semantischen Segmentierung wird zwischen verschiedenen Objektklassen wie Auto, Fußgänger oder Fahrradfahrer unterschieden, jedoch nicht zwischen einzelnen Instanzen derselben Klasse. Jede Objektklasse erhält einen eigenen Index in Form einer Intensität in der Segmentierungsmaske. Bei der Instanz-Segmentierung sind die Objektklassen uninteressant und es wird nur zwischen verschiedenen Objektinstanzen unterschieden. Hier erhält jedes Objekt in der Segmentierungsmaske einen eigenen Index. Als dritte, aber selten anzutreffende Art der Segmentierung ist die panoptische Segmentierung, bei welcher sowohl zwischen Objektklassen als auch deren Instanzen unterschieden wird [58]. Für die Ergebnisdarstellung werden für die Instanzen und Klassen jeweils eigene Segmentierungsmasken benötigt.

Deep-Learning-basierte Ansätze zur Segmentierung haben wiederholt gezeigt, dass sie klassischen Methoden insbesondere bei heterogenen Daten überlegen sind [41, 43, 44, 46, 47, 48, 49, 50, 51, 59]. Während klassische Methoden stets eine Anpassung an neue Datensätze erfordern, können komplexe Netzwerkstrukturen die Segmentierung heterogener Daten erlernen und somit eine robustere und übertragbare Lösung bieten.

Eine der wohl bekanntesten Netzwerkarchitekturen im biomedizinischen Bereich ist die U-Net-Architektur [60]. Sie ermöglicht die semantische Segmentierung verschiedener Zellaufnahmen. Ein Problem stellt die fehlende Möglichkeit zur Instanz-Segmentierung dar. Dies wird durch die Verwendung unterschiedlicher Ergebnisrepräsentationen gelöst, welche in Kombination mit einem Nachbearbeitungsschritt die Trennung der Instanzen erlauben. Diese reichen von einfachen



Methoden, bei denen neben dem Vordergrund auch die Ränder der Objekte vorhergesagt werden [61], bis zu komplexeren Distanzrepräsentationen [62, 63]. Auch die Architektur selbst wurde im Laufe der Zeit angepasst, um beispielsweise 3D-Daten verarbeiten zu können [64]. Die U-Net-Architektur findet sich in abgewandelter Form immer noch in aktuellen Methoden wieder. So nutzen Cellpose [65] und StarDist [66], beides bekannte Netzwerke zur Segmentierung von dicht gepackten Objekten, eine abgeänderte Version des U-Nets.

Neben dem U-Net stellt Mask R-CNN [67] eine weitere bekannte Architektur für die Instanz-Segmentierung dar. Hierbei werden zuerst sogenannte Region-Proposals, also Vorschläge für Objektpositionen, in Form von Bounding-Boxen generiert. Anhand dieser wird im Anschluss eine Segmentierung des Sub-Bereichs durchgeführt. Die Struktur erlaubt so die direkte Instanz-Segmentierung von Objekten ohne eine wie beim U-Net benötigte Nachbearbeitung. Vor allem bei stark überlappenden Objekten kann das Mask R-CNN bessere Segmentierungsergebnisse erzielen [68]. Eine auf biologische Fluoreszenzdaten optimierte Version des Mask R-CNN wurde in [69] vorgestellt.

Die Segmentierung von 3D-Daten stellt immer noch ein Spezialgebiet dar. Die meisten Deep-Learning-Methoden sind für die Segmentierung von 2D-Daten ausgelegt. Zwar existieren für einzelne Modelle wie Cellpose und StarDist angepasste Strukturen für die native Prozessierung von 3D-Daten [70, 71], allerdings kann hier nicht mehr auf vortrainierte Modelle zurückgegriffen werden, da keine diversen annotierten Datensätze für das Training zur Verfügung stehen, mit denen diese trainiert werden können. Daher wurden auch Methoden entwickelt, um mit 2D-Segmentierungsmodellen 3D-Daten zu prozessieren. Die Bilder werden dazu in 2D-Schichten aufgeteilt, welche anschließend einzeln verarbeitet und mithilfe eines Nachverarbeitungsschrittes zu einer 3D-Segmentierungsmaske zusammengefügt werden [65, 72]. Dieser Ansatz besitzt im Vergleich zu einer nativen 3D-Segmentierung einige Nachteile. Der wohl größte Nachteil stellt der Verlust der räumlichen Information dar. Gerade bei sich berührenden Objekten ist es in 2D-Schnitten häufig schwer zu erkennen, ob es sich um ein oder mehrere Objekte handelt. Weiterhin können bei der Verrechnung der einzelnen Ergebnisse Objekte

fälschlicherweise getrennt oder zusammengefügt werden. Zudem kann es aufgrund der ebenenweisen Prozessierung zu Inkonsistenzen zwischen den Ebenen kommen. Mit einer nativen 3D-Verarbeitung lassen sich, unter der Voraussetzung geeigneter Trainingsdatensätze, in der Regel bessere Ergebnisse erzielen.

Für die Erstellung solcher annotierten 3D-Datensätze eignet sich Software, die eine synchronisierte Darstellung von Schnittbildern in den drei Hauptebenen sowie eine 3D-Visualisierung der annotierten Objekte ermöglicht. Bekannte Beispiele hierfür sind *MITK* [73], *3D-Slicer* [74] oder *Napari* [75]. Aufgrund des Aufwands werden meist statt vollständiger Bilder nur kleinere 3D-Bildausschnitte annotiert. Dabei ist zu beachten, dass viele Deep-Learning-basierte Modelle eine Mindestanforderung an die Größe der Bildausschnitte haben. Die Ursache hierfür liegt in den verwendeten Pooling- und Faltungsoperationen, bei denen die Bildgröße mehrfach reduziert wird.

### 1.3.3 Reduktion des Annotationsaufwands

Bevor sich die Arbeit den synthetischen Daten als Alternative zur händischen Annotation widmet, sind Methoden zu erwähnen, mit denen der Annotationsaufwand reduziert werden kann. Dabei gibt es vier wesentliche Ansätze: semi-automatische Annotation, schwach überwachtes Lernen, semi-überwachtes Lernen und selbst-überwachtes Lernen.

Bei der semi-automatischen Annotation wird der Nutzer entlastet, indem automatisierte Algorithmen beispielsweise zur Vorsegmentierung eingesetzt werden. Die Ergebnisse werden anschließend nur noch überprüft und gegebenenfalls korrigiert [76, 77, 78, 79, 80, 81, 82, 83]. Ein Beispiel für Modelle zur Vorsegmentierung sind Cellpose [65] oder das Segment Anything Model (SAM) [84]. Die beiden Modelle wurden auf einem großen diversen Datensatz trainiert, um verschiedene Zellen bzw. Objekte zu segmentieren. Erfolgt dieser Prozess iterativ, wird von einem Human-in-the-Loop-Ansatz gesprochen. Ein klassisches Beispiel hierfür ist Ilastik [85], bei dem der Nutzer vereinzelt Vorder- und Hintergrundbereiche markiert, anschließend eine Vorhersage generiert wird, welche

wiederum an den fehlerhaften Stellen korrigiert werden kann, um die Prognose weiter zu verbessern. Weitere Beispiele hierfür sind [65, 86, 87, 88]. Verbreitet sind auch Algorithmen, welche aus vereinfachten Annotationen, z. B. Punkt- oder Bounding-Box-Annotationen, vollständige Annotationen der markierten Objekte generieren [89, 90, 91, 92, 93].

Schwach überwachtes Lernen (engl.: weakly-supervised learning) umfasst Trainingsmethoden, die ohne vollständige oder pixelgenaue Annotationen auskommen. Beispielsweise reicht es in manchen Ansätzen aus, nur jede zweite Bildebene oder die enthaltenen Objekte lediglich durch Punktmarkierungen zu annotieren. Dadurch wird der Annotationsaufwand für das Modelltraining erheblich reduziert. Beispiele für die Verwendung von spärlichen Annotationen sind [61, 64, 72, 94, 95], bei denen nur Ausschnitte einzelner Bildebenen annotiert werden. In [96, 97, 98, 99, 100] kann anhand von Punkt-, Linien- oder Bounding-Box-Annotationen eine vollständige Segmentierung gelernt werden. In [101] ist sogar die Nennung der in den Bildern vorhandenen Objektklassen ausreichend. Eine klare Abgrenzung zu den zuvor besprochenen semi-automatischen Annotationsverfahren ist nur schwer möglich, da die schwach überwachten Lernmethoden auch für die semi-automatische Annotation eingesetzt werden können.

Die Lernparadigmen des semi- und selbst-überwachten Lernens (engl.: semi-supervised und self-supervised) stellen ein großes Forschungsfeld dar. Bei diesen Methoden werden während des Trainingsprozesses eine Mischung aus annotierten und nicht annotierten Daten oder ausschließlich nicht annotierte Daten genutzt. Eine Abgrenzung von semi-überwachtem zu dem zuvor beschriebenen schwach-überwachtem Lernen stellt die aktive Nutzung von nicht annotierten Daten für die Verbesserung des Trainings dar. So werden im semi-überwachten Lernen beispielsweise Pseudoannotationen verwendet, die das Modell während des Trainings selbst generiert [56, 102, 103, 104]. Bei selbst-überwachten Methoden wird beispielsweise die Eigenschaft genutzt, dass ein leicht verändertes Bild dem Original ähnlicher ist als ein beliebiges anderes Bild, wodurch es in der internen Repräsentation des Netzwerks näher beieinander liegen sollte [105, 106]. Ein Überblick über das Gebiet wird in [107, 108, 109] gegeben.

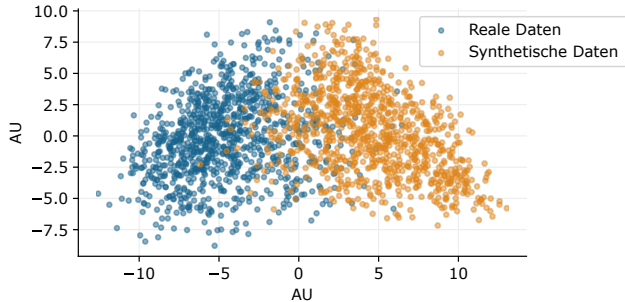
Allen Ansätzen ist gemein, dass sie den Annotationsaufwand verringern. In vielen Fällen erfolgt dies zu Lasten der erreichbaren Modellgüte. Ein alternativer Ansatz zur händischen Annotation von Trainingsdaten stellt die Synthese von Aufnahmen dar, bei der die Annotationen durch die Art der Erstellung bereits bekannt sind.

## 1.4 Synthese von experimentellen Daten

### 1.4.1 Übersicht

Eine große Herausforderung für das Training von Deep-Learning-basierten Modellen stellt die Annotation einer ausreichenden Menge an Trainingsdaten dar. Um eine gute Segmentierungsleistung und eine gute Generalisierungsfähigkeit der Modelle zu erhalten, wird eine große und diverse Menge an annotierten Daten benötigt. Die kontinuierlich zunehmende Leistungsfähigkeit der Rechenhardware sorgt für eine zunehmende Komplexität der eingesetzten Deep-Learning-Modelle. Durch die gesteigerte Anzahl an Parametern geht aber auch eine Steigerung in der benötigten Trainingsdatenmenge einher. Problematisch hierbei ist, dass die händische Erstellung von pixelgenauen Annotationen eine subjektive, zeitaufwendige, repetitive, ermüdende und fehleranfällige Aufgabe darstellt. Während für 2D-Daten der Aufwand einer händischen Annotation häufig noch vertretbar ist, stellt die Erstellung eines ausreichend großen Trainingsdatensatzes für 3D-Aufnahmen eine nicht mehr praktikable Aufgabe dar. Wie in Abschnitt 1.3 dargestellt, führt die zusätzliche Dimension zu einer drastischen Erhöhung der Komplexität und des Zeitaufwands der Annotationsaufgabe. Eine zunehmend beliebte Alternative zur händischen Annotation stellt die synthetische Generierung von Aufnahmen dar. Durch die Art der Erstellung sind hier die Annotationen direkt verfügbar.

Neben dem Wegfall der händischen Annotation bieten synthetische Trainingsdaten weitere Vorteile. Bei einer reinen Verwendung von synthetischen Daten entfallen sowohl Annotationsfehler als auch die Variabilität der Annotationen, welche sowohl zwischen den Annotationen mehrerer Personen als auch innerhalb der Annotationen einer einzelnen Person auftreten können [110]. Synthetische Daten



**Abbildung 1.2:** Visualisierung von Datenpunkten der realen und synthetischen Domäne. Die Datenpunkte wurden auf Grundlage echter realer und synthetischer Bilddaten ermittelt. Die Zentren der Domänen sind zueinander verschoben und die Verteilung der Domänen unterscheidet sich. Die Unterschiede zwischen den Domänen werden als Domänenlücke bezeichnet. AU stellt eine willkürliche Einheit dar (engl. arbitrary unit, AU).

werden auch dann genutzt, wenn die Aufnahme einer ausreichenden Menge an Daten zu aufwendig ist [111, 112, 113]. Mit synthetischen Daten lassen sich je nach Methode Datensätze mit einer beliebig großen Varianz erzeugen, womit die Generalisierungsfähigkeit von Modellen gesteigert werden kann [114, 115, 116]. Weiterhin kann einer Imbalance von Objekten mit ungleicher Auftretenswahrscheinlichkeit entgegengesteuert werden [117].

Ein weiterer Vorteil von synthetischen Daten ergibt sich bei Bereichen, in denen personenbezogene Daten verarbeitet werden. Gerade in medizinischen Anwendungen liegen häufig hochsensitive Daten vor, welche einen Aufschluss über den Gesundheitszustand des Patienten geben und so nicht ohne Weiteres zum Training von Modellen verwendet werden können.

Aufgrund der hohen Komplexität der realen Welt stellen die nachgebildeten synthetischen Versionen immer nur eine Annäherung der realen Daten dar. Es verbleibt eine sogenannte Domänenlücke (engl. Domain-Gap), also eine Lücke zwischen der echten und synthetischen Domäne. Die Domänenlücke zeigt sich häufig nicht nur in Form leichter Abweichungen von den realen Daten, sondern auch durch eine unterschiedliche Verteilung der synthetischen Daten (vgl. Abbildung 1.2). Die Domänenlücke führt oft zu einem Leistungsabfall des Modells bei der Anwendung auf realen Daten.

Neben der Steigerung des Realismus synthetischer Daten durch neue und verbesserte Methoden existiert ein alternativer Ansatz, um die Übertragbarkeit auf reale Daten zu optimieren: der Domain-Randomization-Ansatz. Hierbei wird versucht, die Domänenlücke zu überbrücken, indem die Variabilität der synthetischen Daten gezielt stark erhöht wird. Diese bewusst überhöhte Variabilität steigert die Generalisierungsfähigkeit des Modells und verbessert so dessen Leistung bei der Anwendung auf realen Daten [116, 118, 119, 120]. So werden beispielsweise die Rauschanteile und Objektgrößen stark variiert oder sogar der Hintergrund unrealistisch eingefärbt. Nachteilig bei diesem Ansatz ist, dass das spätere Netzwerk mit vielen für die Aufgabe nicht relevanten und teilweise unrealistischen Daten trainiert wird. Das kann die Generalisierungsfähigkeit des Modells zwar steigern, allerdings können die nicht relevanten Daten die Spezifität senken und unrealistische Daten das Training negativ beeinflussen [120, 121]. Weiterhin ist nicht sichergestellt, dass die gesteigerte Generalisierungsfähigkeit des trainierten Modells tatsächlich die Übertragbarkeit auf reale Daten verbessert [118, 120, 122].

Eine weitere wichtige Problematik bei der Datensynthese stellen seltene Elemente oder Ereignisse in den realen Daten dar [123]. Ohne spezielle Beachtung und geeignete Synthesemethoden werden diese nicht in den synthetischen Daten abgebildet. Solche Elemente sind in der Auswertung jedoch meist von besonderer Bedeutung. Aufgrund ihres Fehlens in den Trainingsdaten werden solche Ereignisse und Elemente häufig nicht erkannt oder fehlerhaft segmentiert.

Die Verfahren zur Generierung von synthetischen Daten lassen sich grob in simulationsbasierte und datengetriebene Ansätze aufteilen [124], wobei auch hybride Ansätze existieren, die beide Ansätze in unterschiedlichen Schritten vereinen [125]. Bei simulationsbasierten Ansätzen werden, wie auch bei digitalen Zwillingen, reale Eigenschaften mittels parametrischer Modelle nachgebildet. Bei datengetriebenen Ansätzen lernen Algorithmen anhand von realen Aufnahmen, diese authentisch nachzubilden.

## 1.4.2 Simulationsbasierte Methoden

Bei simulationsbasierten Methoden werden Modelle verwendet, um die Realität und die darin auftretenden Effekte möglichst genau nachzubilden. Für die Erstellung solcher Modelle wird in der Regel ein hohes Maß an Fachwissen benötigt. Zum einen ist ein Verständnis der abzubildenden physikalischen Zusammenhänge erforderlich und zum anderen auch die Fähigkeit, diese rechnerisch zu modellieren.

Videospiele sind ein passendes Beispiel für simulationsbasierte Methoden. Hier werden Modelle für Objekte erstellt, welche in Kombinationen mit anderen Objekten eine virtuelle Welt ergeben, welche anschließend gerendert wird, was mit einer Aufnahme einer virtuellen Kamera verglichen werden kann. Ein Vorteil bei der Nutzung von fertigen Spielen liegt darin, dass kein Aufwand für die Modellierung zu betreiben ist. Nachteilig ist, dass die Modelle nicht auf das genaue Szenario ausgerichtet sind. Aufnahmen aus dem Spiel *GTA V* werden aufgrund der realitätsnahen Straßenszenen gerne als Trainingsdaten für den CityScapes-Datensatz verwendet [126]. Um die Daten näher an die realen Vorbilder zu bringen, wird meist eine Deep-Learning-basierte Optimierung durchgeführt (siehe Abschnitt 1.4.4).

Die Simulation von Mikroskopaufnahmen unterscheidet sich grundlegend von der Simulation herkömmlicher Kameraaufnahmen, da der Aufnahmeprozess deutlich komplexer ist und daher einen größeren Stellenwert im Simulationsprozess einnimmt.

Ein verbreitetes simulationsbasiertes Programm zur Erzeugung von synthetischen Zelldaten stellt CytoPacq [127, 128] dar. In drei Modulen wird eine digitale Probe erstellt, die optische Verfremdung durch ein Mikroskop simuliert und die Eigenschaften des digitalen Bildsensors nachgebildet. Das erste Modul dient der Erstellung einer *digitalen Probe*, welche in den weiteren Modulen verfremdet wird. Mit *CytoGen*, *MitoGen* und *FiloGen* [127, 129, 130] stehen drei Möglichkeiten zur Verfügung, um verschiedene Arten von Zellen zu generieren. Aufgrund

der aufwendigen Simulation ist die Generierung von großen Aufnahmen zeitaufwendig und die Parametrierung der Module komplex. In der vereinfachten Web-Version lassen sich hingegen nur eine begrenzte Anzahl an Zellen simulieren. Da das Programm nicht für die Erstellung von großen 3D-Zellstrukturen konzipiert wurde, werden die hier relevanten Eigenschaften wie die Verschlechterung der Bildqualität in tieferen Schichten nicht abgebildet.

Stegmaier et al. [131] zeigen eine einfachere Simulation von synthetischen 3D-Zellkulturaufnahmen, bei denen reale Zellkerne zufällig in Bildern eingefügt werden und anschließend eine Simulation des Aufnahmeprozesses erfolgt. Bei dieser Methode wird auch ein einfacher linearer Helligkeitsabfall in der Tiefe modelliert.

Weigert et al. [132] stellen 2018 mit *biobeam* eine Software vor, welche sich auf die Simulation der optischen Verfremdung bei der Aufnahme von Proben mittels Light-Sheet-Mikroskopie spezialisiert. Hier wird eine synthetische Punktspreizfunktion mit einer regionalen Abhängigkeit berechnet. Die Parametrierung der Software ist allerdings nicht trivial und erfordert Expertenwissen. Zudem wird eine digitale Probe mit bekannten Stoffeigenschaften als Grundlage der Simulation benötigt.

Daneben gibt es weitere Tools für die Generierung von Bilddaten eines Mikroskops [133, 134, 135, 136, 137]. In [138] findet sich eine ausführliche Beschreibung der Simulation von biologischen Zelldaten.

Nachteilig bei diesen Ansätzen ist das benötigte Fachwissen für die Erstellung, Anwendung und die Parametrierung der Modelle. Aufgrund der teilweisen detailreichen Modellierung der Effekte wird speziell bei 3D-Daten eine gute Rechenhardware vorausgesetzt. Weiterhin benötigt die Berechnung der Effekte Rechenzeit. Da selbst bei komplexeren Ansätzen immer noch Vereinfachungen wie eine rauminvariante Punktspreizfunktion genutzt werden, entsteht eine Domänenlücke zwischen den realen und den simulierten Daten. Auch sind simulationsbasierte Ansätze stark auf einen Anwendungsfall spezialisiert und können über die Parametrierungen nur begrenzt an verschiedene Gegebenheiten angepasst werden, wodurch eine Domänenlücke zwischen realen und synthetischen Daten verbleibt.



Eine tiefgreifende Anpassung durch eine Abänderung des Programmcodes erfordert aufgrund der komplexen Struktur gute Programmierkenntnisse und Einarbeitungszeit.

### 1.4.3 Datengetriebene Methoden

Datengetriebene Methoden verzichten auf eine händische Nachbildung physikalischer Effekte und extrahieren diese stattdessen direkt aus den realen Daten. Beliebte sind generative Netzwerke, welche es ermöglichen, anhand von realen Vorbildern synthetische Bilddaten zu generieren. Um mit den synthetischen Daten später Deep-Learning-Modelle zu trainieren, ist es erforderlich, dass die generativen Modelle neben der Bildsynthese auch die Möglichkeit bieten, korrespondierende Annotationen zu erzeugen. Eine etablierte Klasse generativer Netzwerke, welche diese Anforderung erfüllt, sind Generative Adversarial Networks (GANs), wie das CycleGAN. Eine ausführliche Einführung in GANs findet sich im Anhang A.1.

Fu et al. [139] nutzen 2018 das CycleGAN-Modell, um damit synthetische Trainingsdaten für ein 3D-Segmentierungsnetzwerk zu generieren. Dazu werden zuerst synthetische Annotationen generiert, indem zufällig Ellipsoide in einem Bild platziert werden. Anschließend wird das 2D-CycleGAN trainiert, um zwischen realen Zellaufnahmen und den synthetischen Annotationen zu transformieren. Die dabei genutzten Bild- und Annotationsdaten stellen kleine Ausschnitte dar. Es wird beschrieben, dass die räumliche Übereinstimmung zwischen Annotationen und synthetischen Bildern nicht immer stimmig ist, weshalb ein erweitertes CycleGAN, das spatially constrained CycleGAN (SpCycleGAN) eingeführt wird. In [140] (2019) wird das Vorgehen auf unterschiedliche 3D-Zelldaten angewandt und die Segmentierungsqualität mit verschiedenen herkömmlichen Methoden verglichen. Dabei zeigt sich, dass das mit synthetischen Daten trainierte Netzwerk besonders bei heterogenen Bedingungen Vorteile bietet. Ein Nachteil der Methode liegt bei der 2D-basierten Transformation. Das kleine 3D-Annotationsbild wird dabei ebenenweise transformiert und anschließend wieder zusammengesetzt, wodurch es zu Inkonsistenzen zwischen den Ebenen kommen kann. Um solche

Inkonsistenzen zu vermeiden, wird 2020 in [141] der Einsatz eines Recycle-GANs [142] getestet, welches ursprünglich für die Synthese von Videodaten entwickelt wurde. Ein Nachteil dieser Methode ist die deutlich erhöhte Komplexität. So kommen zwei weitere Netzwerke und weitere Parameter dazu. Zudem zeigen die vorgestellten Ergebnisse nur eine geringe Übereinstimmung zwischen den synthetischen Bildern und den Annotationen, insbesondere in Bezug auf die Form und Ausrichtung der Objekte.

Eschweiler et al. [143] nutzen 2019 ein 2D-CycleGAN, um damit Zellmembranen von Pflanzen zu synthetisieren. Dabei wird eine Methode vorgestellt, um membranartige Annotationen auf Basis einer vorgegebenen Zellkulturform zu generieren. Um die Übereinstimmung der Membran-Positionen zwischen generierten Annotationen und daraus erstellten Bildern zu gewährleisten, wird beim Training ein gewichteter Cyclic-Loss verwendet, bei dem die Gewichtung mit dem Abstand zur Membran zunimmt. Die verwendeten Pflanzendaten besitzen sehr gleichmäßig angeordnete Zellen und gerade Membranen, was die Generierung von synthetischen Annotationen deutlich vereinfacht.

2021 stellen Eschweiler et al. [144] eine auf dem cGAN-basierte Methode vor, um große 3D-Zellkulturen zu synthetisieren. Durch den begrenzten Grafikspeicher wird das 3D-Bild in einzelne Ausschnitte unterteilt, welche dann einzeln transformiert und wieder zusammengesetzt werden. Um auch globale Positionsinformationen zu erhalten, wird ein zusätzlicher Kanal genutzt, bei dem die Position der Pixel in Relation zum Rand der Kultur encodiert wird. Die Übereinstimmungen zwischen den Annotationen und den daraus generierten Strukturen werden bei dieser Methode nur durch den Identity-Loss (vgl. Gleichung A.14) von realen Bildern erreicht. Auch hier scheint es bei Betrachtung der generierten Daten zu kleineren Abweichungen in der Form der Strukturen zu geben, was sich bei dem späteren Segmentierungstraining nachteilig auswirkt. Weiterhin wird das generative Modell, welches zur Auswertung der Segmentierungsleistung genutzt wird, anhand von manuell korrigierten Annotationsdaten der realen Daten trainiert. Das verbessert die Übereinstimmung der generierten Daten zu den Annotationen. Allerdings können die manuell korrigierten Annotationen auch direkt für das Training von Segmentierungsnetzwerken genutzt werden.

Bahr et al. [145] stellt 2021 eine Methode vor, um einzelne Zellkerne in vorgegebenen Zellzyklusstadien zu generieren. Dabei wird die Form und das Stadium der Zelle parametrisch modelliert und die Textur anschließend mit einem cGAN generiert.

Daneben gibt es eine Vielzahl von weiteren Methoden, welche GANs nutzen, um synthetische Zelldaten zu generieren [146, 147, 148, 149].

Die Ähnlichkeit der Annotationen zu den realen Bildern ist wichtig, da es sonst bei der GAN-basierten Transformation zu Unstimmigkeiten zwischen den Annotationen und den generierten Bildern kommen kann [144, 150]. Beispielsweise verändert das GAN die Formen der Objekte, wenn diese nicht mit den realen Daten übereinstimmen, oder fügt zusätzliche Zellen hinzu, wenn die Zelldichte in den realen Daten deutlich höher ist als in den zu transformierenden Annotationsdaten. Solche Diskrepanzen zwischen den generierten Bildern und zugehörigen Annotationsdaten wirken sich auch auf das Segmentierungsnetzwerk aus, welches auf diesen Daten trainiert wird. Wenn das GAN neue Zellen hinzufügt, die nicht in den Annotationsdaten vorhanden sind, neigt das Segmentierungsnetzwerk dazu, falsch-negative Segmentierungsfehler zu erzeugen. Ein bekanntes Problem bei der Annäherung von Zellformen ist, dass diese oft durch einfache Ellipsoide dargestellt werden. Ellipsoide sind zwar einfach zu generieren, stellen jedoch für manche Zelltypen mit komplexen Zellformen eine erhebliche Vereinfachung dar. Um realistischere Formen zu erhalten, nutzen Wiesner et al. [151] 3D-GANs, um anhand von zufälligen Vektoren Annotationen von einzelnen Zellkernen zu generieren. Weiterhin stellt Chen et al. [152] eine Methode vor, bei der Annotationsmasken einzelner Zellkerne mit Bézierkurven generiert werden. Auch die Generierung von Zell- und Kernformen mittels Fourier-Form-Deskriptoren ist erprobt [135].

Während diffusionsbasierte Modelle bereits für die Generierung von Bilddaten verbreitet sind, gibt es noch wenig Ansätze, welche die Generierung von Bilddaten mit passenden Annotationen erlauben. Eine kurze Einführung von Denoising-Diffusion-Probabilistic-Models (DDPM) findet sich im Anhang A.1.2. 2023 stellen Wu et al. [153] eine Methode vor, welche auch die zusätzliche Generierung

von Annotationen ermöglicht. Eschweiler et al. [154] stellen 2024 eine weitere diffusionsbasierte Methode vor, welche auf Basis synthetischer Annotationsdaten realistische Bilddaten generiert. Diffusionsbasierte Modelle besitzen allerdings den Nachteil, dass eine große Menge an Trainingsdaten benötigt wird [155]. Weiterhin erfordern diffusionsbasierte Modelle im Vergleich zu GAN-Modellen aufgrund des Diffusionsprozesses einen erhöhten Rechenaufwand, was die Trainings- und Inferenz-Zeit erhöht [156].

Bei den vorgestellten datengetriebenen Ansätzen bestehen immer noch Limitierungen. So besitzen 2D-basierte Modelle das Problem, dass es bei der Generierung von 3D-Bildern zu Inkonsistenzen zwischen den Ebenen kommen kann. Weiterhin zeigen die meisten Modelle Schwächen sowohl in der Qualität der generierten Daten als auch in der Übereinstimmung zwischen Bild und Annotationen. Das beeinträchtigt die Segmentierungsqualität von Segmentierungsnetzwerken, welche auf solchen synthetischen Daten trainiert werden. Zudem bieten die derzeitigen Modelle keine Möglichkeit, seltene Elemente und Objektklassen gezielt in die synthetischen Daten einzubringen. Das kann dazu führen, dass diese nicht in den synthetischen Daten vorhanden sind und deren Segmentierung nicht explizit gelernt werden kann. Zudem gibt es keine Möglichkeit, neben den Instanzannotationen auch Klassenannotationen zu erhalten.

### 1.4.4 Hybride Methoden

Simulierte Daten weisen meistens eine Domänenlücke zu den realen Daten auf, da nicht alle Feinheiten der realen Welt abgebildet werden [157]. Um die Daten realistischer zu machen, werden Deep-Learning-basierte Nachbearbeitungsmethoden entwickelt <sup>1</sup>. Eine der ersten hybriden Ansätze, welcher simulierte Daten anhand von GAN-basierten Modellen verbessert, ist SimGAN [163] (2017). Die

---

<sup>1</sup> Gerade für Straßenszenen ist die Transformation von Computerspieldaten in die reale Welt beliebt. So werden in [158, 159, 160, 161] Bildaufnahmen aus einem Computerspiel an die Daten des CityScapes-Datensatzes [162] angeglichen.

dabei verwendeten Daten sind simulierte Aufnahmen von Augen, welche verwendet werden, um ein Modell zur Bestimmung der Blickrichtung zu trainieren. Um die Übereinstimmung der Blickrichtung zwischen den simulierten und den optimierten Bildern sicherzustellen, wird ein zusätzlicher Regularisierungsterm genutzt, der die pixelweise Differenz zwischen den beiden Bildern minimiert. Die Regularisierung steht damit in Konkurrenz zu der eigentlichen Aufgabe des Modells, die Bilder anzupassen, was sich in einem instabilen Training äußert und zu einer geringen Varianz der Bilder führt [163]. Mit dem GazeGAN [164] (2017) wird für die Optimierung eine auf dem CycleGAN-basierende Struktur eingesetzt, welche durch die Cycle-Consistency eine Sicherstellung der Blickrichtung erzielt.

### **1.4.5 Synthese von Bildmodalitäten**

Neben der Datensynthese zur Generierung von Trainingsdaten können die lernbasierte Methoden auch für die Übersetzung zwischen verschiedenen Aufnahmemodalitäten oder Kanälen verwendet werden. Im medizinischen Bereich ist die Übersetzung zwischen CT- und MRT-Aufnahmen interessant. Für die Übersetzung der Modalitäten wird unter anderem die CycleGAN Struktur eingesetzt [165, 166, 167, 168, 169]. Daneben werden die Techniken auch im biologischen Bereich für virtuelle Färbungen, bei denen Signale von ungefärbten Proben rekonstruiert werden, und die Übersetzung zwischen verschiedenen Färbungen eingesetzt. [170] gibt hierbei einen Überblick der Methoden und Anwendungen für histopathologische Aufnahmen. Im Bereich der Mikroskopie ist auch die Übersetzung von unmarkierten Hellfeldmikroskopaufnahmen in Fluoreszenzaufnahmen interessant. Durch die Übersetzung kann auf die zeitintensive Färbung der Proben mit fluoreszierenden Markern verzichtet werden, welche teilweise toxische Wirkung auf die lebenden Zellen haben. Allerdings bestehen weiterhin die Limitierungen des Hellfeldmikroskops hinsichtlich der Auflösung, im Speziellen der Tiefenauflösung. In [171, 172, 173] (2018-2020) werden dazu Methoden vorgestellt, um anhand von gepaarten Daten ein Transformationsmodell zu trainieren, welches Hellfeldmikroskopaufnahmen zu Fluoreszenzaufnahmen eines

Konfokalmikroskops übersetzt. Chen et al. [174] präsentiert 2023 einen speziellen Mikroskopaufbau, mit dem solche gepaarten Aufnahmen von quantitativer Phasenkontrastmikroskopie und Fluoreszenzmikroskopie auf einfache Art und Weise erstellt werden können.

## 1.5 Offene Probleme

Dunn et al. [140] zeigt zwar vielversprechend, dass mit synthetischen Daten Deep-Learning-Netzwerke für die Segmentierung von realen Aufnahmen trainiert werden können. Daneben wurden viele weitere Methoden vorgestellt, welche die Qualität der Synthese verbessern oder die Synthese von größeren Strukturen ermöglichen. Trotz dieser Fortschritte scheitert der derzeitige Stand der Technik an folgenden Problemen:

### **Bild-Synthese 3D-Mikroskopie**

- Fehlende Einbringung biophysikalischer Modellierung. Die Anordnung der Zellen spielt in den Zellkulturen eine wichtige Rolle. Bisherige Methoden verwenden eine Platzierung, welche nur teilweise an der Realität orientiert ist. Während Simulationsmethoden für die Modellierung der Entwicklung von Zellkulturen existieren, gibt es bislang keine Verbindung zwischen den beiden Bereichen, um eine biophysikalisch realistische Anordnung der Zellen und Kerne zu erhalten.
- Fehlende Einbringung von statistisch unterrepräsentierten Elementen. Verfügbare Modelle vernachlässigen selten auftretende Elemente, was dazu führt, dass diese nicht in den synthetischen Trainingsdaten vorhanden sind. Um die Segmentierung von seltenen Elementen zu verbessern, ist es notwendig, diese in die Trainingsdaten mitaufzunehmen. Das optimale Verhältnis ist dabei nicht bekannt.
- Fehlende Möglichkeit zur gezielten Einbringung mehrerer Objektklassen. Durch die fehlende Möglichkeit, gezielt Objekte aus bestimmten Klassen in die Daten einzubringen, enthalten die generierten Daten keine expliziten Informationen zu den Objektklassen. Somit lassen sich beispielsweise keine synthetischen Trainingsdaten für eine Klassifikation oder der simultanen Segmentierung und Klassifikation (auch panoptische Segmentierung genannt) generieren.

- Fehlende Möglichkeit zur Generierung kombinierter Signalmodalitäten. Es fehlen Methoden, um synthetische Aufnahmen mit zueinander passenden Nukleus- und Membransignalen zu generieren. Ebenso wurde bisher nicht untersucht, inwieweit die herausfordernde Membransegmentierung durch die zusätzliche Einbeziehung von Nukleussignalen verbessert werden kann.

## **Synthese von Bildmodalitäten**

- Fehlende Möglichkeit zur Synthese von zusätzlichen Bildmodalitäten durch die Transformation von Fluoreszenzmarkern. Die Anzahl von fluoreszierenden Markern, welche gleichzeitig aufgenommen werden können, ist aufgrund des begrenzten Fluoreszenz-Spektrums limitiert. Es fehlt an Methoden, um die Signale von Markern anhand von anderen Markern zu präzisieren.
- Mangel an Verfahren zur Überprüfung und Bewertung von Methoden zur Synthese von Bildmodalitäten. Es fehlen Bewertungsmaße, um die Qualität der Synthese von Bildmodalitäten bezüglich relevanter Eigenschaften zu bewerten. Weiterhin fehlen Ansätze, mit denen die Methoden hinsichtlich ihrer Übertragbarkeit auf Anwendungsdaten überprüft werden können. Die Schwierigkeit ist hierbei, dass keine perfekte Fixierung der Probe zwischen den Färbungen möglich ist, wodurch die Zell-Positionen zwischen den Aufnahmen leicht unterschiedlich sind.
- Fehlende Untersuchung für relevante Merkmale für eine erfolgreiche Synthese. Bislang ist nicht bekannt, welche Marker sich für die Transformation eignen. Hier ist zu untersuchen, welche Eigenschaften relevant sind.
- Fehlende Analyse des Potenzials synthetischer Bildmodalitäten in der Datenauswertung. Es ist bisher unklar, ob sich synthetische Bildmodalitäten dazu eignen, die Leistung von Auswertemethoden zu steigern.



## 1.6 Zielsetzung und Struktur der Arbeit

In dieser Arbeit werden die offenen Probleme in den Bereichen der Trainingsdatensynthese und der Synthese von Bildmodalitäten adressiert und innovative Lösungen gefunden. Die spezifischen Zielstellungen für die Trainingsdatensynthese sind:

1. Konzeption einer Strategie zur erstmaligen Integration von seltenen Elementen und multipler Objektklassen in synthetische Trainingsdaten. Eine wichtige Anforderung ist, dass neben Instanz- auch erstmals Klassenannotationen verfügbar sind.
2. Erarbeitung eines neuen Konzepts für die erstmalige Integration von biophysikalischen Simulationen in den Syntheseprozess zur Erzielung eines neuen Levels an Realismus in der Zellanordnung von synthetischen Daten.
3. Umfängliche Analyse der neu entwickelten Methoden zur Datensynthese bezüglich ihres Aufwandes und ihrer Eignung zur Generierung von Trainingsdaten für die Segmentierung von realen Daten.
4. Umfangreiche Untersuchung des Einflusses von seltenen Elementen in synthetischen Daten auf die Segmentierungsleistung von regulären und seltenen Elementen.
5. Entwicklung eines neuen Verfahrens zur Generierung synthetischer Aufnahmen mit gepaarten Membran- und Nukleussignalen, um die bislang herausfordernde Segmentierung von Membransignalen zu verbessern.

Für die Synthese von Bildmodalitäten ergeben sich folgende Zielstellungen:

1. Konzeption eines Frameworks zur erstmaligen Transformation zwischen fluoreszierenden Markern, um die einschränkende Limitierung an maximal gleichzeitig möglichen Fluoreszenzfärbungen zu umgehen.
2. Entwicklung einer Bewertungsmethodik zur Bestimmung der Transformationsqualität auf realen Anwendungsdaten.

3. Umfassende Untersuchung zur Identifikation relevanter Merkmale für die erfolgreiche Transformation zwischen Bildmodalitäten.
4. Erprobung der entwickelten Transformationsmethoden auf einem realen Anwendungsbeispiel zur Verdeutlichung der praktischen Relevanz.

Die Schriftfassung gliedert sich wie folgt: Kapitel 2 beginnt mit einem Überblick über die entwickelte Methodik und deren Einordnung in den Prozess der Datenauswertung. Darauf folgend bietet das Kapitel eine Beschreibung der neuen Methodik zur Generierung von Trainingsdaten und zur Synthese von Bildmodalitäten sowie der entwickelten Bewertungsmethoden. Kapitel 3 gibt eine Übersicht über die Implementierung der generierten Software. Kapitel 4 stellt die Datensätze vor, die später zur Auswertung der eingeführten Methodik verwendet werden. Kapitel 5 präsentiert die Ergebnisse und bietet eine Diskussion und deren Einordnung. Kapitel 6 gibt abschließend eine Zusammenfassung der Arbeit sowie einen Ausblick.

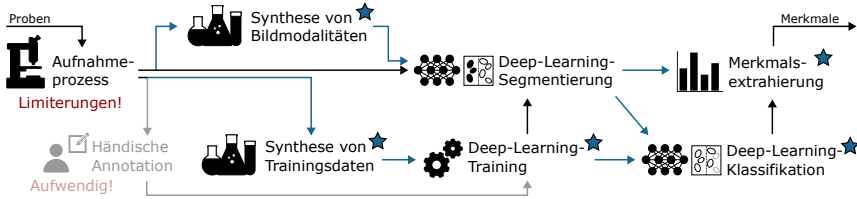
## 2 Neue Methodik zur Datensynthese

### 2.1 Übersicht

Generative Deep-Learning-Modelle eröffnen neue Möglichkeiten in der Datenverarbeitung. Die Anwendungsgebiete sind dabei vielfältig. Sie ermöglichen unter anderem die Generierung von Daten, die Verbesserung bestehender Bildaufnahmen oder auch die Überführung zwischen Domänen. Diese Arbeit konzentriert sich auf zwei Gebiete: die Generierung von Informationen aus vorhandenen Daten und die Generierung synthetischer Daten für das Training von Deep-Learning-Modellen. Da sich die eingesetzten Modelle in den beiden Bereichen ähneln und in Teilen auch Methoden zur Synthese von Bildmodalitäten für die Trainingsdatensynthese eingesetzt werden, ist eine konsistente Abgrenzung der Gebiete nicht immer möglich.

Abbildung 2.1 zeigt eine Einordnung der in dieser Arbeit erstmalig untersuchten und neu entwickelten Methoden in den Auswerteprozess von Bilddaten. Die Grundlage für die Datenauswertung stellt die Aufnahme der Proben dar. Bei dem *Aufnahmeprozess* erfolgt eine Abbildung der realen analogen Welt in die diskrete digitale Welt. Dabei unterliegt die Aufnahme physikalischen und technischen Grenzen, woraus sich verschiedene Qualitätsverluste und auch Limitierungen ergeben.

Eine entscheidende Einschränkung optischer Instrumente stellt die maximal mögliche Anzahl an Kanälen dar. So lassen sich nur eine bestimmte Anzahl an Kanälen kombinieren, ohne dass es zu einer Überlagerung und damit zu unspezifischen



**Abbildung 2.1:** Übersicht der Datenauswertung mittels Deep-Learning-Algorithmen. Blaue Sterne markieren eigene Zielstellungen der vorliegenden Arbeit gegenüber dem Stand der Technik. Schwarze Pfeile stellen Pfade dar, welche mit den Methoden des Stand der Technik gangbar sind. Blaue Pfeile markieren Pfade, welche durch die entwickelten Methoden eröffnet werden. Ausgegraute Elemente stellen den Stand der Technik ohne Trainingsdatensynthese dar.

Signalen kommt. Unter dem Stichwort *Synthese von Bildmodalitäten* führt diese Arbeit daher ein Framework ein, welches es anhand von aufgenommenen Markern erstmalig ermöglicht, Signale anderer Marker zu synthetisieren. Beispielsweise lassen sich so anhand von Zellkernsignalen weitere Marker für Zellteilung oder Zelltod vorhersagen. Das Framework basiert auf generativen Deep-Learning-Modellen und geht auf den Prozess der Erstellung von optimalen Trainings- und Anwendungsdaten ein. Zudem wird ein neues Protokoll entwickelt, um eine quantitative Bewertung der Synthesequalität unter anwendungsnahen Bedingungen zu ermöglichen. Weiterhin wird in dieser Arbeit untersucht, welche Merkmale für die erfolgreiche Generierung weiterer Marker relevant sind. Hierzu werden Anforderungshypothesen aufgestellt, welche mit ausgesuchten Markerkombinationen überprüft werden.

Nach der Erweiterung der aufgenommenen Daten (*Synthese von Bildmodalitäten*) um synthetisierte Daten erfolgt die Datenauswertung. Ein häufig elementarerer Schritt ist die Segmentierung der Aufnahmen, bei der sich aufgrund der Komplexität und Heterogenität Deep-Learning-Modelle etabliert haben (*Deep-Learning-Segmentierung*). Für eine optimale Leistung benötigen die Modelle allerdings eine ausreichende Menge an Trainingsdaten. Wie in Abschnitt 1.3.1 beschrieben, ist der manuelle Aufwand zur Erstellung von Annotationen für 3D-Daten enorm. Daher wird eine händische Annotation in den meisten Fällen als nicht praktikabel erachtet.

Um die aufwendige händische Annotation zu umgehen, werden Methoden zur Synthese von Trainingsdaten eingesetzt (*Synthese von Trainingsdaten*). Hierbei werden Bilddaten inklusive deren Annotationen synthetisch generiert, welche anschließend für das Training von Segmentierungsmodellen verwendet werden können. Aktuelle Verfahren nutzen generative Modelle, die anhand realer Daten lernen, diese nachzubilden. Ein bedeutendes Problem dieser Ansätze besteht jedoch darin, dass selten auftretende Elemente von den Modellen häufig vernachlässigt werden. Gerade diese seltenen Elemente sind in der Analyse oft von besonderem Interesse. Diese Arbeit zielt daher darauf ab, neue Ansätze für die Synthese großer 3D-Strukturen zu entwickeln, die es erstmals ermöglichen, seltene Elemente gezielt in synthetische Daten einzubringen. Hierzu wird eine neue hybride Methodik eingeführt, die simulationsbasierte mit datengetriebenen Ansätzen kombiniert. Da weiterhin das optimale Verhältnis von seltenen zu regulären Elementen im *Deep-Learning-Training* unbekannt ist, ist die Segmentierungsleistung in Abhängigkeit von verschiedenen Verhältnissen in den Trainingsdaten zu untersuchen. Hierzu ist die Generierung synthetischer Trainingsdaten erforderlich, welche gezielt vorgegebene Anteile seltener Elementen enthalten. Damit wird unter anderem geklärt, ob eine überproportionale Einbringung zu einer Verbesserung der Segmentierungsleistung führt.

Während es bereits Ansätze gibt, um physikalische Gesetze in Form von Differenzialgleichungen in den Trainingsprozess zu integrieren [175], stellt die Einbringung von physikalischen Modellierungen in den Deep-Learning-basierten Syntheseprozess ein recht unerforschtes Feld dar. Häufig werden Strukturen nur zufällig angeordnet oder anhand einfacher Heuristiken platziert. Für die Einbringung von physikalischen Simulationen in synthetische Aufnahmen wird in dieser Arbeit ein parametrierbarer Methodenpool entwickelt. Die entwickelten Methoden ermöglichen es, bereits etablierte Ansätze für die Simulation von Zellstrukturen einzusetzen. Im Rahmen der Methodenentwicklung zur Integration biophysikalischer Modellierung wird auch ein neuartiger Trainingsansatz für GANs entworfen, der es ermöglicht, neben Bilddaten auch korrespondierende Annotationen zu erzeugen. Weiterhin wird mit den Methoden erstmals untersucht, welche Auswirkungen eine physikalisch motivierte Modellierung der Strukturen auf die Qualität

der Synthese und auf die anschließende Leistung eines Segmentierungsmodells hat. Zudem wird eine Vorgehensweise entwickelt, die beschreibt, wie bei neuen Daten und Anforderungen vorzugehen ist.

Mit den entwickelten Synthesemethoden können Daten in theoretisch unbegrenzter Menge generiert und direkt für das Training von Segmentierungsnetzwerken verwendet werden (*Deep-Learning-Training*). Zu beachten ist, dass die synthetischen Daten die realen Daten so gut wie möglich nachbilden. Anhand von zwei Bewertungsmethoden werden die neu entwickelten Methoden untereinander und mit dem Stand der Technik verglichen. Eine Bewertungsmethode basiert auf der Segmentierungsqualität, die mit synthetischen Trainingsdaten auf realen Daten erreicht werden kann. Hier werden die Methoden auch mit vortrainierten Segmentierungsmodellen und mit händisch generierten Trainingsdaten verglichen. Weiterhin wird eine Vorgehensweise entwickelt, die aufzeigt, welche Methoden sich unter den gegebenen Rahmenbedingungen am besten eignen.

Die mit synthetischen Aufnahmen trainierten Segmentierungsmodelle erlauben nun die Instanz-Segmentierung der Daten, welche die Grundlage für die Extraktion von Merkmalen einzelner Objekte bildet. Für eine umfängliche Auswertung der Daten wird daher eine neue Pipeline zur *Merkmalsextrahierung* entwickelt. Die Pipeline vereint dabei etablierte mit neu entwickelten Methoden, um eine große Menge an Informationen aus Datensätzen zu extrahieren. Dabei werden unter anderem Informationen zur Intensität und Morphologie auf globaler (gesamte Struktur bzw. Unterbereiche) als auch lokaler Ebene (einzelne Objekte) extrahiert und mit anderen Aufnahmen verglichen. Ziel der Pipeline ist es, eine benutzerfreundliche, grafisch geführte und umfassende Auswertung von Daten zu ermöglichen, die in komplexen Strukturen angeordnet sind.

Weiterhin kann aufbauend auf den Segmentierungsergebnissen eine *Deep-Learning-Klassifikation* der in der Aufnahme vorhandenen Objekte durchgeführt werden. Es wird eine grafische Benutzeroberfläche entworfen, die eine effiziente händische Klassenannotation von Objekten in 3D-Bilddaten ermöglicht. Zur Verbesserung der Klassifikation werden in dieser Arbeit erstmals die entwickelten Methoden zur Synthese von Bildmodalitäten eingesetzt.

## 2.2 Synthese von Trainingsdaten

Teile der folgenden Sektion basieren auf den im Rahmen der vorliegenden Arbeit veröffentlichten Publikationen von Bruch et al. [176] und Bruch et al. [177].

### 2.2.1 Übersicht

Das Ziel der Synthese von Trainingsdaten stellt die Generierung neuer Aufnahmen inklusive deren Annotationen dar, um damit Segmentierungsmodelle zu trainieren. Wie in Kapitel 1 vorgestellt, sind mehrere Deep-Learning-Methoden verfügbar, welche die Generierung von synthetischen Daten mit dazugehörigen Annotationen ermöglichen. Allen gemein ist, dass dabei Netzwerke verwendet werden, welche Bilder aus einer Domäne  $\mathcal{X}$  in eine andere Domäne  $\mathcal{Y}$  überführen. Für die Transformation können verschiedene Netzwerkarchitekturen eingesetzt werden, solange sie das Training mit ungepaarten Daten erlauben. Ungepaart bedeutet, dass keine direkte Zuordnung, z. B. in Position und Ausrichtung der Objekte, zwischen den Bildern der beiden Domänen bestehen. Während die Zieldomäne ( $\mathcal{Y}$ ) immer aus realen Aufnahmen besteht, ist die Art der Eingangsdomäne ( $\mathcal{X}$ ) nicht direkt vorgegeben. Wichtig ist nur, dass anhand der Eingangsdomäne Rückschlüsse auf die Annotationen der zu generierenden Bilder zu ziehen sind. Konkret werden bei den Methoden künstliche erzeugte binäre Annotationen (Domäne  $\mathcal{X}$ ) in reale Bilder (Domäne  $\mathcal{Y}$ ) überführt. Auf Grundlage dieser künstlichen Annotationen werden Anordnung, Pose und Morphologie der zu generierenden Objekte vorgegeben, wobei diese beliebig komplex modelliert werden können.

Die Verwendung von binären Annotationen als Eingangsdaten hat den entscheidenden Nachteil, dass keine Informationen über verschiedene Objektklassen in die Transformation eingebracht werden können. Ein generatives Modell kann aus den binären Annotationen zwar Objekte verschiedener Klassen generieren, wenn diese in den realen Trainingsdaten vorhanden sind, allerdings erfolgt dies nicht

gezielt. Aus einer binären Annotation kann bei ähnlicher Morphologie jede Klasse generiert werden. Damit kann nach der Transformation nicht nachvollzogen werden, welches Objekt welcher Klasse angehört. Problematisch wird es bei seltenen Elementen. Da diese, wie der Name impliziert, nur vereinzelt in den realen Daten auftreten, ist nicht sichergestellt, dass ein generatives Modell solche Elemente berücksichtigt. Das kann dazu führen, dass seltene Elemente nicht in dem synthetischen Trainingsdatensatz eines Segmentierungsnetzwerks vorhanden sind. Unterscheiden sich seltene Elemente von regulären Elementen beispielsweise in der Textur oder in der Morphologie, kommt es zu Fehlern in der Segmentierung. Doch gerade seltene Elemente sind häufig bei einer Auswertung von großem Interesse. So stellen Zellen in Teilungsphasen eine wichtige Kenngröße für den Zustand einer Zellkultur dar.

Die Ähnlichkeit zwischen den Domänen bzw. deren Unterschied (Domänenlücke) ist entscheidend für die Schwierigkeit und den Erfolg der Transformation. Es ist besonders wichtig, dass die enthaltenen Objekte in ihrer Morphologie und Dichte ähnlich sind, da andernfalls die Modelle möglicherweise neue Objekte hinzufügen oder bestehende entfernen [144, 150]. Eine geringere Domänenlücke erleichtert zudem das Training der Transformationsmodelle, was in der Regel zu einem effizienteren Trainingsprozess und besseren Ergebnissen führt.

Anhand der genannten Limitierungen derzeitiger Ansätze wurden Anforderungen an die zu entwickelnde Methode zur Generierung synthetischer Trainingsdaten herausgearbeitet, welche im Folgenden beschrieben werden.

1. Die zentrale Anforderung besteht darin, synthetische Daten mit mehreren Objektklassen zu erstellen, inklusive der zugehörigen Klassenannotationen.
2. Eine weitere Anforderung stellt die Möglichkeit zur gezielten Einbringung von seltenen Elementen dar.
3. Die Qualität der generierten Daten muss mindestens dem Stand der Technik entsprechen oder diesen übertreffen.



4. Die Methode muss zudem in der Lage sein, verschiedene Strukturen zu simulieren, die sich sowohl in Form als auch in ihrer Zusammensetzung deutlich unterscheiden.
5. Um unterschiedliche Mikroskope und deren Einstellungen nachbilden zu können, muss die Parametrisierung von Faktoren wie Auflösung, Rauschanteil und Rauschart flexibel möglich sein.
6. Darüber hinaus ist es wichtig, verschiedene Punktspreizfunktionen (PSF) zu verwenden, um theoretisch oder experimentell bestimmte Bildverzerrungen zu berücksichtigen.
7. Ein weiterer Aspekt betrifft die Nachbildung der variierenden Effizienz von Probenklärungen. Hierfür ist ein parametrierbarer Helligkeitsabfall zu integrieren, um realistische Verläufe der Bildintensität abzubilden.
8. Schließlich muss die Methode eine Möglichkeit bieten, die generierten Annotationen an den menschlichen Annotationsstil anzupassen. Aufgrund der Lichtstreuung erscheinen die erfassten Objekte auf Mikroskopaufnahmen größer als ihre tatsächliche Größe, was bei händischen Annotationen oft unbewusst nicht berücksichtigt wird. Da die synthetischen Annotationen auf den Rohdaten basieren, in denen die Lichtstreuung bislang nicht simuliert wurde, fallen diese typischerweise kleiner aus als die menschlichen Annotationen. Bei einer Kombination von synthetischen und händisch generierten Trainingsdaten kann dieser Unterschied in der Annotationsgröße die erzielbare Segmentierungsgüte einschränken.

Anhand dieser Anforderungen wurde in dieser Arbeit eine neue Deep-Learning-basierte Methode entwickelt (SimOptiGAN), welche im folgenden Abschnitt vorgestellt wird. SimOptiGAN kombiniert Simulation und Deep Learning für die Generierung von 3D-Daten. Dies ermöglicht es zum ersten Mal bei Deep-Learning-basierten Methoden gezielt mehrere Objektklassen und daher auch seltene Elemente in synthetische Daten einzubringen. Neben den bekannten Instanzannotationen können mit dieser Methode auch erstmalig Klassenannotationen generiert werden. Im Gegensatz zu den bekannten Methoden eignen sich die mit

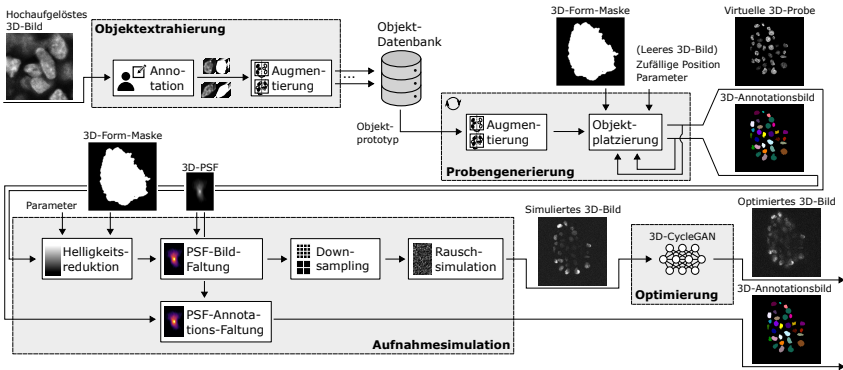
der entwickelten Methode erstellten Daten auch für das Training von panoptischen Segmentierungsnetzwerken, bei denen in einem Schritt Instanz- und Klassenannotationen vorhergesagt werden.

### **2.2.2 SimOptiGAN – Eine neue Methode zur Symbiose von Simulation und Deep Learning**

Mit SimOptiGAN wird eine neue Methode für die Synthese großer 3D-Strukturen vorgestellt. Die Pipeline beruht auf der Verbindung von klassischer Simulation und datengetriebener Optimierung zu einem hybriden Ansatz (siehe Abbildung 2.2). In dieser Arbeit wird der Begriff Synthese für die allgemeine Erstellung synthetischer, d. h. nicht realer Daten verwendet. Im Gegensatz dazu bezieht sich der Begriff Simulation ausschließlich auf klassische, parameterbasierte Simulationen, welche ohne den Einsatz datengetriebener Algorithmen arbeiten.

Weder simulationsbasierte noch Deep-Learning-basierte Verfahren sind allein ausreichend, um die oben genannten Anforderungen zu erfüllen. Simulationsbasierte Ansätze sind oft schwer zu parametrieren und neigen aufgrund von benötigten Abstrahierungen zu einer suboptimalen Qualität der generierten Bilder. Deep-Learning-basierte Verfahren, welche auf der Transformation von binären Annotationen beruhen, sind weniger flexibel und erlauben keinen direkten Einfluss auf die Generierung der Daten. Für die Generierung von Bildern in anderen Aufnahmequalitäten, wie ein anderes Signal-Rausch-Verhältnis, ist ein erneutes Training erforderlich. Zudem können keine Objekte generiert werden, welche nicht in den Trainingsdaten vorhanden sind. Weiterhin verläuft das Training dieser Modelle aufgrund der komplexen Aufgabestellung nicht immer stabil.

Der hybride Ansatz aus Simulation und Deep Learning bringt gegenüber einer klassischen Simulation den Vorteil, dass Effekte wie der Einfluss der Lichtbrechung oder die Art des Rauschens stärker abstrahiert werden können und somit die Parametrierung vereinfacht werden kann. Durch die nachgeschaltete Deep-Learning-Optimierung werden die daraus resultierenden Verluste in der Bildqualität ausgeglichen. Weiterhin bietet der hybride Ansatz Vorteile gegenüber einem



**Abbildung 2.2:** Konzept von SimOptiGAN zur Synthese von Aufnahmen. Die Pipeline kann in vier Blöcke unterteilt werden: Im ersten Block *Objektextrahierung* werden Objektprototypen aus realen Bildern extrahiert und in einer Datenbank platziert. In dem Block *Probengenerierung* wird mittels der extrahierten Objektprototypen eine virtuelle Probe erstellt, welche im Block *Aufnahmesimulation* mit den bei einer Mikroskopaufnahme auftretenden Effekten beaufschlagt wird. Zuletzt wird das simulierte 3D-Bild in dem Block *Optimierung* anhand eines Deep-Learning-Modells optimiert. Übernommen und übersetzt aus [176].

reinen Deep-Learning-basierten Ansatz: Zum einen wird, wie später gezeigt, die Domänenlücke verkleinert, was die Transformation zwischen den beiden Domänen erleichtert. Dies führt zu einem stabileren Trainingsprozess, einer verkürzten Trainingszeit und, wie nachgewiesen wird, zu besseren Ergebnissen. Zum anderen bietet der hybride Ansatz den Vorteil, dass Objekte und deren Ausprägungen gezielt eingebracht werden können. So kann etwa die Anzahl und die Positionen von Objekten einer Klasse genau festgelegt werden. Das ist auch dann noch möglich, wenn solche Objekte nicht in den Trainingsdaten der Deep-Learning-basierten Optimierung vorhanden sind.

Abbildung 2.2 zeigt eine grafische Übersicht der einzelnen Schritte von SimOptiGAN. Die Pipeline lässt sich in die vier Blöcke *Objektextrahierung*, *Probengenerierung*, *Aufnahmesimulation* und *Optimierung* aufteilen. In dem ersten Block *Objektextrahierung* werden aus realen Aufnahmen mittels händischer Annotation sogenannte Objektprototypen extrahiert und in einer Objekt-Datenbank

platziert. In dem folgenden Block *Probengenerierung* werden die Objektprototypen in einem Bild platziert, um damit die virtuelle Probe zu generieren. In dem nachfolgenden Block *Aufnahmesimulation* wird der Mikroskopaufnahmeprozess der virtuellen Probe simuliert. Im letzten Block *Optimierung* erfolgt dann eine Deep-Learning-basierte Optimierung, um die Qualität der simulierten Bilder zu steigern. Im Folgenden wird genauer auf die einzelnen Blöcke eingegangen.

Für die *Objektextrahierung* werden idealerweise hochaufgelöste Bilder mit einer hohen Qualität angefertigt. Eine hohe Auflösung, nahe an der Nyquist-Grenze, ermöglicht die Dekonvolution der Aufnahmen und damit die Reduktion der optischen Einflüsse des Mikroskops. Durch eine hohe Qualität werden zusätzlich die Effekte des Sensors wie das Rauschen gering gehalten. Nach der Aufnahme der Bilder findet eine händische *Annotation* von wenigen Objekten statt, um diese aus den Bildern auszuschneiden. In der folgenden *Augmentierung* werden für jedes Objekt  $n_{\text{aug}}$  augmentierte Versionen erzeugt, um so die Anzahl der Objekte weiter zu erhöhen. Da diese Augmentierung nicht zeitkritisch ist, können auch aufwendige Verfahren wie die elastische Transformation eingesetzt werden. Die originalen und augmentierten Objekte werden zuletzt in der *Objekt-Datenbank* platziert. Wichtig ist hierbei, dass die Datenbank eine Vielzahl von verschiedenen Objektklassen enthalten kann. Dabei können auch Objekte aus vorherigen Läufen wiederverwendet werden, wenn die Objekte hinreichend ähnlich sind. Neben der händischen Extraktion von Objekten können diese auch synthetisch generiert werden.

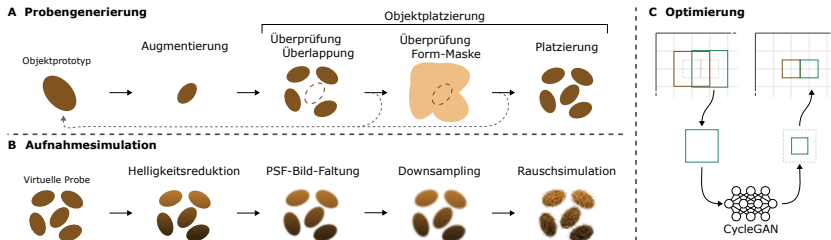
Der Block *Probengenerierung* stellt einen iterativen Prozess dar. Zu Beginn werden zufällig Objektprototypen aus der *Objekt-Datenbank* gezogen und augmentiert, um die Variabilität weiter zu steigern. Aufgrund der vorherigen komplexeren Augmentierung werden hier nur Spiegelungen und Rotationen durchgeführt. Damit lässt sich die Anzahl der Prototypen in der Objekt-Datenbank gering halten, ohne dabei die Varianz in der Orientierung einzuschränken. In diesem Zuge findet auch eine Anpassung der Auflösung der Objektprototypen an die Simulationsauflösung statt. Die augmentierten Objektprototypen werden anschließend in einem anfangs leeren Bild platziert. Die Positionierung der augmentierten Objektprototypen erfolgt anhand von zufällig bestimmten Positionen, welche anschließend auf

verschiedene Bedingungen überprüft werden. Abbildung 2.3a gibt hierzu eine detailliertere Darstellung des Prozesses. So sind keine oder nur leichte Überlappungen mit anderen bereits platzierten Objektprototypen zugelassen. Weiterhin können die möglichen Positionen anhand einer binären Form-Maske eingeschränkt werden. Dabei stellt der Vordergrund der Form-Maske mögliche Objektpositionen dar, während eine Platzierung im Hintergrund unterbunden wird. Mit der Form-Maske lassen sich verschiedene Formen von Objektstrukturen nachbilden. Wenn die Platzierung des Objektes aufgrund der Randbedingungen nicht möglich ist, wird so lange eine neue zufällige Position bestimmt, bis die Platzierung erfolgreich ist oder eine maximale Anzahl an Positions-Iterationen  $N_{\text{Pos}}$  überschritten ist. Tritt der zweite Fall ein, wird die Platzierung als fehlgeschlagen eingestuft und das Objekt wird verworfen. In beiden Fällen erfolgt anschließend eine weitere Objekt-Iteration, bestehend aus der Ziehung eines Objektprototyps, dessen Augmentierung und Platzierung. Ist eine Platzierung in  $N_{\text{Obj}}$  aufeinanderfolgenden Objekt-Iterationen fehlgeschlagen, wird die *Probengenerierung* gestoppt. Das Ergebnis dieses Blocks stellt die virtuelle Probe mit den dazugehörigen Annotationen dar. Über die Abbruchbedingungen  $N_{\text{Pos}}$  bzw.  $N_{\text{Obj}}$  kann auf die resultierende Objektdichte der generierten Struktur Einfluss genommen werden. Werden die Grenzwerte höher gesetzt, ergibt sich eine größere Dichte, da so auch kleinere Lücken noch gefüllt werden.

In biologischen Zellkulturen sind teilweise geordnete Strukturen zu erkennen. Fibroblasten-Zellen bilden etwa häufig Subkulturen, in denen nur Fibroblasten vorhanden sind. Die vorgestellte Methode erlaubt es, solche Subkulturen nachzubilden. Dazu wird vor der eigentlichen Probengenerierung die Generierung der Subkultur durchgeführt. Die Erstellung beruht auf dem gleichen Prinzip. Für die Platzierung der Objekte aus der Subkultur wird eine zusätzliche runde Form-Maske generiert. Dazu wird zufällig eine Position und ein Radius bestimmt. Liegt die Form-Maske der Subkultur nicht vollständig in der Form-Maske der Zellkultur, wird erneut zufällig eine Position und ein Radius bestimmt. Liegt die Maske innerhalb der Form-Maske der Hauptkultur, wird die Form-Maske der Subkultur für die weitere Platzierung verwendet. Die Platzierung der Objektprototypen erfolgt auf gleiche Weise wie die reguläre Probengenerierung. Wenn keine weiteren

Objektprototypen in der Subkultur platziert werden können, dann wird mit der Probengenerierung der Hauptkultur fortgefahren.

Das Ziel des Blocks *Aufnahmesimulation* ist es, möglichst authentische Bilder mit bereits realistischen Eigenschaften bei moderatem Aufwand zu generieren. Ein komplexer physikalischer Effekt wird so eher approximiert, als ihn in aller Detailtreue zu modellieren. Zuerst erfolgt die durch die Lichtbrechung verursachte *Helligkeitsreduktion*. Weiterhin wird durch die Lichtbrechung innerhalb der Optik eine Verzerrung des Signals verursacht, was durch die Punktspreizfunktion (PSF) beschrieben wird. Dieser Effekt wird durch die Faltung des Bildes mit einer gemessenen oder geschätzten PSF abgebildet (*PSF-Bild-Faltung*). Weiterhin können optional auch die Annotationen mittels der PSF, passend zu den Bilddaten, vergrößert werden (*PSF-Annotations-Faltung*). Das Ziel ist, die Größe der Annotationen an den Stil menschlicher Annotationen anzupassen. Aufgrund der mit der PSF beschriebenen Signalstreuung wirken Objekte in realen Aufnahmen



**Abbildung 2.3:** Detailsichten der in Abbildung 2.2 dargestellten Blöcke Probengenerierung (A), Aufnahmesimulation (B) und Optimierung (C). **A** Probengenerierung: Ein Objektprototyp aus der Objekt-Datenbank wird augmentiert. Für eine zufällig bestimmte Position wird die Überlappung des Objektprototyps mit bereits platzierten Objekten und der Form-Maske geprüft. Ist eine der beiden Bedingungen nicht erfüllt, wird der Prozess durch die Ziehung eines neuen Objektprototyps wiederholt, andernfalls wird das Objekt platziert. **B** Aufnahmesimulation: Die virtuelle Probe unterläuft einer Helligkeitsreduktion, bei der die Intensität in tieferen Schichten reduziert wird, gefolgt von der PSF-Bild-Faltung, bei der die Verzerrung des Bildes nachgebildet wird. Anschließend erfolgt ein Downsampling, bei dem die Auflösung angepasst wird, und die Rauschsimulation, bei der verschiedene Rauschmodalitäten simuliert werden. **C** Optimierung: Das Bild wird in Ausschnitte unterteilt, welche einzeln durch ein CycleGAN transformiert werden. Durch die Überlappung der Bildausschnitte können die oft Artefakt-behafteten äußeren Regionen bei der erneuten Zusammensetzung des Bildes verworfen werden.

größer, was somit bei einer händischen Annotation auch zu größeren Annotationen führt. Bei der Kombination von realen und synthetischen Annotationen kann durch die *PSF-Annotations-Faltung* die Größe der synthetischen Annotationen an die der menschlich erstellten Annotationen angeglichen werden, um so die Konsistenz zu steigern. Anschließend erfolgt in dem *Downsampling* Schritt die Anpassung der Auflösung an die gewünschte Ausgabeauflösung <sup>1</sup>. Zuletzt wird bei der *Rauschsimulation* das Rauschen des Bild-Sensors durch die Modellierung von Poisson- und Gaußschem-Rauschen simuliert. Die Stärke der Rauschanteile kann dabei über Parameter eingestellt werden <sup>2</sup>. Ein Anhaltspunkt für die Stärke des Rauschens kann über Histogramme von Hintergrundregionen in realen Bildern erhalten werden. Zusätzlich kann der Dunkelstrom eines Sensors durch einen konstanten Offset simuliert werden. Bei der alleinigen Verwendung von Poisson-Rauschen stellt der Dunkelstrom einen wichtigen Faktor dar, da die Stärke des Rauschens von der Bildintensität abhängig ist. In Bereichen mit einer Intensität von null ist die Varianz des Poisson-Rauschens gleich null. Abbildung 2.3b zeigt eine Detailansicht der beschriebenen Schritte mit einer Visualisierung der Effekte.

Aufgrund der bewusst vereinfachten Simulation resultieren leichte Unterschiede zwischen realen und simulierten Daten <sup>3</sup>. Mithilfe der Deep-Learning-basierten *Optimierung* werden Differenzen zwischen der simulierten und realen Domäne weiter reduziert. Genutzt wird dazu ein CycleGAN (siehe Abschnitt 1.4.3), welches eigens für die native Verarbeitung von 3D-Daten angepasst wurde. GANs ermöglichen die Transformation von Bildern einer Domäne  $\mathcal{X}$  in Bilder einer anderen Domäne  $\mathcal{Y}$ . In dem speziellen Anwendungsfall stellt die Domäne  $\mathcal{X}$  die simulierten Aufnahmen dar, während die Domäne  $\mathcal{Y}$  die realen Aufnahmen umfasst. Eine Besonderheit von CycleGANs ist, dass die Transformation anhand von ungepaarten Bildern gelernt werden kann. In der Literatur ist für die Synthese von

<sup>1</sup> Da sich die über die PSF beschriebene Signalstreuung auf einen kleinen Raum beschränkt, führt eine Synthese, welche direkt in der gewünschten Ausgabeauflösung durchgeführt wird, zu einem starken Qualitätsverlust. Deswegen wird, wenn es die Hardware zulässt, die Synthese bis nach der Faltung mit der PSF in einer höheren Auflösung durchgeführt. Dabei wird deutlich mehr RAM-Speicher und auch eine längere Rechenzeit benötigt.

<sup>2</sup> In Versuchen hat sich gezeigt, dass die Einbringung von Poisson-Rauschen ausreichend ist.

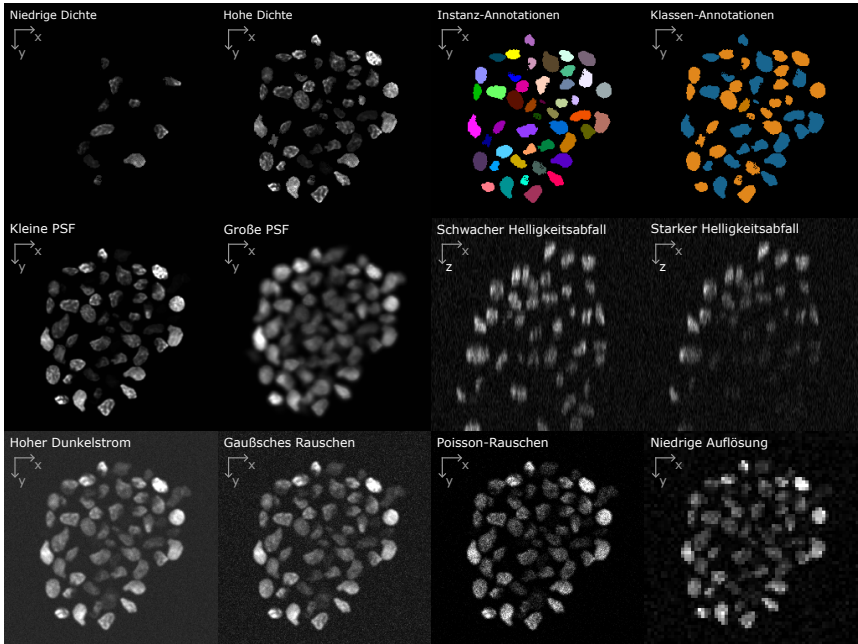
<sup>3</sup> Zum Beispiel wird keine Simulation autofluoreszierender Signale durchgeführt.

Trainingsdaten eine Transformation auf Basis von binären Annotations-Daten üblich. Bei der hier vorgestellten Methode werden allerdings Bilder der simulierten Domäne  $\mathcal{X}$  in die reale Domäne  $\mathcal{Y}$  überführt. Dadurch wird, wie später nachgewiesen, die Distanz zu der realen Domäne reduziert. Zum einen wird durch die vereinfachte Aufgabe die benötigte Trainingsdauer reduziert und der Trainingsprozess stabiler. Zum anderen wird das Risiko gesenkt, dass durch die Domänen-Transformation neue, unerwünschte Objekte platziert werden [144, 150]. Solche unerwünschten Objekte führen zu Diskrepanzen zwischen den generierten Bildern und deren Annotationen.

Im Allgemeinen wird für das Training von GANs und im Speziellen bei 3D-GANs eine große Menge an Grafikspeicher benötigt. Daher ist die direkte Prozessierung von großen 3D-Bildern nicht möglich. Folglich wird auf eine Patch-basierte Transformation zurückgegriffen, bei der die Bilder in kleine Bildausschnitte aufgeteilt, transformiert und wieder zu einem vollständigen Bild zusammengesetzt werden (vgl. Abbildung 2.3c). Für den Erhalt der 3D-Information wird in diesem Ansatz eine Zerlegung in allen Bild-Achsen  $(x, y, z)$  vorgenommen. Da häufig eine geringere Auflösung in der  $z$ -Achse vorliegt, wird hier die Größe des Ausschnitts kleiner gewählt. Eine übliche Größe für Bildausschnitte ist  $32 \times 256 \times 256 \text{ px}^3$   $(z, y, x)$ . Ein Problem des Patch-basierten Ansatzes ist, dass globale Informationen durch die unabhängige Prozessierung der Ausschnitte nicht mehr vorhanden sind. Bei dem nachträglichen Zusammensetzen der Ausschnitte können unstetige Übergänge entstehen (e.g. Helligkeitssprünge) [178]. Hier liegt ein weiterer Vorteil der vorgestellten Methode. So sind globale Informationen wie die Helligkeitsverteilung bereits in den simulierten Daten enthalten und müssen nicht wie bei anderen Methoden in zusätzlichen Kanälen kodiert werden. So nutzt Eschweiler et al. [144] etwa einen zusätzlichen Kanal, um darin Positionsinformationen in Form einer Distanzkarte zu erhalten. Weiterhin führt die geringere Distanz zwischen den Domänen zu einer erhöhten Konsistenz zwischen den ursprungs- und transformierten Bildern [144, 150]. Durch die bereits vorhandenen Textur- und Helligkeitsinformationen können unstetige Übergänge bei der Patch-basierten Transformation reduziert werden [178]. Die Ausschnitte werden so gewählt, dass



diese sich überlappen. Dies ermöglicht, die Randbereiche beim späteren Zusammensetzen zu exkludieren. Für Pixel am Rand des Bildes liegen die in Faltungsnetzwerken genutzten Filterkerne teilweise außerhalb des Bildbereiches, wodurch die Faltungsoperation hier nicht direkt durchführbar ist. Die fehlende Information kann aber beispielsweise durch eine Spiegelung des Bildes am Rand erzeugt werden. Da die extrapolierten Signale von den echten Signalen abweichen, kann es bei diesen Bereichen nach der Transformation zu Artefakten kommen.



**Abbildung 2.4:** Veranschaulichung der Wirkung verschiedener Parametereinstellungen auf die mit SimOptiGAN generierten 3D-Aufnahmen. Zusätzlich sind auch Instanz- und Klassenannotationen abgebildet. Mit Ausnahme der *niedrigen Dichte* basieren alle Aufnahmen auf derselben generierten Probe und weisen damit die gleiche Zellanordnung auf. Die Aufnahme *Hohe Dichte* zeigt Ergebnisse ohne Simulation der PSF, womit die korrespondierenden *Annotationen* eins zu eins übereinstimmen. Durch die Signalspreizung der PSF (vgl. *große PSF*) sind Objekte aus angrenzenden  $z$ -Ebenen sichtbar, wodurch der Eindruck einer höheren Zelldichte entsteht.

Die Abbildung 2.4 zeigt mit der hybriden Synthesepipeline generierte synthetische Aufnahmen inklusive Instanz- und Klassenannotationen. Die Aufnahmen basieren mit Ausnahme der *niedrigen Dichte* auf der identischen Probe und weisen so die gleiche Zellanordnung auf. Die Abbildung veranschaulicht weiterhin die Auswirkung verschiedener Parametereinstellungen und verdeutlicht damit die generierbare Varianz. So kann mit der maximalen Anzahl an aufeinanderfolgenden Platzierungsversuchen auf die *Dichte* der Objekte Einfluss genommen werden. Mit unterschiedlichen *PSF* ergeben sich Unterschiede in der Stärke der Signalspreizung. Gerade in der *z*-Achse werden die Signale stark gestreut, wodurch sich die Objektgrenzen erweitern und sich der optische Effekt einer höheren Objektdichte ergibt. Weiterhin entsteht durch die Signalspreizung der Eindruck einer inkonsistenten Annotation. Das verdeutlicht den Unterschied zwischen synthetischer und menschlicher Annotationsweise. Durch die Signalspreizung der *PSF* werden auch die unrealistisch harten Kanten der platzierten Objekte gedämpft (vgl. *Hohe Dichte* und *Große PSF*). Idealerweise wird eine gemessene *PSF* genutzt, um eine ähnliche Signalstreuung in den generierten Aufnahmen zu erhalten. Weiterhin kann durch die Rauschsimulation Einfluss auf die Art des Rauschens und dessen Stärke genommen werden. Dabei ist gut zu erkennen, dass *Gaußsches-Rauschen* unabhängig vom Bildinhalt ist, während die Varianz des *Poisson-Rauschens* von der Intensität der Bildinhalte abhängt. Somit liegt beim *Poisson-Rauschen* im Hintergrund kaum Rauschen vor, während das Rauschen in den Objekten gut zu erkennen ist. Der durch einen konstanten Offset simulierte *Dunkelstrom* ist daher bei einer ausschließlichen Nutzung von *Poisson-Rauschen* wichtig, um auch im Hintergrund einen Rauschanteil zu erzielen. Durch das Downsampling können auch Aufnahmen mit geringen Auflösungen generiert werden. Zudem kann die Stärke des *Helligkeitsabfalls* in tieferen Schichten individuell angepasst werden, um beispielsweise unterschiedliche Clearing-Effizienzen zu simulieren. Es wird deutlich, dass kein linearer Abfall verwendet wird, sondern die darüberliegenden Vordergrundregionen in der Simulation berücksichtigt werden.

Wie anhand der gezeigten Abbildungen deutlich wird, eignet sich die vorgestellte Methode zur Erstellung synthetischer Mikroskopaufnahmen biologischer Zellproben. Allerdings ist SimOptiGAN nicht auf diesen Anwendungsbereich beschränkt,

sondern kann auch zur Generierung anderer Probenarten eingesetzt werden. Eine direkte Übertragbarkeit ist insbesondere bei Strukturen gegeben, die aus vielen Instanzen mit einer eher kompakten Anordnung bestehen. Da SimOptiGAN modular aufgebaut ist, lassen sich auch einzelne Komponenten der Methode flexibel anpassen. So kann etwa das Modul zur Probengenerierung modifiziert werden, um synthetische Aufnahmen anderer Strukturen zu erzeugen. Beispielsweise können mit entsprechenden Anpassungen schwammartige Strukturen generiert werden, wie sie in der Materialwissenschaft untersucht werden.

### **2.2.3 Einbringung biophysikalischer Modellierung in den Syntheseprozess**

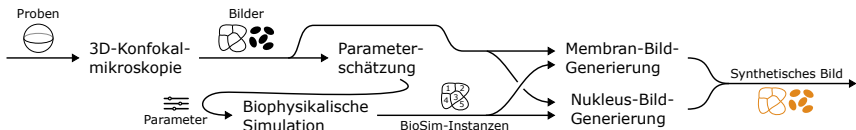
Die Anordnung von Zellen in einer Zellkultur ist nicht willkürlich, sondern ist ein Ergebnis komplexer biophysikalischer Wechselwirkungen. Zum Beispiel sind in Krebszell-Sphäroiden teilende Zellen oft stärker zum Rand der 3D-Kultur hin konzentriert, während nekrotische oder apoptotische Zellen häufiger im Zentrum des Sphäroids vorkommen. Weiterhin zeigen Zellen oft eine Präferenz für eine tangential Ausrichtung relativ zur äußeren Hülle [179]. Mit der oben vorgestellten Methode SimOptiGAN können bereits synthetische Bilder von 3D-Zellkulturen mit hoher Qualität generiert werden, wobei die Platzierung der Zellkerne in einer zufälligen Art und Weise erfolgt. Daraus resultieren unweigerlich Diskrepanzen in der Anordnung von Nuklei zwischen der echten und synthetischen Bild-Domäne. Inwieweit sich die Diskrepanzen in der Anordnung auf die Segmentierungsqualität auswirken, ist in dieser Arbeit zu untersuchen.

Durch die gesteigerte Kohärenz in der Anordnung wird das Risiko verringert, dass durch Deep-Learning-basierte Transformationen unerwünschte Strukturen generiert werden [144, 150]. Solche Strukturen sind unerwünscht, da sie zu Inkonsistenzen zwischen Bild- und Annotations-Domäne führen. Die korrekte physikalische Modellierung von Zellstrukturen stellt allerdings eine komplexe und rechenintensive Aufgabe dar. Bis heute nutzt keine Methode zur Synthese

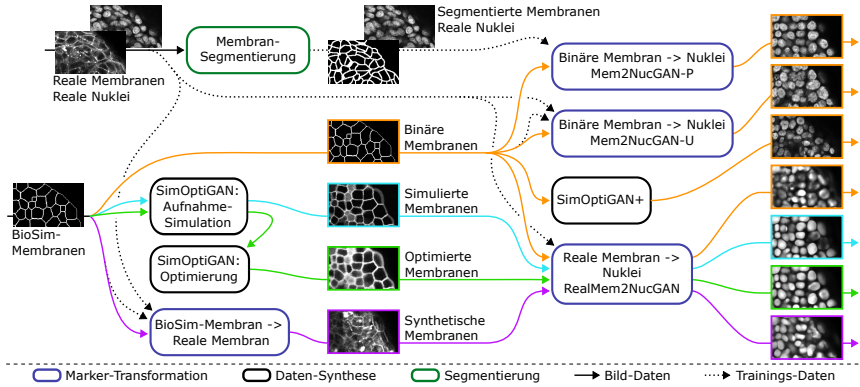
von 3D-Zell-Bilddaten eine biophysikalisch motivierte Zellanordnung, um die Organisation von Zell- und Kernstrukturen zu verbessern.

Vorhandene physikalisch motivierte Zellmodelle ermöglichen zwar die Simulation von ganzen Zellstrukturen, allerdings berechnen diese nur Positionen und Morphologie der Zellen. Die Ergebnisse können als Annotations-Daten angesehen werden, da sie keinerlei Informationen bezüglich der Textur enthalten. Für den Einsatz als Trainingsdaten für Segmentierungsmodelle werden zusätzlich Bilddaten mit Texturen benötigt. Eine weitere Limitierung ist, dass mit vorhandenen Simulationsmodellen meistens keine den Zellen zugehörigen Zellkerne generiert werden. Allerdings stellt die Segmentierung der Nuklei einen grundlegenden Schritt für weitere Analysen dar. Deshalb werden in dieser Arbeit neue Ansätze entwickelt, um anhand von biophysikalisch motivierten Zellmodell-Simulationen synthetische Bilder von Nukleus- und Membransignalen zu generieren.

Abbildung 2.5 zeigt eine Übersicht der Vorgehensweise. Für die im Rahmen dieser Arbeit mitentwickelte *biophysikalische Simulation* [180] werden *Proben* erstellt und mittels *3D-Konfokalmikroskopie* aufgenommen. Anhand der resultierenden *Bilder* erfolgt eine *Parameterschätzung* von biophysikalischen Eigenschaften der Zellen und deren Grenzflächen. Die so bestimmten *Parameter* werden in der anschließenden biophysikalischen Simulation genutzt, um damit Zellstrukturen zu simulieren. Diese liegen in Form einer Instanzannotationskarte vor (folgend als *BioSim-Instanzen* bezeichnet), bei der die ID einer Zelle durch einen Graustufenwert abgebildet wird. Anschließend erfolgt die *Membran- und Nukleus-Bild-Generierung*, bei der die *BioSim-Instanzen* genutzt werden, um die Anordnung von Zellen in *synthetischen Bildern* zu verbessern.



**Abbildung 2.5:** Konzept zur Einbringung von biophysikalischen Simulationen in den Syntheseprozess. Übernommen und übersetzt aus [177].



**Abbildung 2.6:** Entwickelte Möglichkeiten zur Generierung von Nukleusaufnahmen auf Basis bio-physikalisch simulierter binärer Membransignale. Die farbigen Pfeile und die farbig umrandeten Bilder markieren zusammenhängende Pfade. Übernommen und angepasst aus [177].

Wie in Abbildung 2.6 gezeigt, wurde im Rahmen dieser Arbeit eine Vielzahl von Ansätzen entworfen, um die Ergebnisse der biophysikalischen Simulation in den Generierungsprozess synthetischer Aufnahmen einzubringen. Neben der Synthese von Nukleussignalen wird die biophysikalische Simulation auch dazu genutzt, zu den Nuklei passende Membransignale zu generieren. Ausgangspunkt sind die mithilfe der biophysikalischen Simulation generierten BioSim-Instanzen, auf deren Basis synthetische Aufnahmen von Zellkern- und dazu passenden Membransignalen zu generieren sind. Allen Ansätzen ist gemein, dass für die Synthese von Nukleus- und Membranaufnahmen die BioSim-Instanzen zuerst in ein binäres Membranbild umgewandelt werden, was im Folgenden als BioSim-Membranen bezeichnet wird. Dabei werden die Voxel an jedem Zellübergang als Vordergrund und der Rest als Hintergrund gesetzt. Eine Herausforderung der Modelle zur *Marker-Transformation* (blau umrandet), die Membransignale in Nukleussignale überführen, besteht darin, korrespondierende Nukleusannotationen zu erhalten. Nach aktuellem Kenntnisstand bieten bisherige GAN-Methoden keine Möglichkeit, zugehörige Annotationen zu generieren. Daher wird in dieser Sektion mit DualGenGAN eine neue Trainingsstruktur eingeführt, welche die gleichzeitige Generierung von Bildern und deren Annotationen ermöglicht.

Die verschiedenen Ansätze zur Generierung von synthetischen Membransignalen bzw. Nukleussignalen werden nachfolgend erläutert. Weiterhin werden die Ansätze im Zuge dieser Arbeit miteinander verglichen und eine Variantendiskussion durchgeführt.

### Von Zellannotationen zum synthetischen Membransignal

Um authentische Membransignale aus den binären BioSim-Membranen zu erzeugen, wird die CycleGAN-Architektur verwendet (vgl. Abbildung 2.6, lila Pfad, *BioSim-Membran*  $\rightarrow$  *Reale Membran*). Dabei können ungepaarte Bilder von BioSim-Membranen und realen Membranaufnahmen für das Training der Transformation eingesetzt werden. Abgesehen von der 3D-Anpassung der Generator- und Diskriminator-Modelle, bei denen 2D-Operationen wie Faltung und Pooling mit den äquivalenten 3D-Versionen ausgetauscht wurden, gleicht die CycleGAN-Architektur der ursprünglichen in [181] vorgestellten. Nach dem Training der Modelle kann mittels des Generators die Transformation von BioSim-Membranen zu authentisch wirkenden synthetischen Membransignalen durchgeführt werden.

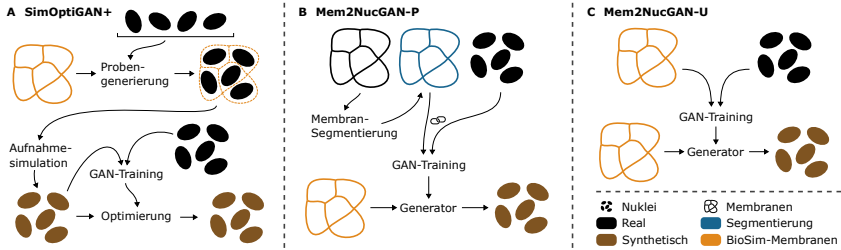
### Von Zellannotationen zu Nukleussignalen

Die Segmentierung von Nukleussignalen stellt eine wichtige Grundlage für weitere Analysen dar. Um die BioSim-Membranen zu nutzen und damit synthetische Nukleusaufnahmen mit authentischer Anordnung zu generieren, wurden verschiedene Methoden entwickelt und miteinander verglichen (siehe Abbildung 2.6). Alle Ansätze bieten den Vorteil, dass sich die synthetischen Nukleussignale mit den synthetischen Membransignalen kombinieren lassen. Das ist möglich, da beide auf Basis der gleichen BioSim-Membranen generiert werden und damit die gleiche Anordnung aufweisen. Durch die Kombination der Signale können Segmentierungsnetzwerke mit beiden Signalen trainiert werden, was, wie später gezeigt, zu einer Verbesserung der Segmentierungsqualität führen kann.

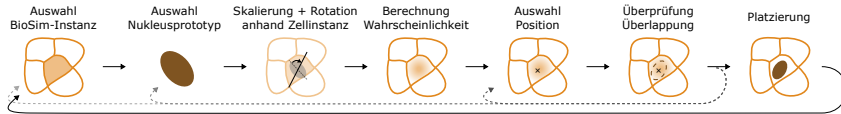
Die erste Methode wird als *SimOptiGAN+* bezeichnet (orangefarbener Pfad). Sie stellt eine Weiterentwicklung von *SimOptiGAN* dar, bei der die Platzierung der Objekte durch die BioSim-Membranen verbessert wird. Die zweite Methode (*Mem2NucGAN-P*, orangefarbener Pfad) nutzt ein Transformationsmodell *Binäre Membran to Nuklei* welches anhand gepaarter Aufnahmen von realen Nuklei und segmentierten Membransignalen trainiert wird. Nach dem Training des Modells wird dieses eingesetzt, um die binären BioSim-Membranen in synthetische Nukleussignale zu transformieren. Eine genaue Beschreibung findet sich im Abschnitt *Mem2NucGAN-P*. Die dritte Methode (*Mem2NucGAN-U*, orangefarbener Pfad) nutzt im Gegensatz zur vorherigen Methode ungepaarte Daten bestehend aus realen Nukleussignalen und BioSim-Membranen, um damit das Transformationsmodell *Binäre Membran to Nuklei* zu trainieren. Nach dem Training können mit dem Modell wiederum BioSim-Membranen in synthetische Nukleussignale überführt werden. Eine genaue Beschreibung findet sich im Abschnitt *Mem2NucGAN-U*. Die vierte Methode (*RealMem2NucGAN*, mehrere Pfade) nutzt ein Transformationsmodell *Reale Membran to Nuklei*, welches mit gepaarten Aufnahmen von realen Membran- und Nukleussignalen trainiert wird. Da sich die binären BioSim-Membranen zu stark von den eigentlichen Trainingsdaten unterscheiden, ist eine Prä-Prozessierung für die Inferenz notwendig, bei der die BioSim-Membranen an die realen Membransignale angeglichen werden. Für die Prä-Prozessierung werden wiederum verschiedene Ansätze getestet. Eine genaue Beschreibung findet sich im Abschnitt *RealMem2NucGAN*.

### **Verbesserte Kernplatzierung für den simulationsbasierten Ansatz:**

**SimOptiGAN+** Bei dem bereits vorgestellten Ansatz *SimOptiGAN* erfolgt die Anordnung der Zellkerne anhand eines zufälligen Musters. Die Positionen und Orientierungen der Nuklei sind nicht physikalisch motiviert. Bei *SimOptiGAN+* wird der Prozess der Nukleus-Platzierung angepasst, um eine physikalisch motivierte Anordnung und Orientierungen zu erhalten (siehe Abbildung 2.7a). Dazu wird der Teil *Probengenerierung* angepasst, bei dem die digitale Probe einer Zellstruktur generiert wird. Anstatt zufällig *Objektprototypen* aus einer Datenbank



**Abbildung 2.7:** Detailsansicht der drei weiter untersuchten Methoden zur Generierung von Nukleusaufnahmen, basierend auf den aus biophysikalischen Simulationen resultierenden BioSim-Membranen. SimOptiGAN+ nutzt die BioSim-Membranen, um die Platzierung der Nuklei in einem zuvor vorgestellten simulationsbasierten Ablauf zu verbessern. Mem2NucGAN-P und Mem2NucGAN-U setzen auf eine GAN-basierte Transformation, um anhand der BioSim-Membranen synthetische Nukleusaufnahmen zu generieren. Übernommen und übersetzt aus [177].



**Abbildung 2.8:** Detailsansicht der adaptierten Probengenerierung von SimOptiGAN+. Gestrichelte Linien zeigen den Weg bei fehlgeschlagenen Platzierungsversuchen.

zu ziehen und an zufälligen Positionen zu platzieren, erfolgt die Auswahl und Platzierung auf Basis der BioSim-Instanzen.

Eine Detailsansicht der adaptierten Probengenerierung ist in Abbildung 2.8 gezeigt. Für jede Zelle der *BioSim-Instanzen* wird ein zufälliger *Nukleusprototyp* aus der Datenbank ausgewählt. Der Prototyp wird anschließend durch eine *Rotation* an der Orientierung der Zelle ausgerichtet. Anschließend erfolgt eine *Skalierung*, um ein festgelegtes Verhältnis zwischen Nukleus und Zellvolumen zu erreichen.

In eigenen Untersuchungen wurde festgestellt, dass Zellkerne selten exakt im Zentrum der Zellen vorzufinden sind. Die Positionierung der Nukleusprototypen erfolgt daher anhand einer im Folgenden beschriebenen *Wahrscheinlichkeitsverteilung* (siehe Abbildung 2.8). Die Wahrscheinlichkeiten für die Ziehung einer bestimmten Position orientieren sich an der Distanz der Position zu



dem Rand der Zelle. Punkte, welche sich näher am Zentrum der Zelle befinden, weisen eine höhere Wahrscheinlichkeit auf. Sei  $d_{\text{edt}}(x, y, z)$  die euklidische Distanz eines Punktes zum Rand der Zelle. Dann ergibt sich mit  $r(x, y, z) = \max_{x, y, z} d_{\text{edt}}(x, y, z) - d_{\text{edt}}(x, y, z)$  die Distanz eines Punktes zum Zentrum der Zelle. Bei 3D-Rastergrafiken und kugelförmigen Objekten nimmt die Anzahl der Voxel, welche den gleichen Abstand zum Zentrum aufweisen, mit größeren Radien kubisch zu. Daher wird eine Normierung durchgeführt, welche sich an dem Volumen einer hohlen Kugel orientiert. Der Normierungsfaktor  $k_{\text{norm}}$  und die normierte Distanz  $d_{\text{norm}}$  berechnet sich wie folgt:

$$k_{\text{norm}} = \frac{4}{3}\pi * \left( (r(x, y, z) + 1)^3 - r(x, y, z)^3 \right), \quad (2.1)$$

$$d_{\text{norm}}(x, y, z) = \frac{d_{\text{edt}}(x, y, z)}{k_{\text{norm}}}. \quad (2.2)$$

Damit die Summe aller Wahrscheinlichkeiten eins ergibt, wird  $d_{\text{norm}}$  nochmals normiert. Die finale Wahrscheinlichkeit für einen Punkt  $(x, y, z)$  ergibt sich damit zu:

$$p_{\text{position}}(x, y, z) = \frac{d_{\text{norm}}(x, y, z)}{\sum_{x, y, z} d_{\text{norm}}(x, y, z)}. \quad (2.3)$$

Anhand dieser Verteilung wird anschließend zufällig eine *Position* für die Platzierung des Nukleus bestimmt. Für diese Position wird dann anhand der prozentualen *Überlappung* zwischen Nukleusprototyp und Zelle geprüft, ob der Nukleusprototyp in die Zelle passt (siehe Abbildung 2.8). Ist die Überlappung größer als ein spezifischer Schwellenwert  $k_{\text{vol}}$  des Nukleus Volumens, wird der Nukleus an dieser Stelle *platziert*, wobei nicht überlappende Bereiche des Nukleus abgeschnitten werden. Ist die Überlappung kleiner als der Schwellenwert  $k_{\text{vol}}$ , wird eine neue Position ausgewählt und die Platzierung an dieser Stelle mit der gleichen Ausrichtung erneut geprüft.

Dieser Vorgang wird wiederholt, bis eine valide Position gefunden wurde, bis alle möglichen Positionen in der Zelle getestet wurden oder eine maximale Anzahl an Versuchen  $N_{\text{pos}}$  erreicht ist. In den beiden letzten Fällen wird ein neuer Objektprototyp ausgewählt, passend zur Zelle rotiert und skaliert, und anschließend

die Platzierung geprüft. Für jede Zelle werden maximal  $N_{\text{Obj}}$  Objektprototypen geprüft, bevor die Zelle übersprungen wird und somit leer bleibt. Nach einer erfolgreichen Platzierung oder nach einem Überspringen der Zelle wird das Prozedere für die nächste Zelle wiederholt, bis alle Zellen der BioSim-Instanzen bearbeitet wurden.

Gerade wenn sich die Morphologie von Zellkernen und den BioSim-Instanzen zu stark voneinander unterscheiden, kann die Platzierung der Zellkerne häufig fehlschlagen. Konkrete Beispiele sind ungleiche Rundheiten zwischen Zellkernen und BioSim-Instanzen oder auch Fehler in der Simulation, bei denen schlauchartige Zellen entstehen. Durch starke morphologische Unterschiede schlagen die Platzierungsversuche fehl, was durch die wiederholten Platzierungsversuche zu einer langen Rechenzeit und unter Umständen zu einer Vielzahl von nicht befüllten BioSim-Instanzen führt. In solchen Fällen kann das Verhältnis zwischen Nukleus- und Zellvolumen angepasst werden, um mit kleiner skalierten Nukleusprototypen eine Platzierung zu begünstigen. Allerdings führt dies auch zu kleineren Zellkernen im generierten Bild.

Nachdem die Objektprototypen in die simulierten Zellen platziert wurden, ist der Schritt *Prototypengenerierung* abgeschlossen. Die weitere Verarbeitung erfolgt analog zur ursprünglichen Pipeline SimOptiGAN (vgl. Abschnitt 2.2.2). Die Optimierung kann dabei mit dem gleichen bereits trainierten Netzwerk erfolgen, wie es bei SimOptiGAN verwendet wurde. Am Ende stehen sowohl die synthetischen Bilder der Nuklei zur Verfügung, als auch die dazugehörigen Annotationen.

Dieser Ansatz bietet den Vorteil, dass durch die systematische Auswahl und Platzierung von Objektprototypen in den simulierten Zellen auch gezielt seltene Elemente eingebracht werden können. Weiterhin stehen, wie auch beim ursprünglichen Ansatz, direkt Instanzannotationen und gegebenenfalls auch Klassenannotationen der Objekte zur Verfügung. Es besteht zudem die Möglichkeit, zusätzliche Modalitäten gemeinsam mit den Nukleussignalen zu synthetisieren, sofern deren Signale innerhalb der Zellkerne auftreten. So werden diese gemeinsam mit den Nukleusprototypen aus Aufnahmen extrahiert, in der Datenbank gespeichert und können anschließend gemeinsam platziert werden. In diesem Zusammenhang

können die Zellkerne anhand der zusätzlichen Modalität in Klassen eingeteilt werden, wodurch nach der Synthese direkt die Klassenannotationen zur Verfügung stehen. Weiterhin ist SimOptiGAN+ nicht auf ein bestimmtes biophysikalisches Simulationsmodell beschränkt, sondern kann mit Zellinstanzen bzw. Membranen beliebiger physikalischer Modelle genutzt werden.

### **BioSim-Membranen zu Nuklei mit gepaarten Daten:**

**Mem2NucGAN-P** Die Platzierung von Zellkernen in BioSim-Instanzen (SimOptiGAN+) besitzt den Nachteil, dass nur eine begrenzte Anzahl an Zellprototypen für die Platzierung in den BioSim-Instanzen vorliegt. Damit ergeben sich zwangsläufig suboptimale Zell-Nukleus-Paare, welche in ihrer Morphologie nicht optimal zusammenpassen. Daher wird eine weitere Methode entwickelt, bei der keine Nukleusprototypen für die Synthese von Nukleussignalen notwendig sind. Eigene Untersuchungen bezüglich der Synthese von Bildmodalitäten (Abschnitt 2.3) zeigten, dass GANs die Generierung von Kernsignalen auf Basis von realen Membransignalen erlernen können, wobei die Übereinstimmung zwischen echten und generierten Signalen zu prüfen ist. Die Verwendung eines solchen, auf echten Daten trainierten Transformationsmodells kann damit die Generierung von Kernsignalen mit realistischen Eigenschaften entsprechend den BioSim-Membranen ermöglichen. Eine Herausforderung stellt allerdings die Diskrepanz zwischen den im Training verwendeten realen Membransignalen und den in der Inferenz-Phase verfügbaren binären BioSim-Membranen dar. Diese Domänenlücke zwischen Trainings- und Anwendungsdaten lässt die Transformation scheitern. Daher ist entweder eine Anpassung der Trainingsdaten oder eine Anpassung der späteren Anwendungsdaten erforderlich. Beide Varianten wurden untersucht. Bei der Methode Mem2NucGAN-P (vgl. Abbildung 2.7b) werden die Trainingsdaten den späteren Anwendungsdaten angepasst, indem die realen Membransignale zuerst automatisch segmentiert und binarisiert werden (siehe Abbildung 2.7b).

Die automatische Segmentierung der Membransignale erfordert das Training eines Netzwerks *Membran-Segmentierung*. Da die manuelle Annotation einer ausreichenden Menge an Membrandaten für das Training mit einem unvermeidbar großen Aufwand verbunden ist, werden auch hier synthetische Trainingsdaten

eingesetzt. Für deren Erstellung kann das bereits eingeführte Transformationsmodell *Synthetische Zellgrenzen*  $\rightarrow$  *Reale Membran* verwendet werden, welches BioSim-Membranen in realistische Membranen überführt (vgl. Abschnitt 2.2.3). Dabei stellen die BioSim-Membranen die Annotationsdaten und die daraus generierten synthetischen Membranen die Bilddaten des Trainingsdatensatzes für das Segmentierungsnetzwerk *Membran-Segmentierung* dar.

Mit dem trainierten Segmentierungsnetzwerk (*Binäre Membran*  $\rightarrow$  *Nuklei*) können die Membransignale von Aufnahmen realer Membran- und Nukleussignale segmentiert und binarisiert werden. Die daraus resultierenden Bilderpaare von segmentierten binären Membranen und realen Nukleussignalen bilden den Trainingsdatensatz des Transformationsmodells. Da hier gepaarte Daten vorliegen, kann als Modell ein conditional GAN (cGAN) genutzt werden (siehe Abschnitt 1.4.3). Nach dem Training lassen sich aus den BioSim-Membranen passende realistische Nukleusaufnahmen generieren.

### **Reale Membranen zu Nuklei mit gepaarten Daten:**

**RealMem2NucGAN** Generative Modelle sind in der Lage, auf Basis von realen Membranaufnahmen Nukleussignale zu generieren. Allerdings verhindert die Domänenlücke zwischen realen Membranaufnahmen und den zur Verfügung stehenden BioSim-Membranen eine direkte Anwendung eines solchen Modells. Der vorherige Ansatz erfordert eine Anpassung der Trainingsdaten in Form der Segmentierung von Membransignalen. Die Segmentierung von Membransignalen stellt aufgrund der dünnen Strukturen allerdings eine besondere Herausforderung dar. Daher kommt es oft zu Fehlern in der Segmentierung, welche sich negativ auf das Training des Transformationsmodells auswirken. Daher wird eine weitere Methode entwickelt, welche keine vorherige Segmentierung von realen Daten erfordert, sondern eine Anpassung der Anwendungsdaten in Form der BioSim-Membranen an die realen Trainingsdaten vornimmt. Für die Anpassung werden drei Verfahren untersucht (siehe Abbildung 2.6, türkis, grün und lila Pfade):

1. Bei dem ersten Ansatz *SimOptiGAN Aufnahmesimulation* (türkiser Pfad) wird der Teil „Aufnahmesimulation“ der SimOptiGAN-Pipeline genutzt,

um die optischen Effekte des Mikroskopes zu modellieren. In der Simulation werden die Punktspreizfunktion, der Helligkeitsabfall und die Rauschanteile modelliert und auf das Bild angewendet. Daraus ergibt sich ein Bild (*Simulierte Membran*, türkis umrandet), was aufgrund von weicheren Übergängen und vorhandenem Rauschen besser zu realen Membranaufnahmen passt.

2. In dem zweiten Verfahren *SimOptiGAN Optimierung* (grüner Pfad) wird der Teil „Optimierung“ der SimOptiGAN-Pipeline genutzt, um das simulierte Bild *Simulierte Membran*, mittels eines Deep-Learning-Netzwerks zu optimieren. Dazu wird das CycleGAN der SimOptiGAN-Optimierung verwendet, welches mittels ungepaarter Daten von simulierten Nuklei und realer Nuklei trainiert wurde. Da sich die Daten stark von den ursprünglichen Trainingsdaten unterscheiden, ist nicht garantiert, dass diese DL-Optimierung zufriedenstellende Ergebnisse liefert. Allerdings wird durch das Modell keine komplexe Transformation zwischen verschiedenen Strukturen, sondern eher eine Optimierung der Texturen durchgeführt, weshalb davon auszugehen ist, dass die Optimierung auch für Membrandaten funktioniert. Das resultierende Bild *Optimierte Membran* ist mit einer grünen Umrandung dargestellt.
3. Bei dem dritten Verfahren wird das bereits beschriebene Transformationsmodell zur Überführung der BioSim-Membranen in reale Aufnahmen verwendet. Das trainierte Modell *BioSim-Membran*  $\rightarrow$  *Reale Membran* (lila Pfad) wird genutzt, um die BioSim-Membranen in realistische synthetische Membranen zu transformieren (lila umrandetes Bild). Bei einem qualitativen Vergleich wird mit diesem Ansatz die größte Ähnlichkeit zu den realen Membranen erreicht, allerdings können durch das GAN Strukturen verändert, weggelassen oder hinzugefügt werden.

Allen drei Verfahren ist gemein, dass die vorverarbeiteten BioSim-Membranen anschließend mit dem Modell *Reale Membranen*  $\rightarrow$  *Nuklei* in Nukleusaufnahmen transformiert werden. Dieses auf einem cGAN-basierende Transformationsmodell wird zuvor mit realen gepaarten Aufnahmen von Membranen und Nuklei

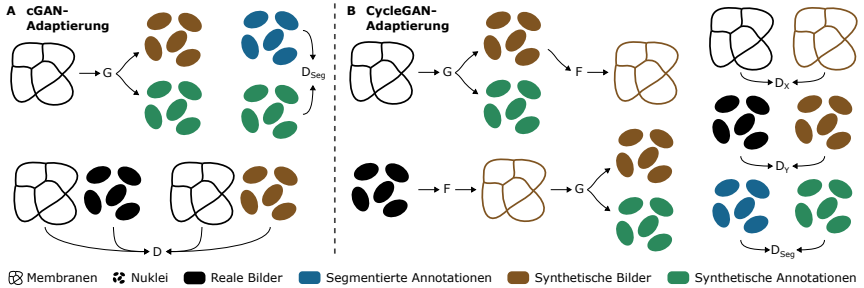
trainiert. Zusätzlich wird noch ein vierter, naiver Ansatz getestet, bei dem keine Vorverarbeitung der BioSim-Membranen vor der Transformation durchgeführt wird (orange dargestellt).

### **BioSim-Membranen zu Nuklei mit ungepaarten Daten:**

**Mem2NucGAN-U** Während die Transformationsmodelle der beiden vorherigen Ansätze auf echten Daten trainiert werden und daher morphologische Korrelationen zwischen Membran und Kernen erlernen können, handelt es sich um komplexe Verfahren. Das für Mem2NucGAN-P erforderliche Segmentierungsmodell und sein Training stellen potenzielle Fehlerquellen dar. Weiterhin können Unterschiede zwischen segmentierten Membranen, die während des Trainings verwendet werden, und BioSim-Membranen, die während der Inferenz verwendet werden, die Qualität der generierten Kernsignale verringern. Daher wird eine weitere Methode getestet, die auf einem CycleGAN-basiert und daher keine gepaarten Daten für das Training erfordert (siehe Abbildung 2.7c). Damit kann die im vorherigen Ansatz benötigte Segmentierung realer Membransignale entfallen. Stattdessen werden als Trainingsdaten direkt biophysikalisch simulierte binäre Membranen in Kombination mit realen Nukleussignalen für das Training des Transformationsmodells genutzt. Damit existieren keine Unterschiede zwischen den Trainings- und den späteren Anwendungsdaten. Im Vergleich zu dem vorherigen Ansatz ist dieser deutlich weniger komplex, da keine zusätzlichen Modelle für das Training erforderlich sind.

### **DualGenGAN: Einführung einer neuen GAN-Struktur für die zusätzliche Generierung von Annotationsdaten**

Die drei vorgestellten Methoden nutzen GANs, um auf Basis von BioSim-Instanzen bzw. den daraus abgeleiteten BioSim-Membranen synthetische Nukleusaufnahmen zu generieren. Für das Training von Segmentierungsmodellen werden allerdings neben den synthetischen Nukleusaufnahmen auch die zugehörigen Annotationen benötigt. Keine der bekannten GAN-Trainingsstrukturen ermöglicht neben der Erstellung der Bildinformationen auch die Generierung von dazugehörigen Annotationen. Um diese



**Abbildung 2.9:** Überblick der eingeführten Trainingsstruktur *DualGenGAN* für die zusätzliche Generierung von Annotationsdaten. **A** und **B** zeigen die angepassten Trainingsstrukturen für ein cGAN bzw. CycleGAN.  $G$  und  $F$  repräsentieren Generatormodelle, während  $D$  Diskriminatoremodelle darstellt. Die Indizes geben die Richtung der Transformation (Generatoren) bzw. die Art oder Domäne der zu bewertenden Daten an (Diskriminatoren). Beide Methoden verwenden einen extra Diskriminator  $D_{Seg}$ , um zusätzliche Annotationen zu erzeugen. Übernommen und übersetzt aus [177].

Einschränkung zu umgehen, greifen bestehende Ansätze zur Trainingsdatensynthese [140, 144, 148, 182] darauf zurück, zuerst synthetische Annotationen zu generieren, beispielsweise durch zufällige Ellipsenplatzierung, und diese anschließend mit GANs in Bilddaten zu transformieren. Dieser Ansatz ist jedoch für die vorliegende Aufgabe nicht anwendbar, da hier zwischen zwei verschiedenen Bildmodalitäten (Membran – Nukleus) transformiert wird, und hierbei bloß Membrannotationen über biophysikalisch simulierte Zellgrenzen zur Verfügung stehen. Daher wird eine GAN-Struktur benötigt, welche neben den synthetischen Nukleussignalen auch direkt passende Annotationen generiert.

Nach aktuellem Kenntnisstand ermöglicht kein bestehendes GAN-Trainingsverfahren die zusätzliche Generierung entsprechender Annotationen. Daher wird mit *DualGenGAN* eine neuartige GAN-Struktur vorgestellt, die es erstmals ermöglicht, Nukleusaufnahmen zusammen mit passenden Annotationen zu generieren. Die neue Struktur kann sowohl mit der cGAN als auch mit der CycleGAN Struktur kombiniert werden (siehe Abbildung 2.9).

Zur Generierung der Annotationsinformationen wird das cGAN so parametrisiert, dass ein Bild mit zwei Ausgabekanälen generiert wird. Der erste Ausgabekanal des Generators  $G$  wird für die Ausgabe des Nukleus-Bildes und der zweite für

die dazugehörige binäre Annotationsmaske verwendet (vgl. Abbildung 2.9a). Der Generator  $G$  bildet also die Bedingung  $x$  und einen Rauschvektor  $z$  auf ein synthetisches Bild  $\hat{y}$  und dessen Annotation  $\hat{v}$  ab:  $G : \{x, z\} \rightarrow \{\hat{y}, \hat{v}\}$ . Da hier die Auftrennung in die beiden Kanäle erst in der letzten Schicht des Generators erfolgt und somit für beide Kanäle auf die gleichen Featuremaps zurückgegriffen wird, wird die Übereinstimmung zwischen den Kanälen begünstigt.

Bei dem regulären Training des cGANs werden gepaarte Bilder verwendet, um die Konditionierung zu ermöglichen. So wird dem Diskriminator  $D$  neben dem Bild der Nukleus-Domäne auch das dazugehörige Bild der Membran-Domäne als Konditionierung bereitgestellt. Damit wird gewährleistet, dass ein generiertes Nukleus-Bild auch konsistent zu dem Ausgangsbild der Membran-Domäne ist. Während des Trainings erhält der Diskriminator  $D$  gefälschte Proben, die aus der Bedingung  $x$  (den Membransignalen) und dem ersten Kanal der Generator-Ausgabe  $\hat{y}$  (den Kernsignalen) bestehen, sowie echte Proben, die aus der Bedingung  $x$  (den Membransignalen) und echten Kernsignalen  $y$  bestehen.

Während der Standard-Diskriminator  $D$  lediglich die Qualität des Nukleus-Kanals bewertet, wird eine zusätzliche Komponente benötigt, um die Qualität der generierten binären Annotationen zu beurteilen und die Gewichte des Generators  $G$  entsprechend zu optimieren. Ohne eine explizite Bewertung der generierten Annotationen wird der Generator nicht dazu angeregt, realistische und plausible Annotationen zu erstellen. Daher wird das herkömmliche Trainingsverfahren um einen zusätzlichen Diskriminator  $D_{\text{Seg}}$  erweitert, der speziell darauf abzielt, die Qualität der erzeugten Annotationen zu bewerten. Das Ziel des Diskriminators  $D_{\text{Seg}}$  ist es, zwischen echten ( $v$ ) und gefälschten ( $\hat{v}$ ) Annotationen zu unterscheiden. Ein wesentliches Merkmal ist, dass  $D_{\text{Seg}}$  im Gegensatz zum regulären Diskriminator  $D$  keine zusätzliche Konditionierung erhält. Dies ermöglicht die Verwendung beliebiger, nicht gepaarter binärer Annotationsmasken für das Training von  $D_{\text{Seg}}$  und verringert so den Bedarf an hochwertigen Annotationsmasken in zwei wesentlichen Aspekten: Erstens sind einzelne falsch-negative Segmentierungsfehler ohne die korrespondierenden Nukleussignale wenig auffällig. Damit wirkt die Segmentierung nicht fehlerhaft, wenn ein Nukleus nicht



erkannt wurde. Zweitens beseitigt die Verwendung einer binären Annotationsmaske Unter- und Übersegmentierungsfehler. Beides trägt dazu bei, dass Fehler in den Echtwelt-Annotationsmasken den Trainingsprozess weniger negativ beeinflussen. Zur Generierung von Segmentierungsmasken eignen sich daher einfache lokale Grenzwertverfahren oder aber auch das vortrainierte *nuclei*-Modell von Cellpose.

Die ursprüngliche Optimierungsfunktion des cGAN (vgl. Gleichung A.4) wird um einen zusätzlichen GAN-Loss  $\mathcal{L}_{GAN_S}(G, D_{\text{Seg}})$  ergänzt, welcher sich aus dem eingeführten Diskriminator  $D_{\text{Seg}}$  ergibt:

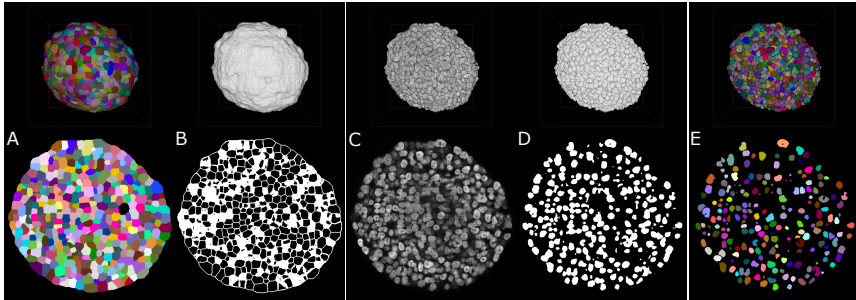
$$\min_G \max_{D, D_{\text{Seg}}} \mathcal{L}_{cGAN}(G, D) + \mathcal{L}_{GAN_S}(G, D_{\text{Seg}}) + \lambda_{L1} * \mathcal{L}_{L1}(G), \text{ mit} \quad (2.4)$$

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y} [\log D(y, x)] + \mathbb{E}_{x,z} [\log(1 - D(\hat{y}, x))], \quad (2.5)$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z} [\|y - \hat{y}\|_1] \text{ und} \quad (2.6)$$

$$\mathcal{L}_{GAN_S}(G, D_{\text{Seg}}) = \mathbb{E}_v [\log D_{\text{Seg}}(v)] + \mathbb{E}_{x,z} [\log(1 - D_{\text{Seg}}(\hat{v}))]. \quad (2.7)$$

Da der Generator  $G$  nur binäre Nukleusannotationen generiert, ist ein Vorverarbeitungsschritt erforderlich, um die für das Training der Segmentierung erforderlichen Instanzannotationen daraus abzuleiten (vgl. Abbildung 2.10). Dabei sind den Nukleusannotationen, welche sich innerhalb der BioSim-Instanzen befinden, die gleichen IDs zuzuordnen wie den BioSim-Instanzen. Weiterhin sind die prognostizierten Annotationen nicht frei von Artefakten. So können Annotationen vorhergesagt werden, an denen sich keine Zellen befinden. Daher sind zudem Maßnahmen zu ergreifen, um Artefakte aus den Annotationen zu entfernen. Der mit den BioSim-Instanzen bzw. den daraus abgeleiteten BioSim-Membranen (Abbildung 2.10a bzw. b) trainierte Generator generiert synthetische Nukleussignale und dazugehörige Annotationen (Abbildung 2.10c und d). In einem ersten Schritt wird aus den generierten Nukleusannotationen mittels eines Schwellwertverfahrens eine binäre Annotationsmaske erstellt (vgl. Abbildung 2.10d). Anschließend wird das binäre Membransignal, das als Eingabe für den Generator verwendet



**Abbildung 2.10:** Generierungsprozess von synthetischen Nukleusaufnahmen mit zugehörigen Instanzannotationen. Die obere Reihe zeigt 3D-Visualisierungen, während die untere Reihe 2D-Schnitte der in der oberen Reihe dargestellten Strukturen präsentiert. **A** zeigt BioSim-Instanzen, aus denen die binären BioSim-Membranen (**B**) generiert werden. Auf Basis von **B** wird mit einem GAN die synthetische Nukleusaufnahme (**C**) und die dazugehörige binäre Annotationsmaske (**D**) generiert. Durch einen Nachbearbeitungsschritt wird auf Basis der Informationen aus **A** und **B** Instanzannotationen (**E**) generiert. Übernommen aus [177].

wird, von der binären Annotationsmaske abgezogen, um sich berührende Instanzen zu trennen. Eine binäre morphologische Öffnen-Operation sorgt anschließend für die Entfernung potenzieller Artefakte. Um die binären Annotationen nun in Instanzannotationen zu überführen, erfolgt eine Zuordnung der Voxel auf Basis der BioSim-Instanzen: Jedem Vordergrund-Voxel in der binären Annotationsmaske wird anhand seiner Position die entsprechende Instanz der BioSim-Instanzen zugewiesen. Voxel, für die keine Instanzen in den BioSim-Instanzen vorhanden sind, werden als Hintergrund definiert und damit weitere Artefakte in den Annotationen beseitigt. Damit ergeben sich aus den generierten binären Annotationen die für die Segmentierung benötigten Instanzannotationen (siehe Abbildung 2.10e).

Die CycleGAN-Struktur wird für die Generierung von zusätzlichen Annotationen in ähnlicher Weise angepasst (vgl. Abbildung 2.9b). Der Generator  $G$  wird darauf trainiert, ein Bild mit zwei Kanälen auf Basis von Membransignalen zu generieren:  $G : x \rightarrow \{\hat{y}, \hat{v}\}$ . Der erste Kanal  $\hat{y}$  wird auch hier für die Ausgabe des Nukleussignals verwendet, während der zweite Kanal  $\hat{v}$  für die Ausgabe der dazugehörigen binären Annotationen genutzt wird. Da für realen Nukleusaufnahmen keine passende Segmentierung verfügbar ist, erfolgt die Rücktransformation in

Membransignale  $\tilde{x}$  mit dem Generator  $F$  nur auf Basis des ersten Kanals  $\hat{y}$ , also des Nukleussignals. Ebenso erhält der Diskriminator  $D_Y$  nur Nukleusaufnahmen, um zwischen echten  $y$  und gefälschten  $\hat{y}$  Daten zu unterscheiden. Die Aufgabe des Diskriminators  $D_X$  ist es, zwischen realen  $x$  und synthetisch generierten  $\hat{x}$  Membransignalen zu unterscheiden. In der entgegengesetzten Transformationsrichtung wird der Generator  $F$  eingesetzt, um reale Nukleussignale  $y$  in synthetische Membransignale  $\hat{x}$  zu übersetzen  $F : y \rightarrow \hat{x}$ . Bei der Rücktransformation wird aus den synthetischen Membransignalen mit  $G$  wiederum synthetische Nukleussignale  $\tilde{y}$  und deren Annotationen  $\tilde{v}$  generiert. Die Konsistenz zwischen den Domänen wird durch die sogenannte Cycle-Consistency erreicht, bei der Differenzen zwischen  $x$  und  $\tilde{x}$  bzw.  $y$  und  $\tilde{y}$  bestraft werden. Um die zusätzlich generierten Annotationen zu verbessern, wird wieder ein zusätzlicher Diskriminator  $D_{Seg}$  eingeführt, welcher zwischen echten  $v$  und synthetischen  $\hat{v}$  Annotationen unterscheidet. Da auch in diesem Fall lediglich ungepaarte binäre Annotationsmasken für das Training erforderlich sind, ergeben sich dieselben Vorteile bezüglich eventueller Fehler in den Annotationen wie beim angepassten cGAN.

Die ursprüngliche Optimierungsfunktion des CycleGAN, bestehend aus dem GAN-Loss  $\mathcal{L}_{GAN_X}$  bzw.  $\mathcal{L}_{GAN_Y}$  und dem Cycle-Consistency-Loss  $\mathcal{L}_{cyc}$ , wird mit dem aus dem Diskriminator  $D_{Seg}$  resultierendem GAN-Loss  $\mathcal{L}_{GAN_S}$  ergänzt:

$$\min_{G,F} \max_{D_X,D_Y,D_{Seg}} \mathcal{L}_{GAN_X}(F, D_X) + \mathcal{L}_{GAN_Y}(G, D_Y) + \mathcal{L}_{GAN_S}(G, D_{Seg}) + \lambda_{cyc} \mathcal{L}_{cyc}(G, F), \text{ mit} \quad (2.8)$$

$$\mathcal{L}_{GAN_X}(F, D_X) = \mathbb{E}_x [\log D_X(x)] + \mathbb{E}_y [\log(1 - D_X(\hat{x}))], \quad (2.9)$$

$$\mathcal{L}_{GAN_Y}(G, D_Y) = \mathbb{E}_y [\log D_Y(y)] + \mathbb{E}_x [\log(1 - D_Y(\hat{y}))], \quad (2.10)$$

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_x [\|x - \tilde{x}\|_1] + \mathbb{E}_y [\|y - \tilde{y}\|_1] \text{ und} \quad (2.11)$$

$$\mathcal{L}_{GAN_S}(G, D_{Seg}) = \mathbb{E}_v [\log D_{Seg}(v)] + \mathbb{E}_x [\log(1 - D_{Seg}(\hat{v}))]. \quad (2.12)$$

Neben der hier vorgestellten Anwendung zur Generierung von Nukleussignalen samt zugehörigen Annotationen kann die angepasste GAN-Modellstruktur

auch für andere Markerkombinationen eingesetzt werden. So kann beispielsweise anhand der Nukleussignale der Proliferationsmarker Ki-67 mit zugehörigen Annotationen vorhergesagt werden. Weiterhin bietet die Methode auch Potenzial für den Einsatz in anderen Anwendungsfeldern. Im medizinischen Kontext werden etwa GAN-Modelle eingesetzt, um MRT- und CT-Aufnahmen ineinander zu transformieren. Mit der angepassten Modellstruktur können diese Modalitäten jedoch nicht nur ineinander überführt, sondern zugleich die entsprechenden Annotationen generiert werden.

### 2.2.4 Bewertung von synthetischen Trainingsdaten

#### Übersicht

Um die Qualität der synthetischen Daten bzw. deren Ähnlichkeit mit den realen Vorlagen zu untersuchen, werden quantitative Bewertungsmaße benötigt. Die Bewertung von synthetischen Aufnahmen stellt eine nicht triviale Aufgabe dar [183]. Eine Schwierigkeit besteht in der nicht vorhandenen Übereinstimmung von Objektpositionen zwischen synthetischen und realen Daten. Somit eignen sich die aus der Informationstheorie bekannten Maße wie der Root Mean Squared Error (RMSE), das Structural Similarity Index Measure (SSIM) oder die Kreuzkorrelation nicht für die Bewertung der synthetischen Trainingsdaten.

Durch die fehlende Übereinstimmung zwischen den Daten sind daher Maße interessant, bei denen die Bilder nicht auf Pixelebene verglichen werden. So existieren für die Bewertung synthetischer Aufnahmen von natürlichen Szenen bereits Metriken, bei denen trainierte Deep-Learning-Modelle genutzt werden, um den Vergleich auf Basis damit extrahierter Merkmalskarten durchzuführen. Dazu zählt unter anderem der Inception-Score (IS) [184], die Fréchet-Inception-Distance (FID) [185] oder die Kernel-Inception-Distance (KID) [186]. Standardmäßig wird als Grundlage bei allen drei Maßen das Klassifikationsnetzwerk Inception v3 [187] verwendet, welches auf 2D-Aufnahmen natürlicher Szenen (z. B. Bilder von Hunden, Fahrrädern oder Autos) trainiert wurde. Die Grundlage der

Bewertung ist, dass ähnliche Aufnahmen, welche etwa das gleiche Objekt zeigen, ähnliche Antworten innerhalb des Netzwerks erzeugen. Je geringer die Distanz zwischen den vom Netzwerk extrahierten Merkmalen ist, desto ähnlicher sind die Bilder.

Die Verwendung eines zweidimensionalen Inception Netzwerkes als Grundlage für die Distanzbestimmung stellt allerdings eine Limitierung bezüglich der verwendbaren Daten-Dimensionalität dar. 3D-Daten können lediglich schichtweise verglichen werden, wodurch globale Eigenschaften, wie ein Helligkeitsverlauf, der orthogonal zu den Ebenen verläuft, nicht in die Bewertung einfließen. Die Umstellung der KID-Metrik auf ein dreidimensionales Netzwerk würde allerdings das Training eines 3D-Bildklassifikators erfordern, was wiederum einen großen, vielfältigen und annotierten Trainingsdatensatz voraussetzt. Da ein solcher Datensatz nicht verfügbar ist und dessen Erstellung mit einem immensen Aufwand verbunden wäre, wird in dieser Arbeit auf das Training eines eigenen Netzwerks verzichtet. Stattdessen wird eine ebenenweise Prozessierung angewendet, bei der drei unterschiedliche Ebenenausrichtungen genutzt werden, um globale Eigenschaften zu berücksichtigen. Die Methodik wird im folgenden Abschnitt detailliert beschrieben.

Eine weitere wichtige Bewertungsmethode, insbesondere in Hinblick auf den späteren Hauptanwendungsfall der Trainingsdatensynthese, ist die Analyse der synthetischen Daten basierend auf der erzielbaren Segmentierungsgüte bei realen Daten. Diese Art der Bewertung ist jedoch mit erheblichem Aufwand verbunden und widerspricht dem Ziel der Trainingsdatensynthese. So wird zur Bestimmung der Leistung der Segmentierungsmodelle eine Ground-Truth der realen Daten benötigt, deren manuelle Erstellung durch die Nutzung synthetischer Daten eigentlich vermieden werden soll. Die Erstellung einer solchen Ground-Truth ist insbesondere für 3D-Daten, wie in Abschnitt 1.3.1 beschrieben, komplex und sehr zeitintensiv. Weiterhin sind für jede zu bewertende Datensynthesemethode separate Segmentierungsmodelle zu trainieren. Um statistisch belastbare Aussagen zu treffen, sind aufgrund der Trainingsvariabilität zudem mehrere Wiederholungen pro Modelltraining erforderlich. Im Gegenzug zu dem Aufwand liefert die Bewertung der synthetischen Daten anhand der Segmentierungsgüte aussagekräftige

Ergebnisse über die Eignung der Daten für das Training von Segmentierungsmodellen. Im Gegensatz dazu bieten die oben beschriebenen Bewertungsmaße, die allein die Ähnlichkeit zwischen realen und synthetischen Daten untersuchen, keinen direkten Hinweis auf die tatsächlich erzielbare Segmentierungsleistung.

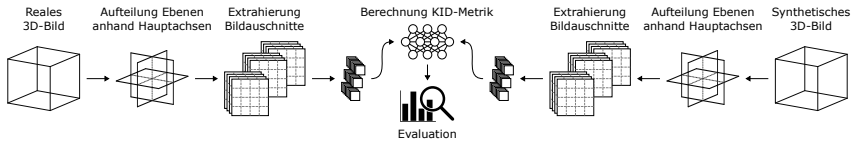
In Fällen, bei denen beispielsweise neu eingeführte Synthesemethoden mit etablierten Verfahren oder realen Trainingsdaten zu vergleichen sind, eignet sich die direkte Bewertung anhand der Segmentierungsgüte. Ist hingegen lediglich der optimale synthetische Datensatz aus einem Pool vorhandener Datensätze zu bestimmen, sind ähnlichkeitsbasierte Metriken wie die KID meist ausreichend. Im Rahmen der vorliegenden Arbeit wird auch eine Bewertung auf Basis der Segmentierungsgüte durchgeführt. Die Methodik hierzu wird im entsprechenden Abschnitt beschrieben.

### Kernel-Inception-Distance (KID) für 3D-Daten

Wie eingangs beschrieben verhindert die Verwendung des vortrainierten zweidimensionalen Inception-Netzwerkes zur Berechnung der Distanzen eine direkte Anwendung auf 3D-Daten. Die Prozessierung von 3D-Daten kann allerdings ebenenweise erfolgen. Für einen Vergleich von nicht nur der  $x - y$  Ebene werden die Ebenen von allen drei Hauptachsen miteinander verglichen (vgl. Abbildung 2.11). Da das Inception-Netzwerk für die Verarbeitung von  $128 \times 128 \text{ px}^2$  großen Bildern ausgelegt ist, werden die Bilder für jede Hauptachse  $(x, y, z)$  in  $128 \times 128 \text{ px}^2$  große 2D-Schnitte aufgeteilt. Diese werden anschließend einzeln prozessiert und die Ergebnisse für jede Hauptachsrichtung gemittelt. Zuletzt erfolgt die Bestimmung des finalen Ergebnisses über die Mittelung der Hauptachsen-Ergebnisse.<sup>4</sup>

---

<sup>4</sup> Die Prozessierung der Ausschnitte wird mit der KID-Klasse des TorchMetrics-Pakets durchgeführt (<https://lightning.ai/docs/torchmetrics/v1.6.0/>).



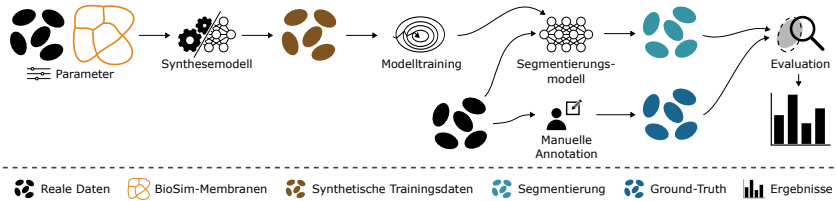
**Abbildung 2.11:** Berechnung der KID-Metrik zwischen realen und synthetischen 3D-Aufnahmen anhand von Bildausschnitten orthogonaler Ebenen.

## Bewertung synthetischer Trainingsdaten auf Basis der Segmentierungsgüte

Ein zentrales Ziel der Trainingsdatensynthese stellt das Training von Segmentierungsmodellen dar, welche anschließend auf reale Daten angewendet werden. Die Leistung von Segmentierungsmodellen, die mit synthetischen Daten trainiert wurden, eignet sich daher als ideale Bewertungsmetrik für die Qualität der synthetischen Daten. In dieser Arbeit wird für den Vergleich das 3D-StarDist [71] Segmentierungsmodell<sup>5</sup> genutzt. Die Bewertung der Segmentierungsleistung erfolgt über die zwei Maße SEG (Segmentierung) und DET (Detektion) für deren Berechnung ein Kommandozeileninterface von den Organisatoren der Cell-Tracking-Challenge [41] bereitgestellt wird. Das auf dem Jaccard-Koeffizienten basierende SEG-Maß bewertet die Genauigkeit der Segmentierung, indem sie die Übereinstimmung zwischen den vorhergesagten und den tatsächlichen Segmentierungsmasken quantifiziert. Im Gegensatz dazu konzentriert sich die DET-Metrik auf die Bewertung der Erkennung von Objekten, indem sie Faktoren wie falsch-positive, falsch-negative Erkennungen und Untersegmentierungen von Objekten berücksichtigt. Beide Metriken können Werte von null bis eins annehmen, wobei null das schlechteste und eins das beste Ergebnis darstellt.

<sup>5</sup> Mit dem Begriff Modell ist nur die Segmentierungsmethode gemeint. Vor der Verwendung des Modells ist ein Training anhand von annotierten Trainingsdaten erforderlich. Aus Mangel an öffentlichen 3D-Trainingsdatensätzen existiert keine vortrainierte Version des Modells.

Der Ablauf der Bewertung ist in Abbildung 2.12 dargestellt. Mit den zu vergleichenden *Synthesemodellen* zur Datensynthese werden zuerst synthetische Trainingsdaten generiert. Je nach Methode werden dazu reale Daten für das Training der GAN-Modelle, BioSim-Membranen für die Verbesserung der Zellanordnung und Parameter für die Modelle benötigt. Für jede Synthesemethode erfolgt mit den generierten Trainingsdaten ein *Modell-Training* des 3D-StarDist-Segmentierungsmodells. Die *Evaluation* der Segmentierungen erfolgt anschließend auf einem realen Testdatensatz, für den mittels *manueller Annotation* eine Ground-Truth-Segmentierung zur Verfügung steht.



**Abbildung 2.12:** Ablauf der Bewertung synthetischer Trainingsdaten auf Basis der Segmentierungsgüte auf realen Daten. Zunächst werden reale Vorbilder sowie gegebenenfalls benötigte BioSim-Membranen und Parametrierungen genutzt, um mit den zu bewertenden Synthesemodellen synthetische Trainingsdaten zu generieren. Anschließend wird mit diesen jeweils ein Segmentierungsmodell trainiert, das dann zur Segmentierung eines realen Datensatzes verwendet wird. Die Segmentierungsgüte der Modelle wird schließlich anhand einer durch händische Annotation generierten Ground-Truth evaluiert und die Ergebnisse der verschiedenen Synthesemodelle miteinander verglichen.



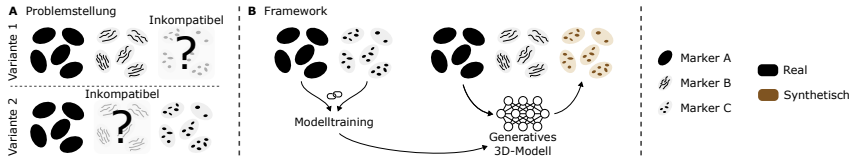
## 2.3 Synthese von Bildmodalitäten

Teile der folgenden Sektion basieren auf der im Rahmen der vorliegenden Arbeit veröffentlichten Publikation von Bruch et al. [188].

### 2.3.1 Übersicht

In der Fluoreszenzmikroskopie werden Proben, die mit fluoreszierenden Markern gefärbt sind, durch Licht einer spezifischen Wellenlänge angeregt, wodurch sie Licht mit einer längeren Wellenlänge emittieren. Solche Marker binden in Zellkulturen beispielsweise an Proteinstrukturen, wodurch Nuklei oder Zellmembranen sichtbar gemacht werden können. Da fluoreszierende Substanzen nicht nur auf einer einzigen Wellenlänge antworten, sondern ein breites Spektrum emittieren, ist die Anzahl von Markern und deren Kombinationsmöglichkeiten zur gleichzeitigen Abbildung mehrerer Modalitäten begrenzt. So führen überlappende Antwort-Spektren zu nicht eindeutigen Signalen, wodurch keine aussagekräftige Analyse mehr möglich ist (siehe Abbildung 2.13a). Diese Limitierungen sorgen in der Praxis dafür, dass bestimmte Markerkombinationen nicht genutzt, oder aber für die inkompatiblen Marker zusätzliche Zellkulturen erstellt werden. Beide Ansätze sind suboptimale Kompromisse: Im ersten Fall wird die mögliche Analyse eingeschränkt. Im zweiten Fall, bei dem die Marker auf verschiedene Zellkulturen aufgeteilt sind, ist eine Kolo-kalisierung der Markersignale nicht mehr möglich, da die Proben unterschiedlich sind. In beiden Fällen kommt es somit zu einem Informationsverlust.

Zwischen einigen Markern existieren Zusammenhänge, welche Rückschlüsse auf die Signale eines Markers zulassen. Beispielsweise besitzt die Form der Zelle einen Einfluss auf die Form des darin enthaltenen Zellkerns. Um die Limitierungen bezüglich Anzahl und Kombinationsmöglichkeiten von Markern zu umgehen und damit eine vollständige Auswertung zu ermöglichen, führt diese Arbeit ein



**Abbildung 2.13:** **A** Aufgrund Überlagerungen in den Antwortspektren und den damit verbundenen uneindeutigen Signalen sind verschiedene Marker inkompatibel zueinander. **B** Mit dem in dieser Arbeit vorgestellten Framework lassen sich auf Grundlage eines Basismarkers weitere Marker vorhersagen. Dazu werden gepaarte Aufnahmen des Basis- und Zielmarkers erstellt, was das Training eines generativen Modells ermöglicht. In der Anwendungsphase kann mittels des trainierten Modells anhand des Basismarkers der inkompatible Zielmarker synthetisch generiert werden.

Framework ein, um die Signale von fluoreszierenden Markern auf Basis von anderen Markern zu synthetisieren (vgl. Abbildung 2.13). Im Folgenden wird der Marker, auf dessen Grundlage die Prognose erfolgt, als Basismarker und der Marker, welcher vorherzusagen ist, als Zielmarker bezeichnet. Zur Abgrenzung der Methoden zur Trainingsdatensynthese wird in dieser Arbeit von der Synthese von Bildmodalitäten gesprochen.

Wie im Stand der Technik beschrieben, wurde die Synthese von Bildmodalitäten bereits für Hellfeld-Daten durchgeführt [171, 173]. Im Unterschied dazu werden in dieser Arbeit Methoden für die Synthese von Bildmodalitäten von 3D-Fluoreszenzmikroskopieaufnahmen vorgestellt. Die Anwendungsfälle unterscheiden sich drastisch, da bei Hellfeld-Daten alle Strukturen zu erkennen sind, welche für die sichtbaren Frequenzen nicht vollständig durchsichtig sind. Es handelt sich also mehr um eine Filterung der richtigen Signale. Bei der Fluoreszenzmikroskopie sind die Signale des Zielmarkers nicht vorhanden, weshalb eine Prognose notwendig ist. Weiterhin erlaubt die Hellfeld-Technik nur die Aufnahme geringer Probendicken, während bei der Fluoreszenzmikroskopie die mögliche Aufnahmetiefe von der Effektivität des Clearings bestimmt wird und damit deutlich größere Probendicken erlaubt. Allerdings treten damit auch Effekte wie eine variierende Helligkeit auf, was eine Prognose erschwert.

### 2.3.2 Neues Framework zur Synthese von Bildmodalitäten

Um die beschriebenen Einschränkungen der limitierten Anzahl an Markern und der Inkompatibilität bestimmter Markerkombinationen zu umgehen, wird in dieser Arbeit neues ein Framework, *IMS* (Imaging Modality Synthesis), zur erstmaligen Synthese von Bildmodalitäten großer 3D-Zellkulturen entwickelt (vgl. Abbildung 2.13). Das Ziel ist es, anhand eines Basismarkers (*Marker A*) die Signale eines Zielmarkers (*Marker C*) vorherzusagen, um so Einschränkungen bezüglich inkompatibler Markerkombinationen (*Marker B* und *Marker C*) zu umgehen (vgl. Abbildung 2.13). Das Ziel ist es, die Vorhersage für große 3D-Aufnahmen zu ermöglichen, die sowohl innerhalb einer Aufnahme als auch zwischen verschiedenen Aufnahmen heterogene Bedingungen wie Helligkeits- und Qualitätsniveaus sowie eine anisotrope Auflösung aufweisen.

Für die Prognose des Zielmarkers wird ein *generatives 3D-Modell* genutzt. Für das Training des Modells sind gepaarte Aufnahmen des Ziel- und Basismarkers nötig. Gepaart bedeutet, dass die Signale der gleichen Struktur entspringen, also die Position der Probe identisch ist und sich nur der Fluoreszenz-Kanal unterscheidet. Um selbst bei einem Drift der Probe während einer Aufnahme nur einen geringen Versatz zwischen den Markern zu erhalten, ist der zeitliche Abstand zwischen den Kanalwechseln so kurz wie möglich zu halten. So ist es vorteilhafter, zwischen jeder Bildebene den Kanal zu wechseln, als die Kanäle vollständig, aber nacheinander aufzunehmen.

Durch den erhöhten Speicherbedarf von 3D-Daten ist für das Training und auch für die spätere Anwendung nur eine Prozessierung von Bildausschnitten möglich. Daher werden die Trainingsdaten vor bzw. während des Trainings in  $32 \times 256 \times 256$  px ( $z, y, x$ ) große Ausschnitte aufgeteilt. Aufgrund der meist geringeren  $z$ -Auflösung ist der Bildausschnitt in dieser Richtung kleiner gewählt. Auch in der späteren Anwendung wird das zu prozessierende Bild wieder in Ausschnitte aufgeteilt, welche sich dieses Mal aber überlappen (vgl. Abbildung 2.3c). Die Überlappung ermöglicht es, das gesamte Bild zu verarbeiten und zusätzlich die Randregionen

der Bildausschnitte nach der Prozessierung zu verwerfen<sup>6</sup>. Zu beachten ist, dass durch eine stärkere Überlappung die Anzahl der Bildausschnitte steigt und damit auch die zur Prozessierung benötigte Zeit. In Experimenten erwies sich eine Überlappung von 50% der Kantenlänge als guter Kompromiss zwischen Qualität und Geschwindigkeit. Nach der Aufteilung werden die Bildausschnitte dann einzeln prozessiert und wieder zu einem vollständigen Bild zusammengesetzt (vgl. Abbildung 2.3c). Die Übereinstimmung der Ausschnittgrößen zwischen Training und Anwendung ist wichtig, da sich andernfalls die Bildstatistik ändern kann, was meist mit einer Verschlechterung der Qualität einhergeht.

Da unbekannt ist, welche generativen Netzwerkarchitektur bessere für die Prognose unter der oben genannten Zielstellung geeignet ist, werden zwei bekannte Architekturen miteinander verglichen: In Anlehnung an die Prognosemethode für Hellfeld-Daten wird zum einen eine U-Net basierte Architektur verwendet. Im Gegensatz zu [173] wird allerdings nicht das klassische U-Net verwendet, sondern ein ResUnet [189], was eine Kombination aus U-Net und ResNet darstellt. Die Residual-Blöcke des ResNets helfen unter anderem, die Konvergenz des Trainings zu steigern [190]. Zum anderen wird für die Prognose das GAN-basierte Modell pix2pix [191] getestet, welches für die Bild-zu-Bild-Transformation entwickelt wurde. Der größte Unterschied zwischen den Modellen liegt in dem zusätzlichen Adversarial Loss des pix2pix-Modells, welcher sich aus dem zusätzlichen Diskriminatornetzwerk berechnet. Sowohl in der Literatur [191] als auch bei eigenen Versuchen hat sich gezeigt, dass die generierten Bilder aufgrund des zusätzlichen Adversarial Losses eine höhere Schärfe ausweisen als bei der alleinigen Verwendung eines intensitätsbasierten Distanz-Losses wie MAE oder MSE. Beide Modelle sind ursprünglich für die Verarbeitung von 2D-Bilddaten ausgelegt.

---

<sup>6</sup> Diese Regionen können eine schlechtere Qualität als die innere Region aufweisen oder sogar Artefakte beinhalten. Mitverantwortlich hierfür sind fehlende Informationen bei den Faltungsoperationen innerhalb der Netzwerke. Für Pixel am Rand liegt der Faltungskern teilweise außerhalb des Bildbereiches, weshalb die Faltungsoperation hier nicht direkt durchführbar ist. Die fehlende Information kann aber beispielsweise durch eine Spiegelung am Rand generiert werden, wobei dies zu unnatürlichen geformten Objekten führen kann. Daher werden die Randregionen nach der Prozessierung verworfen.

Daher erfordert die native Verarbeitung von 3D-Daten eine Anpassung sowohl der Netzwerkarchitektur als auch der Implementierung.

Im Folgenden werden beide Methoden im Hinblick auf Trainingsgeschwindigkeit und -stabilität, Modellkapazität, Generalisierungsfähigkeit sowie Bildqualität miteinander verglichen und diskutiert (vgl. Tabelle 2.1).

**Tabelle 2.1:** Vergleich der Vor- und Nachteile des ResUnet im Vergleich zu pix2pix.

Kriterium	ResUnet	pix2pix
Trainingszeit	+	-
Trainingsstabilität	+	-
Modellkapazität	-	+
Generalisierungsfähigkeit	+	-
Bildqualität	-	+

Bezogen auf die Trainingszeit liegt das ResUnet aufgrund des fehlenden Diskriminatormodells und der allgemein geringeren Parameteranzahl im Vorteil. Zudem ermöglicht das ResUnet ein Early-Stopping, bei dem das Training beendet wird, sobald sich der Loss auf den Validierungsdaten nicht weiter verbessert. Die Trainingsdauer beim pix2pix-Modell wird üblicherweise über Erfahrungswerte festgelegt, da der Loss von Generator und Diskriminator nicht konvergiert. Dies führt beim pix2pix-Modell zu längeren Trainingszeiten, da diese zur Sicherheit großzügiger bemessen werden. Das Training des ResUnets verläuft zudem stabiler, da ein klar definiertes Optimierungskriterium vorliegt. Im Gegensatz dazu kann das Training des pix2pix-Ansatzes aufgrund des MiniMax-Spiels zwischen Generator und Diskriminator instabil werden, was sich unter anderem in der ausschließlichen Generierung von schwarzen Bildern äußern kann. Hinsichtlich der Modellkapazität und der damit verbundenen Fähigkeit, komplexe Muster in den Daten zu erlernen, stellt die geringere Parameteranzahl des ResUnets einen Nachteil dar. Gleichzeitig fördert die geringere Parameteranzahl jedoch die Generalisierungsfähigkeit, da das Modell sich weniger auf die Trainingsdaten spezialisieren kann. Vorteile des pix2pix-Modells liegen, aufgrund des Adversarial Losses, hingegen

bei einer höheren Bildqualität, was sich in einer höheren Bildschärfe und in einem größeren Detailreichtum der Prädiktion äußert. Generierte Strukturen weisen so klarere Kanten und Texturschärfen auf, welche den realen Daten ähneln. Dies kann unter anderem wichtig sein, wenn auf Basis der synthetisch generierten Marker weitere Netzwerke, z. B. für die Klassifikation, trainiert werden.

### 2.3.3 Bewertung von synthetischen Markersignalen

#### Qualitätsbewertung

Der Vergleich der Netzwerkarchitekturen und die Bewertung der synthetisch generierten Bilder erfordert objektive Maße, mit denen die Bilder auf Basis der realen Marker bewertet werden können. Da die Prognose der Modelle anhand des Basismarkers durchgeführt wird, liegen gepaarte Daten zwischen realem und generiertem Zielmarker vor, bei denen die Strukturen sich an den gleichen Stellen befinden. Das erlaubt die Verwendung von Maßen, welche die Bilder auf Voxelbasis vergleichen. Für die Bewertung der generierten Markersignale wird daher der Pearson-Korrelationskoeffizient  $r$  genutzt, bei dem die Voxelintensitäten des realen  $z_i$  und des prognostizierten Markers  $\hat{z}_i$  miteinander verglichen werden:

$$r_{z,\hat{z}} = \frac{\sum_{i \in I} (z_i - \bar{z})(\hat{z}_i - \bar{\hat{z}})}{\sqrt{\sum_{i \in I} (z_i - \bar{z})^2 \sum_{i \in I} (\hat{z}_i - \bar{\hat{z}})^2}}. \quad (2.13)$$

Dabei beschreibt  $\bar{z}$  bzw.  $\bar{\hat{z}}$  die mittlere Intensität der Voxel  $i$  aus der Menge aller Voxel eines Bildes ( $I$ ). Die Nutzung des Korrelationskoeffizienten hat den Vorteil, dass dieser unabhängig gegenüber Unterschieden in der mittleren Intensität und dem Kontrast ist und so der Fokus so auf dem Bildinhalt liegt. Eine Betrachtung der absoluten Intensitäten ist für Auswertungen nicht zielführend, da sich diese auch zwischen realen Aufnahmen unterscheidet. Zudem ist eine nachträgliche Anpassung der Helligkeit und des Kontrastes einfach möglich.

Bei Mikroskopaufnahmen ist häufig ein großer Anteil an Hintergrundbereichen vorhanden. Diese Bereiche lassen sich aufgrund der übereinstimmend geringen

Intensitäten im Basis- und Zielkanal leicht vorhersagen. Bei der Betrachtung der Maße entsteht also eine ungewünschte Abhängigkeit zu dem Verhältnis zwischen der Anzahl an Vorder- und Hintergrundvoxeln. Um diese Abhängigkeit zu umgehen, werden Maße benötigt, welche sich nur auf die Vordergrundregionen beziehen. Daher wird das Maß  $r_{z,\hat{z};\text{VG}}$  eingeführt, bei dem der Vergleich der Intensitäten nur für Voxel im Vordergrund ( $I_{\text{VG}}$ ) anstatt des gesamten Bildes ( $I$ ) durchgeführt wird. Die Definition des Vordergrundes ist dabei sorgfältig zu wählen. Wird der Vordergrund anhand des Zielmarkers bestimmt, verhindert dies, dass falsch-positive Signale in die Bewertung einfließen, da diese in den maskierten Bereichen auftreten. Daher werden idealerweise die Vordergrundregionen anhand eines Markers bestimmt, welcher die gesamte Struktur bzw. die Objekte anfärbt, in denen die Signale des Zielmarkers auftreten können. Bei Zellkulturen eignen sich hierzu beispielsweise Nukleusmarker. Die Segmentierung zur Bestimmung des Vordergrundbereiches kann mit einfachen Methoden wie automatischen Grenzwertverfahren (z. B. Verfahren von Otsu) durchgeführt werden, da hier keine Unterscheidung zwischen den Instanzen notwendig ist.

Bei der Reduzierung auf die Vordergrundbereiche ist mit einer Verschlechterung der Metrikergebnisse zu rechnen, da die leicht vorherzusagenden Hintergrundregionen entfallen. Um schwerwiegende Fehler in den Hintergrundbereichen auszuschließen, ist eine Überprüfung mittels des generellen Maßes ( $r_{z,\hat{z}}$ ) weiterhin sinnvoll. Insbesondere wenn das Vordergrundmaß entgegen den Erwartungen bessere Ergebnisse zeigt als das generelle Maß, sind die Prognosen visuell zu überprüfen.

Bei einer späteren Auswertung der Markersignale sind häufig nur die mittleren Intensitäten innerhalb einzelner Objekte von Interesse, um Zellen in Positiv und Negativ einzuteilen. Um für diesen Anwendungsfall ein spezialisiertes Maß zu haben, wird der Korrelationskoeffizient  $r_{o,\hat{o};\text{Obj}}$  zwischen den Objekten eingeführt. Hierbei wird nicht wie zuvor zwischen Intensitäten von einzelnen Voxeln ( $z$ ) verglichen, sondern die mittlere Intensität von ganzen Objekten ( $o$ ). Neben der Relevanz für einfache Klassifikationsaufgaben ergeben sich auch weitere Vorteile. So verliert der Einfluss des Bildrauschens, welcher von den Modellen nicht

vorhergesagt werden kann, an Bedeutung, da gemittelte Intensitäten betrachtet werden.

Für die Bestimmung der einzelnen Objekte ist eine Instanz-Segmentierung notwendig. Liegen Daten mit dicht beieinanderliegenden Objekten vor, erfordert dies die Verwendung von aufwendigeren Segmentierungsverfahren, um die einzelnen Instanzen aufzulösen. Allerdings erfordert die weitere Analyse der Objekte ohnehin in den meisten Fällen eine Instanz-Segmentierung. Anhand der Segmentierung wird für jedes Objekt die mittlere Intensität des Zielmarkers mit der des prognostizierten Markers verglichen. Im Vergleich zu dem Vordergrundmaß ist mit einer leichten Verbesserung der Ergebnisse zu rechnen, da nun nicht mehr einzelne Pixel, sondern mittlere Intensitäten verglichen werden. Folgend verliert die exakte Position der Signale innerhalb der Objekte an Bedeutung und weiterhin wird der Einfluss des Bildrauschens verringert, was beides die Prognoseaufgabe vereinfacht.

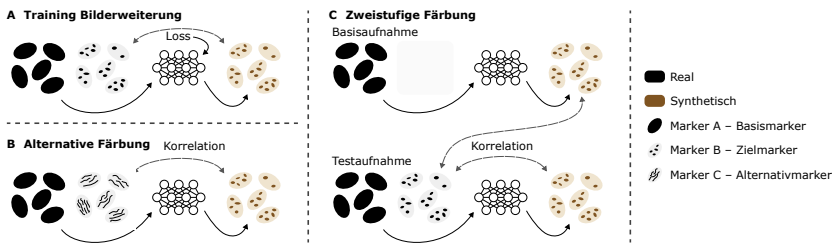
### **Bleed-Trough-Test**

Die optischen Filter eines Mikroskops, die zur Abgrenzung der fluoreszierenden Signale eingesetzt werden, weisen keine perfekte Trennung der Signale auf. Eher besitzen diese einen fließenden Übergang zwischen Sperr- und Transmissionsbereich. Damit ist es möglich, dass Signale von benachbarten Kanälen in den betrachteten Kanal durchscheinen. Der auch als Bleed-Through bekannte Effekt kann in Betracht der Synthese von Bildmodalitäten dazu führen, dass das Signal des Zielmarkers schwach in dem Kanal des Basismarkers zu sehen ist. Sind solche Signale vorhanden, kann ein Netzwerk während des Trainings lernen, diese Informationen für die Prognose des Zielmarkers zu nutzen. Während des Trainingsprozesses ist dies nicht direkt ersichtlich, da auch bei den zur Evaluierung und zum Testen genutzten Daten die Zielmarker gefärbt sind. In einer späteren Anwendungsphase der trainierten Modelle ist hingegen keine Färbung des Zielmarkers mehr vorhanden, womit ausgeschlossen ist, dass es zu einem Bleed-Through des Zielmarkers kommt. Haben die Transformationsmodelle nun gelernt, solche Signale für eine Prognose zu nutzen, ist mit einer Abnahme der



Prognosegüte in der Anwendungsphase zu rechnen. Im Extremfall kann dies sogar dazu führen, dass die Modelle vollständig versagen.

Im Folgenden führt diese Arbeit daher zwei neue Methoden (*Alternative Färbung* und *Zweistufige Färbung*) ein (siehe Abbildung 2.14), um erstmalig den Einfluss der beschriebenen Effekte zu analysieren. Die erste Methode, *Alternative Färbung* (Abbildung 2.14b), beruht auf der Erstellung von zusätzlichen Proben, welche anstatt des Zielmarkers mit einem anderen Marker gefärbt werden. Wichtig dabei ist, dass hierfür ein Marker genutzt wird, welcher im gleichen Bereich des Farbspektrums liegt wie der Zielmarker. In den meisten Fällen stehen solche Aufnahmen zur Verfügung, da die Synthese von Bildmodalitäten eingesetzt wird, um Signale eines Markers zu generieren, welcher aufgrund eines anderen Markers mit ähnlicher Emissionswellenlänge nicht gleichzeitig aufgenommen werden kann.



**Abbildung 2.14:** Überblick über das Training der Modelle zur Synthese von Bildmodalitäten **A** und zwei Evaluierungsmethoden **B** und **C** zur Prüfung, ob sich Transformationsmodelle an eventuell vorhandenen durchscheinenden Signalen orientieren. Die erste Methode **B** *Alternative Färbung* prüft, ob eine Korrelation des prognostizierten Zielmarkers mit dem gefärbten Alternativmarker besteht. Die zweite Methode **C** *Zweistufige Färbung* ermöglicht zum einen die Bestimmung der Prognosegüte für eine Probe ohne vorhandene Zielmarkerfärbung. Zum anderen erlaubt sie die Prüfung, ob sich die Prognosegüte bei Vorhandensein der Zielmarkerfärbung verbessert. Dazu wird die Probe zuerst ohne die Färbung des Zielmarkers aufgenommen. Anschließend wird zusätzlich der Zielmarker gefärbt und die Probe erneut aufgenommen. Anhand der Korrelationen zwischen den, anhand des Basismarkers der ersten und zweiten Aufnahme, prognostizierten Zielmarkers und des realen Zielmarkers der zweiten Aufnahme kann geprüft werden, ob sich die Prognosegüte durch das Vorhandensein der Zielmarkerfärbung verbessert.

Für die Bewertung anhand der Färbung eines Alternativmarkers (mit ähnlichem Antwortspektrum) wird das bereits trainierte Modell genutzt, um anhand des Basismarkers eine Prognose des Zielmarkers zu erstellen. Anschließend wird die Korrelation des prognostizierten Zielmarkers mit dem Alternativmarker untersucht (vgl. Abbildung 2.14b). Weisen die beiden Marker eine hohe Korrelation auf, ist dies ein Hinweis darauf, dass das Netzwerk die durchscheinenden Signale für die Prognose nutzt. Ein Nachteil der Methode ist, dass eine hohe Korrelation auch durch ähnliche Bindungsstellen des Alternativ- und Zielmarkers hervorgerufen werden kann. Das ist bei der Auswahl des Alternativmarkers zu berücksichtigen. Beispielsweise ist die Kombination des Nukleusmarker DAPI und des Proliferationsmarker Ki-67 ungeeignet, da beide innerhalb der Zellkerne zu sehen sind. Alternativ kann in diesem Fall auch eine visuelle Überprüfung Klarheit schaffen, bei der die generierten Signale mit denen des Alternativmarkers und realen Signalen des Zielmarkers verglichen werden. Vorteile dieser Methode ist der geringe Aufwand, da häufig die benötigten Daten bei der Anwendung zur Verfügung stehen und so keine zusätzlichen Proben, Färbungen und Aufnahmen zu erstellen sind. Weiterhin lässt sich auch schon visuell überprüfen, ob Modelle auf eventuell vorhandene durchscheinende Signale zurückgreifen.

Die zweite Methode, *Zweistufige Färbung* (Abbildung 2.14c), beruht auf einer zweistufigen Färbung, mit je einer Aufnahme und Prognose. Hierzu werden Proben erstellt, die zunächst nur mit dem Basismarker gefärbt sind. Bei der als *Basisaufnahme* ( $y$ ) bezeichneten Aufnahme ist durch den fehlenden Zielmarker sichergestellt, dass keine durchscheinenden Signale auftreten. Nach der Bildaufnahme der Probe wird diese in einem zweiten Schritt zusätzlich mit dem Zielmarker gefärbt und anschließend unter den gleichen Einstellungen erneut aufgenommen. Diese zweite Aufnahme wird im Folgenden als *Testaufnahme* ( $z$ ) bezeichnet. Da trotz einer Fixierung der Probe kleine Verschiebungen sowohl im gesamten als auch in lokalen Bereichen der Probe zwischen den Aufnahmen nicht ausgeschlossen werden kann, ist ein Vergleich zwischen den beiden Aufnahmen nicht direkt möglich. Um diese Verschiebungen zu kompensieren und so den pixelweisen Vergleich zu ermöglichen, ist eine elastische Bildregistrierung anzuwenden. Dabei wird der reale Zielmarker der Testaufnahme anhand der Basismarker an

die Basisaufnahme angeglichen. Genauer gesagt, erfolgt die Berechnung des Verschiebungsfeldes anhand der Basismarker aus der Basis- und Testaufnahme. Die berechnete Verschiebung wird anschließend auf den Zielmarker der Testaufnahme angewendet.

Basierend auf dem Basismarker  $y_b$  der Basisaufnahme wird mit dem zu bewerteten Modell eine Prognose des Zielmarkers  $\hat{y}_z$  erstellt. Diese kann bereits mit dem realen und registrierten Zielmarker  $z_z$  der Testaufnahme verglichen werden, um die Prognosegüte in der Anwendung zu bestimmen. Diese Auswertung wird im Folgenden als *Basisprädiktion* bezeichnet und ermöglicht bereits ein für den späteren Anwendungsfall realistisches Testszenario. Für die Analyse, ob das Modell durch eventuell vorhandene durchscheinende Signale einen Vorteil erzielt, ist eine zweite Auswertung (*Testprädiktion*) notwendig. Dazu wird dieses Mal anhand des Basismarkers  $z_b$  der Testaufnahme die Prognose des Zielmarkers  $\hat{z}_z$  erstellt und mit dem realen Zielmarker  $z_z$  verglichen. Ist eine deutliche Verbesserung in der Prognosegüte zwischen den beiden Experimenten festzustellen, ist dies ein Hinweis darauf, dass das Modell die Prognose auf durchscheinende Signale stützt. Als Grundlage für die Bewertung wird das Verhältnis  $p_r$  der Korrelationskoeffizienten zwischen der Basis- und Testprädiktion herangezogen:

$$p_r = \frac{r_{\hat{y}_z, z_z}}{r_{\hat{z}_z, z_z}}. \quad (2.14)$$

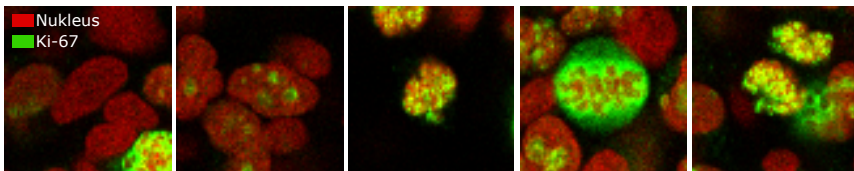
Die Definition eines generellen Grenzwertes für  $p_r$ , der festlegt, dass sich ein Modell nicht auf durchscheinende Signale stützt, ist nur schwer möglich, da dieser Wert von der Qualität der Registrierung abhängig ist. Es ist zu erwarten, dass durch die benötigte Bildregistrierung eine Verringerung der Korrelation einhergeht. Aus eigenen Untersuchungen mit visuellen Vergleichen wurde sich auf einen Grenzwert von  $p_r \geq 0,9$  festgelegt. Ist diese Bedingung erfüllt, wird davon ausgegangen, dass sich ein Modell für die Prädiktion des Zielmarkers nicht auf durchscheinende Signale stützt.

Die Methode der zweistufigen Färbung erlaubt im Gegensatz zur Methode der alternativen Färbung die direkte Bestimmung der Prädiktionsgüte des Modells auf

anwendungsnahen Daten. Allerdings ist die zweistufige Färbung deutlich komplexer und mit einem einmalig höheren Zeitaufwand in der Probengenerierung und der Aufnahme verbunden. Wenn die einmalige Bearbeitungszeit keine Rolle spielt, ist die Methode der zweistufigen Färbung zu präferieren. Andernfalls ist die Methode der alternativen Färbung zu wählen, die mit nur mit einem sehr geringen Mehraufwand verbunden ist, da die benötigten Daten meist schon für die spätere Auswertung benötigt werden.

### 2.3.4 Zellzyklus-Klassifikation

Eine in der Zellbiologie relevante Analyse stellt die Klassifikation der Zellzyklusstadien von einzelnen Zellen dar. Zur Unterstützung wird dabei häufig der Proliferationsmarker Ki-67 eingesetzt, bei dem die markierten Zellen je nach Zyklusstadium unterschiedliche Intensitäten und Texturen aufweisen (vgl. Abbildung 2.15). Die Synthese dieses Markers stellt daher eine ideale Möglichkeit dar, die Qualität und Eignung der Synthese von Bildmodalitäten in einem realen Anwendungsfall zu testen.



**Abbildung 2.15:** Darstellung verschiedener Zellzyklusstadien. Gezeigt sind Nukleussignale in Rot und Ki-67-Signale in Grün. Je nach Stadium weist der Marker Ki-67 unterschiedliche Texturen und Helligkeiten auf.

Der Zellzyklus einer Zelle lässt sich in verschiedene Phasen aufteilen. Ein Überblick über die einzelnen Phasen findet sich im Anhang A.2. Da die Zellteilung einen kontinuierlichen Prozess darstellt, gibt es auch Übergangszustände zwischen den einzelnen Phasen. Die eindeutige Zuordnung einzelner Zellen zu festen

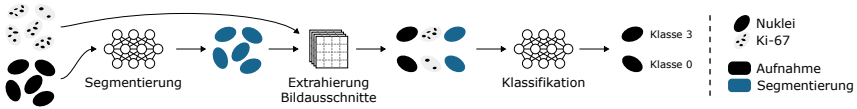
Klassen gestaltet sich daher schwierig und kann bei verschiedenen biologischen Experten zu unterschiedlichen Einschätzungen führen.

Die Zellteilung stellt daher auch eine besondere Anforderung an die Bewertung von Klassifikationsmodellen dar. Klassische Metriken bewerten fälschliche Zuweisungen unabhängig von der Nähe der Klasse stets gleich schlecht. In der Praxis sind Fehler jedoch weniger gravierend, wenn Zellen einer benachbarten Klasse zugeordnet werden. Zwar existiert eine Bewertungsmetrik, die die natürliche Reihenfolge der Klassen durch eine distanzbasierte Metrik berücksichtigt [192], doch wird dabei die periodische Natur der Zyklusstadien nicht beachtet. Um diese periodische Eigenschaft in die Bewertung einzubeziehen, wird eine Metrik mit dem Namen *Klassendistanz* (KD) eingeführt. Bei dem Maß wird die Distanz zwischen der Vorhersage  $\hat{y}_i$  und der Ground-Truth  $y_i$  eines Samples  $i$  genutzt. Das Maß berechnet sich mit  $N_k$  Klassen und  $N_s$  Samples wie folgt:

$$KD = 1 - \left\lfloor \frac{N_k}{2} \right\rfloor \frac{1}{N_s} \sum_{i=0}^{N_s-1} \min(|\hat{y}_i - y_i|, N_k - |\hat{y}_i - y_i|). \quad (2.15)$$

Der zweite Term  $N_k - |\hat{y}_i - y_i|$  der Minimum-Operation ist notwendig, da der Zellzyklus einen sich wiederholenden Prozess darstellt. Durch den Term werden Fälle abgedeckt, bei denen Klassendistanzen über den Anfang und das Ende hinweg geringer sind. Die Normierung mit  $N_k$  beschränkt den Wertebereich von null bis eins, wobei ein Wert von 0 das schlechteste und ein Wert von 1 das beste Ergebnis darstellt.

Für die Klassifikation der Zyklusstadien einzelner Zellen wird eine eigens entwickelte Methode verwendet Abbildung 2.16. Eine Herausforderung besteht neben der Synthese von Ki-67-Signalen in der kombinierten Aufgabe der Segmentierung, um die einzelnen Zellkerne zu lokalisieren, und der anschließenden Klassifikation. Zwar können panoptische Segmentierungsmethoden genutzt werden, die beide Aufgaben mit einem Modell bewältigen, jedoch erfordern diese Methoden Trainingsdaten, die sowohl Instanz- als auch Klassenannotationen umfassen. Die manuelle Erstellung solcher Daten ist aufgrund des hohen Aufwands nicht praktikabel. Daher wird ein zweistufiger Ansatz aus Segmentierung und Klassifikation



**Abbildung 2.16:** Konzept zur Klassifikation von Zellzyklusstadien. Die Klassifikation erfolgt anhand von auf den Zellkern zentrierten Bildausschnitten, die drei Kanäle umfassen: Nukleussignal, Membransignal und Segmentierung.

genutzt: Zunächst erfolgt eine *Segmentierung* der Aufnahmen, um die darin enthaltenen Objektinstanzen zu bestimmen. Darauf aufbauend wird in *Extrahierung Bildausschnitte* für jede Objektinstanz ein Bildausschnitt extrahiert. In Fällen, in denen umliegende Bildbereiche nützliche Informationen enthalten, ist die Ausschnittgröße so zu wählen, dass mehr als nur das Objekt sichtbar ist. Bei der Bestimmung des Zellzyklusstadiums sind beispielsweise die umliegenden Bereiche wichtig, da sie die Abschätzung der globalen Helligkeit und die Identifikation der zweiten Tochterzelle bei kürzlich geteilten Zellkernen ermöglichen. Um im Falle der Sichtbarkeit weiterer Objekte jedoch den Fokus auf das zu klassifizierende Objekt zu gewährleisten, wird dem Netzwerk die Segmentierungsmaske des Objekts in einem zusätzlichen Kanal übergeben. Von einem Schwärzen der nicht zu dem Objekt gehörenden Bildbereiche wird abgesehen, da die umliegenden Bereiche eine wichtige Vergleichsmöglichkeit der Helligkeit bieten.

Da klassische Klassifikationsnetzerke aufgrund der vollständig verbundenen Schichten eine einheitliche Bild-Eingangsgröße erfordern, ist entweder eine Skalierung der extrahierten Bildausschnitte oder die Wahl einheitlicher Ausschnittgrößen notwendig. In Fällen, in denen die Objektgröße ein wichtiges Klassenmerkmal darstellt, ist eine Skalierung der Ausschnitte zu vermeiden. Ein Beispiel hierfür ist der Zellzyklus, bei dem sich die Größe der Zellen zwischen den verschiedenen Stadien unter anderem aufgrund der DNA-Replikation sowie der Chromosomenkondensation und -dekondensation unterscheidet [193, 194]. In solchen Fällen ist eine einheitliche Ausschnittgröße zu wählen, bei der sichergestellt ist, dass nahezu alle Objekte vollständig abgebildet sind.

Anhand der Bildausschnitte, welche Marker und Segmentierungsmaske des Objekts enthalten, wird in der Deep-Learning-basierten *Klassifikation* die Klasse

der Objekte bestimmt. Das Segmentierungsmodell wird mithilfe synthetischer Nukleusaufnahmen trainiert, während die Trainingsdaten für die Klassifikation durch manuelle Annotation einzelner Zellkerne erstellt werden (siehe Abschnitt 4.5).

Für einen Vergleich des synthetisch generierten Ki-67-Signals wird das Training des Klassifikationsnetzwerks auf drei Versuchsanordnungen durchgeführt (vgl. Abbildung 2.17). Die erste Versuchsanordnung *DS-Klassifikation-Real* dient als Benchmark, um die mit echten Markern erreichbare Klassifikationsgüte zu bestimmen. Hier stehen dem Netzwerk Bildausschnitte von einzelnen Zellen mit drei Markern zur Verfügung: DAPI (Zellkern), Ki-67 (Proliferation) und eine Segmentierungsmaske. In der Segmentierungsmaske ist ausschließlich der Nukleus der zu klassifizierenden Zelle markiert. Um die Leistung der vorgestellten Markersynthese zu testen, wird bei der zweiten Versuchsanordnung *DS-Klassifikation-Syn* (vgl. Abbildung 2.17b) das reale Ki-67-Signal durch ein synthetisiertes Signal ersetzt, welches mit den beiden zuvor vorgestellten Modellen ResUnet und pix2pix generiert wird. Um eine untere Abschätzung für die Leistung des Klassifikationsnetzwerks ohne den Proliferationsmarker zu erhalten, wird bei der dritten Versuchsanordnung *DS-Klassifikation-Ohne* (vgl. Abbildung 2.17c) das Ki-67-Signal entfernt.



**Abbildung 2.17:** Darstellung der drei Versuchsanordnungen zum Testen der Methoden zur Synthese von Bildmodalitäten anhand der Klassifikation von Zellzyklusstadien. Die Datensätze bestehen aus, auf einzelne Nuklei zentrierte Bildausschnitte. Gemein ist den Datensätzen, dass der erste bzw. der letzte Kanal immer das Nukleussignal bzw. die Segmentierungsmaske des zu klassifizierenden Nukleus zeigt. Während beim Datensatz *DS-Klassifikation-Real* (A) reales Ki-67-Signal zur Verfügung steht, wird dies in Datensatz *DS-Klassifikation-Syn* (B) durch ein synthetisches Signal ersetzt und in Datensatz *DS-Klassifikation-Ohne* (C) steht zur Kontrolle kein Ki-67 zur Verfügung.

Jeweils wird ein Klassifikationsnetzwerk anhand der oben beschriebenen drei Versuchsanordnungen trainiert und anschließend auf dem jeweiligen Testdatensatz evaluiert. So kann analysiert werden, wie gut sich synthetisch generierte

Markersignale in einem realen Anwendungsfall nutzen lassen. Für die Bewertung werden die Genauigkeit, der F1-Score und die oben eingeführte Klassendistanz verwendet.

### **2.3.5 Voraussetzungen für eine erfolgreiche Synthese von Bildmodalitäten**

Die Synthese von Bildmodalitäten setzt voraus, dass Zusammenhänge zwischen Basis- und Zielmarker bestehen. Ohne solche Zusammenhänge ist es für ein Modell unmöglich, basierend auf dem Basismarker vorherzusagen, wie die Signale des Zielmarkers aussehen. Welche spezifischen Zusammenhänge hierfür erforderlich sind, ist bislang unbekannt. Um diese zu identifizieren, werden drei zentrale Fragestellungen formuliert:

1. Müssen die Signale des Basismarkers flächendeckend vorhanden sein, oder kann das Modell kleinere signalfreie Regionen anhand benachbarter Bereiche ableiten?
2. Ist es erforderlich, dass sich die Signale von Basis- und Zielmarker in ihrer Position überlagern? Kann ein Modell beispielsweise Signale außerhalb des Zellkerns korrekt prognostizieren?
3. Müssen Veränderungen der Morphologie von Objekten im Basismarker spezifisch für den Zielmarker sein? Was geschieht, wenn beispielsweise eine Fragmentierung eines Objektes durch mehrere Ursachen hervorgerufen werden kann, von denen jedoch nur eine durch den Zielmarker visualisiert wird?

Für die Beantwortung der Fragestellungen sind Basis- und Zielmarker-Kombinationen zu identifizieren, welche die in den jeweiligen Fragen genannten Zusammenhänge nicht erfüllen. Kann trotz fehlender Zusammenhänge eine erfolgreiche Transformation erzielt werden, erweist sich der untersuchte Zusammenhang als ein nicht notwendiges Kriterium.



Für die erste Fragestellung eignet sich die Synthese von Nukleussignalen auf Basis von Ki-67-Signalen. Interessant ist hierbei, dass Ki-67-Protein in Zellen, die sich in der G0-Phase befinden (außerhalb eines Zellzyklusstadiums), nicht exprimiert wird. Daher treten in solchen Zellen, genauer gesagt deren Zellkernen keine fluoreszierenden Signale auf. Anhand von Zellkernen in der G0-Phase kann geprüft werden, ob die Prognose von Nukleussignalen allein auf Basis der umliegenden Signale dennoch möglich ist. Die zweite Fragestellung lässt sich mithilfe der Markerkombination aus Membran- und Nukleussignalen analysieren. Zwischen der Morphologie der Membran und des Nukleus bestehen Zusammenhänge, die eine Transformation ermöglichen können. Allerdings befinden sich die Membransignale zwangsläufig außerhalb des Nukleus, was eine potenzielle Herausforderung darstellt. Eine geeignete Markerkombination zur Beantwortung der dritten Fragestellung ist die Kombination aus Nukleusmarker und Caspase-3. Caspase-3 ist ein Marker für Apoptose, eine spezifische Form des Zelltods. Während der Apoptose kommt es zur Fragmentierung des Nukleus, was sich mit dem Nukleusmarker beobachten lässt. Allerdings kann die Fragmentierung des Nukleus auch durch andere Arten des Zelltods verursacht werden, die nicht mit Caspase-3 visualisiert werden.

## 3 Implementierung

### 3.1 Übersicht

Die automatisierte Auswertung hochdimensionaler Mikroskopdaten ist ein wesentlicher Bestandteil der modernen biologischen Forschung. Die in dieser Arbeit entwickelten Methoden ermöglichen nicht nur die Erstellung von Datensätzen für das Training präziser Segmentierungsmodelle, sondern auch die Generierung zusätzlicher Bildmodalitäten, die den Informationsgehalt der Daten weiter erhöhen. Um die Anwendung dieser Methoden zu erleichtern, ist es entscheidend, dass sie als frei zugängliche Implementierungen verfügbar sind. Daher werden die im Rahmen dieser Arbeit entwickelten Softwarepakete in öffentlichen Code-Repositorys bereitgestellt. Dies gewährleistet nicht nur eine einfache Anwendung und Reproduzierbarkeit der Ergebnisse, sondern fördert auch die Weiterentwicklung der Methoden durch andere Forschende.

Für die Implementierung wurde die Programmiersprache Python gewählt, da hier bereits eine breite Palette an Bibliotheken für die Bilddatenverarbeitung zur Verfügung steht und es eine bevorzugte Sprache für die Entwicklung und Anwendung von Deep-Learning-Modellen ist. Python ermöglicht weiterhin einen einfachen Einstieg in die Programmierung, was es fachfremden Anwendern erleichtert, Anpassungen an der Software vorzunehmen. Für die Implementierung Deep-Learning-bezogener Anwendungen wurde das Framework PyTorch [195] verwendet. In den folgenden Abschnitten wird auf die Implementierung der Methoden und deren Veröffentlichungen in öffentlichen Repositorys eingegangen. Weiterhin werden auch grafische Benutzerschnittstellen erläutert.

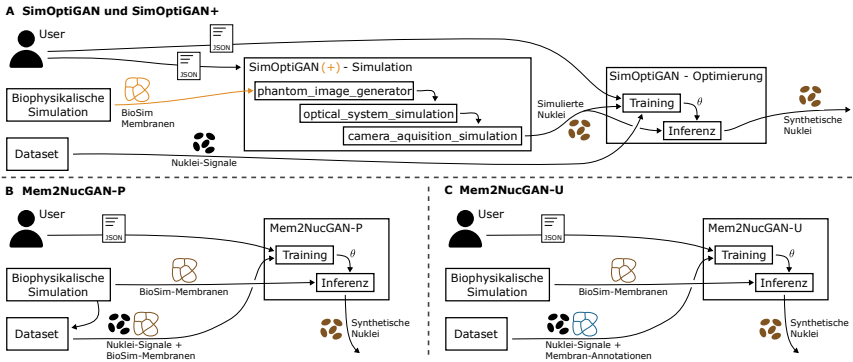
## 3.2 Implementierung der entwickelten Methoden

### 3.2.1 Synthese von Trainingsdaten

Der frei verfügbare Quellcode der Methoden zur Trainingsdatensynthese (SimOptiGAN, SimOptiGAN+, Mem2NucGAN-P, Mem2NucGAN-U) inklusive einer genauen Beschreibung der Parametrierung und des Ablaufs finden sich in dem öffentlichen Repository: [https://github.com/bruchr/cell\\_synthesis](https://github.com/bruchr/cell_synthesis).

Eine Übersicht über die erstellten Softwarepakete zur Trainingsdatensynthese und deren Zusammenspiel ist in Abbildung 3.1 gegeben. Die Implementierung der Synthesemethode SimOptiGAN besteht aus zwei Hauptkomponenten (vgl. Abbildung 3.1a). Die erste Komponente *SimOptiGAN-Simulation* umfasst den Simulationsprozess, der sowohl die Probengenerierung als auch die Aufnahmesimulation abdeckt. Die zweite Komponente beinhaltet die GAN-basierte *SimOptiGAN-Optimierung*. Der Simulationsprozess von SimOptiGAN ist weiterhin modular aufgebaut. Die Hauptmodule bestehen aus *phantom\_image\_generativ*, *optical\_system\_simulation* und *camera\_acquisition\_simulation*. Während das erste Modul die Probengenerierung abdeckt, simulieren die anderen beiden Module den Prozess der Aufnahmesimulation.

Ein wesentlicher Vorteil dieses zweigeteilten Ansatzes aus Simulation und Optimierung ist Flexibilität, da das verwendete GAN-Modell bei Bedarf ausgetauscht werden kann. Angesichts der rasanten Weiterentwicklungen im Bereich der Bild-zu-Bild-Transformationsmodelle ermöglicht dies, zukünftig neue und leistungsfähigere Modelle zu integrieren. Der modulare Aufbau der Simulationskomponente erlaubt es zudem, die einzelnen Komponenten flexibel an neue Anforderungen oder unterschiedliche Probenstrukturen anzupassen. Die Parametrierung des Simulations- und Optimierungsteils erfolgt über je eine JSON-Datei. In diesen Dateien werden die Parameter durch Schlüssel-Werte-Paare definiert, was eine flexible und benutzerfreundliche Anpassung der Einstellungen ermöglicht.



**Abbildung 3.1:** Übersicht der erstellten Softwarepakete zur Synthese von Trainingsdaten und deren Zusammenspiel. Gezeigt sind die Pakete für SimOptiGAN und SimOptiGAN+ (A), Mem2NucGAN-P (B) und Mem2NucGAN-U (C). In der Übersicht von SimOptiGAN(+) (A) sind die zusätzlichen Schritte zur Integration biophysikalischer Simulationen (SimOptiGAN+) in Gelb hervorgehoben.

Die Erweiterung SimOptiGAN+ (vgl. Abbildung 3.1a), welche SimOptiGAN um die Integration von biophysikalisch simulierten Zellanordnungen erweitert, nutzt den modularen Aufbau der SimOptiGAN Implementierung. So erfolgt nur eine Anpassung des *phantom\_image\_generator* Moduls zur Einbringung der *Biophysikalischen Simulation*, während der restliche Aufbau erhalten bleibt.

Mem2NucGAN-P und Mem2NucGAN-U sind eigenständige Methoden, welche ausschließlich GAN-Modellen zur Synthese verwenden und keinen Simulationsanteil aufweisen (vgl. Abbildung 3.1b bzw. c). Die Implementierungen der beiden Methoden basieren auf einer von [181] abgeleiteten Implementierung des pix2pix bzw. des CycleGANs. Diese wurden vereinfacht und angepasst, um sowohl 2D- als auch 3D-Daten verarbeiten zu können.

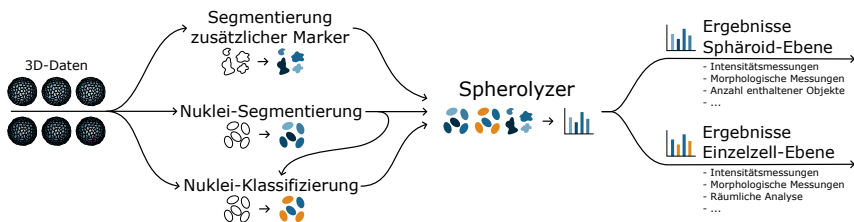
### 3.2.2 Synthese von Bildmodalitäten

Die Implementierung der entwickelten Methoden zur Synthese von Bildmodalitäten ist im frei zugänglichen GitHub-Repository verfügbar: <https://github.com>

om/bruchr/marker\_synthesis. Dort findet sich auch eine detaillierte Ablaufbeschreibung für das Training und die Anwendung der Modelle. Es wurde sowohl die auf dem ResUnet basierende Methode als auch die auf dem pix2pix-Modell basierende Methode implementiert. Für die Prozessierung von 3D-Daten wurden die Netzwerke so abgeändert, dass diese nativ 3D-Daten verarbeiten können. Für das pix2pix-Modell wird bei dem Generator zudem auf ein ResUnet statt eines U-Nets gesetzt.

### 3.2.3 Spherolyzer - Eine neue Software für die Analyse von Drug-Screenings

Die Segmentierung der Daten stellt den ersten Schritt in der Auswertung von Zellaufnahmen dar. In einem zweiten Schritt können anhand der segmentierten Instanzen verschiedene Parameter extrahiert und komplexere Analysen durchgeführt werden. Für eine effiziente Datenauswertung komplexer 3D-Sphäroid-Experimente wird die grafisch geführte Software *Spherolyzer* entwickelt. Abbildung 3.2 gibt einen groben Überblick über den Ablauf der Auswertung und die dabei extrahierbaren Informationen.



**Abbildung 3.2:** Veranschaulichung des Ablaufs zur Auswertung von 3D-Sphäroid-Daten mittels Spherolyzer. In aufgenommenen 3D-Daten werden Nuklei segmentiert und gegebenenfalls klassifiziert. Weiterhin können auch weitere Marker segmentiert werden. Anhand der Ergebnisse erfolgt mit Spherolyzer die Auswertung der Experimente. Dabei werden sowohl Informationen auf Sphäroid-Ebene als auch auf Einzelzell-Ebene extrahiert.

Die Software erfordert bereits segmentierte Aufnahmen von 3D-Zellkulturen. Weiterhin kann eine Klassifikation der Objektinstanzen genutzt werden, um zusätzliche Klassenmerkmale zu bestimmen. Auch können Segmentierungen von weiteren Markern genutzt werden. Die Segmentierung und Klassifikation wurden bewusst von der Auswertesoftware ausgelagert, um die Nutzung beliebiger Segmentierungs- und Klassifikationsalgorithmen zu ermöglichen. Anhand der Roh-Daten und den bereitgestellten Annotationen erfolgt die Auswertung der Daten mit der entwickelten Software. Dabei lässt sich die Analyse in zwei Bereiche aufteilen:

1. Analyse auf Sphäroid-Ebene: Diese umfasst unter anderem die Messung von Intensitäten und morphologischen Eigenschaften der gesamten Zellstruktur sowie die Zusammenfassung der enthaltenen Objektmerkmale. Weiterhin können verschiedene Bereiche innerhalb der Zellkultur miteinander verglichen werden
2. Analyse auf Einzelzell-Ebene: Diese liefert Eigenschaften bezüglich der Intensitäten, Morphologie, Klasse und Lage von einzelnen Zellen.

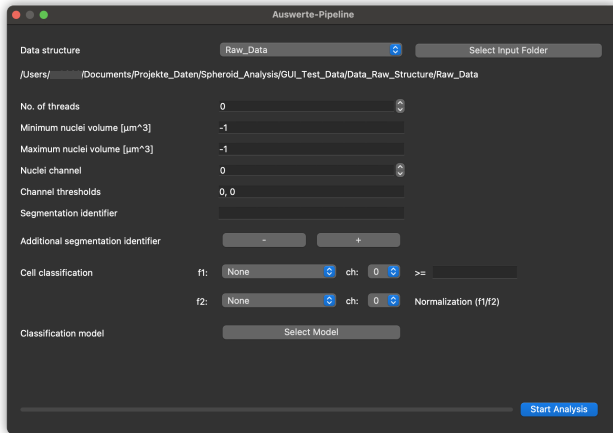
Die Software speichert die Ergebnisse in zwei Tabellen: Eine, die die Merkmale der einzelnen Zellkerne innerhalb einer einzelnen Zellkultur beschreibt, und eine andere, die die kollektiven Ergebnisse über eine Gruppe hinweg zusammenfasst. Neben den Tabellen werden zudem Python-Datendateien generiert, welche auch zusätzliche Merkmale und Zwischenergebnisse enthalten. Diese Dateien können genutzt werden, um spezifischere oder komplexere Merkmale, wie die Aggregation von Daten über mehrere Bilder hinweg oder spezifische Merkmalskombinationen zu berechnen oder um Visualisierungen zu erstellen. Die Software arbeitet im Batch-Modus, was es ermöglicht, mehrere Verzeichnisse und deren Unterverzeichnisse über die GUI zu verarbeiten. Für Bilder im selben Ordner wird jeweils eine Zusammenfassung in Form einer Tabelle erstellt, die den Mittelwert und die Standardabweichung für die Merkmale angibt. Um die Effizienz zu steigern, kann optional eine parallele Prozessierung durchgeführt werden und so mehrere Bilder gleichzeitig verarbeitet werden. Die parallele Prozessierung erfordert im Vergleich zur seriellen Prozessierung eine größere Menge an RAM-Speicher.

Die entwickelte Software ist frei auf GitHub verfügbar: [https://github.com/bruchr/spheroid\\_analysis](https://github.com/bruchr/spheroid_analysis). Zudem wurde die Software bereits erfolgreich in einem Drug-Screening eingesetzt [59].

## 3.3 Grafische Benutzeroberflächen

### 3.3.1 Spherolyzer – Zugehörige Benutzerschnittstelle

Die GUI ermöglicht die einfache Parametrierung der Analysepipeline (siehe Abbildung 3.3). Zu Beginn wird das Hauptdatenverzeichnis ausgewählt (*Select Input Folder*), woraufhin einige Felder anhand der dort enthaltenen Bilder mit Standardwerten vorausgefüllt werden. Das Tool ermöglicht anschließend die Anpassung zahlreicher Parameter, einschließlich Zellkerngrößenfiltern zur Ausschließung von Ausreißern (*nuclei volume*) und kanalspezifischen Intensitätsschwellen zur Abgrenzung von Vordergrundsignalen (*Channel thresholds*). Die Dateien der Zellkernsegmentierungen werden automatisch gefunden. Falls es mehrere Segmentierungen gibt, können diese über einen Suchstring gefiltert werden (*Segmentation identifier*). Zudem können weitere Segmentierungsmasken von anderen Signalen hinzugefügt werden. Eine einfache Klassifizierung kann durch die Definition eines Merkmals und dessen Schwellenwert erreicht werden, wodurch Zellkerne als positiv oder negativ eingestuft werden (*Cell classification*). Eine Normalisierung der Merkmale wird ebenfalls unterstützt, um Schwankungen, wie abnehmende Helligkeit, zu korrigieren, indem das Merkmal mit der Zellkernintensität korreliert wird. Für fortgeschrittenere Klassifizierungsaufgaben können Benutzer Modelle wie Support Vector Machines (SVMs) in Python entwickeln und trainieren und über die GUI für die Software bereitstellen (*Classification model*). Das Skript ermöglicht zudem die optionale Integration vortrainierter Klassifizierungsmodelle oder die Nutzung eines benutzerdefinierten Merkmalschwellenwerts, um Zellkerne beispielsweise als positiv oder negativ in Bezug auf einen spezifischen Fluoreszenzmarker zu klassifizieren.



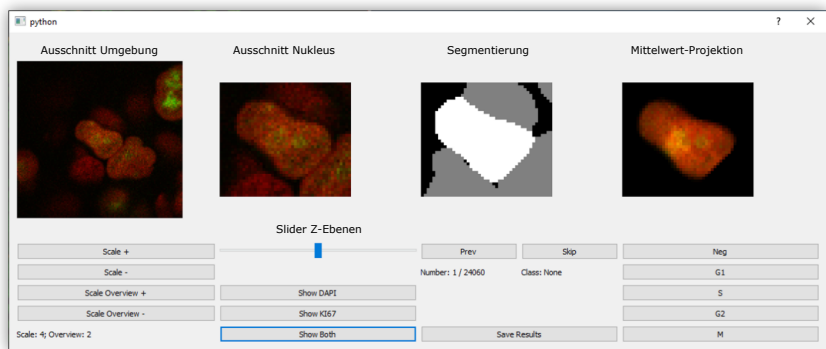
**Abbildung 3.3:** Grafische Benutzeroberfläche für die Parametrierung der Software zur Auswertung von 3D-Zellkulturaufnahmen.

### 3.3.2 3D-ClassAnnotator – Grafische Oberfläche zur Klassenannotation in 3D-Daten

Für die effiziente Klassenannotation von Objekten in 3D-Daten wurde eine grafische Benutzeroberfläche entworfen (*3D-ClassAnnotator*). Diese Schnittstelle erlaubt es für bereits segmentierte Objekte, deren Klasse anzugeben. Dazu stehen mehrere Visualisierungen und Einstellmöglichkeiten zur Verfügung (vgl. Abbildung 3.4). So wird anhand der Segmentierung ein größerer und kleinerer 2D-Ausschnitt gezeigt (*Ausschnitt Umgebung* bzw. *Ausschnitt Nukleus*), welcher in der  $z$ -Achse verschoben werden kann (*Slider Z-Ebenen*). Der große Ausschnitt ermöglicht es dem Nutzer, Umgebungsinformationen wie die generelle Helligkeit der Signale in den Entscheidungsprozess einfließen zu lassen, während der kleine Ausschnitt es erlaubt, mehr Details des zu klassifizierenden Objektes zu erkennen. Weiterhin wird die Segmentierung angezeigt, welche Aufschluss über das Objekt von Interesse gibt. Um einen schnellen Überblick über den 3D-Ausschnitt zu erhalten, wird zudem eine Mittelwert-Projektion über die  $z$ -Ebenen angezeigt. Neben einer überlagerten Darstellung können die Marker auch einzeln angezeigt



werden, um mehr Details zu erkennen. Um eine starke Imbalance zwischen der Anzahl der annotierten Klasseninstanzen zu vermeiden, besteht auch die Möglichkeit, Objekte unannotiert zu überspringen. Am Ende erhält der Nutzer sowohl eine CSV-Datei, die Objekt-IDs und die zugehörigen Klassen enthält, als auch eine mit dem Python pickle-Modul<sup>1</sup> generierte Datei. Letztere umfasst eine Liste von RegionProperties-Objekten<sup>2</sup>, bei denen die entsprechenden Klasseninformationen ergänzt ist.



**Abbildung 3.4:** 3D-ClassAnnotator: Entwickelte grafische Oberfläche zur Erstellung der Klassenannotationen. Mit der Oberfläche kann der Benutzer Klassen von Objekten in 3D-Daten annotieren. Das Tool zeigt dem Nutzer einen Ausschnitt der Umgebung des Objektes, einen Ausschnitt des Objektes, dessen Segmentierung und eine Mittelwert-Projektion von mehreren  $z$ -Ebenen. Über einen Slider kann der Nutzer sich verschiedene  $z$ -Ebenen anzeigen lassen. Zusätzlich bietet die Oberfläche die Möglichkeit, die Ausschnitte zu vergrößern oder zu verkleinern. Weiterhin können die Kanäle separat als auch zusammen angezeigt werden. Über Schalter kann zum nächsten bzw. zum vorherigen Objekt gewechselt werden.

<sup>1</sup> <https://docs.python.org/3/library/pickle.html>

<sup>2</sup> <https://scikit-image.org/docs/0.24.x/api/skimage.measure.html>

## 4 Datensätze

### 4.1 Übersicht

Um die Leistung von Algorithmen zu evaluieren und vergleichbar zu machen, sind standardisierte Datensätze unerlässlich. Aufgrund der Anforderungen der in dieser Arbeit entwickelten Methoden existieren derzeit keine geeigneten Datensätze, die eine solche Evaluierung ermöglichen. Daher werden im Rahmen dieser Arbeit neue Datensätze erstellt, die in diesem Kapitel beschrieben werden.

Um erstmalig den Einfluss seltener Elemente in synthetischen Trainingsdaten für Segmentierungsmodelle zu untersuchen, wird ein Datensatz mit unterschiedlichen Verhältnissen an seltenen Elementen benötigt. Daher wird in Abschnitt 4.2 ein neuer 3D-Datensatz *DS-Seltene-Elemente* vorgestellt, der mehrere synthetische 3D-Datensätze mit variierenden Anteilen von Objekten zweier Klassen enthält. *DS-Seltene-Elemente* ermöglicht damit die Analyse, wie seltene Elemente die Segmentierungsgüte beeinflussen.

Für die Anwendung der in Abschnitt 2.2 vorgestellten Methoden zur Trainingsdatensynthese werden reale Bilder benötigt, anhand derer die GAN-basierten Modelle trainiert werden können. Weiterhin sind Ergebnisse biophysikalischer Simulationen nötig, die in manchen Modellen für eine realistischere Anordnung der Zellen eingesetzt werden. Ein anwendungsnaher Vergleich der Synthesemethoden erfordert zudem einen händisch annotierten Datensatz, um die Leistung der auf synthetischen Daten trainierten Segmentierungsmodelle zu evaluieren. Die notwendigen 3D-Datensätze werden in Abschnitt 4.3 eingeführt.

Weiterhin wird ein heterogener Datensatz benötigt, um das in Abschnitt 2.3 vorgestellte Framework IMS zur Synthese von Bildmodalitäten zu testen. In Abschnitt 4.4 wird dazu ein heterogener 3D-Datensatz *DS-Marker-Syn* vorgestellt, welcher sowohl Färbungen des Basis- als auch des Zielmarkers enthält. Anhand dieses Datensatzes wird zudem die Leistung von zwei Modellen verglichen.

Schließlich wird zur Testung des entwickelten Frameworks IMS unter realen Anwendungsbedingungen ein händisch annotierter Datensatz benötigt, welcher die Klassifikation verschiedener Zellstadien ermöglicht und sowohl die Signale des Basis- als auch des Zielmarker enthält. Der entsprechende 3D-Datensatz (*DS-Klassifikation*) wird in Abschnitt 4.5 eingeführt.

## 4.2 Datensatz zur Untersuchung des Einflusses seltener Elemente

Bislang ist unklar, wie die Einbringung seltener Elemente in synthetische Trainingsdatensätze die allgemeine Segmentierungsqualität sowie die Qualität der Segmentierung dieser seltenen Elemente beeinflusst. Hier ist zu klären, in welchem Verhältnis eine Einbringung vorteilhaft ist und ob eine überproportionale Einbringung zu einer Verschlechterung der Segmentierungsgüte regulärer Elemente führt. Zur Untersuchung dieser Fragestellungen werden mehrere synthetische 3D-Datensätze mit unterschiedlichen Verhältnissen erstellt, um damit die Segmentierungsqualität über den Anteil an seltenen Elementen zu untersuchen. Die Datensätze werden im Folgenden unter dem Namen *DS-Seltene-Elemente* zusammengefasst. Eine solche Untersuchung ist besonders interessant, wenn die seltenen Objekte sich von den regulären Objekten hinsichtlich ihrer Morphologie und/oder ihrer Textur unterscheiden. Als Grundlage zur Untersuchung werden zwei synthetische Objektklassen genutzt, welche sich in ihrer Morphologie, Helligkeit und Textur unterscheiden. Insgesamt werden jeweils sechs dreidimensionale Trainings- und Testdatensätze mit spezifisch festgelegten Klassenverhältnissen (0 %, 1 %, 5 %, 20 %, 50 % und 100 %) generiert. Die Trainingsdatensätze werden

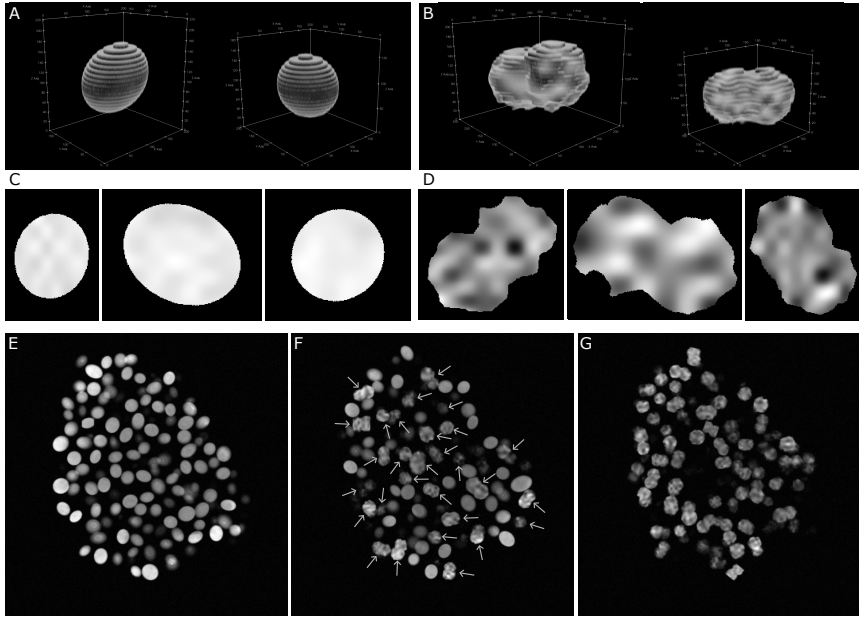
verwendet, um jeweils ein 3D-Segmentierungsmodell zu trainieren. Anschließend wird die Leistung dieser Modelle auf den verschiedenen Testdatensätzen bestimmt. Damit wird analysiert, welche Anteile an seltenen Elementen in den Trainingsdaten die beste Segmentierungsleistung bei gegebenen Verhältnissen in den Testdaten erzielen. Insbesondere wird mit den Testdatensätzen 0% bzw. 100% die Segmentierungsleistung gezielt für die Objektklassen regulär bzw. selten analysiert.

Die dreidimensionalen Datensätze werden mit der in Abschnitt 2.2.2 vorgestellten Methode SimOptiGAN erstellt. Mit dieser Methode kann das Verhältnis zwischen den Klassen in den generierten Bildern durch die Auswahlwahrscheinlichkeit der Objektprototyp-Klassen gesteuert werden. Für die Generierung der synthetischen Aufnahmen mit SimOptiGAN wurden für jede Klasse 159 synthetische Objektprototypen erstellt und in die Objekt-Datenbank platziert<sup>1</sup>. Ein Teil der generierten Prototypen ist in einer 3D-Darstellung bzw. 2D-Schnittdarstellung in Abbildung 4.1a und b bzw. c und d gezeigt. Da keine realen Daten als Vorbilder für die Optimierung der synthetischen Aufnahmen existieren, wird dieser Schritt von SimOptiGAN übersprungen.

Die bei der Erstellung angegebene Auswahlwahrscheinlichkeit gilt zwar für die Auswahl der Klasse, allerdings wird ein ausgewählter Objektprototyp dieser Klasse nicht zwingend platziert. Gerade wenn die Instanzen einer Objektklasse kleiner sind als die der anderen, können diese in Bereiche noch eingesetzt werden, an denen für die andere Klasse kein Platz mehr vorhanden ist. Das hat zur Folge, dass die gemessenen Verteilungen der zwei Objektklassen leicht von den vorgegebenen Verteilungen abweichen. Im Mittel über Trainings- und Testdaten ergeben sich die gemessenen Verhältnisse zu 0 %, 0,5 %, 3 %, 14 %, 38 % und 100 %. Für die Erstellung der Trainings- und Testdatensätze werden jeweils sechs bzw. zwei Bilder generiert, welche eine Auflösung von  $128 \times 512 \times 512 \text{ px}^3$  ( $z, y, x$ ) besitzen. Die Anzahl und Größe der Bilder bieten eine ausreichende Datenmenge

---

<sup>1</sup> Um die Varianz zu erhöhen, wurde auch die Generierung synthetischer Prototypen während der Probengenerierung getestet, anstatt zuvor erstellte und in einer Datenbank abgelegte Prototypen zu verwenden. Dieser Ansatz erwies sich jedoch als zu zeitaufwendig.



**Abbildung 4.1:** 3D-Darstellungen (A, B) und 2D-Schnitte (C, D) synthetischer 3D-Objektprototypen der ersten Klasse mit regulären Objekten (A, C) und der zweiten Klasse mit seltenen Objekten (B, D). Die beiden Klassen unterscheiden sich in der Morphologie, Helligkeit und Textur. (E-G) 2D-Schnitte von synthetischen 3D-Aufnahmen mit einem unterschiedlichen Verhältnis an seltenen Elementen: 0% E, 50% F und 100% G. In F sind die seltenen Elemente mit Pfeilen markiert.

für das Training der Segmentierungsmodelle, ohne dabei eine zu lange Rechenzeit zu erfordern. Abbildung 4.1(e-g) zeigen beispielhaft 2D-Schnitte von drei synthetischen Aufnahmen mit einem Verhältnis von 0% (e), 50% (f) und 100% (g) an seltenen Objekten.

Die Objekte der ersten Klasse repräsentieren reguläre Zellkerne und werden durch Ellipsoide modelliert (vgl. Abbildung 4.1a und c). In Bezug auf die Länge der ersten Hauptachse werden die Längen der anderen beiden Hauptachsen bei diesen Objekten zufällig im Bereich von 0,5 bis 1,5 gewählt, um sowohl kugelförmige als auch leicht ellipsoide Formen zu erzeugen. Anschließend wird die Größe

des Ellipsoids so skaliert, dass sein Volumen mit einem zufällig ausgewählten Volumen aus einer vorgegebenen Liste<sup>2</sup> übereinstimmt.

Die Objekte der zweiten Klasse weisen, wie Zellen im mitotischen Zustand, abweichende Formen zu Objekten der ersten Klasse auf. In Anlehnung an die Metaphase werden gestreckte Formen mit einer nicht vollständig konvexen Oberfläche verwendet (vgl. Abbildung 4.1b und d). Für deren Generierung werden zwei überlappende Sphären mit identischem Volumen genutzt. Das Volumen der beiden Sphären wird ebenfalls zufällig aus derselben Liste wie für die erste Objektklasse ausgewählt. Die Überlappung der Sphären wird über das Übereinanderschieben der Bilder in  $x$ -Richtung erzielt. Die prozentuale Überlappung wird zufällig aus einer gleichverteilten Verteilung zwischen 25% und 60% gezogen, woraus sich eine ausreichende Überlappung ergibt, um sie von nah beieinander liegenden Objekten auseinanderzuhalten, aber immer noch zwei Kugeln statt eines Ellipsoids erkennbar sind. Wie beim Vorbild sind die Objekte der zweiten Klasse in der Regel größer als die der ersten Klasse (Faktor 1,6). Mitotische Zellen bzw. deren Zellkerne zeichnen sich auch durch eine unruhige Oberfläche aus. Daher werden die Formen der Prototypen mittels einer elastischen Transformation verzerrt. Dazu wird ein  $10 \times 10 \times 10 \times 3 \text{ px}^4$  großes dreidimensionales Vektorfeld generiert, wobei jeder Wert des Feldes anhand der Normalverteilung  $\mathcal{N}(0, 16)$  bestimmt wird. Mit der gewählten Standardabweichung der Normalverteilung ergibt sich zusammen mit der Auflösung der Objektbilder eine ausreichend große Verzerrung. Die letzte Dimension des Feldes stellt den Verschiebungsvektor dar. Anschließend wird das Vektorfeld auf die Größe des Objektbildes skaliert, wobei sich durch die Vergrößerung glatte Übergänge zwischen den anfangs zufälligen Voxelwerten ergeben. Das transformierte Bild ergibt sich schließlich aus der Abbildung des Eingangsbilds auf die neuen verschobenen Koordinaten, wobei die exakten Voxelwerte an den Rasterpunkten durch Interpolation bestimmt werden.

---

<sup>2</sup> Diese Liste basiert auf den Volumina realer Zellkerne, die mit dem Faktor 8 skaliert wurden, um realistische Volumenverhältnisse zwischen den Objekten zu erzielen. Die Skalierung ist notwendig, da ansonsten bei der späteren Aufnahmesimulation die Texturen der Objekte durch die mit der PSF beschriebenen Signalspreizung nicht mehr erkennbar sind.

Da die Formen der Objekte beider Klassen noch an den Hauptachsen ausgerichtet sind, werden die Objekte anhand von drei zufällig bestimmten Winkeln um die Hauptachsen gedreht. Durch die noch isotrope Auflösung der Objektbilder kann die Drehung um alle Rotationsachsen ohne vorherige Angleichung der Auflösung durchgeführt werden. Anschließend werden die Texturen der Objekte generiert. Für beide Klassen wird dazu anhand der Normalverteilung  $\mathcal{N}(230, \sigma_t^2)$  ein  $10 \times 10 \times 10 \times 3 \text{ px}^4$  großes Zufallsfeld generiert. Da mitotische Zellkerne eine unruhigere Textur aufweisen, wird für die Textur der zweiten Klasse eine höhere Standardabweichung von  $\sigma_t = 80$  gewählt, während für die erste Klasse eine Standardabweichung von  $\sigma_t = 10$  genutzt wird. Das Zufallsfeld wird wie zuvor kleiner als das Objektbild gewählt, da sich so bei der späteren Skalierung glatte Übergänge zwischen den zufälligen Werten ergeben. Im Gegensatz zur elastischen Transformation stellen die Werte dieses Mal die Intensität der Voxel dar. Der Mittelwert von 230 wird aufgrund des verwendeten Datentyps (unsigned integer 8-Bit) mit einem Wertebereich von  $[0, 255]$  gewählt. Um glatte Übergänge zu erzielen, wird das Feld wiederum auf die Größe der Objektbilder skaliert. Die so erstellte Textur wird zuletzt auf die Vordergrundbereiche der Objektbilder appliziert. In einem letzten Schritt wird die Auflösung der Objektbilder und deren Annotationen an die später verwendete Simulationsauflösung angepasst, woraufhin auch die isotrope Auflösung verloren geht.

## 4.3 Datensätze zum Methodenvergleich der Trainingsdatensynthese

### 4.3.1 Übersicht

Zur Generierung synthetischer Trainingsdaten sind je nach verwendeter Methode unterschiedliche Daten erforderlich: Allen ist gemein, dass reale Bilder zum Training der Methoden benötigt werden. Für die Methoden, die auf den Ergebnissen von biophysikalischen Simulationen beruhen, sind die Simulationsergebnisse in Form der Zellannotationen (BioSim-Instanzen) erforderlich. Schließlich

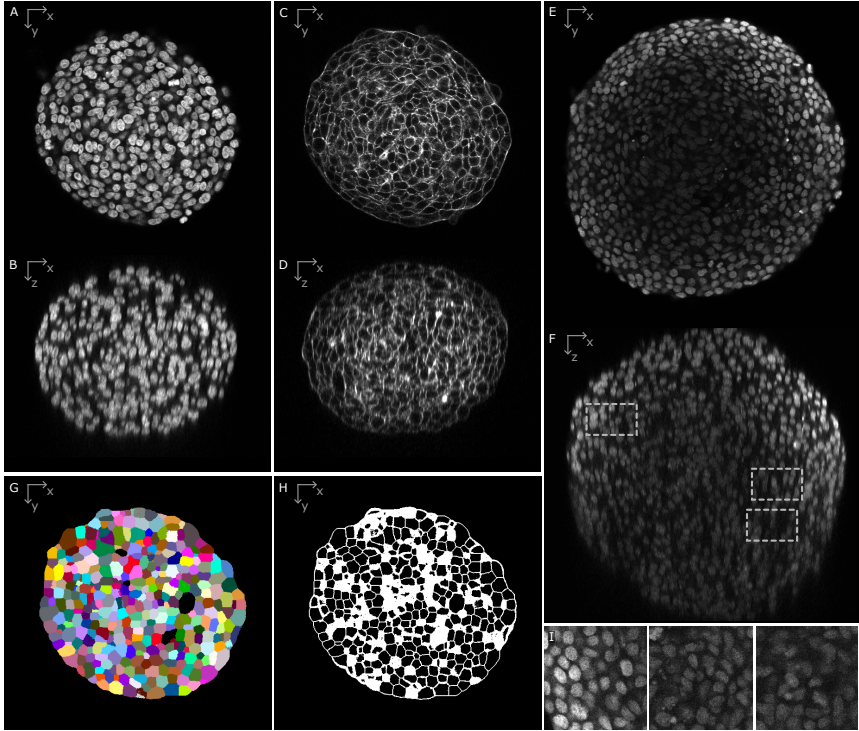
ist eine Ground-Truth eines realen Datensatzes erforderlich, um den Vergleich auf Basis der mit den synthetischen Daten trainierten Segmentierungsnetzwerken auszuwerten. Die genannten Datensätze werden unter dem Namen *DS-Train-Syn* zusammengefasst und im Folgenden beschrieben.

### 4.3.2 Reale Bilder zum Training der GAN-Modelle

Für das Training der Mem2NucGAN-Methoden werden Aufnahmen benötigt, welche sowohl mit einem Nukleusmarker als auch mit einem Membranmarker gefärbt sind. Dazu stehen vier Aufnahmen von HT29 Zellkulturen zur Verfügung. Ein 2D-Schnitt und die seitliche Sicht einer dieser Aufnahmen sind in Abbildung 4.2a-d dargestellt. Als Nukleusmarker wurde DAPI (a und b) und als Membranmarker SiR-Actin (c und d) verwendet. Die Bilder besitzen im Mittel eine Größe von  $133 \times 541 \times 548 \text{ px}^3 (z, y, x)$  und sind mit  $2 \times 0,568 \times 0,568 \mu\text{m}^3 / \text{px}^3 (z, y, x)$  aufgelöst. Aufgrund einer guten Clearing-Effizienz weisen die Proben über die 136 Ebenen nur einen sehr geringen Helligkeitsabfall auf (siehe Abbildung 4.2b und d).

Die beiden SimOptiGAN-Methoden setzen GAN-Modelle nur zur Optimierung der simulierten Nukleussignale ein. Dabei transformiert das GAN-Modell simulierte Nukleussignale in realistische Nukleussignale. Für das Training der GAN-basierten Optimierung kann daher als reales Bild direkt die Aufnahme verwendet werden, die auch zur Erstellung der Segmentierungs-Ground-Truth genutzt wurde. Diese Aufnahme ist in Abbildung 4.2e und f zu sehen. Zu beachten ist, dass diese Aufnahme aufgrund eines anderen Clearing-Protokolls einen höheren Helligkeitsabfall über die 180 Ebenen aufweist (vgl. Abbildung 4.2f). Eine detaillierte Beschreibung der Aufnahme erfolgt bei der Beschreibung der Ground-Truth Erstellung (Abschnitt 4.3.4).





**Abbildung 4.2:** A-F Beispielhafte 2D-Schnitte von realen 3D-Aufnahmen, welche als Bilder für das Training der Transformationsmodelle genutzt werden. In **A**, **C** und **E** sind  $xy$ -Schnitte dargestellt, während in **B**, **D** und **F** die entsprechenden  $xz$ -Schnitte visualisiert werden, anhand derer der Helligkeitsabfall ersichtlich ist. **A** und **B** zeigen den Nukleusmarker DAPI und **C** und **D** den Membranmarker SiR-Actin. **E** und **F** zeigen die für die Ground-Truth Erstellung genutzte Aufnahme, welche mit dem Nukleusmarker DAPI gefärbt ist. Aus dieser Aufnahme wurden drei 3D-Ausschnitte extrahiert, welche anschließend händisch annotiert wurden. Die ungefähren Positionen der drei Ausschnitte sind durch grau gestrichelte Boxen in **F** angedeutet. **I** zeigt  $xy$ -Schnitte der drei extrahierten Ausschnitte. **G** zeigt das Ergebnis einer biophysikalischen Simulation (BioSim-Instanzen) und **H** die daraus generierten synthetischen Membranen (BioSim-Membranen).

### 4.3.3 Biophysikalisch simulierte Zellgrenzen

Die in Abschnitt 2.2.3 vorgestellten Methoden verwenden die Ergebnisse von biophysikalischen Simulationen [180], um eine realistischere Anordnung der Nuklei zu erzielen. Die Ergebnisse stehen dabei in Form von Bildern zur Verfügung, welche die Zellannotationen (BioSim-Instanzen) bzw. die daraus berechneten Zellgrenzen (BioSim-Membranen) zeigen (vgl. Abbildung 4.2g).

Der *DS-BioSim*-Datensatz umfasst insgesamt vier BioSim-Instanz-Bilder. Die Simulationen wurden mit unterschiedlichen Parametersätzen durchgeführt, basieren jedoch auf derselben Startkonfiguration, die von einer realen Aufnahme einer HT29-Zellkultur abgeleitet wurde. Die Simulationsergebnisse sind damit untereinander ähnlich, zeigen aber Unterschiede in den Verläufen der Zellgrenzen. Die BioSim-Instanzen besitzen eine isotrope Auflösung von  $0,568 \times 0,568 \times 0,568 \mu\text{m}^3/\text{px}^3$ . Für die Weiterverarbeitung unterlaufen alle BioSim-Instanzen eine eigens entwickelte und im Folgenden beschriebene Vorverarbeitung, bei der kleine Annotationsfehler korrigiert werden und die Auflösung an die spätere Zielauflösung angepasst wird. Bei den unbearbeiteten BioSim-Instanzen sind häufig einzelne Inselexel zu finden, welche nicht mit der Instanz der benachbarten Pixel übereinstimmen. Daher werden Objekte, welche aus weniger als fünf Pixeln bestehen, mit der Instanz fusioniert, welche die größte angrenzende Fläche aufweist. Anschließend werden die BioSim-Instanzen skaliert, um die Zielauflösung von  $2 \times 0,568 \times 0,568 \mu\text{m}^3/\text{px}^3$  ( $z, y, x$ ) zu erhalten. Abbildung 4.2g zeigt einen 2D-Schnitt eines resultierenden Bilds von BioSim-Instanzen. Ein Teil der Methoden benötigt die BioSim-Instanzen in Membranform. Diese ergeben sich aus den Übergängen zwischen den Annotationen (vgl. Abbildung 4.2h).

### 4.3.4 Händische Ground-Truth zur Bewertung der Segmentierungsleistung

Ein wichtiges Bewertungsmaß für Methoden zur Trainingsdatensynthese stellt die Evaluierung der damit trainierten Segmentierungsmodelle auf realen Daten dar.

Hierzu wird ein händisch annotierter Ground-Truth-Datensatz benötigt. Da, wie bereits beschrieben, der händische Annotationsaufwand und dessen Komplexität für 3D-Daten enorm sind, wurde die händische Annotation auf drei Bildausschnitte begrenzt. Als Ausgangsbild wird dazu eine  $180 \times 681 \times 706 \text{ px}^3 (z, y, x)$  große Aufnahme mit einer Auflösung von  $2 \times 0,568 \times 0,568 \mu\text{m}^3/\text{px}^3 (z, y, x)$  genutzt, welche eine HT29 Zellkultur zeigt. Ein 2D-Schnitt und eine seitliche Ansicht ist in Abbildung 4.2 e und f gegeben. Die daraus extrahierten Ausschnitte weisen eine Größe von  $25 \times 128 \times 128 \text{ px}^3$  bis  $51 \times 128 \times 128 \text{ px}^3 (z, y, x)$  auf und enthalten insgesamt 1001 Zellkerne. Der Datensatz wird im Folgenden als *DS-Seg-GT* bezeichnet. Für die Annotation der Daten wurden ca. 37,5 h benötigt.

Die Größe der Ausschnitte wurde so gewählt, da sie unter anderem bei vielen Segmentierungsmodellen die Mindestbildgröße für das Training darstellt. Diese resultiert aus den bei Netzwerken verwendeten Pooling- oder Faltungsoperationen, bei denen die Bildgröße über die Schichten mehrfach verringert wird. Größere Ausschnitte sind in der Regel vorteilhafter, da sie das Verhältnis von Objekten im Inneren zu denen im Randbereich erhöhen. Objekte am Rand der Ausschnitte können durch den Zuschchnitt abgeschnitten werden, was die Segmentierung der Objekte und die Berechnung der Segmentierungsmetriken beeinträchtigen kann. Bei nur teilweise vorhandenen Objekten sind geringere Abweichungen in Position und Größe ausreichend, um zu einem Segmentierungsfehler zu führen. Die Größe wurde dennoch so gewählt, um anstelle eines großen Ausschnitts drei kleinere zu erhalten und damit unterschiedliche Bildqualitäten abzudecken. Das Problem mit den Randobjekten wird in Teilen kompensiert, indem die Segmentierung auf dem ganzen Bild ausgeführt wird und erst anschließend die relevanten Bereiche ausgeschnitten werden. Mehrere annotierte Bildausschnitte erlauben zudem einen Vergleich der synthetischen Trainingsdaten mit realen Trainingsdaten, da für das Training und das Testen unterschiedliche Bildausschnitte verwendet werden können.

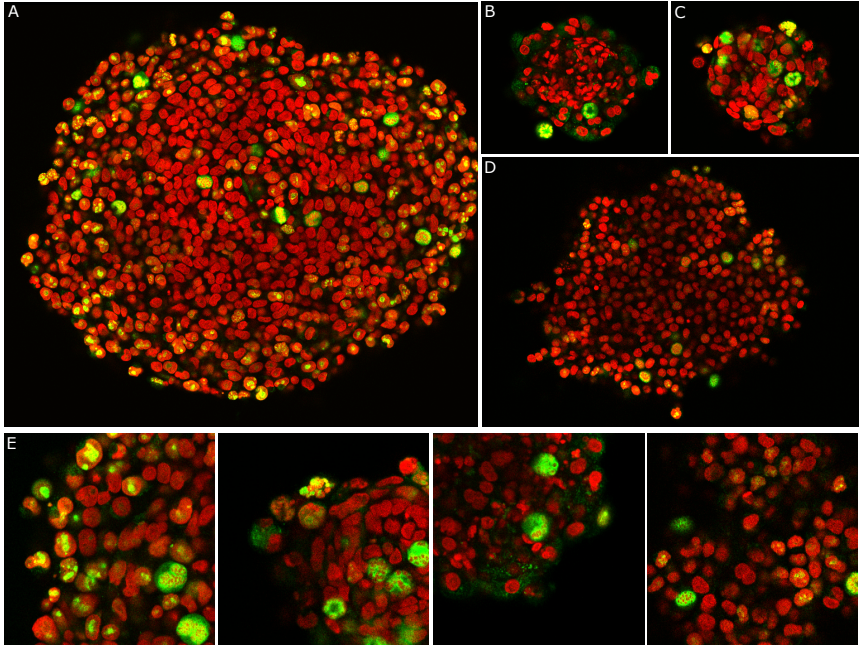
Die drei Bildbereiche wurden aus der oberen, mittleren und unteren Region des Sphäroids ausgewählt (vgl. Abbildung 4.2i), wobei die ungefähren Positionen der Ausschnitte in Abbildung 4.2f mit grau gestrichelten Kästen angedeutet sind. So weisen die Ausschnitte ein abnehmendes Signal-Rausch-Verhältnis auf (vgl.

Abbildung 4.2i). Zu beachten ist, dass die Aufnahme des Sphäroids im Vergleich zu den oben beschriebenen realen Aufnahmen eine geringere Clearing-Effektivität zeigt. Daher weist die für die Ground-Truth-Erstellung genutzte Aufnahme einen stärkeren Helligkeitsabfall auf (vgl. Abbildung 4.2b und f).

Um die Erstellung der Ground-Truth für die Zellkernsignale zu beschleunigen, wird auf eine Vorsegmentierung zurückgegriffen, welche anschließend durch einen biologischen Experten mit Annotationserfahrung korrigiert wird. Die Vorsegmentierung wurde mit dem *nuclei*-Modell von Cellpose erstellt, bei der das Bild im Gesamten segmentiert und erst danach die Segmentierung anhand der drei vorher ausgewählten Bildbereiche ausgeschnitten wird. Durch die Segmentierung des gesamten Bildes wird sichergestellt, dass keine abgeschnittenen Zellkerne vorliegen, welche von den Segmentierungsmodellen schlecht erkannt werden. Im Anschluss wurden die so erstellten Segmentierungen mittels der Segmentor-Software [196] von einem biologischen Experten händisch korrigiert. Dabei ist zu beachten, dass diese Vorgehensweise zu einem Bias bezüglich des zur Vorsegmentierung verwendeten Modells führen kann, da die korrekten Annotations-Grenzen unverändert übernommen werden.

## 4.4 Datensatz zur Untersuchung der Synthese von Bildmodalitäten

Zum Vergleich der in Abschnitt 2.3.2 vorgestellten Methoden zur Synthese von Bildmodalitäten wird ein heterogener Datensatz *DS-Marker-Syn* erstellt, welcher aus Aufnahmen von unterschiedlichen Zellkulturen besteht (siehe Abbildung 4.3a-d). Die Proben sind alle mit dem Zellkernmarker DRAQ5 (rot) und dem Proliferationsmarker Ki-67 (grün) gefärbt. Die Aufnahmen stammen aus einem Experiment, bei dem die Effektivität von zwei Wirkstoffen untersucht wurde. Die in den Zellkulturen verwendeten Zellarten sind KP4 und A549 Zellen, welche entweder als Monokultur oder zusammen mit Fibroblasten Zellen als Kokultur kultiviert wurden. Einige der Kulturen wurden mit den Wirkstoffen Paclitaxel



**Abbildung 4.3:** A-D 2D-Schnitte von beispielhaft ausgewählten 3D-Aufnahmen, welche für die Erstellung des Datensatzes *DS-Marker-Syn* genutzt wurden. Gefärbt ist der Nukleusmarker Draq5 (rot) und der Proliferationsmarker Ki-67 (grün). Für das Training werden die Aufnahmen in ihrer Intensität normiert und anschließend in 3D-Bildausschnitte aufgeteilt, von denen 2D-Schnitte in E gezeigt sind. Die verwendeten Daten sind stark heterogen, da sich unter anderem die Zelltypen unterscheiden (KP4 A, A549 D und Fibroblasten B (Mitte)) und weiterhin Wirkstoffbehandlungen Einfluss auf die Zellen nehmen (B und C).

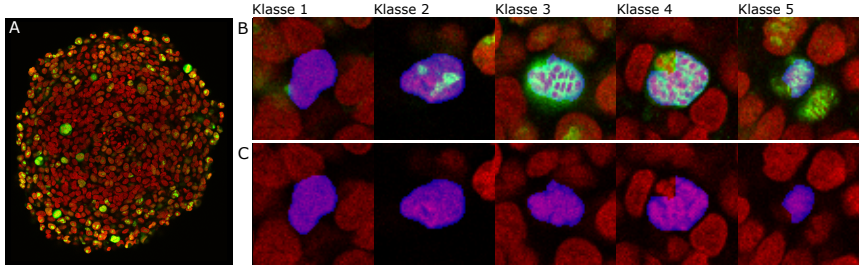
bzw. Doxorubicin behandelt, welche zu einer merklichen Reduktion der Zellanzahl führten. Insgesamt betrachtet ist der Datensatz *DS-Marker-Syn* mit den verwendeten vier unterschiedlichen Zelltypenkombinationen, den unterschiedlichen Clearing-Effizienzen und daraus resultierenden Helligkeitsunterschieden und den teilweise zusätzlichen Behandlungen der Proben stark divers. Damit stellen die Daten die ideale Grundlage für das Testen und den Vergleich der Methoden auf einem herausfordernden Datensatz dar.

Die Zellproben wurden mittels eines invertierten Leica TCS SP8 Konfokalmikroskops in einer Auflösung von ungefähr  $1 \times 0,568 \times 0,568 \mu\text{m}^3/\text{px}^3$  ( $z, y, x$ ) aufgenommen. Insgesamt besteht der Trainings- und Validierungsdatensatz aus 19 bzw. 9 Aufnahmen. Für das Training der Modelle wurden diese Aufnahmen in  $32 \times 256 \times 256 \text{ px}^3$  große Ausschnitte aufgeteilt (vgl. Abbildung 4.3e), woraus sich für Training und Validierung eine Anzahl von 2265 bzw. 1161 Ausschnitten ergibt. Vor der Aufteilung in die Ausschnitte wurden die Intensitäten der Aufnahmen pro Kanal normiert. Im späteren Experiment wird der Zielmarker Ki-67 (grün) anhand des Basismarkers Draq5 (rot) prognostiziert. Zum Testen der Modelle stehen weitere 9 Aufnahmen zur Verfügung, welche im Ganzen verarbeitet und nur intern von den Modellen in Ausschnitte aufgeteilt werden.

## 4.5 Datensatz zur Zellzyklus-Klassifikation anhand von synthetischen Markern

Wie in Abschnitt 2.3.4 beschrieben, wird in dieser Arbeit geprüft, ob die eingeführten Methoden zur Synthese von Bildmodalitäten in einem realen Anwendungsfall Vorteile bringen. Als geeignete Anwendung wird die Klassifikation von Zellzyklus-Stadien gewählt, bei der sowohl Nukleusmarker als auch der Proliferationsmarker Ki-67 verwendet werden. Für den Test wird das reale Ki-67-Signal durch ein synthetisch generiertes Signal ersetzt. Hierbei wird erwartet, dass die Klassifikationsleistung durch den Einsatz des synthetischen Signals im Vergleich zum echten Signal abnimmt. Weiterhin wird überprüft, ob die Verwendung des synthetischen Signals dennoch einen Vorteil gegenüber dem vollständigen Weglassen des Signals bietet.

Für die Untersuchungen wird ein neuer, händisch annotierter Datensatz *DS-Klassifikation* erstellt. *DS-Klassifikation* basiert auf einer realen Aufnahme, welche eine unbehandelte, also mit keinen zusätzlichen Wirkstoffen versehene Kultur, aus KP4- und Fibroblastenzellen zeigt (Abbildung 4.4a und b). Gefärbt ist die Kultur mit dem Nukleusmarker Draq5 (rot) und dem Proliferationsmarker Ki-67 (grün). Aufgenommen wurde die Probe mit einem invertierten Leica TCS SP8



**Abbildung 4.4:** A 2D-Schnitt des Sphäroids, welcher für die Erstellung des Trainingsdatensatzes (*DS-Klassifikation*) genutzt wurde. Der Nukleusmarker DraG5 ist in Rot und der Proliferationsmarker Ki-67 in Grün dargestellt. Zum Testen der Markersynthese werden drei Teildatensätze erstellt (mit realem, mit synthetischem und ohne Ki-67-Signal). B und C zeigen beispielhaft 2D-Schnitte der extrahierten Bildausschnitte zweier Teildatensätze. Zu sehen ist für jede Klasse jeweils ein Ausschnitt. B zeigt den ersten Teildatensatz (*DS-Klassifikation-Real*), bei dem die Ausschnitte den Nukleusmarker (rot), den realen Ki-67-Marker grün und die Segmentierung des zu klassifizierenden Nukleus enthalten blau. C zeigt den dritten Teildatensatz (*DS-Klassifikation-Ohne*), bei dem kein Ki-67-Signal zur Verfügung steht.

Konfokalmikroskop in einer Auflösung von ungefähr  $1 \times 0,568 \times 0,568 \mu\text{m}^3/\text{px}^3$  ( $z, y, x$ ). Die Aufnahme stammt aus der gleichen Untersuchung wie *DS-Marker-Syn*, allerdings ist diese Zellkultur nicht in *DS-Marker-Syn* vorhanden.

Um einen Vergleich zwischen dem realen und einem synthetisch generierten Ki-67-Signal zu ermöglichen, werden aus *DS-Klassifikation* drei Teildatensätze abgeleitet (*DS-Klassifikation-Real*, *-Syn* und *-Ohne*) (vgl. Abbildung 4.4c-f). Bei allen drei Teildatensätzen stehen immer das Nukleussignal und die Segmentierungsmaske der zu klassifizierenden Zelle zur Verfügung. Der erste Teildatensatz *DS-Klassifikation-Real* beinhaltet zusätzlich noch das echte Ki-67-Signal. Bei *DS-Klassifikation-Syn* wird das echte Ki-67-Signal durch eine synthetisch generierte Variante ersetzt. Beim dritten Datensatz *DS-Klassifikation-Ohne* steht kein Ki-67-Signal zur Verfügung.

Wie in Abschnitt 2.3.4 beschrieben, erfolgt die Klassifikation auf Basis von Bildausschnitten einzelner Zellkerne durchgeführt. Dazu wird für jede annotierte Zelle ein  $48 \times 64 \times 64 \text{ px}^3$  großer Bildausschnitt aus der Aufnahme extrahiert, welcher auf den Zellkernmittelpunkt zentriert ist. Die Größe wurde gewählt, da in



den Ausschnitten fast alle Zellkerne vollständig abgebildet sind. Auch andere Ausschnittgrößen und die Skalierung der Bounding Boxen einzelner Zellkerne wurden untersucht. Beides brachte allerdings keine Verbesserung bezüglich der Klassifikationsleistung. Die Helligkeit der Bildausschnitte wurde anhand der maximal bzw. minimal auftretenden Intensitäten auf einen Bereich zwischen null bis eins normiert. Auch hier wurde ein Vergleich mit einer über die Ausschnitte einheitlichen Normierung, für den Erhalt der Helligkeitsinformation, durchgeführt, wodurch sich die Prognosegüte allerdings verschlechterte.

Die händische Annotation der Klasseninformationen wurde mit der neuen grafischen Benutzerschnittstelle *3D-ClassAnnotator* (vgl. Abschnitt 3.3.2) durchgeführt. Basierend auf dem Nukleus- und Ki-67-Signal wurden mit *3D-ClassAnnotator* die Klassen von insgesamt 508 Zellkernen durch einen Zellbiologen annotiert. Die Zellkerne wurden dem Biologen dabei zufällig präsentiert, konnten aber auch übersprungen werden, um eine Überrepräsentation häufig auftretender Klassen zu vermeiden. Aufgrund der stark unterschiedlichen Auftretenswahrscheinlichkeit einzelner Zellstadien liegt dennoch eine hohe Klassenimbalance im erstellten Datensatz vor. Annotiert wurden fünf Klassen<sup>3</sup>: G0 (99 Zellen), G1/S/G2 (215 Zellen), Metaphase (55 Zellen), Anaphase (114 Zellen) und Telophase (25 Zellen). Eine solche Imbalance beeinflusst das Training in negativer Weise, da der Einfluss der seltenen Klassen auf den Loss gering ist und damit die Vorhersagegenauigkeit dieser Klassen schlechter ist. In extremen Fällen ist es für das Netzwerk bezüglich des Losses sogar effizienter, seltene Klassen vollständig zu vernachlässigen, als ihre Erkennung zu erlernen und damit die Klassifikationsgüte der anderen Klassen zu verschlechtern. Um daher den Einfluss der ungleichen Klassenverteilung während des Trainings entgegenzuwirken, werden vor dem Training die Bilder der weniger vertretenen Klassen zufällig dupliziert, um so pro Klasse die gleiche Anzahl an Bildern zu erhalten. Damit nimmt jede Klasse den gleichen Einfluss auf das Training. Für den Vergleich der Klassifikationsgüte wird der Datensatz in Trainings-, Validierungs- und Testdatensatz aufgeteilt, mit einem Anteil von jeweils 323, 77 und 102 Zellen.

---

<sup>3</sup> Eine Übersicht der einzelnen Phasen der Zellteilung ist im Anhang A.2 zu finden.



# 5 Anwendung

## 5.1 Übersicht

Zuvor wurden in Kapitel 2 Methoden für die Datensynthese vorgestellt, die sowohl neue Ansätze zur Generierung von Trainingsdaten für die Deep-Learning-basierte Segmentierung als auch zur Synthese zusätzlicher Bildmodalitäten umfassen. In diesem Kapitel wird nun eine umfassende Auswertung der präsentierten Methoden anhand der in Abschnitt 4 beschriebenen Datensätze vorgenommen.

Zunächst erfolgt die Evaluierung der Methoden zur Trainingsdatensynthese. Dabei wird analysiert, inwiefern der vorgestellte hybride Ansatz aus Simulation und GAN-basierter Transformation die Domänenlücke der Transformation verringert. Weiterhin wird aufgrund der erstmaligen Möglichkeit zur präzisen Steuerung von seltenen Elementen in synthetischen Daten eine Untersuchung durchgeführt, welche den Einfluss dieser auf die Segmentierungsqualität untersucht. Im Anschluss erfolgt ein Vergleich der vorgestellten Synthesemodelle, sowohl mit als auch ohne die erstmalige Einbringung biophysikalischer Simulationen, um deren Mehrwert im Vergleich zum Aufwand zu beurteilen. Weiterhin wird kurz auf die Übertragbarkeit der entwickelten Methodik auf andere Forschungsbereiche eingegangen.

Im zweiten Teil erfolgt die Evaluierung der Methoden zur Synthese von Bildmodalitäten. Begonnen wird mit einem Methodenvergleich zweier vorgestellter Ansätze, gefolgt von einer Validierung der Modelle bezüglich der Nutzung

von ungewünschten durchscheinenden Signalen. Daraufhin erfolgt eine Machbarkeitsanalyse für die Übertragbarkeit der Methode auf weitere Markerkombinationen. Schließlich wird ein Anwendungsszenario aus der biologischen Wirkstoffforschung vorgestellt und untersucht, in dem synthetisch generierte Signale zur Klassifikation von Zellzyklusstadien verwendet werden.

## 5.2 Synthese von Trainingsdaten

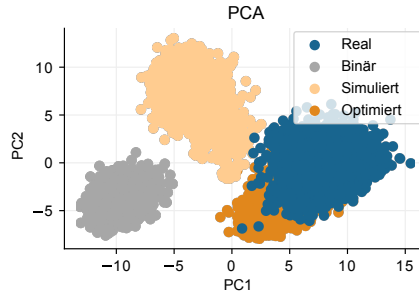
Teile der folgenden Sektion basieren auf der im Rahmen der vorliegenden Arbeit veröffentlichten Publikation von Bruch et al. [177].

### 5.2.1 Übersicht

In Abschnitt 2.2 wurden Methoden für die Synthese von Aufnahmen von großen 3D-Strukturen vorgestellt. Nachfolgend werden zuerst die Vorteile des hybriden Ansatzes der SimOptiGAN Methode aufgezeigt. Im Anschluss wird der Einfluss von seltenen Elementen in den Trainingsdaten analysiert. Im weiteren Verlauf werden die neu entwickelten Methoden zur Trainingsdatensynthese miteinander und mit dem Stand der Technik verglichen. Dabei werden Vor- und Nachteile hinsichtlich der visuellen Qualität, der Eignung als Trainingsdaten sowie des Rechen- und Ressourcenaufwands gegenübergestellt.

### 5.2.2 SimOptiGAN reduziert die Domänenlücke durch hybriden Ansatz

Bisherige Deep-Learning-basierte Methoden nutzen GAN-Modelle, um damit synthetisch generierte Annotationen in reale Bilddaten zu überführen. Die synthetischen Annotationen weisen weder Textur- noch Helligkeitsinformationen auf, was zu einer großen Lücke zur Domäne der realen Daten führt. So ist von dem



**Abbildung 5.1:** A Visualisierung der Domänenlücke anhand der Hauptkomponentenanalyse (PCA). Die für die PCA genutzten Datenpunkte stellen die anhand des Inception v3 extrahierten Features der Bildausschnitte dar.

GAN Modell eine vollständige Textursynthese durchzuführen, welche bei den heterogenen Bedingungen keine einfache Aufgabe darstellt. Eine große Domänenlücke verlängert die Trainingszeit, kann zu einem instabilen Training und auch Inkonsistenzen zwischen den transformierten und den realen Bildern führen. Die in dieser Arbeit vorgestellte Methode SimOptiGAN nutzt hingegen einen hybriden Ansatz aus einfacher Simulation und anschließender GAN-basierter Optimierung. Die vorgelagerte Simulation verringert die Domänenlücke zwischen den zu transformierenden und den realen Aufnahmen. Abbildung 5.1 zeigt einen Vergleich der Ähnlichkeit zwischen binären Daten (welche den Stand der Technik darstellen), den simulierten, optimierten und realen Daten. Anhand des Inception v3 werden äquivalent zu der KID-Metrik Merkmale extrahiert, welche mittels einer Hauptkomponentenanalyse auf zwei Dimensionen reduziert werden (siehe Abbildung 5.1a). Die Transformationsmatrix der PCA wird dabei anhand der Merkmale aller Klassen bestimmt.

Die Analyse zeigt, dass die vorherige Simulation dazu beiträgt, die Domänenlücke zu den realen Daten zu verkleinern. Während der Clustermittelpunkt der binären Annotationen eine euklidische Distanz<sup>1</sup> von 19,1 zu den realen Daten aufweist,

<sup>1</sup> Aufgrund der hohen Dimensionalität der extrahierten KID-Merkmale und der damit verbundenen Auswirkungen des Fluchs der Dimensionalität wird die Distanz basierend auf den ersten beiden PCA-Komponenten berechnet.

wird diese Distanz durch die Verwendung von simulierten Daten auf 14 reduziert. Es bleibt jedoch erkennbar, dass trotz der einfachen Simulation weiterhin Differenzen zu den realen Daten bestehen. Diese können durch die nachgelagerte Optimierung deutlich verringert werden. So reduziert die Optimierung die euklidische Distanz zu den realen Daten auf lediglich 4, 8. Ein weiterer Vorteil des hybriden Ansatzes besteht darin, dass globale Informationen, wie die Helligkeitsverteilung, bereits in den simulierten Daten enthalten sind und nicht wie bei anderen Methoden in zusätzlichen Kanälen zu kodieren sind.

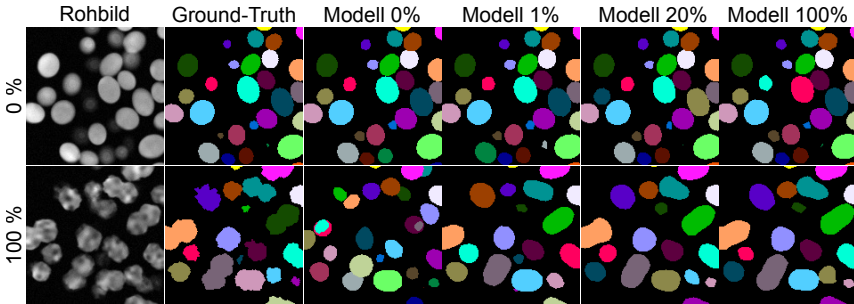
### 5.2.3 Verbesserte Segmentierung durch Einbringung seltener Elemente

Um den Einfluss von seltenen Elementen in Trainingsdaten auf die Segmentierungsqualität zu untersuchen, wurde der synthetische Datensatz *DS-Seltene-Elemente* generiert (vgl. Abschnitt 4.2). *DS-Seltene-Elemente* besteht aus sechs Datensätzen, welche unterschiedliche Verhältnisse an seltenen Elementen aufweisen: 0%, 1%, 5%, 20%, 50% und 100%. Das grundlegende Ziel dieser Untersuchung ist es herauszufinden, ob und in welchem Umfang die Einbringung von seltenen Elementen eine Verbesserung der Segmentierungsgüte ergibt. Weiterhin wird auch geprüft, ob sich durch die Einbringung seltener Elemente eine Verschlechterung der Segmentierungsgüte bei regulären Elementen ergibt. Die Segmentierungsgüte wird mit den in der Cell-Tracking-Challenge verwendeten SEG und DET Maßen festgestellt.

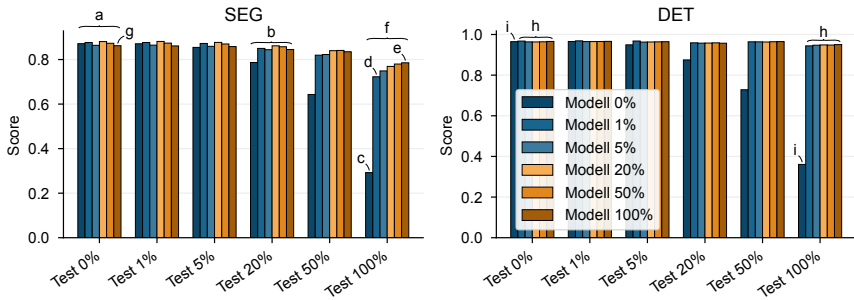
Um diese Fragestellungen zu beantworten, wird für jeden der sechs Datensätze ein Segmentierungsmodell trainiert, welches im Anschluss auf den Test-Daten aller sechs Datensätze ausgewertet wird. Zur Vereinfachung werden die Segmentierungsmodelle entsprechend den in den Trainingsdaten enthaltenen Anteilen seltener Elemente benannt. Beispielsweise wird das Segmentierungsmodell, das mit einem Datensatz trainiert wird, der einen Anteil von 5% an seltenen Elementen aufweist, als Modell 5% bezeichnet. In Abbildung 5.2 sind exemplarisch

2D-Ausschnitte der Test-Datensätze 0% und 100% (linke Spalte), die dazugehörige Ground-Truth und die Segmentierungsergebnisse von vier Modellen (0%, 1%, 20% und 100%) zu sehen. Bei dem Datensatz 0% sind die Ergebnisse der Modelle visuell fast identisch. Alle Objekte in der dargestellten Bildebene werden erkannt und korrekt segmentiert. Objekte, welche nahe der Bildebene beginnen oder enden, werden vereinzelt nicht erkannt. Dabei ist zu beachten, dass solche Fehler bei 2D-Ausschnitten von 3D-Bilddaten wenig Aussagekraft haben, da diese Objekte in den benachbarten Ebenen erkannt werden können. Eine Abweichung von einer Ebene führt zu dem Eindruck, dass ein Objekt nicht erkannt wird. Bei den Datensätzen mit seltenen Objekten (50% und 100%) unterscheiden sich die Ergebnisse der Modelle deutlicher. So ist bei den Ergebnissen des Modells 0% zu erkennen, dass viele Fehler bei den seltenen Elementen gemacht werden. Die seltenen Elemente werden häufig übersegmentiert, also fälschlicherweise in mehrere Objekte unterteilt. Das Modell 1% führt zu einer deutlichen Qualitätsverbesserung bei der Segmentierung von seltenen Elementen. Der geringe Anteil an seltenen Elementen in den Trainingsdaten ist bei diesen Daten bereits ausreichend, um eine Übersegmentierung dieser Objekte zu vermeiden. Die Ergebnisse der Modelle 20% und 100% unterscheiden sich nur in der Form der segmentierten Objekte. Der steigende Anteil in den Trainingsdaten führt zu einer geringfügigen Verbesserung in der Objektgröße und Form. Allerdings werden bei allen Modellen die ungleichmäßigen Ränder von seltenen Elementen nicht erkannt. Stattdessen besitzen alle segmentierten Objekte glatte Ränder. Da dies alle Modelle betrifft, ist hier von einer Limitierung des Segmentierungsmodells auszugehen.

Die SEG und DET Ergebnisse dieser Analyse sind in Abbildung 5.3 dargestellt. Die beschriebenen Ergebnisse sind mit korrespondierenden Buchstaben in der Abbildung referenziert. Auf dem ersten Testdatensatz 0% zeigen alle Modelle eine ähnliche Güte bei der SEG-Metrik (a). Der Durchschnittswert über alle Modelle beträgt  $0,871(\sigma = 0,007)$ . Ab einem Verhältnis von 20% an seltenen Elementen in den Test-Daten ist ein genereller Abfall in der Güte zu erkennen (b). Das Modell 0% zeigt dabei den stärksten Abfall. Bei dem Datensatz 100% wird mit diesem Modell nur noch ein Wert von  $0,292$  (c) erzielt. Interessant ist, dass schon eine kleine Anzahl an seltenen Elementen in den Trainingsdaten ausreichend ist



**Abbildung 5.2:** Exemplarische 2D-Bildausschnitte des Datensatzes mit zugehörigen Segmentierungen der StarDist-Modelle, welche auf unterschiedlichen Datensätzen mit verschiedenen Verhältnissen an seltenen Elementen trainiert wurden.



**Abbildung 5.3:** Ergebnisse mehrerer StarDist-Segmentierungsmodelle, die auf Datensätzen mit unterschiedlichen Anteilen seltener Elemente trainiert und getestet wurden. Für jeden Trainingsdatensatz wurden zwei StarDist-Modelle trainiert. Die Markierungen mittels Buchstaben dienen als Referenzen für die Beschreibung im Text.

(Modell 1%), um eine deutliche Verbesserung zu erzielen. So liegt der Wert für den Datensatz 100% bei 0,723 (d). Konträr zu den Erwartungen zeigt auch das Modell 100%, das ausschließlich auf seltenen Elementen trainiert wurde, einen abfallenden SEG-Wert bei steigenden Verhältnissen an seltenen Elementen. Bei diesem Modell (e) ist der Abfall mit einem Wert von 0,785 auf dem Datensatz 100% am geringsten.

Abgesehen von dem Testdatensatz 100% (f) scheint es keinen eindeutigen Zusammenhang zwischen der Segmentierungsgüte und der Verhältnisdifferenz zwischen

Trainings- und Testdaten zu geben. Die Modelle, welche auf den Datensätzen mit den gleichen Verhältnissen wie die späteren Testdaten trainiert wurden, zeigen in keinem der Fälle deutlich bessere Ergebnisse als die anderen Modelle. Die Ergebnisse führen zu dem Eindruck, dass die Segmentierungsqualität von regulären Objekten nicht negativ von einem steigenden Verhältnis an seltenen Elementen in den Trainingsdaten beeinflusst wird. So zeigt das Modell 100% auf dem Datensatz 0% (g) eine ähnlich gute Leistung wie die anderen Modelle. Ein Grund für dieses Ergebnis kann die unterschiedliche Komplexität der Objektklassen darstellen. Während die regulären Elemente glatte, konvexe Ränder und eine Textur mit einer minimalen Helligkeitsvariation aufweisen, besitzen die seltenen Elemente unruhige, nicht konvexe Ränder sowie eine Textur mit starken Helligkeitsunterschieden. Die reguläre Objektklasse kann somit als eine vereinfachte Version der seltenen Elemente gesehen werden. Bei komplexeren regulären Elementen ist anzunehmen, dass für eine gute Segmentierungsleistung auch eine ausreichende Repräsentation dieser Elemente im Trainingsdatensatz erforderlich ist.

Bei Betrachtung der DET Metrik zeigt sich ein ähnlicher Trend wie bei der SEG-Metrik. Allerdings ist, bis auf das Modell 0%, der Abfall bei steigendem Verhältnis an seltenen Elementen sehr gering. Das Modell 0% ausgenommen liegt der Unterschied zwischen Datensatz 0% und Datensatz 100% bei den Modellen im Mittel bei 0,017 (h). Das Modell 0% zeigt zwischen diesen Datensätzen einen Abfall von 0,604 (i). Insgesamt decken sich die Ergebnisse der Maße SEG und DET mit den visuellen Eindrücken.

### **5.2.4 Eingeführte Methoden zeigen Vorteile in Aufwand und Segmentierungsgüte**

In diesem Abschnitt werden die in Abschnitt 2.2 eingeführten Methoden zur Synthese von Mikroskopiedaten verglichen. Das Training und die Auswertung der Modelle werden auf dem in Abschnitt 4.3 vorgestellten Datensatz *DS-Train-Syn* durchgeführt.

**Tabelle 5.1:** Vergleich der zur Generierung von 3D-Trainingsdaten benötigten Zeit. Die Spalte *Manuelle Annotation* beschreibt den zeitlichen Aufwand, der zur manuellen Generierung oder zur Korrektur von Annotationsmasken erforderlich ist. Die Spalte *Rechenzeit Training* und *Rechenzeit Anwendung* umfassen die Rechenzeit, die für die Generierung synthetischer Bilder erforderlich ist. Während das Training einmalige Berechnungen wie die Parameterabschätzung für die biophysikalische Simulation und das Training von GAN-basierten Modellen enthält, besteht die Anwendung aus der verbleibenden Rechenzeit, welche für die Anwendung der Modelle für die Generierung von vier synthetischen 3D-Bildern von Zellkulturen erforderlich ist.

Methode	Man. Annotation	Rechenzeit Training	Rechenzeit Anwendung
Manuelle Annotation	37,5 h	-	-
SimOptiGAN	0,4 h	99,5 h	<0,1 h
SimOptiGAN+	4,4 h	112,5 h	148 h
Mem2NucGAN-P	4 h	33,5 h	148 h
Mem2NucGAN-U	4 h	72,5 h	148 h

### Aufwandsvergleich

Durch die Nutzung der vorgestellten Methoden für die Generierung synthetischer Daten können Nutzer den manuellen Annotationsaufwand erheblich reduzieren, der für die Erzeugung einer ausreichenden Menge an Trainingsdaten erforderlich ist. Ganz nach dem no-free-lunch-Theorem wird der Zeitaufwand für die manuelle Annotation verringert, allerdings ist dafür GPU-Hardware und Rechenzeit erforderlich. Die Rechenzeit ist in vielen Fällen vernachlässigbar, da die Berechnungen zu beliebigen Zeiten und ohne menschliches Zutun durchgeführt werden können. Ein Vergleich der Zeiten, die für die Generierung synthetischer Trainingsdaten mit den vorgeschlagenen Methoden benötigt werden, ist in Tabelle 5.1 gegeben <sup>2</sup>.

Für die Erzeugung realistischer Zellgrenzen durch biophysikalische Simulation wurde zunächst eine einmalige Segmentierung der Membransignale durchgeführt.

<sup>2</sup> Die initiale Membransegmentierung sowie die nachfolgende Parameterabschätzung und biophysikalische Simulation wurden auf einer Workstation durchgeführt, die mit einem AMD Ryzen 9 5950X Prozessor, 128 GB RAM und einer NVIDIA GeForce RTX 3060 Ti 8 GB GPU ausgestattet ist. Alle übrigen Berechnungen wurden auf einem Servercomputer durchgeführt, der mit einem AMD EPYC 7252 Prozessor, 64 GB RAM und einer NVIDIA A6000 48 GB GPU ausgestattet ist.



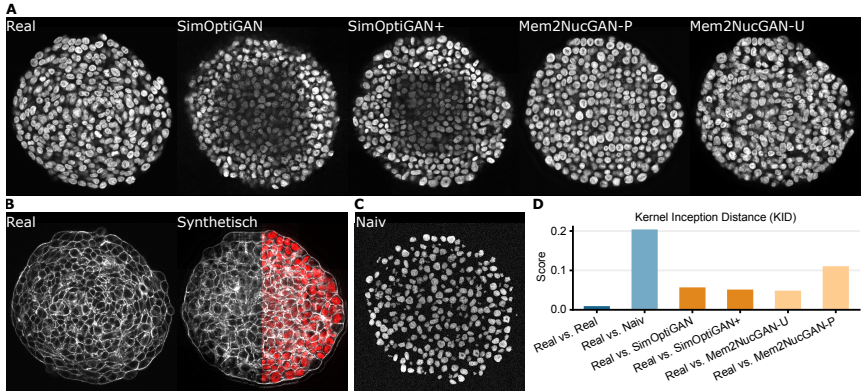
Hierfür wurde das vortrainierte *cyto2*-Modell von Cellpose mithilfe eines Human-in-the-Loop-Annotationsverfahrens optimiert, wofür 4 Stunden erforderlich waren. Auf Basis der segmentierten Membranen ist im Anschluss eine automatische Parameterschätzung für die Parametrierung der biophysikalischen Simulation erforderlich. Hierfür wurde eine Rechenzeit von 13 Stunden benötigt. Sowohl die Membransegmentierung als auch die Abschätzung der Simulationsparameter stellen einmalige Prozesse dar, die für jeden Zelltyp einmalig durchzuführen sind. Nach diesen anfänglichen Schritten kann der Benutzer mit den extrahierten Parametern eine beliebige Anzahl von simulierten Zellgrenzen erzeugen. Die Berechnung der eigentlichen biophysikalischen Simulation benötigte eine Rechenzeit von 148 Stunden für die Generierung von vier 3D-Zellgrenzbildern. Diese Zellgrenzbilder dienen als Grundlage für die anschließende Generierung von realistischen Kern- und Membransignalen. Da SimOptiGAN keine biophysikalisch simulierten Zellgrenzen verwendet, sind die zugehörigen Annotations- und Rechenzeiten für diese Methode nicht relevant.

Für die Generierung von Kernsignalen mit SimOptiGAN und SimOptiGAN+ ist eine manuelle Annotation einiger weniger Zellkerne erforderlich, um die Objektdatenbank zu befüllen. Für die Auswertung der beiden Methoden wurden nur 10 Nukleusprototypen annotiert, was 22,5 min erforderte. Die meisten Ressourcen für SimOptiGAN und SimOptiGAN+ wurden für das Training des Optimierungnetzwerks benötigt (99,5 h und 42 GB GPU-Speicher). Für die anderen Teile des SimOptiGAN- und SimOptiGAN+-Pipelines (Probengenerierung und Aufnahmesimulation) waren nur 0,5 GB RAM und eine Rechenzeit von weniger als zwei Minuten erforderlich. Für den Vergleich der Methoden wurde auf eine höher aufgelöste Simulation verzichtet und die Berechnungen von SimOptiGAN (und SimOptiGAN+) direkt in der gewünschten Ausgabeauflösung durchgeführt. Dadurch wird die benötigte Rechenzeit und der Speicherverbrauch drastisch reduziert. Am stärksten wird dadurch die Qualität der PSF Faltung beeinträchtigt, was allerdings durch die anschließende GAN-basierte Optimierung kompensiert wird.

Für Mem2NucGAN-P und Mem2NucGAN-U ist keine zusätzliche manuelle Annotation erforderlich, jedoch sind auch bei diesen beiden Methoden GAN-basierte Modelle zu trainieren. Das Training des Transformationsnetzwerks Mem2NucGAN-P erforderte 20,5 h und 16,5 GB GPU-Speicher, während Mem2NucGAN-U 59,5 h und 31 GB GPU-Speicher benötigte.

## Optische Eigenschaften

Ein menschlicher Ansatz zur Bewertung synthetischer Bilddaten ist der visuelle Vergleich mit den realen Vorbildern. Abbildung 5.4a zeigt 2D-Ausschnitte der mit den verschiedenen Methoden generierten synthetischen Aufnahmen im Vergleich zu dem Ausschnitt einer realen Aufnahme. Die Synthesemethoden zeigen Unterschiede in der Morphologie und Anordnung der Zellkerne sowie Unterschiede in Helligkeit und Textur. Die Zellkerne in den mit SimOptiGAN generierten Daten, also ohne biophysikalische Modellierung, wirken weniger scharf als die realen Zellkerne und weisen zudem eine gleichmäßigere Textur auf. Durch die Einbeziehung der biophysikalischen Modellierung (SimOptiGAN+) ergibt sich eine stärker strukturierte Anordnung der Kerne. So existieren etwa kleinere Bereiche, an denen keine Zellkerne vorhanden sind. Die Bilder beider SimOptiGAN Methoden zeigen zudem eine geringere Helligkeit in den mittleren Regionen. Der Helligkeitsabfall wurde in der Aufnahmesimulation gezielt eingebracht, da auch die reale Aufnahme, die zur Bestimmung der Segmentierungsleistung verwendet wird, einen entsprechenden Helligkeitsabfall aufweist. SimOptiGAN+ zeigt dabei einen harten Übergang in der Helligkeit. Dies ist auf die patchbasierte Verarbeitung zurückzuführen. Dabei ist zu erwähnen, dass für SimOptiGAN+ kein zusätzliches Optimierungsnetzwerk trainiert, sondern das bestehende von SimOptiGAN genutzt wurde. Es ist zu erwarten, dass mit einem eigenen Optimierungsnetzwerk ein gleichmäßigerer Übergang erzielt werden kann. Die mit Mem2NucGAN-P erzeugten Daten weisen eher große und rundliche Zellkerne auf. Zudem ist ein leichtes Schachbrettmuster in der Textur erkennbar. Bei den Bildausschnitten der Mem2NucGAN-U Methode sind bei einigen Zellkernen keine klaren Umrisse zu erkennen, wodurch diese mit benachbarten Kernen verschmelzen.



**Abbildung 5.4:** Vergleich von realen und synthetisch generierten Bildern. **A** zeigt repräsentative Bildausschnitte von realen und synthetischen 3D-Nukleusaufnahmen. Bei SimOptiGAN erfolgt die Anordnung der Zellkerne zufällig, während SimOptiGAN+, Mem2NucGAN-P und Mem2NucGAN-U eine biophysikalische Modellierung zur realistischen Anordnung der Kerne genutzt wird. Im Gegensatz zu den anderen Methoden weisen die Bildausschnitte der SimOptiGAN(+) Methoden einen Helligkeitsabfall zur Mitte der Zellkultur auf, welcher bei der Generierung gezielt eingebracht wird. Bei SimOptiGAN+ ergibt sich aufgrund der Nutzung eines vortrainierten Optimierungsmodells ein harter Helligkeitsübergang. **B** zeigt Bildausschnitte von realen und synthetischen Membransignalen. Das synthetische Membransignal wird auf Basis der gleichen simulierten Zellgrenzen erzeugt, welche auch in SimOptiGAN+, Mem2NucGAN-P und Mem2NucGAN-U verwendet werden. Dadurch weist das Membransignal und das Nukleussignal eine konsistente Zellanordnung auf, wie das Overlay von synthetischen Membranen mit durch SimOptiGAN+ generierten Zellkernen (rot) zeigt. **C** zeigt einen Ausschnitt von naiv erzeugten Daten<sup>4</sup>, welche als Referenzbeispiel für die KID-Bewertung in D dienen. **D** zeigt die Ergebnisse des Vergleichs von synthetischen Nukleussignalen mit den realen Vorbildern anhand der Kernel-Inception-Distance (KID). Niedrigere Werte deuten auf eine größere Ähnlichkeit zwischen realen und synthetischen Signalen hin. Übernommen und übersetzt aus [177].

In Abbildung 5.4b sind zudem Bildausschnitte einer realen und einer synthetisch generierten<sup>3</sup> 3D-Membranaufnahme gezeigt. Da die Methode die gleichen biophysikalisch simulierten Zellgrenzen nutzt, wie die anderen Methoden zur Generierung von Nukleussignalen stimmen die Strukturen überein, was eine Kombination

<sup>3</sup> Die synthetische Aufnahme wurde wie in Abschnitt 2.2.3 beschrieben auf Basis der BioSim-Membranen generiert.

der beiden Signale ermöglicht. Veranschaulicht wird dies durch die Überlagerung der synthetischen Membran mit den mit SimOptiGAN+ synthetisch generierten Zellkernen (rot) (vgl. Abbildung 5.4b).

Die visuellen Unterschiede, die ein Mensch wahrnimmt, korrelieren nicht zwangsläufig mit denen, die für ein Segmentierungsmodell wichtig sind. Daher wird als zusätzlicher Vergleich die Kernel-Inception-Distance (KID) angegeben (vgl. Abbildung 5.4d), welche auf Differenzen im Merkmalsraum eines Netzwerks basiert und sich damit auf Merkmale konzentriert, welche eine größere Relevanz für die Segmentierungsaufgabe besitzen. Niedrigere Distanzwerte stehen für eine größere Ähnlichkeit zwischen realen und synthetischen Bildern und somit für ein besseres Ergebnis. Zur besseren Einordnung der Ergebnisse wird zusätzlich noch der KID-Score zwischen zwei realen Aufnahmen bzw. zwischen realen und naiv<sup>4</sup> (vgl. Abbildung 5.4c) generierten Aufnahmen angegeben, was als bestmögliches Ergebnis bzw. als suboptimales Ergebnis angesehen werden kann. Die naiven Daten sind dahingehend suboptimal, da hier keine Signalstreuung der PSF simuliert und keine Hintergrundsignale enthalten sind. Bei dem Vergleich anhand der KID zwischen den generierten Aufnahmen und den realen Vorbildern ergeben sich nur geringe Unterschiede zwischen SimOptiGAN, SimOptiGAN+ und Mem2NucGAN-U (vgl. Abbildung 5.4d). Hingegen zeigt die Methode Mem2NucGAN-P einen fast doppelt so hohen KID-Wert. Im Vergleich zu dem naiven Ansatz zeigen allerdings alle Methoden deutlich geringere Distanzen zu den realen Vorbildern. Abgesehen von den zentralen dunklen Bereichen in den Zellkernbildern von SimOptiGAN und SimOptiGAN+, scheint das KID-Maß gut mit dem visuellen Eindruck zu korrelieren.

---

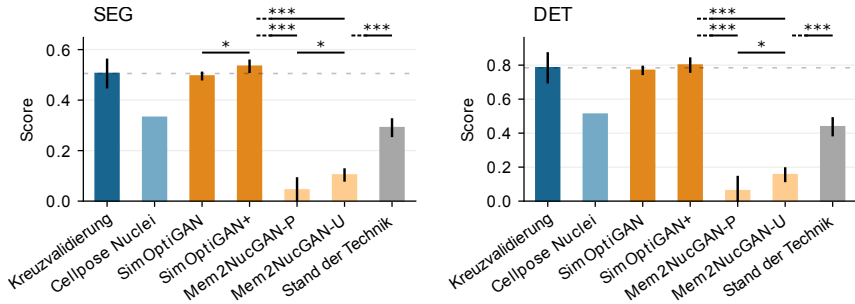
<sup>4</sup> Die naiven Bilder stellen Zwischenergebnisse von SimOptiGAN dar, welche die Aufnahmesimulation ohne die Simulation der PSF durchlaufen haben und zudem nicht optimiert wurden. Sie stellen damit Daten dar, welche mit einfachen Simulationsmethoden generiert werden können.

## Segmentierungsvergleich

Der zentrale Anwendungsfall für synthetische Daten besteht darin, das Training von 3D-Segmentierungsmodellen mit einer großen Menge an Daten zu ermöglichen. Daher werden die eingeführten synthetischen Methoden sowie eine dem Stand der Technik entsprechende Methode in diesem Abschnitt auf ihre Eignung als Trainingsdaten für Segmentierungsmodelle hin verglichen. Zusätzlich wird ein Vergleich mit händisch annotierten Daten und dem verbreiteten vortrainierten Segmentierungsmodell Cellpose *nuclei* durchgeführt. Die händisch generierten Trainingsdaten stellen eine obere Grenze dar. Das vortrainierte Cellpose-Modell stellt eine Referenz dar, die bereits ohne zusätzlichen Aufwand erreicht werden kann.

Die Bewertung der Methoden erfolgt anhand der in Abschnitt 2.2.4 beschriebenen Vorgehensweise. Eine detaillierte Beschreibung der Datengenerierung mittels der Modelle inklusive einer Beschreibung des Vorgehens für die Segmentierung findet sich im Anhang A.3. Als Ground-Truth wird der Datensatz *DS-Seg-GT* verwendet, welcher aus drei händisch annotierten Ausschnitten einer realen Mikroskopaufnahme von HT29 Zellkernen besteht und insgesamt 1001 annotierte Nuklei umfasst. Für die Erstellung der Ground-Truth durch einen Biologen wurden ca. 37,5 h benötigt. Das Training der Segmentierungsmodelle, ausgenommen des vortrainierten Cellpose *nuclei*-Modells, wird sechsmal wiederholt und die mittlere Leistung der Modelle angegeben.

Abbildung 5.5 zeigt die Segmentierungsqualität der Modelle, welche auf verschiedenen Trainingsdaten trainiert wurden. *Kreuzvalidierung* bezeichnet die Ergebnisse, welche durch die Verwendung eines Leave-One-Out-Ansatzes auf dem Ground-Truth-Datensatz erzielt wurden. Das heißt, jeder der drei Ausschnitte wurde einmal als Testset verwendet, während das Modell auf den verbleibenden zwei Ausschnitten trainiert wurde. Die mit der *Kreuzvalidierung* erhaltenen Ergebnisse dienen als Referenz für synthetische Trainingsdaten und sind mit einer gestrichelten horizontalen Linie gekennzeichnet. Cellpose *Nuclei* [65] stellt ein vortrainiertes Segmentierungsmodell für Zellkerne dar, welches auf öffentlich verfügbaren Ground-Truth-Datensätzen trainiert wurde. Für die Inferenz des Modells



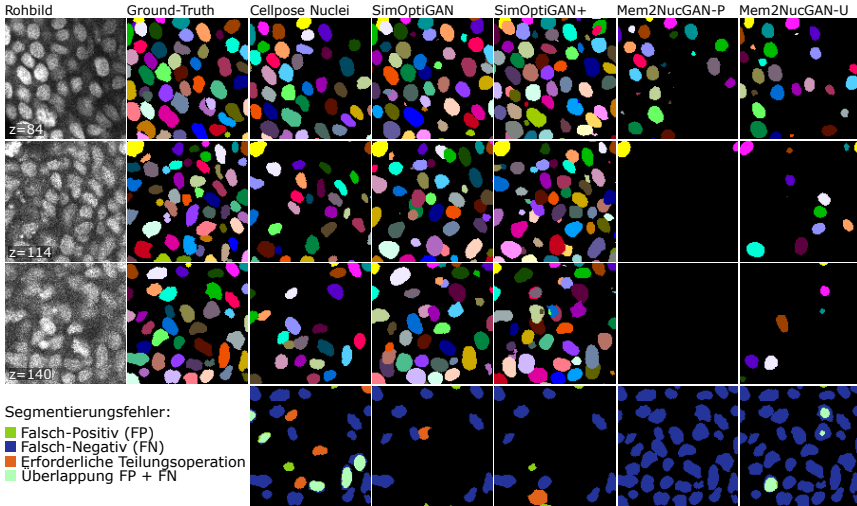
**Abbildung 5.5:** Ergebnisse der Nuklei Segmentierungsmodelle bezüglich der SEG und DET Metrik (Wertebereich: 0-1). Gegeben ist der Mittelwert und die Standardabweichung von sechs Trainingswiederholungen. Drei händisch korrigierte Segmentierungen von Bildausschnitten dienen als Testdaten. Die Bezeichnungen geben die Methode der Trainingsdatengenerierung bzw. den Namen des vortrainierten Segmentierungsmodells an. Die Balkenfarbe kennzeichnet Methodengruppen: Dunkelblau - händische Annotation, Hellblau - vortrainiertes Segmentierungsmodell, Dunkelorange - hybride Synthesemethode, Hellorange - GAN-basierte Synthesemethode und Grau - Stand der Technik. Orange Balken stellen die eingeführten Methoden zur Generierung von Trainingsdaten dar. *Kreuzvalidierung* beschreibt die Leave-One-Out-Kreuzvalidierung auf den Testdaten. *Cellpose Nuclei* stellt ein vortrainiertes Segmentierungsmodell dar. Bei *Stand der Technik* werden Annotationsdaten direkt mit einem GAN transformiert. Eine Beschreibung der statistischen Signifikanzanalyse findet sich im Anhang A.5. Übernommen und übersetzt aus [177].

wurde nur der Durchmesser an die Größe der Nuklei in den Ground-Truth-Daten angepasst. *Stand der Technik* stellt die Methode zur Trainingsdatengenerierung dar, bei der Daten durch die direkte GAN-basierte Transformation von Annotationsdaten generiert werden. Die dazu genutzten Annotationen stellen die Annotationsdaten der SimOptiGAN Methode dar, welche mit Rauschen und dem gleichen Helligkeitsabfall beaufschlagt wurden. Die übrigen Modelle wurden alle mit synthetischen Datensätzen trainiert. *SimOptiGAN* wurde mit der eingeführten Synthesemethode generiert, die keine biophysikalische motivierte Zellanordnung nutzt. Im Gegensatz dazu enthalten die Datensätze mit den Bezeichnungen *SimOptiGAN+*, *Mem2NucGAN-P* und *Mem2NucGAN-U* die biophysikalische motivierte Modellierung der Zellanordnung.

Der Vergleich zeigt deutlich, dass die Qualität der Segmentierung signifikant von den verwendeten Trainingsdaten abhängig ist. Der unter Verwendung der *Kreuzvalidierung* erzielte Wert liegt bei 0,505. *Cellpose Nuclei* erreicht einen SEG-Wert von 0,332. Das *SimOptiGAN*-Modell erzielt einen SEG-Wert von 0,495, welcher durch die Integration einer biophysikalischen Modellierung (*SimOptiGAN+*) auf 0,534 verbessert wird, was das beste Ergebnis darstellt. Sowohl *Mem2NucGAN-P* als auch *Mem2NucGAN-U* zeigen deutlich schlechtere SEG-Ergebnisse mit Werten von 0,044 bzw. 0,103. Bei Verwendung der *SimOptiGAN*-Annotationen als Grundlage für die Transformation (*Stand der Technik*) wird ein Wert von 0,291 erreicht. Der Vergleich mit dem Stand der Technik verdeutlicht die Überlegenheit der *SimOptiGAN(+)*-Ansätze, welche einen hybriden Ansatz aus Bildsimulation und GAN-basierter Optimierung nutzen.

Die DET-Werte aller Modelle zeigen im Vergleich zu den SEG-Werten einen leichten Anstieg, da hier nur Detektionsfehler bestraft werden. Die Unterschiede zwischen den Modellen bleiben allerdings konstant. Das DET Ergebnis der *Kreuzvalidierung* beträgt 0,785. *Cellpose Nuclei* und der *Stand der Technik* erreichen einen DET-Wert von 0,512 bzw. 0,404. *SimOptiGAN+* verbessert das Ergebnis erneut von 0,769 (*SimOptiGAN*) auf 0,8. Sowohl *Mem2NucGAN-P* als auch *Mem2NucGAN-U* zeigen wiederum die schlechtesten Ergebnisse mit Werten von 0,061 bzw. 0,155. Der Stand der Technik erzielt ein Ergebnis von 0,437.

In Abbildung 5.6 werden die Segmentierungsergebnisse im Vergleich zum *Rohbild* und der *Ground-Truth* dargestellt. Gezeigt wird ein 2D-Schnitt für jeden der drei *Ground-Truth*-Ausschnitte. Die Positionen der drei *Ground-Truth* Ausschnitte sind so gewählt, dass diese unterschiedliche Bildbereiche zeigen: den oberen, mittleren und unteren Bereich der 3D-Aufnahme. Die Helligkeit und das Signal-Rausch-Verhältnis nehmen ab, was eine Herausforderung für die Segmentierungsmodelle darstellt. Um die Bildinhalte besser erkennen zu können, sind die gezeigten Bilder aufgehellt dargestellt. Auf den ersten Blick ist zu erkennen, dass die Modelle *Cellpose Nuclei*, *Mem2NucGAN-P* und *Mem2NucGAN-U* Schwierigkeiten haben, Kerne in dunkleren Bildbereichen zu erkennen. Mit dem Segmentierungsmodell *Mem2NucGAN-P* werden nur zwei Kerne im zweiten Ausschnitt und keine im



**Abbildung 5.6:** Qualitativer Vergleich der Segmentierungsergebnisse. Zur besseren Visualisierung wird nur ein repräsentativer optischer Schnitt der drei Ground-Truth-Ausschnitte gezeigt. Die erste und zweite Spalte visualisiert das *Rohbild* und die entsprechende *Ground-Truth*, während die angrenzenden Spalten die Segmentierungsergebnisse der Modelle zeigen. In der letzten Zeile sind die DET-bezogenen Fehler der dritten Zeile markiert, einschließlich falsch-negativer, falsch-positiver und erforderlicher Teilungsoperationen. Eine Visualisierung der DET-Fehler über alle Ground-Truth-Ausschnitte hinweg findet sich im Anhang Abbildung A.4. Übernommen und übersetzt aus [177].

dritten Ausschnitt erkannt. Dies entspricht den Ergebnissen des DET-Maßes. Die Modelle *SimOptiGAN* und *SimOptiGAN+* erkennen die meisten Kerne.

## Diskussion

Die eingeführten Methoden ermöglichen die Generierung von synthetischen Membran- und Nukleussignalen und erlauben auch erstmals die Integration von biophysikalischer Simulation in den Syntheseprozess. Diese Basis ermöglicht es, zueinander passende Membran- und Nukleussignale zu generieren und so synthetische Trainingsdatensätze mit zwei Markern zu erstellen. Die Kombination beider Signale im Training von Segmentierungsnetzwerken kann, wie die Experimente zeigen, die Segmentierungsqualität von Membransignalen steigern.



Mit der erweiterten GAN-Struktur ist es zudem bei der Bild-zu-Bild-Transformation erstmals möglich, neben dem Zielsignal auch die dazu passenden Annotationen zu generieren. Damit lassen sich auf Basis von Membransignalen zum ersten Mal direkt mit GANs Nukleussignale und deren Annotationen generieren, welche sich für das Training von Segmentierungsmodellen eignen. In einer durchgeführten Ablationsstudie (vgl. Anhang A.4), bei der einzelne Komponenten des Ansatzes zur Prüfung ihres Einflusses weggelassen werden, zeigte sich, dass ohne den zusätzlichen Diskriminator keine nutzbaren Annotationen generiert werden. Weiterhin wird in diesem Abschnitt auch ein zuvor entwickelter Ansatz diskutiert.

Die Generierung von Markersignalen und deren Annotationen kann potenziell auch in die andere Richtung angewendet werden, um aus Nukleussignalen Membransignale und deren Annotationen zu generieren. Diese Neuheit ist auch für andere Anwendungsbereiche, wie das virtuelle Staining, von potenziellem Interesse. In diesem Bereich ermöglicht die eingeführte GAN-Struktur die Erzeugung synthetischer Markersignale, wie dem Proliferationsmarker Ki-67 inklusive der korrespondierenden Annotationsmasken.

Die erstmalige Einbringung der biophysikalischen Modellierung in die Synthese von Zellkernaufnahmen (SimOptiGAN+) führte im Vergleich zu einer zufälligen Platzierung (SimOptiGAN) zu einer visuellen Verbesserung der Zellkernanordnung (vgl. Abbildung 5.4). Die visuelle Inspektion der synthetischen Bilder offenbart auch deutliche Unterschiede in der Helligkeitsverteilung. Synthetische Aufnahmen, welche mit SimOptiGAN und SimOptiGAN+ generiert wurden, weisen eine abnehmende Helligkeit in tieferen Schichten des Sphäroids auf. Hingegen zeigen die mit den Mem2NucGAN-Modellen generierten Aufnahmen eine nahezu konstante Helligkeitsverteilung über den gesamten Bildinhalt. Eine mögliche Erklärung für diese Diskrepanz kann in der Nutzung unterschiedlicher Trainingsdatensätze für das Optimierungsmodell von SimOptiGAN(+) und den Transformationsmodellen von Mem2NucGAN liegen. Die realen Aufnahmen in den Trainingsdatensätzen der Mem2NucGAN-Modelle zeigen nur eine minimale Abnahme in der Intensität in tieferen Schichten. Zusätzlich führt das Fehlen von Helligkeitsinformationen in den zur Transformation genutzten simulierten Membransignalen zu einer Herausforderung. In Kombination mit dem patchbasierten

Transformationsansatz führt dies dazu, dass die Mem2NucGAN-Modelle nicht in der Lage sind, eine abnehmende Helligkeit zu berücksichtigen. In einer im vorherigen Abschnitt beschriebenen zusätzlichen Untersuchung wurde getestet, ob die Einbringung des Helligkeitsabfalls in die simulierten Membransignale zu einer abnehmenden Helligkeit in den transformierten Daten führt. Dies zeigte allerdings nicht den gewünschten Effekt. Zu vermuten ist, dass der bei den realen Vorbildern fehlende Helligkeitsabfall dazu führt, dass der Generator diese Information ignoriert, da dunklere Bilder vom Diskriminator zu einfach als synthetisch erkannt werden können. Dadurch stehen beim Training der Segmentierung keine dunkleren Bildbereiche mit schlechterer Signalqualität zur Verfügung, was die Segmentierung solcher Bereiche in realen Bildern beeinträchtigt.

Die Diskrepanzen in der Helligkeitsverteilung der synthetischen Aufnahmen beeinflussen zweifellos die Segmentierungsqualität von Zellkernen in dunkleren Bildregionen. Wie bei der Begutachtung der Segmentierungsmasken (vgl. Abbildung 5.6) zu erkennen ist, zeigen die Modelle, welche mit den synthetischen Daten der Mem2NucGAN-Modelle trainiert wurden, Schwierigkeiten bei der Segmentierung von dunklen Bildbereichen. Die Modelle, welche mit generierten Daten der SimOptiGAN Methoden trainiert wurden, schnitten in den dunkleren Bildausschnitten wesentlich besser ab, was darauf hindeutet, dass die Helligkeitsreduktion die Segmentierungsleistung verbessert. Das KID-Maß scheint von Unterschieden in der Bildhelligkeit unbeeinflusst zu sein, was einen Teil der Diskrepanz zwischen KID und den Segmentierungsergebnissen erklären kann. Ein Grund für die Helligkeitsinvarianz von KID kann das zugrunde liegende Inception-Netz sein, das auf einem vielfältigen Klassifizierungsdatensatz trainiert wurde, bei dem Helligkeitsinformationen für das Klassenergebnis irrelevant sind.

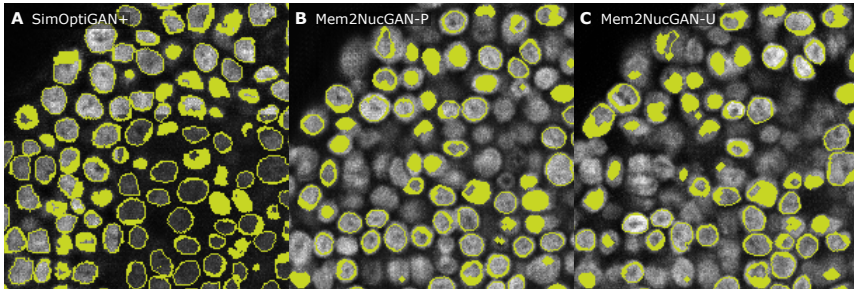
Ein weiterer Faktor für das schlechte Abschneiden der Mem2NucGAN-Modelle kann auf den Annotationsgenerierungsprozess zurückzuführen sein. Da die binären Annotationen nur eine zusätzliche Ausgabe der Generatornetzwerke sind, ist möglicherweise nicht in allen Fällen eine strikte Übereinstimmung zwischen Bild- und Annotationsdaten gegeben. Obwohl der Nachbearbeitungsschritt, der



**Abbildung 5.7:** 2D-Visualisierung der mittels Mem2NucGAN-P generierten Aufnahme (A) und generierten binären Annotationen (B) inklusive der mittels des Nachbearbeitungsschritts bestimmten Instanzannotationen (C). Der markierte Bereich zeigt, dass durch die Prozessierung der Annotationen Inkonsistenzen zwischen der generierten Aufnahme und den Instanzannotationen entstehen können.

zur Ableitung von Instanzannotationen verwendet wurde, die Qualität der Annotationsdaten visuell verbessert, kann er potenziell auch zusätzliche Diskrepanzen zwischen Annotationen und generierten Bilddaten einführen (vgl. Abbildung 5.7). Die Tatsache, dass trotz der schlechten Segmentierungsergebnisse der Mem2NucGAN-U Methode die KID-Ergebnisse insgesamt am besten sind, kann auf eine mögliche Inkonsistenz der Annotationen hinweisen. Da KID die Konsistenz der Annotationen nicht berücksichtigt, kann ein Modell gute KID-Ergebnisse und schlechte Segmentierungsergebnisse erzielen, wenn die generierten Beschriftungen nicht mit den entsprechenden synthetischen Bildern übereinstimmen. Zur visuellen Überprüfung der Konsistenz der Annotationen findet sich in Abbildung 5.8 eine Darstellung der generierten Nukleussignale bei denen die korrespondierenden Annotationen überlagert dargestellt sind. In der Abbildung zeigt sich, dass bei den Mem2NucGAN-Methoden vereinzelt Annotationen für generierte Nuklei fehlen. Allerdings ist bei der Abbildung zu berücksichtigen, dass es sich um 3D-Daten handelt, bei denen die Annotationen auch erst in angrenzenden Bildebenen beginnen können.

Auch die zusätzliche Membransegmentierung, welche für die Methode Mem2NucGAN-P erforderlich ist, kann die Aufgabe der Transformation aufgrund von Segmentierungsfehlern beeinträchtigen. Es wurden auch Ansätze getestet, welche keine Segmentierung der Membran erfordern und stattdessen die simulierten binären Membransignale an die realen Signale anpassen. Damit kann das Training des GANs auf unbearbeiteten realen Signalen erfolgen. Dabei wurden mehrere Varianten der Signalvorverarbeitung getestet, darunter die Bildsimulation und ein



**Abbildung 5.8:** Vergleich der Konsistenz zwischen generierten Bildern und zugehörigen generierten Annotationen. **A-C** zeigen Bildausschnitte der erzeugten Nukleussignale mit den entsprechenden Annotationen (gelbe Umrisse) **A** SimOptiGAN, **B** Mem2NucGAN-P und **C** Mem2NucGAN-U. Während die Annotationen von SimOptiGAN+ konsistent zu den Bilddaten sind, scheinen bei Mem2NucGAN-P und Mem2NucGAN-U einige Inkonsistenzen zwischen Bilddaten und Annotationen vorhanden zu sein. Dabei ist zu beachten, dass die Annotationen der Mem2NucGAN Methoden, insbesondere entlang der  $z$ -Achse, kompakter sind als die der SimOptiGAN Methode. Das kann den Eindruck erwecken, dass bei einigen Kernen die Annotation fehlt, welche jedoch in den benachbarten Schichten sichtbar wird. Übernommen aus [177].

GAN-basiertes Transformationsmodell zur Generierung realistischer Membransignale, allerdings konnten keine Zellkerne mit realistischen Texturen erhalten werden.

Der Vergleich des hybriden Ansatzes von SimOptiGAN, der eine Kombination aus Bildsimulation und GAN-basierter Optimierung nutzt, mit dem Stand der Technik, bei der synthetische Annotationsdaten direkt mit einem GAN in die reale Domäne überführt werden, zeigte, dass der SimOptiGAN Ansatz signifikant bessere Segmentierungsergebnisse erzielte. Der Ansatz der direkten Transformation der Annotationsdaten kann auch als Alternative zu Mem2NucGAN-U betrachtet werden, bei dem die Aufgabe der Transformation durch die vordefinierte Form und Position der Zellkerne vereinfacht wird. Dabei geht jedoch das grundlegende Konzept der Mem2NucGAN-Methoden verloren, nämlich die Fähigkeit des GAN-Modells, realistische Zellkernmorphologien basierend auf den bereitgestellten Membransignalen zu generieren.

Die Segmentierungsergebnisse zeigen, dass synthetische Trainingsdaten nicht nur das vortrainierte Cellpose-Modell übertreffen können, sondern auch händisch generierte Trainingsdaten. Es kann argumentiert werden, dass durch eine größere Menge an manuell annotierten Trainingsdaten die Ergebnisse des Kreuzvalidierung-Modells verbessert werden, was allerdings auch den Zeitaufwand für die Annotation erheblich erhöht. Idealerweise werden kleinere Patch-Größen verwendet, um bei gleichem Zeitaufwand eine größere Anzahl von Patches zu erhalten und dadurch die Heterogenität zu erhöhen. Allerdings erfordern die meisten Segmentierungsmodelle eine Mindestgröße für das Training der Modelle, wodurch die Heterogenität in den Trainingsdaten beschränkt ist. Im Falle seltener Objekte kann eine reduzierte Anzahl von Patches zu einer Unterrepräsentation oder sogar zur vollständigen Abwesenheit solcher Objekte im Trainingsdatensatz führen. Synthetische Trainingsdaten können hingegen in großen Mengen, mit minimaler Benutzerinteraktion und somit mit minimalem Zeitaufwand generiert werden. Hervorzuheben ist auch, dass mit beiden SimOptiGAN(+) Methoden deutlich bessere Segmentierungsergebnisse erzielt werden können als mit der Stand-der-Technik-Methodik. Da auch beim Stand der Technik Rauschen und Helligkeitsabfall in die binären Annotationen eingebracht werden, liegt der Vorteil vermutlich im hybriden Ansatz von SimOptiGAN(+). Durch die stärkere Ähnlichkeit zu den realen Daten wird die Transformationsaufgabe vereinfacht und so auch die Konsistenz während der Transformation gesteigert. SimOptiGAN und SimOptiGAN+ ermöglichen zusätzlich eine Überrepräsentation seltener Objekte und verbessern so die Segmentierungsleistung für diese Objekte. Sowohl SimOptiGAN als auch SimOptiGAN+ erfordern zwar ebenfalls die Annotation einer kleinen Menge an Zellkernen zur Extraktion von Nukleusprototypen. Der Zeitaufwand hierfür ist jedoch deutlich geringer. Für die Generierung synthetischer Daten wurden beispielsweise nur 20 Kerne markiert, im Vergleich zu den 1001 Kernen, die in den drei Ground-Truth-Patches enthalten sind.

Es ist darauf hinzuweisen, dass die Modelle zur Datensynthese nur einmalig trainiert wurden. Der Grund hierfür liegt bei der Trainings- und Inferenz-Zeit der Modelle, welche aufsummiert über 850 h ( $\sim 35$  Tage) an reiner Rechenzeit

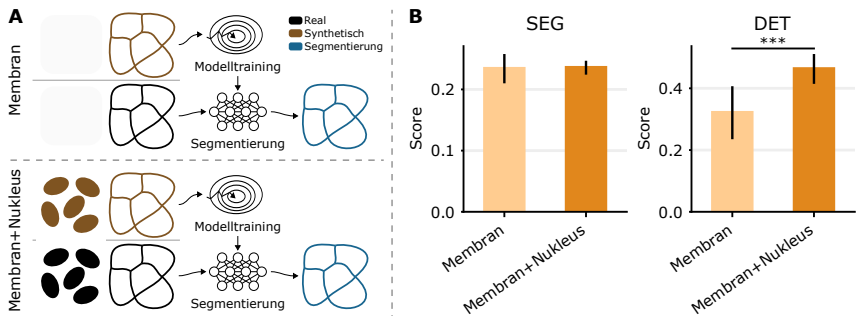
beträgt. In kleineren Voruntersuchungen hat sich allerdings gezeigt, dass vor allem die SimOptiGAN Methoden robust in der Qualität der generierten Daten ist.

Eine klare Aussage, warum die in der Synthese integrierte Biophysik die Ergebnisse der Segmentierung verbessert hat, kann ohne weitere Untersuchungen nicht sicher getroffen werden. Abgesehen vom Platzierungsprozess der Kerne sind die Pipelines SimOptiGAN und SimOptiGAN+ identisch, einschließlich der verwendeten Parameter. So wurden dieselben Nukleusprototypen, Bildsimulations-einstellungen und sogar dasselbe Optimierungsmodell verwendet. Eine Verbesserung kann daher auf eine besser passende Kerngröße oder eine realistischere Ausrichtung und Anordnung der Kerne zurückzuführen sein. Beides kann auch zu einer besseren Transformation während des GAN-basierten Optimierungsschritts führen.

Die eingeführten Synthesemethoden sind nicht nur auf den gezeigten Datensatz beschränkt, sondern sind auch auf andere Datensätze übertragbar, bei denen viele Objekte in einer zusammenhängenden Struktur angeordnet sind. So hat die Methodik das Potenzial, auf andere komplexe Strukturen wie Organoide ausgeweitet zu werden, welche in der biomedizinischen Forschung, insbesondere für Krankheitsmodelle und die Arzneimittelentwicklung, zunehmend an Bedeutung gewinnen. Es ist jedoch wichtig zu beachten, dass die Modellleistung je nach Stärke der Unterschiede in Zellmorphologie, -orientierung und -dichte zwischen den Datensätzen variieren kann. Je größer die Unterschiede sind, desto wahrscheinlicher ist es, dass die Segmentierungsleistung abnimmt. Bei erheblichen Unterschieden empfiehlt es sich, neue, spezifische Trainingsdaten für den zu segmentierenden Datensatz zu erstellen. Bei SimOptiGAN kann die Annotation neuer Nukleusprototypen entfallen, wenn die neuen Daten ähnliche Zellkernmorphologien und -texturen aufweisen. In einem solchen Fall kann eine Reparametrisierung der Simulationspipeline und ein erneutes Training des Optimierungsnetzwerks ausreichen. Da SimOptiGAN+ auf den Ergebnissen einer biophysikalischen Simulation basiert, erfordert eine Anpassung der Zellanordnung die Generierung neuer Zellgrenzen. Für die biophysikalische Simulation kann es genügen, die Ausgangskonfiguration und Parameter wie das erwartete Zellvolumen anzupassen. Sind die Unterschiede jedoch zu groß, sind erneut Parameter aus den neuen Daten zu extrahieren.

### 5.2.5 Membransynthese

Die biophysikalische Modellierung ermöglicht auch die Synthese von Membranaufnahmen, wie in Abschnitt 2.2.3 beschrieben. Ein wichtiger Vorteil der Methodik ist, dass die kombinierte Generierung von Nukleus- und Membransignalen möglich ist. Dies erlaubt es, Segmentierungsmodelle zu trainieren, welche sowohl das Membran- als auch das Nukleussignal für die Membransegmentierung nutzen. Wie Abbildung 5.9 zeigt, kann mit der Ausnutzung beider Kanäle die Detektionsleistung bezüglich der DET-Metrik hochsignifikant ( $p < 0,001$ ) verbessert werden. Interessanterweise ergibt sich bei diesem Experiment keine Verbesserung bezüglich der Präzision der Segmentierung (SEG-Metrik). Ein möglicher Grund hierfür ist, dass das zusätzliche Nukleussignal zwar bei der Detektion von Zellen helfen kann, allerdings keine weiteren Informationen bezüglich der Ausdehnung und Morphologie der Zelle an sich bringt. Die generell niedrigen Ergebnisse aus Abbildung 5.9 zeigen auch, dass die präzise Segmentierung von Membransignalen weiterhin eine Herausforderung darstellt.



**Abbildung 5.9:** Vergleich des Einflusses des zusätzlichen Nukleusmarkers auf die Segmentierungsleistung von Membransegmentierungsmodellen. **A** Das Training beider Modelle erfolgte auf synthetisch generierten Daten, wobei dem Modell *Membran+Nukleus* im Gegensatz zum Modell *Membran* zusätzliche gepaarte Nukleussignale zur Verfügung standen. **B** Die Qualität der Segmentierung wird anhand der SEG und DET Metrik bestimmt, wobei als Ground-Truth drei händisch annotierte reale Bildausschnitte verwendet wurden. Die Balken repräsentieren die Standardabweichung von sechs Trainingsdurchläufen.

Für die Bestimmung der Signifikanz wurden jeweils 18 Wiederholungen der Trainingsläufe durchgeführt. Bis auf eine Anpassung des Durchmessers wurde für das Training der Modelle die Standardparametrierung verwendet. Die Modelle wurden für 100 Epochen ohne die Nutzung eines vortrainierten Netzwerks als Initialisierung trainiert. Der Vergleich der Mittelwerte wurde mit dem einseitigen Welch-Test durchgeführt.

### 5.2.6 SimOptiGAN ermöglicht Datensynthese auch in anderen Forschungsbereichen

Die Methode SimOptiGAN eignet sich nicht nur für die Synthese von den gezeigten Zelldaten, sondern kann auch für die Synthese von Daten aus weiteren Forschungsbereichen, wie der Materialentwicklung, genutzt werden. Die Übertragbarkeit der Methode wird anhand von Aufnahmen eines Rasterelektronenmikroskops (REM) verdeutlicht. Mit dieser Technik können Strukturen im Submikrometerbereich dargestellt werden, um so die Untersuchung der Morphologie und Zusammensetzung von Feststoffen zu ermöglichen. Die in diesem Abschnitt zur Synthese verwendeten Daten sind von Rettenberger et al. [197] frei zugänglich veröffentlicht<sup>5</sup>. Bei diesen Daten handelt es sich um REM-Aufnahmen von Pulverproben. Um die Daten zu analysieren, ist zuerst eine Segmentierung der Partikel erforderlich. Dies stellt aufgrund von teilweise unscharfen Aufnahmen sowie aggregierten und sich überlagernden Partikel eine Herausforderung dar [197].

In den Aufnahmen der Pulverproben sind die Anzahl, die Verteilung, die Größe und die Form der Partikel inhomogen. Zudem kommt es vor allem bei größeren Partikeln zur Aggregation von weiteren kleineren Partikeln. Weiterhin weisen die Aufnahmen einen inhomogenen Hintergrund auf, welcher zusätzlich in der Nähe der Partikel dunkler wird. Auch ist je nach Aufnahme die Schärfe der Bilder unterschiedlich. Diese Punkte stellen eine Herausforderung bei der Synthese solcher

---

<sup>5</sup> <https://github.com/lrettenberger/Uncertainty-Aware-Particle-Segmentation-for-SEM-Data>

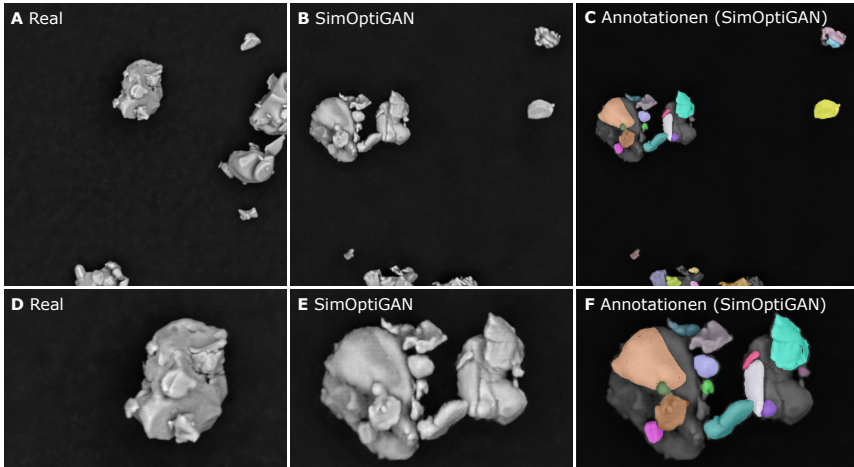


Aufnahmen dar. Zudem unterscheiden sich die REM-Aufnahmen stark von den in dieser Arbeit verwendeten Zellaufnahmen eines konfokalen Mikroskops. Um die Daten mittels SimOptiGAN zu synthetisieren, sind daher kleine Anpassungen notwendig.

Durch die modulare Struktur von SimOptiGAN können diese Anpassungen auf einfache Art und Weise eingebracht werden. Zur Synthese der gezeigten Aufnahmen werden nur die beiden Module *Probengenerierung* und *Aufnahmesimulation* von SimOptiGAN angepasst. Im Teil *Probengenerierung* erfolgt die Platzierung der Objektprototypen zur Generierung der virtuellen Probe. Da bei diesen REM-Daten die Objekte meist spärlich verteilt sind, werden nun lediglich drei bis zehn Objekte im Bild platziert. Im Teil *Aufnahmesimulation* ist aufgrund des unterschiedlichen Aufnahmeverfahrens ausreichend, nur noch eine Rauschsimulation durchzuführen.

Abbildung 5.10 zeigt die Ergebnisse der Synthese der REM-Aufnahmen mittels SimOptiGAN im Vergleich zu einer realen REM-Aufnahme. In der Abbildung sind zudem die synthetischen Annotationen von SimOptiGAN dargestellt. Durch Anpassungen der Platzierung der Objektprototypen kann SimOptiGAN die Variation in der Anzahl und der Verteilung der Partikel nachbilden. Zudem wird durch den Deep-Learning-gestützten Optimierungsschritt von SimOptiGAN der inhomogene Hintergrund sowie die dunkleren Bereiche in der Nähe der Partikel nachgebildet. Im Anhang (Abbildung A.5) ist eine aufgehellte Version von Abbildung 5.10a und b gegeben, bei der sich die Hintergründe besser vergleichen lassen. Bei den Annotationen zeigt sich der Vorteil der SimOptiGAN-Methode, welche es ermöglicht, neben den vollständigen Annotationen nur Annotationen spezieller Klassen zu generieren. So sind in den in Abbildung 5.10c und f gezeigten Annotationen nur die relevanten Partikel annotiert. Diese Annotationen wurden von Experten durchgeführt [197].

Dieses Anwendungsbeispiel zeigt, dass SimOptiGAN nicht nur für die Synthese mikroskopischer Zellaufnahmen geeignet ist, sondern auch in anderen Forschungsbereichen, wie der Materialentwicklung mit REM-Aufnahmen, eingesetzt werden kann.



**Abbildung 5.10:** Vergleich zwischen echten und mittels SimOptiGAN generierten REM-Aufnahmen. A-C zeigen den gesamten Ausschnitt der Aufnahme, während D-F einen vergrößert dargestellten Ausschnitt der darüberliegenden Aufnahmen zeigt. A bzw. D stellen eine repräsentative reale Aufnahme dar und B bzw. E zeigt eine mit SimOptiGAN generierte synthetische Aufnahme. In C bzw. F sind zudem die zugehörigen synthetischen Annotationen mit den Aufnahmen in B bzw. E überlagert dargestellt. Die Annotationen stammen von den extrahierten Objektprototypen und wurden, wie in [197] beschrieben, von Experten durchgeführt, welche nur die relevanten Partikel annotiert haben. Im Anhang (Abbildung A.5) ist zudem jeweils eine zusätzliche Aufnahme mit normaler und erhöhter Helligkeit zu sehen.

## 5.2.7 Zusammenfassung

Für die Trainingsdatensynthese wurde mit SimOptiGAN eine neue Methodik für die Synthese von Trainingsdaten entwickelt. Ein wesentlicher Neuheitswert der Methodik liegt darin, dass Objekte unterschiedlicher Klassen gezielt platziert werden können, wobei neben den Instanzannotationen auch die Klassenannotationen verfügbar sind. Dieses Alleinstellungsmerkmal ermöglicht es erstmalig, synthetische Daten für das Training einer panoptischen Segmentierung zu generieren, bei der Segmentierung und Klassifikation in einem Schritt durchgeführt werden.

Daneben können mit der entwickelten Methode auch seltene Elemente in die Trainingsdaten eingebracht werden, was eine wichtige Zielstellung der Entwicklung

war. Anhand von Experimenten auf Basis eines synthetischen Datensatzes konnte gezeigt werden, dass die Segmentierungsqualität dieser Elemente durch eine überrepräsentative Einbringung deutlich verbessert werden kann.

Ein weiterer Neuheitswert stellt der hybride Ansatz der SimOptiGAN Methode dar. Hier konnte gezeigt werden, dass durch die vorherige Simulation die Domänenlücke zwischen den zu transformierenden Bildern und der realen Domäne gesenkt werden kann, womit potenziell die Trainingsstabilität und -zeit verbessert werden kann.

Die SimOptiGAN Methode ist dabei so allgemein gehalten, dass weitere externe Methoden einfach integriert werden können. So wurde beispielsweise unter dem Namen SimOptiGAN+ ein biophysikalisch motiviertes Modell integriert, um die Anordnung von Zellen und Kernen zu verbessern. Neben dieser Erweiterung von SimOptiGAN wurden zwei weitere Methoden entwickelt, um die Ergebnisse von biophysikalisch motivierten Modellen zur Zellanordnung in den Syntheseprozess zu integrieren. Durch die Beachtung von Oberflächenkräften können so realistischere Anordnungen erzielt werden. Anhand eines händisch annotierten Datensatzes wurde SimOptiGAN mit den drei Methoden zur Integration der biophysikalischen Modellierung verglichen. Hier zeigte sich, dass die Erweiterung von SimOptiGAN (SimOptiGAN+) durch die Integration der biophysikalischen Modellierung signifikant bessere Segmentierungsleistungen erzielt werden können als mit SimOptiGAN und den beiden anderen Methoden. Weiterhin zeigte sich, dass mit den synthetischen Daten Modelle trainiert werden können, welche die Leistung eines etablierten vortrainierten Modells deutlich übersteigen.

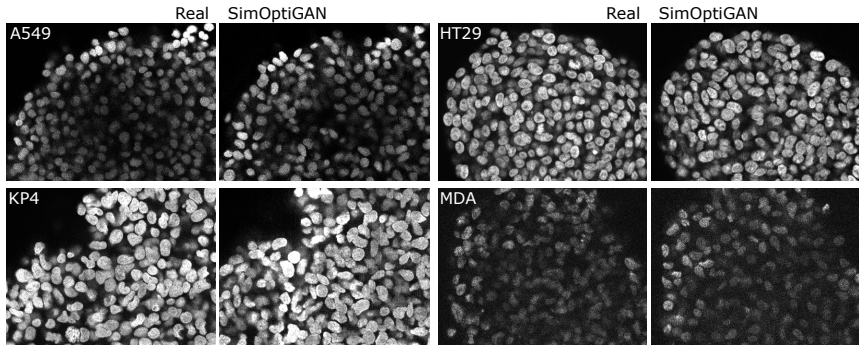
Eine wichtige Erkenntnis dieser Experimente ist weiterhin, dass mit SimOptiGAN und SimOptiGAN+ Segmentierungsleistungen erzielt werden können, welche annähernd vergleichbar sind mit Modellen, welche mit händisch annotierten Daten trainiert wurden. Dabei ist zu beachten, dass zwar nur eine geringe Anzahl an händisch annotierten Daten zur Verfügung standen, aber deren Annotationsaufwand mit 37,5 h deutlich über dem benötigten Segmentierungsaufwand der SimOptiGAN und SimOptiGAN+ Methode (0,4 h und 4,4 h) liegt.

Ein Vergleich der Rechenzeiten und des manuellen Annotationsaufwands zeigte, dass der klassische hybride Ansatz SimOptiGAN den geringsten Aufwand erfordert. Die Einbringung der biophysikalischen Simulation zur Verbesserung der Zellanordnung bringt derzeit noch einen hohen Annotationsaufwand mit sich. Allerdings kann eine Datenbank mit Parametern von verschiedenen Zelltypen aufgebaut werden, welche eine händische Annotation bei neuen Daten überflüssig macht. Dennoch ist abzuwägen, ob der zusätzliche Aufwand für die biophysikalische Modellierung die erzielte Mehrleistung wert ist.

Auch wurde die Ähnlichkeit synthetischer Daten mit realen Daten mittels der für 3D-Daten angepassten KID-Metrik analysiert. Mit allen Methoden können plausibel aussehende 3D-Aufnahmen von Zellkulturen generiert werden. Im Vergleich zu einem naiv generierten Bild, welches mit einer vereinfachten Simulation generiert wurde, zeigten die mit den entwickelten Methoden generierten Daten deutlich geringere Distanzen zu den realen Vorbildern.

Bei den Experimenten zeigte sich zudem, dass die Helligkeit bzw. deren Verteilung in den Trainingsdaten einen großen Einfluss auf die Segmentierungsleistung besitzt. Hier zeigen die beiden Methoden SimOptiGAN und SimOptiGAN+ Vorteile, da hier durch den hybriden Ansatz die Helligkeitsverteilung von den generierten Daten gezielt beeinflusst werden kann. Bei SimOptiGAN+ sind zwar durch die Nutzung des bereits trainierten Optimierungsmodells von SimOptiGAN unsaubere Helligkeitsübergänge entstanden. Diese lassen sich jedoch in der Regel durch ein Training des Optimierungsmodells auf den spezifischen Daten von SimOptiGAN+ vermeiden.

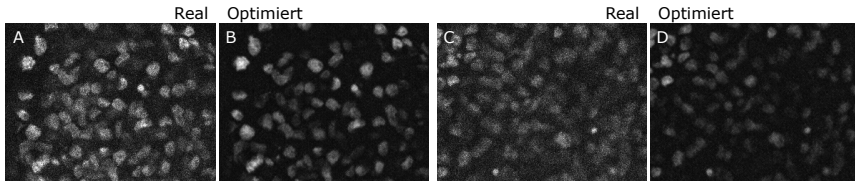
Anhand der Abbildung 5.11 wird weiterhin verdeutlicht, dass die entwickelte Methode SimOptiGAN für die Erstellung von diversen Datensätzen unterschiedlicher Zellkulturen geeignet ist. Die Abbildung zeigt 2D-Schnitte von vier realen und vier mit SimOptiGAN generierten Aufnahmen. Die realen Aufnahmen stammen von verschiedenen Zellkulturen, die sich in den Zelltypen (A549, HT29, KP4 und MDA), den Morphologien und Texturen der Nuklei, den Zelldichten, den Helligkeiten sowie den Rauschanteilen unterscheiden. Wie in der Abbildung zu



**Abbildung 5.11:** Vergleich zwischen realen und synthetisch generierten 3D-Aufnahmen unterschiedlicher Zellkulturen (A549, HT29, KP4 und MDA), die sich durch variierende Anordnungen, Morphologien und Texturen der Zellen auszeichnen. Gezeigt sind 2D-Bildausschnitte der Aufnahmen. Die synthetischen Aufnahmen wurden mithilfe der in dieser Arbeit eingeführten SimOptiGAN Methode generiert.

sehen ist, ermöglicht es SimOptiGAN, diese Variationen abzubilden und dadurch glaubwürdige synthetische Versionen der realen Aufnahmen zu generieren.

Die hybride Methode SimOptiGAN(+) bietet weiterhin den Vorteil, dass bei der CycleGAN-basierten Optimierung automatisch die Rücktransformation von realen zu simulierten Daten gelernt wird. Das Generatormodell zur Rücktransformation kann dazu verwendet werden, die Qualität von realen Aufnahmen zu steigern. Abbildung 5.12 zeigt beispielhaft die Anwendung eines solchen Generators anhand von zwei Bildausschnitten. Es ist erkennbar, dass sowohl das Rauschen als auch die Signalstreuung reduziert werden, wodurch Zellen im Hintergrund verschwinden. Besonders in dunklen Bildbereichen mit schlechtem Signal-Rausch-Verhältnis führt dies zu deutlich schärferen Zellenrändern. Weitere Experimente sind jedoch notwendig, um zu untersuchen, inwieweit sich damit eine Verbesserung bei der Segmentierung erzielen lässt. Ein Vorteil dieses Ansatzes ist, dass im Gegensatz zu bekannten Netzwerken zur Signalaufbereitung (z. B. CARE [198]) keine zusätzlichen Aufnahmen erstellt werden müssen, um gepaarte Daten mit guter und schlechter Qualität zu erhalten.



**Abbildung 5.12:** Bild-Optimierung mittels des trainierten CycleGANs von SimOptiGAN. Der beim CycleGAN-Training für die Rücktransformation benötigte Generator kann eingesetzt werden, um die Qualität von Mikroskopaufnahmen zu steigern. **A** und **C** zeigen 2D-Schnitte von realen 3D-Zellaufnahmen, während **B** und **D** die optimierten Versionen der Zellkulturaufnahmen darstellen.

Die beiden hybriden Methoden SimOptiGAN und SimOptiGAN+ sind zudem nicht auf die Verwendung von GAN Modellen für die ungepaarte Bild-zu-Bild-Transformation limitiert. Neben dem CycleGAN können auch andere Modelle wie das UNIT oder diffusionsbasierte Modelle eingesetzt werden. Somit ist eine Kompatibilität zu zukünftigen Methoden gegeben.

Darüber hinaus kann die modulare SimOptiGAN-Methode, wie gezeigt, flexibel angepasst werden, um Trainingsdaten für andere Forschungsbereiche, wie die Materialentwicklung anhand von REM-Aufnahmen, zu generieren.

## 5.3 Synthese von Bildmodalitäten

Teile der folgenden Sektion basieren auf der im Rahmen der vorliegenden Arbeit veröffentlichten Publikation von Bruch et al. [188].

### 5.3.1 Übersicht

Das Ziel der in Abschnitt 2.3 vorgestellten Methode IMS zur Generierung weiterer Bildmodalitäten ist es, die Limitierungen bezüglich der gleichzeitig aufnehmbaren Färbungen abzuschwächen, indem die Signale von inkompatiblen Markern

synthetisch generiert werden. In diesem Abschnitt erfolgt zuerst ein Methodenvergleich, bei dem die zwei Ansätze ResUnet und pix2pix anhand der benötigten Hardware, der Trainingszeit und ihrer quantitativen und visuellen Prognoseleistung verglichen werden. Weiterhin wird anhand eines realen Anwendungsfalls, der Zellzyklus-Klassifikation, die Effektivität der Methode aufgezeigt. Zudem werden die beiden Ansätze anhand von neu eingeführten Protokollen überprüft. Anschließend werden anhand verschiedener Markerkombinationen die für eine erfolgreiche Transformation benötigten Voraussetzungen und die Grenzen der Methode aufgezeigt.

### 5.3.2 Methodenvergleich

Für die Synthese von Bildmodalitäten mit IMS werden zwei Ansätze verglichen. Zum einen wird, wie in [188] vorgestellt, ein ResUnet mit einem MSE-Loss für die Signaltransformation genutzt. Die mit dem ResUnet generierten Signale sind etwas unscharf, was potenziell durch die Verwendung eines zusätzlichen Adversarial Losses verbessert werden kann. Der ursprüngliche Ansatz wird daher mit der pix2pix Architektur verglichen, welche einen Adversarial Loss im Training verwendet. Der Vergleich der beiden Methoden erfolgt anhand des Datensatzes *DS-Marker-Syn*, bei dem der Nukleusmarker in den Proliferationsmarker Ki-67 transformiert wird.

#### Trainingszeit und Ressourcen

Zuerst erfolgt ein Vergleich der beiden Modelle anhand der Trainingszeit und des Speicherbedarfs. Gerade wenn mehrere spezifische Transformationsmodelle zu trainieren sind, ist die hierfür benötigte Zeit nicht zu vernachlässigen. Durch den zusätzlichen Adversarial Loss des pix2pix-Modells ist das parallele Training eines zusätzlichen Diskriminator-Netzwerkes erforderlich. Bei gleichen Voraussetzungen und Modellgrößen ist damit eine größere Menge an Grafikkartenspeicher und auch eine längere Rechenzeit erforderlich.

**Tabelle 5.2:** Vergleich der zum Training und zur Anwendung benötigten Rechenzeit und Speicherbedarf von zwei Methoden. Die erste Methode ResUnet basiert auf einem klassischen Ansatz, bei dem ein ResUnet für die Transformation verwendet wird. Bei der zweiten Methode pix2pix wird eine GAN-basierte Trainingsstruktur verwendet, was zu einer Erhöhung der benötigten Ressourcen führt.

Methode	Training	Anwendung	GPU Speicherbedarf
ResUnet	25 h	25 min	17,2 GB
pix2pix	98 h	34 min	16,8 GB

Um die nötige Trainingszeit des ResUnets und des pix2pix-Modells für die Transformation von Nukleus- zu Ki-67-Signalen zu vergleichen, wurde ein Laufzeitexperiment durchgeführt. Die beiden Modelle wurden anhand von 20 Bildern trainiert. Bei dem Laufzeitexperiment <sup>6</sup> (vgl. Tabelle 5.2) zeigt sich, dass das Training des ResUnet-Modells mit 25 h fast viermal schneller abgeschlossen ist als das des pix2pix-Modells (98 h) <sup>7</sup>.

Wie in der Methodik (Abschnitt 2.3.2) beschrieben, ist es für das pix2pix-Modell nicht direkt ersichtlich, ob eine Verlängerung der Trainingsdauer zu besseren Ergebnissen führt. Im Gegensatz zu dem ResUnet konvergieren die Lossfunktionen von Generator und Diskriminator nicht. Die Trainingsdauer wird beim pix2pix-Modell im Voraus über die zu trainierenden Epochen explizit festgelegt. Für das Experiment wurde die Epochenanzahl so gewählt, dass sich keine merklichen Verbesserungen mehr bezüglich des L1-Losses ergeben haben. In den Auswertungen der Qualität der generierten Bilder wurden auch vorherige Epochen untersucht, bei denen sich bestätigte, dass der Qualitätszuwachs gegen Ende des Trainings stagniert.

<sup>6</sup> Alle Experimente wurden auf einem Servercomputer mit einem AMD EPYC 7252 Prozessor, 64 GB RAM und einer NVIDIA A6000 48 GB GPU durchgeführt.

<sup>7</sup> Auf die Angaben der trainierten Epochen wird absichtlich verzichtet, da sie aufgrund unterschiedlicher Prä-Prozessierung der Trainingsdaten nicht direkt zu vergleichen sind. So werden für das ResUnet die Aufnahmen bereits vor dem Training in Patches aufgeteilt, während beim pix2pix-Modell pro Epoche zufällig ein Patch pro Aufnahmen extrahiert wird. Damit unterscheiden sich die Anzahl der Patches pro Epoche.



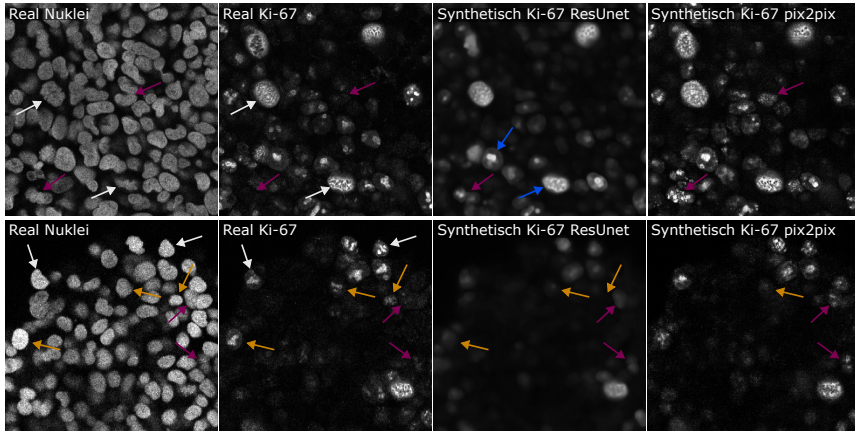
Die Anwendungszeit, also die zur Transformation von neuen Aufnahmen benötigte Zeit, unterscheidet sich nur geringfügig zwischen den beiden Modellen. Für die Inferenz von 9 Aufnahmen werden beim ResUnet 25 min und beim pix2pix-Modell 34 min benötigt. Die 36 % längere Dauer kann durch das größere Transformationsnetzwerk des pix2pix-Modells erklärt werden.

Eine Diskussion der Ergebnisse bezüglich des benötigten Grafikspeichers findet sich im Anhang A.6.

## Qualität der Prognosen

Abbildung 5.13 zeigt in den ersten zwei Spalten Bildausschnitte der Nukleus und Ki-67-Signale von zwei realen Aufnahmen aus *DS-Marker-Syn*. In den letzten beiden Zeilen sind die mit den beiden Ansätzen ResUnet und pix2pix generierten synthetischen Ki-67-Signale dargestellt, welche auf Basis der realen Nukleussignale generiert wurden. Die Aufnahme in der ersten Zeile zeigt eine unbehandelte Kokultur bestehend aus KP4- und Fibroblasten-Zellen. Hier ist die Zelldichte hoch und es befinden sich einige Zellen in der Mitosephase (graue Pfeile), welche sich durch die Größe der Nuklei, der wolkenförmigen Textur des Nukleussignals und der löchrigen Textur des Ki-67-Signals erkennen lassen. In der zweiten Zeile ist eine mit dem Wirkstoff Doxorubicin behandelte Monokultur aus A549 Zellen dargestellt. Hier ist die Zelldichte geringer und es befinden sich die meisten Zellen in der Interphase (graue Pfeile). In dieser Phase lassen sich die Zellen anhand der Nukleussignale kaum noch von Zellen unterscheiden, die sich außerhalb der Zellteilung befinden (G0-Phase). Die Zellen weisen in der Interphase lediglich punktförmige Ki-67-Signale auf, welche innerhalb der Nuklei verteilt sind. Zellen außerhalb der Zellteilung zeigen keine Ki-67-Signale.

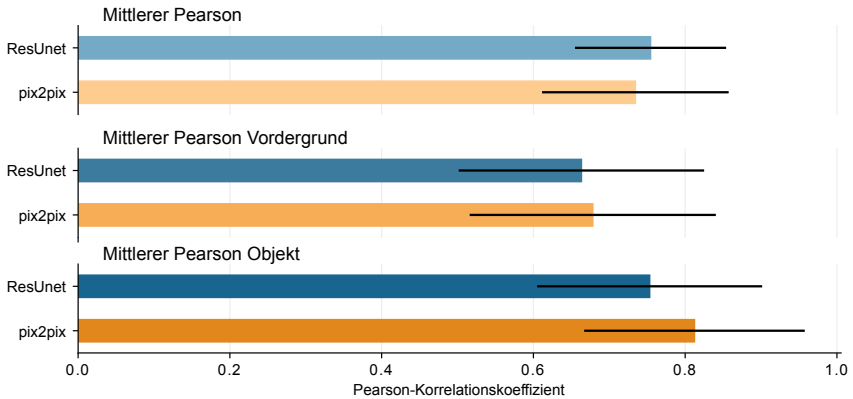
Unabhängig von der Übereinstimmung mit den realen Ki-67-Signalen weisen die mit pix2pix generierten Signale eine höhere Schärfe auf als die mit dem ResUnet generierten. Es kann also argumentiert werden, dass der beim pix2pix verwendete Adversarial Loss zu einer erhöhten Schärfe der generierten Bilder führt. Abgesehen von der Unschärfe besitzen aber auch die mit dem ResUnet



**Abbildung 5.13:** Bildausschnitte zweier 3D-Aufnahmen von Nukleus und dazugehörigen Ki-67-Signalen und den auf Basis der realen Nukleussignale generierten Ki-67-Prädiktionen. Bei der Aufnahme der ersten Reihe handelt es sich um eine unbehandelte Kokultur bestehend aus KP4- und Fibroblasten-Zellen. Die Aufnahme in der zweiten Reihe zeigt eine Monokultur bestehend aus A549-Zellen, welche mit dem Wirkstoff Doxorubicin behandelt wurden. Mit Pfeilen markierte Stellen werden im Text diskutiert.

generierten Signale eine strukturierte Textur, welche den realen Daten ähnelt. So sind beispielsweise die netzförmigen Texturen von Zellen in der Mitosephase oder die punktförmigen Signale von Zellen in der Interphase erkennbar (blaue Pfeile).

Bezüglich der Übereinstimmung der generierten Signale mit der realen Ground-Truth zeigt sich, dass bei beiden Ansätzen die Ki-67-Signale der Zellen in der Mitosephase gut repliziert werden. Diese Zellen sind aufgrund ihrer Größe und der wolkenförmigen Textur der Nuklei auch für den menschlichen Betrachter leicht zu erkennen. Die Nuklei in der Interphase unterscheiden sich nur geringfügig von den Nuklei außerhalb des Teilungsstadiums, was die korrekte Vorhersage der Ki-67-Signalintensität für die Modelle schwierig macht. Hier zeigen sich daher vermehrt falsch-negative Zellen (orangefarbene Pfeile). Daneben werden auch Zellen von den Modellen falsch-positiv prognostiziert (lila Pfeile). Das ResUnet scheint insgesamt zu wenige Signale zu generieren, während das pix2pix-Modell tendenziell zu viele Signale erzeugt.



**Abbildung 5.14:** Pearson-Korrelation zwischen den Intensitäten des realen Ki-67-Signals und des synthetischen Ki-67-Signals. Die Korrelation wird zwischen allen Voxel-Intensitäten (Mittlerer Pearson), allen Vordergrund Voxel-Intensitäten (Mittlerer Pearson Vordergrund) und allen Nukleus-Intensitäten (Mittlerer Pearson Objekt) bestimmt (siehe Abschnitt 2.3.3). Die Vordergrundbereiche und die Objektinstanzen werden dabei über eine vorherige Segmentierung bestimmt. Verglichen werden die Ergebnisse des ResNet mit denen des pix2pix basierten Ansatzes. Die Fehlerbalken geben die Standardabweichung bezüglich neun Aufnahmen an.

Die Abbildung 5.14 zeigt den Vergleich der beiden Ansätze bezüglich der Pearson-Korrelation von verschiedenen Regionen, welche in Abschnitt 2.3.3 eingeführt wurden. Die Korrelationen werden zwischen den Intensitäten von realen und synthetischen Ki-67-Signalen berechnet. Bei einem direkten Vergleich, bei dem die Intensitäten aller Voxel betrachtet werden, ergibt sich ein leicht besseres Ergebnis für den ResNet Ansatz (ResNet:  $r = 0,754$ , pix2pix:  $r = 0,734$ ). Gerade bei Ki-67-Signalen zeigt der Großteil der Voxel niedrige Intensitäten, da sich die Signale meist nicht über den gesamten Nukleus erstrecken und generell viele Hintergrundbereiche in den Aufnahmen vorhanden sind. Die dunklen Hintergrundbereiche besitzen damit ein großes Gewicht und sind zudem auch noch leicht für die Modelle zu erlernen. Ein Modell, welches eine Tendenz zu falsch-negativen Fehlern besitzt, wird bei dieser Bewertung begünstigt. Die interessanten Bereiche des Ki-67-Signals befinden sich hingegen innerhalb der Nuklei.

Daher wird die Korrelation der Intensitäten zusätzlich für Voxel im Vordergrund und weiterhin auch für die mittlere Intensität einzelner Objekte angegeben. Der Vordergrund und die Objekt-Instanzen beruhen auf einer Segmentierung der Nuklei. Das dazu verwendete Segmentierungsmodell wurde mit synthetischen Nukleidaten trainiert. Bei der Betrachtung der beiden spezifischeren Korrelationen zeigt sich, dass der pix2pix Ansatz gegenüber dem ResUnet in beiden Maßen leicht bessere Ergebnisse erzielt. Die Werte für die vordergrundbasierte Korrelation liegen bei  $r = 0,663$  (ResUnet) und  $r = 0,678$  (pix2pix) und bei der objektbasierten Korrelation erreichen die Modelle Werte von  $r = 0,753$  (ResUnet) und  $r = 0,812$  (pix2pix). Dass die Güte der Modelle bei ausschließlicher Betrachtung der Vordergrundregionen abnimmt, ist zu erwarten, da die leicht zu prognostizierenden Hintergrundsignale wegfallen. Die wieder steigende Güte der Modelle bei dem objektbasierten Vergleich ist weniger offensichtlich. Da hier nicht mehr die Intensitäten einzelner Voxel, sondern die mittleren Intensitäten ganzer Nuklei verglichen werden, verliert die Position der Signale an Bedeutung. Es ist so etwa nicht mehr relevant, wo die punktförmigen Signale innerhalb des Nukleus platziert werden, oder ob die Textur mit der des realen Signals übereinstimmt. Je nach späterem Anwendungsfall ist entweder das vordergrundbezogene Maß relevanter, z. B. bei einer anschließenden Klassifikation mit neuronalen Netzen, bei der die Textur eine Rolle spielt, oder das objektbezogene Maß relevanter, z. B. bei einer Auswertung der mittleren Helligkeit der Nuklei.

Auch wenn der pix2pix Ansatz hinsichtlich der Korrelation der relevanten Bereiche etwas bessere Ergebnisse erzielt, zeigt die Standardabweichung eine erhebliche Variation zwischen den einzelnen Aufnahmen. Daher sind keine signifikanten Unterschiede feststellbar. Ein Grund für die hohen Varianzen in den Ergebnissen kann auf die hohe Diversität des verwendeten Datensatzes zurückgeführt werden. Wie in den Bildausschnitten in Abbildung 5.13 dargestellt, unterscheiden sich die Aufnahmen des Trainings- und Testdatensatzes von *DS-Marker-Syn* in mehreren Aspekten. Die Datensätze enthalten Aufnahmen verschiedener Zelltypen (KP4, A549 und Fibroblasten) in Mono- und Kokulturen, die sich merklich in der Morphologie der Nuklei unterscheiden. Zudem variiert die Effektivität der Gewebeklärung zwischen den unterschiedlichen Zelltypen, was sich bei den

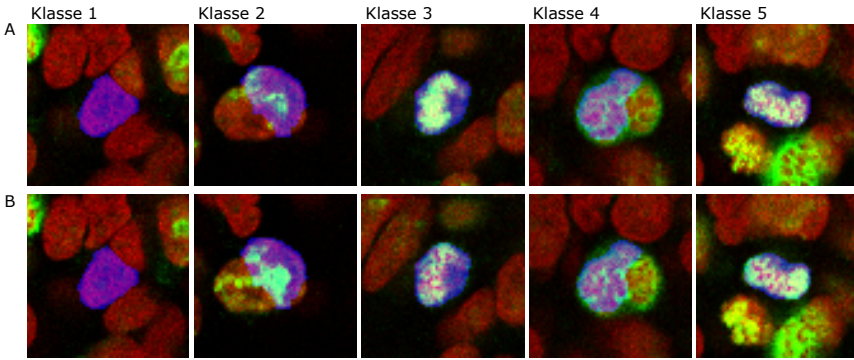
A549 Zellkulturen in einem starken Helligkeits- und Qualitätsabfall in tieferen Schichten der Aufnahmen bemerkbar macht. Die Lichtdurchdringung der Proben ist weiterhin von der Frequenz des Lichts abhängig. So gibt es Bereiche, in denen das Ki-67-Signal nicht mehr erkennbar ist, während die Nukleussignale noch sichtbar sind. Weiterhin wurden die Proben teilweise mit Wirkstoffen (Doxorubicin und Paclitaxel) behandelt, was die Form und Größe der Kulturen, aber auch die Morphologie und Textur der Nuklei beeinflussen kann. Es ist daher anzunehmen, dass ein spezifischerer Datensatz, welcher nur einen einzigen Zelltyp bzw. eine einzelne Kombination von Zelltypen umfasst, die Prädiktionsgüte verbessern kann.

Aufgrund der geringen Unterschiede in den Ergebnissen zwischen ResUnet und pix2pix werden in den folgenden Experimenten zur Steigerung der Übersicht ausschließlich die Resultate des pix2pix-Modells präsentiert.

### 5.3.3 Zellzyklus-Klassifikation mit synthetischem Marker

Ein Anwendungsbeispiel für die Nutzung der synthetisch generierten Marker stellt die Bestimmung des Zellzyklusstadiums von einzelnen Zellen dar. Ki-67, ein Marker für die Zellproliferation, ist hierbei ein wichtiger Indikator für die Bestimmung der einzelnen Zyklus-Stadien. Anhand einer neu entwickelten Methode für die Klassifikation von Zyklus-Stadien von einzelnen Zellen in 3D-Aufnahmen wird untersucht, wie gut der Marker Ki-67 durch seine synthetische Version ersetzt werden kann. Die dazu genutzte Klassifikationsmethode wird in Abschnitt 2.3.4 eingeführt.

Für die Beantwortung dieser Frage wird der in dieser Arbeit eingeführte Datensatz *DS-Klassifikation* genutzt, welcher in drei Teile gegliedert ist. Der erste Teildatensatz (*DS-Klassifikation-Real*) dient als Benchmark, da hier dem Modell das reale Ki-67-Signal für die Klassifikation zur Verfügung steht. Für den zweiten Teildatensatz (*DS-Klassifikation-Syn*) wird das reale Ki-67-Signal durch ein synthetisch generiertes Ki-67-Signal ersetzt. So kann geprüft werden, wie gut sich



**Abbildung 5.15:** Beispielhafte Bildausschnitte der Datensätze *DS-Klassifikation-Real* und *DS-Klassifikation-Syn*. **A** zeigt Bildausschnitte des Datensatzes *DS-Klassifikation-Real* bei dem reale Ki-67-Signale zur Verfügung stehen. **B** zeigt Bildausschnitte des Datensatzes *DS-Klassifikation-Syn*, bei dem das Ki-67-Signal mit dem pix2pix-Modell synthetisiert wurde. Der Nukleusmarker Draq5 ist in Rot, der Proliferationsmarker Ki-67 in Grün und die Nukleus-Segmentierungsmaske in Blau dargestellt.

das synthetische Signal für die Klassifikation eignet. Bei dem dritten Teildatensatz (*DS-Klassifikation-Ohne*) wird dem Modell kein Ki-67-Signal zur Verfügung gestellt. Hierbei wird eine untere Grenze etabliert und so geprüft, ob mithilfe des synthetischen Ki-67-Signals überhaupt eine Verbesserung erzielt werden kann.

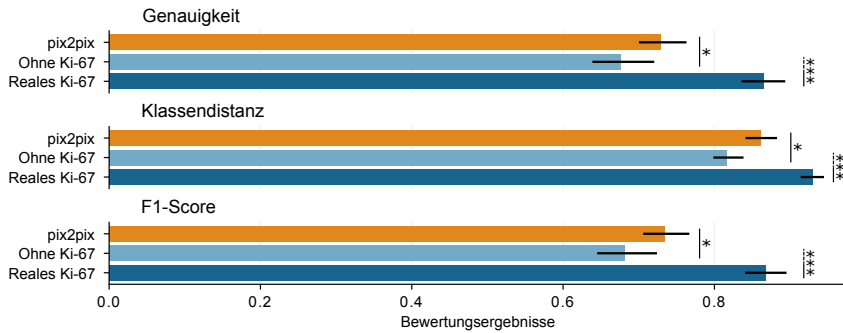
Abbildung 5.15 zeigt einen Vergleich des Datensatzes mit realem Ki-67-Signal (a) und Datensätzen mit synthetischem Ki-67-Signal (b), welches mit dem pix2pix (c) generiert wurde. Hier zeigt sich, dass die synthetischen Ki-67-Signale der einzelnen Klassen insgesamt gut vorhergesagt werden. Schwierigkeiten treten bei der Klasse zwei auf, bei der das Ki-67-Signal sehr lokal ausgeprägt ist. In diesem Fall lassen sich größere Abweichungen in der Position und Menge zwischen realen und synthetischen Markersignalen feststellen.

Als Bewertungskriterien werden die Genauigkeit, die eingeführte Klassendistanz (siehe Abschnitt 2.3.4) und der F1-Score verwendet. Die Genauigkeit wird angegeben, da das Maß einfach zu interpretieren ist. Zudem wird die Klassendistanz verwendet, da Klassengrenzen bei der Zellzyklus-Klassifikation aufgrund des

kontinuierlichen Prozesses oft nicht eindeutig sind. Die Klassendistanz ermöglicht einen faireren Vergleich, da sie im Gegensatz zur Genauigkeit die Stärke des Fehlers berücksichtigt: Falsche Zuweisungen in Nachbarklassen werden weniger stark gewichtet als Zuweisungen in weiter entfernten Klassen. Der F1-Score wird als ergänzendes Maß angegeben, da die Anzahl der Zellen pro Klasse ungleich verteilt ist. Das Maß beruht auf dem harmonischen Mittel von Genauigkeit und Sensitivität, was besonders bei unbalancierten Datensätzen wichtig ist, um eine übermäßige Bevorzugung von häufig auftretenden Klassen zu vermeiden. Bei der verwendeten Version des F1-Scores wird zudem die Klassenimbalance durch eine Gewichtung der Klassenergebnisse basierend auf der Anzahl der darin enthaltenen Zellen berücksichtigt.

Die erzielten Klassifikationsgüten zeigen, dass die Verwendung eines synthetischen Ki-67-Signals zur Bestimmung der Zellzyklus-Klassen einen signifikanten Vorteil gegenüber der Verwendung eines Datensatzes ohne Ki-67-Signale bietet (siehe Abbildung 5.16). Im Schnitt über fünf Trainingswiederholungen kann durch die Verwendung des synthetischen Signals bezüglich der Genauigkeit, der eingeführten Klassendistanz (siehe Abschnitt 2.3.4) und des F1-Scores eine signifikante Verbesserung von 7,8%; 5,4% bzw. 7,8% erzielt werden. Verglichen mit der Genauigkeit zeigt sich, dass die Klassifikationsmodelle bezüglich der Klassendistanz alle besser abschneiden. Anhand der Konfusionsmatrizen (vgl. Abbildung 5.17) ist ersichtlich, dass die meisten falsch klassifizierten Zellen in der Tat der Nachbarklasse zugewiesen wurden. Da sich die Ergebnisse der Modelle im Hinblick auf den F1-Score mit denen der Genauigkeit decken, lässt sich schließen, dass die Modelle nicht stark von der Klassenimbalance beeinflusst werden. Eine Maßnahme dafür war während des Trainings, die weniger vertretenen Klassen durch zufälliges Duplizieren der Proben auszugleichen.

Die Klassifikationsergebnisse auf Basis der synthetischen Ki-67-Signale sind signifikant besser als die, bei denen kein Ki-67-Signal zur Verfügung steht. Im Durchschnitt führt die Verwendung der synthetischen Markervariante zu einer Verbesserung der Modellgüte um 7% im Vergleich zum Modell ohne Ki-67-Signale.



**Abbildung 5.16:** Vergleich der Güte der Klassifikationsmodelle. Als Maße dienen die Genauigkeit, die Klassendistanz (vgl. Abschnitt 2.3.4) und der F1-Score. Die Datensätze unterscheiden sich anhand des verwendeten Ki-67-Signals. Bei *Reales Ki-67* und *Ohne Ki-67* steht das reale bzw. kein Ki-67-Signal zur Verfügung. Bei *pix2pix* wurde das Ki-67-Signal mit dem pix2pix-Modell generiert. Die Fehlerbalken geben die Standardabweichung bezüglich fünf Trainingswiederholungen des Klassifikationsmodells an.

Reales Ki-67							Ohne Ki-67							pix2pix						
		Ground Truth							Ground Truth							Ground Truth				
		0	1	2	3	4			0	1	2	3	4			0	1	2	3	4
Prädiktion	0	93	7	0	0	0	Prädiktion	0	69	30	1	0	0	Prädiktion	0	71	29	0	0	0
	1	20	190	3	0	2		1	54	142	11	6	2		1	56	158	1	0	0
	2	0	0	44	11	0		2	0	2	40	11	2		2	0	0	44	11	0
	3	0	0	22	93	0		3	5	6	19	75	10		3	0	0	28	82	5
	4	0	0	1	3	21		4	2	0	0	4	19		4	1	4	0	3	17

**Abbildung 5.17:** Konfusionsmatrizen der Klassifikationsmodelle. Die Datensätze unterscheiden sich anhand des verwendeten Ki-67-Signals. Bei *Reales Ki-67* und *Ohne Ki-67* steht das reale bzw. kein Ki-67-Signal zur Verfügung. Bei *pix2pix* wurde das Ki-67-Signal mit dem pix2pix-Modell generiert. Die Ergebnisse der fünf Trainingswiederholungen wurden aufsummiert.

Weiterhin kann wahrscheinlich auch hier durch einen spezialisierten Datensatz für das Training der Transformationsmethoden eine Verbesserung bezüglich der Klassifikationsgüte erzielt werden. Der für die Untersuchung verwendete Datensatz umfasst Zellkulturen, welche aus zwei Zelltypen bestehen. Ein spezifischerer Datensatz mit nur einem Zelltyp kann die Transformationsaufgabe erleichtern und damit auch die Klassifikation mit synthetischen Markern verbessern.



Für die statistische Auswertung der Ergebnisse wurde der einseitige t-Test für zwei unabhängige Stichproben verwendet <sup>8</sup>.

### 5.3.4 Bleed-Through-Test

Wie in Abschnitt 1.2.2 beschrieben, weisen fluoreszierende Stoffe sowohl ein Absorptions- als auch ein Emissionsspektrum auf. Fluoreszierende Stoffe können somit auch durch Frequenzen außerhalb der optimalen Anregungsfrequenz angeregt werden, wobei ihre Antwort mit dem Emissionsspektrum des eigentlichen Zielmarkers überlappen kann. Es ist daher nicht auszuschließen, dass Signale von anderen Markern in den Zielkanal durchscheinen. Sind nun durchscheinende Signale des Zielmarkers in dem Kanal des Basismarkers vorhanden, können Transformationsmodelle ihre Prognose fälschlicherweise auf diese Signale stützen. In den Trainingsphasen der genutzten Transformationsmodelle werden Aufnahmen verwendet, bei denen Basis- und Zielmarker gefärbt sind. Im Falle von durchscheinenden Signalen ist dann während der Anwendung ein deutlicher Rückgang in der Prognosegüte zu erwarten, da solche Signale aufgrund der fehlenden Zielfärbung nicht mehr vorhanden sind.

Um zu prüfen, ob sich die Transformationsmodelle bei der Prognose auf eventuell vorhandene durchscheinende Signale stützen, wird ein Bleed-Through-Test anhand einer zweistufigen Färbung durchgeführt. Eine ausführliche Beschreibung der Methodik findet sich in Abschnitt 2.3.3.

Die Qualität der Registrierung der Testaufnahme nimmt einen entscheidenden Einfluss auf die Ergebnisse der beschriebenen Auswertung. Daher erfolgt zuerst eine Evaluierung der Registrierungsqualität anhand der Korrelation zwischen den

---

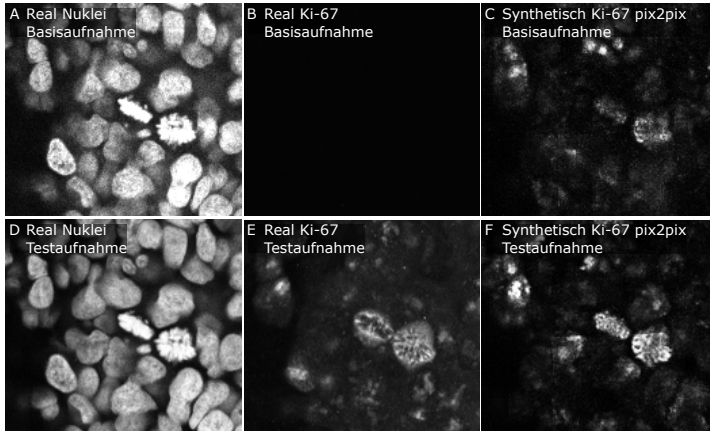
<sup>8</sup> Zuvor wurde untersucht, ob die Ergebnisse normalverteilt sind. Die Prüfung hierfür wurde mit dem Shapiro-Wilk-Test mit einem Signifikanzniveau von 0,05 durchgeführt. Bei allen Untersuchungen kann die Nullhypothese, dass die Daten normalverteilt sind, nicht verworfen werden. Zusätzlich wurde anhand des Levene-Tests geprüft, ob die Ergebnisse Varianzgleichheit aufweisen. Auch hier konnte bei allen Paaren die Nullhypothese nicht verworfen werden, weshalb von einer Varianzhomogenität ausgegangen wird.

realen Nukleussignalen aus der Basis- und Testaufnahme. Idealerweise verbleiben nur geringe Verschiebungen, woraus sich eine hohe Korrelation ergibt. Die Signale weisen bezüglich der Voxel im Vordergrund mit einem Wert von 0,475 nur eine schwache Korrelation auf. Ursachen hierfür können einerseits in verbleibenden Zellverschiebungen nach der elastischen Registrierung und andererseits in unterschiedlichen Helligkeitsverteilungen zwischen Basis- und Testaufnahme liegen. Die Übereinstimmung der Zellkernpositionen ist, visuell bewertet, auf hohem Niveau (vgl. Abbildung 5.18a und e). Größeren Einfluss nehmen daher die unterschiedlichen Helligkeitsverteilungen. In der Testaufnahme sind die Zellkerne, insbesondere in den Randbereichen, deutlich heller als in der Basisaufnahme. Durch technische und biologische Limitierungen ist eine perfekt identische Helligkeitsverteilung der beiden Aufnahmen unmöglich zu gewährleisten<sup>9</sup>. Die niedrige Korrelation zwischen den beiden Nukleussignalen verdeutlicht, dass eine perfekte Korrelation zwischen den auf Grundlage der Basisaufnahme synthetisch generierten und den realen Ki-67-Signalen aus der Testaufnahme kaum zu erreichen ist.

Die Ergebnisse der beiden Auswertungen *Basis-* und *Testprädiktion* sind in Abbildung 5.19 dargestellt. Zur Vereinfachung werden nur die Ergebnisse der objektbasierten Korrelation gezeigt. Die anderen Korrelationen (bezüglich aller Voxel und der Vordergrund-Voxel) und die Ergebnisse des ResUnets finden sich im Anhang (Abbildung A.6). Horizontale Balken geben die Korrelation der Auswertung *Basisprädiktion* an und vertikale Striche die der zweiten Auswertung *Testprädiktion*. Zusätzlich ist das Verhältnis der Korrelationen zwischen der Basis- und Testprädiktion ( $p_r$ , siehe Abschnitt 2.3.3) angegeben. Werte kleiner eins bedeuten, dass in der Testprädiktion bessere Ergebnisse als in der Basisprädiktion erzielt werden. Das verwendete pix2pix-Modell ist dabei identisch zu dem der vorherigen Auswertung (vgl. Abbildung 5.14). Da die Trainingsdaten des pix2pix-Modells sich

---

<sup>9</sup> Zum einen kann die Effektivität des Clearings durch die zweite Färbung beeinflusst werden. Zudem führt die erste Aufnahme zu einem sogenannten Photobleaching, was die allgemeine Fluoreszenz des Markers verringert. Weiterhin kann sich der Sphäroid während der zweiten Färbung leicht im Medium verschieben. Dies kann aufgrund der vom Nutzer eingestellten z-Kompensation, welche die abnehmende Helligkeit durch eine höhere Laserleistung kompensiert, zu einer unterschiedlichen Laserstärke und damit Intensität führen.



**Abbildung 5.18:** Repräsentative Bildausschnitte der zweistufigen 3D-Aufnahme. Die erste bzw. zweite Reihe zeigt reale und synthetische Bildausschnitte der Basis- bzw. Testaufnahme. Bei der Basisaufnahme wurde nur der Nukleusmarker gefärbt (A). Der Kanal für die Ki-67-Färbung (B) ist daher schwarz. Bei der Testaufnahme wurde zusätzlich zum Nukleusmarker (D) der Ki-67-Marker (E) gefärbt. C zeigt die mit dem pix2pix-Modell generierten synthetischen Ki-67-Signale, welche auf Basis von A erstellt wurden. Zusätzlich ist in F die auf Basis von D generierte synthetische Ki-67-Aufnahme dargestellt. Anhand einer Registrierung auf Grundlage der Nukleussignale wurde die Testaufnahme an die Basisaufnahme angeglichen. Alle synthetisch generierten Bilder wurden auf Basis der unregistrierten Aufnahmen erstellt und anschließend registriert.

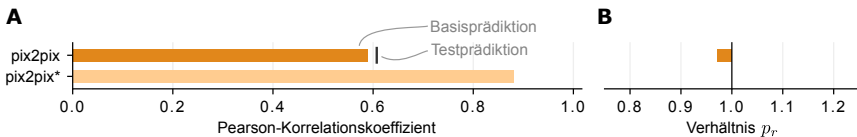
teilweise von der zweistufigen Aufnahme unterscheiden, wurde das Modell für eine Einschätzung über die erzielbare Leistung zudem ausschließlich auf der Testaufnahme trainiert. Das so trainierte Modell ist mit einem Stern gekennzeichnet<sup>10</sup>. Zugehörige repräsentative Bildausschnitte der realen und synthetisch generierten Ki-67-Signale sind in Abbildung 5.18 gezeigt.

<sup>10</sup> Die Auswertung des auf der Testaufnahme trainierten Modells (mit Stern gekennzeichnet) bezüglich der Testprädiktion ist nicht aussagekräftig, da hier die Trainingsdaten den Testdaten entsprechen. Ein Modell mit genügend Parametern kann die Trainingsdaten theoretisch auswendig lernen, was zu unrealistisch guten Ergebnissen führt. Das mit Stern gekennzeichnete Modell erzielt daher bei der Testprädiktion durchweg bessere Ergebnisse als bei der Basisprädiktion. Bei der Basisprädiktion ist die Ähnlichkeit zwischen Trainings- und Testdaten nur gering, da hier die Nuklei zwischen den beiden Aufnahmen elastisch verschoben sind. Weiterhin unterscheiden sich die Aufnahmen, wie zuvor aufgezeigt, in ihrer Helligkeitsverteilung.

Zuerst findet eine Betrachtung der Ergebnisse der *Basisprädiktion* ohne das mit Stern markierte Modell statt. Bei Betrachtung der Objektintensitäten erreicht das pix2pix-Modell eine Korrelation von 0,590. Um zu untersuchen, ob das Fehlen der Ki-67-Färbung bei der Basisprädiktion die Modellgüte negativ beeinflusst, wurde in Abschnitt 2.3.3 das Maß  $p_r$  eingeführt. Das Maß beschreibt das Verhältnis der Korrelationen zwischen Basis- und Testprädiktion. Für das pix2pix-Modell beträgt das Verhältnis  $p_r = 0,97$  und liegt damit oberhalb des festgelegten Grenzwerts von 0,9 (vgl. Abschnitt 2.3.3). Dies zeigt, dass sich das Modell bei der Prädiktion nicht auf durchscheinende Signale stützt.

Interessant ist zudem, dass sich trotz der unterschiedlichen Nuklei-Intensitäten zwischen Basis- und Testaufnahme, die Korrelationen der Basis- und Testprädiktion kaum unterscheiden. Das ist ein starkes Indiz dafür, dass die Prognose des Zielmarkers Ki-67 unabhängig von der Nukleus-Intensität erfolgt.

Beim Vergleich der Korrelation der Basisprädiktion mit den Ergebnissen aus den vorherigen Untersuchungen auf dem Datensatz *DS-Marker-Syn* (siehe Abschnitt 5.3.2) zeigt sich, dass die Korrelation um 27,4% abnimmt. Dies kann



**Abbildung 5.19:** **A** Pearson-Korrelation zwischen den Intensitäten des realen registrierten Ki-67-Signals aus der Testaufnahme und denen des aus den Nukleussignalen der Basisaufnahme generierten synthetischen Ki-67-Signalen (*Basisprädiktion*). Angegeben ist die Korrelation zwischen allen mittleren Nukleus-Intensitäten (Mittlerer Pearson Objekt) bestimmt. Die Auswertung der anderen Bereiche (zwischen allen Voxel-Intensitäten und allen Vordergrund-Voxel-Intensitäten) und des ResUnet-Modells findet sich im Anhang (Abbildung A.6). Die vertikalen Striche zeigen als Vergleich die Korrelation der aus den Nukleussignalen der Testaufnahme generierten Ki-67-Signale (*Testprädiktion*). Die für die Auswertung benötigte Nukleusannotation wird durch eine vorherige Segmentierung ermittelt. Das mit Sternen markierte Modell ist speziell auf der Testaufnahme trainiert worden. **B** Verhältnis der Korrelationen zwischen den generierten Ki-67-Signalen aus Basis- bzw. Testprädiktion ( $p_r$ , siehe Abschnitt 2.3.3). Da bei dem mit einem Stern markierten Modell Trainings- und Testdaten identisch sind, wird auf die Angabe der Testprädiktionsergebnisse und der daraus abgeleiteten Verhältnisse verzichtet <sup>10</sup>.

auf zwei Hauptursachen zurückgeführt werden: Zum einen wirken sich die nach der elastischen Registrierung verbleibenden Verschiebungen zwischen Basis- und Testaufnahme negativ auf die Ergebnisse aus. Zum anderen wurde die Probe der zweistufigen Aufnahme erst deutlich später erstellt und aufgenommen als die für das Modelltraining genutzten Daten (*DS-Marker-Syn*), wodurch sich die Signale bei visueller Betrachtung unterscheiden. Die Gründe hierfür können vielseitig sein und sind nicht zwingend vom Nutzer verursacht. Beispielsweise können verschiedene Batches oder Lagerzeiten der Marker und Zellen zu Veränderungen in den Signalen (z. B. Textur) oder in der Morphologie führen.

Um auch eine Einschätzung der erreichbaren Güte bei optimalen Trainingsdaten zu erhalten, wurde das pix2pix-Modell zusätzlich auf der Testaufnahme der zweistufigen Färbung trainiert. In der Abbildung 5.19 ist das Modell mit einem Stern gekennzeichnet. Durch das Training auf diesen Daten können die erzielten Ergebnisse merklich gesteigert werden. Bezüglich der Korrelation zwischen den mittleren Objektintensitäten ergibt sich eine Verbesserung von 49,2%. Durch den Vergleich mit dem erneut trainierten Modell kann gezeigt werden, dass vor allem bei der Ki-67-Markertransformation die Ähnlichkeit zwischen Trainings- und Anwendungsdaten wichtig ist. Für optimale Ergebnisse sind die Zellkulturen zur Generierung der Trainingsdaten gleichzeitig mit den Zellkulturen für die spätere Anwendung zu kultivieren.

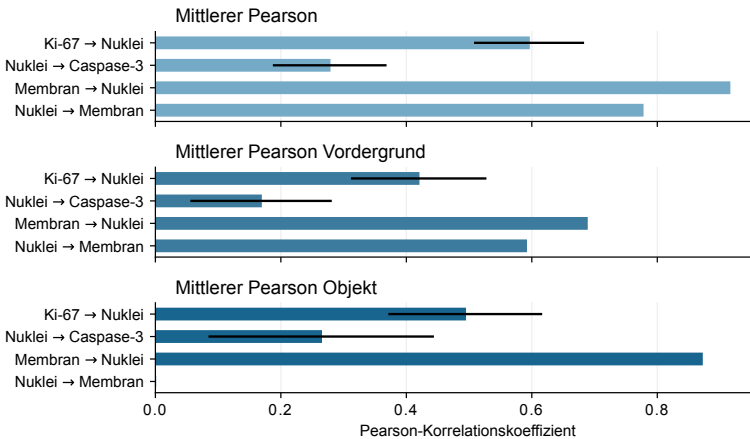
Zusammenfassend kann durch die Untersuchungen anhand der zweistufigen Färbung gezeigt werden, dass das Modell sich nicht auf durchscheinende Signale stützen und damit auch in einer späteren Anwendungsphase funktioniert, bei der keine Färbung des Zielmarkers vorhanden ist.

### 5.3.5 Machbarkeitsanalyse für weitere Marker

Um notwendige Bedingungen für eine erfolgreiche Markertransformation zu identifizieren, werden in Abschnitt 2.3.5 drei zu untersuchende Fragestellungen formuliert. Zur experimentellen Beantwortung dieser Fragen werden in dem genannten

Abschnitt auch geeignete Kombinationen aus Basis- und Zielmarker bestimmt. Im Folgenden werden die Ergebnisse der Experimente vorgestellt und diskutiert.

Für die erste Fragestellung wird die Transformation von Ki-67-Signalen zu Nukleussignalen untersucht. Anhand der Zellen, welche kein Ki-67-Signal zeigen, kann geprüft werden, ob deren Nukleussignale aus den umliegenden Signalen abgeleitet werden können. Im Zuge der zweiten Fragestellung wird die Markerkombination Nukleus- und Membranmarker untersucht. Die Kombination ermöglicht es zu prüfen, ob die Positionen der Signale des Zielmarkers mit denen des Basismarkers überlappen müssen. Die Transformation des Nukleusmarker zu Caspase-3 wird herangezogen, um die dritte Fragestellung zu beantworten. Hier wird geprüft, ob die Veränderungen des Basismarkers spezifisch für den Zielmarker sein müssen.

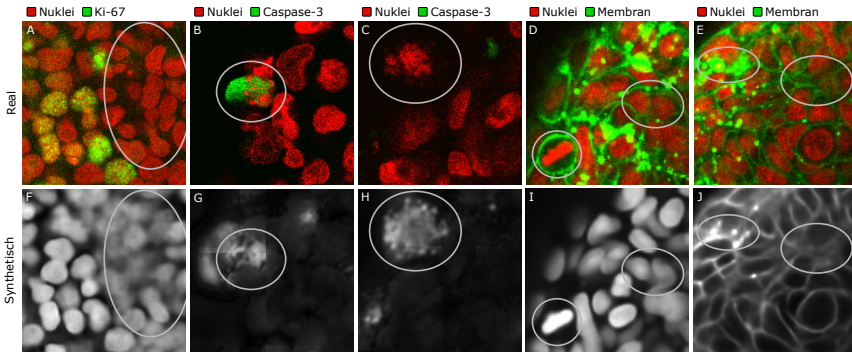


**Abbildung 5.20:** Pearson-Korrelation zwischen den Intensitäten der realen Signale des Zielmarkers und denen anhand des Basismarkers generierten synthetischen Signalen. Die Korrelation wird zwischen allen Voxel-Intensitäten (Mittlerer Pearson), allen Vordergrund Voxel-Intensitäten (Mittlerer Pearson Vordergrund) und allen Nukleus-Intensitäten (Mittlerer Pearson Objekt) bestimmt. Die Vordergrundbereiche und die Objektinstanzen werden dabei über eine vorherige Segmentierung bestimmt. Verglichen werden die Ergebnisse der Transformationen<sup>11</sup> verschiedener Markerkombinationen. Soweit vorhanden, geben die Fehlerbalken die Standardabweichung bezüglich mehrerer Bilder dar.

Abbildung 5.20 zeigt die Ergebnisse der Untersuchung bezüglich der Vorhersage unterschiedlicher Marker.<sup>11</sup> Bei der inversen Transformation des Ki-67-Signals, zeigt sich, verglichen mit der vorherigen untersuchten Richtung, eine leicht schlechtere Leistung. Mit Korrelationen von 0,596, 0,42 bzw. 0,494 für alle Voxel, Vordergrund-Voxel und Objekten sind die Korrelationen im Schnitt um 0.226 geringer als für die zuvor untersuchte Richtung (vgl. Abbildung 5.20 und Abbildung 5.14). Ein Hauptgrund für die Abnahme wird die deutlich schwierigere Aufgabe der Transformation darstellen. Wie beschrieben weisen Zellkerne außerhalb des Zellzyklusstadiums keine Ki-67-Signale auf (vgl. Abbildung 5.21a). In diesem Fall versagt die Prognose der Zellkerne fast vollständig, wobei nur uneindeutige Signale ohne Abgrenzung einzelner Kerne erzeugt werden. Interessanterweise ist das Netzwerk in der Lage, anhand der umliegenden Ki-67-positiven Zellkerne abzuleiten, dass sich zwischen diesen keine Lücke befindet, sondern ein weiterer Zellkern. Da nur ein kleiner Teil der Zellkerne in der Mitte der Zellkultur keine Ki-67-Signale aufweist, sind die Korrelationen vergleichsweise hoch.

Die Generierung des Apoptosemarkers Caspase-3 auf Basis des Nukleussignals stellt eine Herausforderung dar. Hier liegen mit Werten von 0,278, 0,168 bzw. 0,264 eher geringe Korrelationen zwischen dem realen und dem synthetisch generierten Caspase-3-Signal vor (vgl. Abbildung 5.20). Eine Schwierigkeit bei diesen Daten besteht darin, dass nur wenige Caspase-3 positive Zellen in den Daten vorhanden sind. Ein Großteil der Bildinhalte des Zielmarkers ist somit schwarz, woraus sich eine starke Imbalance zwischen Vorder- und Hintergrund ergibt. Weiterhin stellt die mit Caspase-3 markierte Apoptose nur eine von vielen möglichen Zelltodvarianten dar. Das macht es fast unmöglich, allein anhand von fragmentierten Zellkernen auf die genaue Art des Zelltodes und damit auf das Vorhandensein von Caspase-3-Signalen zu schließen. So wird in Abbildung 5.21c und h ein fragmentiert wirkender Zellkern als Caspase-3-Positiv prognostiziert, obwohl in Wirklichkeit kein Signal vorhanden ist. Weiterhin erschwert die Fragmentierung der Zellkerne die positionell korrekte Prognose der Signale, wie es in

<sup>11</sup> Die Untersuchungen zu den Transformationen „Ki-67 → Nuklei“ und „Nuklei → Caspase-3“ wurden im Rahmen der in dieser Arbeit entstandenen Veröffentlichung [188] durchgeführt und nutzen daher das ResUnet statt des pix2pix-Modells zur Transformation.



**Abbildung 5.21:** A-E Repräsentative Bildausschnitte von realen 3D-Aufnahmen, welche mit Basis und Zielmarker gefärbt sind. F-J Zugehörige Bildausschnitte synthetisch generierter 3D-Aufnahmen des Zielmarkers, welche anhand des Basismarkers prognostiziert wurden. A,F Ki-67 als Basismarker und Nuklei als Zielmarker. B,G bzw. C,H Nuklei als Basismarker und Caspase-3 als Zielmarker. D,I Membran als Basismarker und Nuklei als Zielmarker. E,J Nuklei als Basismarker und Membran als Zielmarker. Markierte Bereiche werden im Text diskutiert.

Abbildung 5.21b und g zu sehen ist. Hier wurde zwar korrekterweise Caspase-3-Signal vorhergesagt, allerdings stimmt die Lokalisierung nur teilweise überein. Für das Beispiel der Prognose des Caspase-3-Signals lässt sich keine direkte Beziehung zwischen Zellkernmorphologie und Textur zu dem resultierenden Caspase-3-Signal herstellen. Entsprechend funktioniert auch die Prognose des Markers auf Basis des Nukleusmarkers nur eingeschränkt.

Die Prognose von Nukleussignalen auf Basis von Membransignalen erzielt die höchste Korrelation der in dieser Arbeit analysierten Datensätze (vgl. Abbildung 5.20). Die Korrelation bei Betrachtung aller Voxel liegt bei 0,915. Die auf den Vordergrund bzw. auf die Objekte bezogene Korrelation liegt bei 0,688 bzw. 0,871. Wie schon bei der Prädiktion der Ki-67-Signale nimmt die Korrelation bei der ausschließlichen Betrachtung der Vordergrundregionen ab, da die einfach vorherzusagenden Hintergrundregionen entfallen. Der Anstieg der Korrelation auf Objektebene lässt sich wiederum durch die Vereinfachung der Aufgabe erklären. Während die realen Nukleussignale einen großen Rauschanteil aufweisen, zeigen die prognostizierten Bilder kaum Rauschen und die Zellkerne keine Textur (vgl. Abbildung 5.21i). Da bei einem Vergleich anhand der mittleren Intensität der



Zellkerne das Rauschen und die Textur an Einfluss verlieren, vereinfacht es die Aufgabe.

Bei Betrachtung der prognostizierten Bilder zeigt sich weiterhin, dass die Transformation auch bei Signalartefakten, wie den überbelichteten lokalen Signalanreicherungen, robust ist. Auch in Bereichen, an denen die Signale nur schwach ausgeprägt sind (vgl. Abbildung 5.21d), werden die Nukleusformen stellenweise recht gut prognostiziert. Allerdings gibt es auch Bereiche, an denen eine Abgrenzung der Zellkerne scheitert. Allerdings sind hier auch in den realen Nukleussignalen die Abgrenzungen zwischen den Nuklei nur schwach ausgeprägt. Interessanterweise scheint das Modell anhand der Membransignale auch Zellzyklusstadien zu erkennen und die dazugehörige Morphologie der Nuklei korrekt zu prognostizieren (vgl. Abbildung 5.21d und i).

Auch die Synthese von Membransignalen auf Basis von Nukleussignalen zeigt vielversprechende Ergebnisse. Die Bewertung der synthetischen Membransignale stellt aufgrund ihrer linienförmigen Struktur allerdings keine triviale Aufgabe dar. Kleine Abweichungen in der Position der Signale führen so schon zu großen Änderungen in der Korrelation zwischen den realen und prognostizierten Signalen. Auch stellt die Segmentierung der Membransignale aufgrund der vielen vorhandenen Artefakte, wie den hellen Signalanreicherungen, keine einfache Aufgabe dar. Aus diesen Gründen wird auf eine Berechnung und Angabe der objektbasierten Korrelation verzichtet. Für die vordergrundbezogene Korrelation wird eine anhand der Membransignale durchgeführte Segmentierung der kompletten Zellkultur genutzt. Die Korrelation bei Betrachtung aller Bildbereiche liegt für die prognostizierten Membransignale bei 0,777 und bei Betrachtung des Vordergrundbereiches bei 0,591 (vgl. Abbildung 5.20).

Auffällig bei der Betrachtung der realen Membranaufnahme Abbildung 5.21e sind die hellen Artefakte, welche die eigentlichen Membransignale überdecken. Es ist anzunehmen, dass diese Artefakte durch die übermäßige Aggregation von Fluorophoren entstehen. Wie in Abbildung 5.21j dargestellt, sind solche Artefakte auch in den synthetisch generierten Signalen vorhanden, wenn auch weniger stark ausgeprägt als in den realen Daten. Ohne eine tiefere Analyse ist nicht ersichtlich,

ob diese Artefakte eher zufällig in die synthetischen Daten eingebracht werden, oder ob diese an Stellen auftreten, an denen auch in den realen Daten Artefakte zu erkennen sind. Da nicht zu erwarten ist, dass solche Artefakte genau vorhergesagt werden können, besitzen diese durch die hohen Intensitäten einen starken negativen Einfluss auf die korrelationsbasierte Auswertung. Die synthetisch generierten Membranen scheinen in Position und Signalintensität gut mit den realen Signalen übereinzustimmen. So nimmt etwa die Helligkeit der Membranen in Richtung der Zellkultur ab. Im Vergleich zu den realen Aufnahmen zeigen die synthetisch generierten Membranaufnahmen allerdings kaum Rauschanteile. Probleme bei der Vorhersage der Membranen scheint es an Stellen zu geben, bei denen größere Abstände zwischen den Zellkernen existieren (vgl. Abbildung 5.21e und j). Das Modell hat hier aufgrund fehlender Referenzen Schwierigkeiten, den genauen Verlauf der Membran präzise vorherzusagen. Das äußert sich durch eine verwaschene Prognose in diesem Bereich.

Da keine weitere Optimierung der Modelle und Trainingsparameter für diese Analyse durchgeführt wurde, kann theoretisch eine weitere Verbesserung der Ergebnisse erzielt werden. So kann eine stärkere Gewichtung des Adversarial-Losses die fehlende Textur bei der Prognose der Nukleussignale auf Basis von Membransignalen verbessern.

Insgesamt zeigt die Analyse, dass auch weitere Marker erfolgreich vorhergesagt werden können. Der Erfolg einer Prognose hängt jedoch maßgeblich davon ab, ob im Basismarker relevante Informationen über den Zielmarker enthalten sind. Basierend auf den in Abschnitt 2.3.5 formulierten Fragestellungen und den durchgeführten Experimenten lassen sich folgende Voraussetzungen für eine erfolgreiche Transformation festhalten:

1. Ein eindeutiger Zusammenhang zwischen Basis- und Zielmarker ist erforderlich. Beispielsweise funktioniert die Prognose von Caspase-3-Signalen nur eingeschränkt, da fragmentierte Zellkerne nicht ausschließlich durch Apoptose, sondern auch durch andere Ursachen entstehen können.
2. Marker, bei denen in bestimmten Regionen keine Basismarker-Signale für die Prognose des Zielmarkers vorhanden sind, eignen sich nur bedingt für

eine Transformation. Die Prognose von Nukleussignale auf Basis des Ki-67-Markers hat gezeigt, dass eine korrekte Prognose nicht mehr möglich ist, wenn im Basismarker keine Signale vorhanden sind.

Darüber hinaus konnte gezeigt werden, dass eine positionelle Übereinstimmung zwischen Basis- und Zielmarker nicht erforderlich ist. Die Zusammenhänge zwischen den Markern können also komplexer sein. So erzielten die Transformationen zwischen Membran- und Nukleussignalen gute Ergebnisse, obwohl diese keine örtliche Übereinstimmung aufweisen.

Membran- und Nukleusmarker scheinen ideale Basismarker für die Synthese weiterer Marker darzustellen, da ihre Signale über die gesamte Zellkultur hinweg sichtbar sind. Bei neuen Markerkombinationen können die oben aufgezeigten Bedingungen auf Basis von Vorwissen oder durch visuelle Begutachtung einfach abgeschätzt werden. Ergänzend bietet die Bewertung mithilfe der in dieser Arbeit eingeführten Methoden (vgl. Abschnitt 2.3.3) eine zusätzliche Absicherung, ob eine Transformation erfolgreich ist.

### **5.3.6 Zusammenfassung**

Die Anzahl der gleichzeitig möglichen Färbungen ist bei der Fluoreszenzmikroskopie begrenzt. Die Antworten der Fluorophore erstrecken sich über ein ganzes Spektrum. Überschneidungen in den Antworten führen zu unspezifischen Signalen. Diese Limitierung schränkt den möglichen Informationsgehalt ein. Um diese Limitierungen zu überwinden, wurde eine Methode entwickelt, mit der erstmals die Signale von Fluoreszenzmarkern ineinander überführt werden können. Anhand eines Basismarker kann das Signal eines Zielmarkers mit den Methoden vorhergesagt werden. Bemerkenswert ist zudem, dass die eingeführte Methode auch die Transformation der Marker in großen 3D-Zellkulturen ermöglicht, was angesichts der heterogenen Bedingungen eine anspruchsvolle Aufgabe darstellt.

Die Leistung der zwei entwickelten Methoden wurde anhand der erzielten Korrelation zwischen dem echten Ground-Truth und dem prädierten Signal verglichen.

Da die Aufnahmen große Teile an Hintergrundregionen enthalten, welche für die Modelle einfach vorherzusagen sind, wurden zwei spezialisierte Bewertungsmaße eingeführt: die Korrelation von Vordergrund-Voxel und die Korrelation der mittleren Objektintensitäten. Anhand der Bewertungsmaße konnte gezeigt werden, dass mit beiden Methoden eine hohe Prädiktionsgüte erzielt werden kann. Der GAN-basierte Ansatz zeigte bei den Metriken leicht bessere Ergebnisse. Allerdings erforderte das GAN-basierte Training eine längere Rechenzeit und bei gleichen Bedingungen auch einen erhöhten GPU-Speicherbedarf.

Ein wichtiger Punkt in der Leistungsbewertung stellt zudem die Überprüfung dar, ob sich die trainierten Modelle auf eventuell durchscheinende Signale stützen. Liegen die Frequenzbänder von Basis- und Zielmarker nahe beieinander, kann es zu einem sogenannten Bleed-Through kommen, bei dem Signale des Zielmarkers in den Basismarker durchscheinen. In einer Anwendungsphase entfallen solche Signale, da keine Färbung des Zielmarkers mehr vorliegt. Für die Überprüfung, ob sich die Modelle bei ihrer Prognose auf eventuell vorhandene durchscheinende Signale stützen, wurden zwei Protokolle entwickelt. Die Protokolle unterscheiden sich in ihrem Aufwand und in der Aussagekraft. Das erste Protokoll *Zweistufige Färbung* erlaubt eine direkte quantitative Überprüfung der Leistung in der Anwendungsphase. Das zweite Protokoll *Alternative Färbung* ist deutlich weniger komplex und erfordert in den meisten Fällen keine zusätzliche Probengenerierung. Allerdings besitzt es eine geringere Aussagequalität, da keine Ground-Truth vorliegt. Anhand der zweistufigen Färbung konnte gezeigt werden, dass beide Methoden zur Markertransformation sich nicht auf durchscheinende Signale stützen und in der Anwendungsphase ähnlich gute Ergebnisse liefern wie in der Trainingsphase. Weiterhin konnte mit den Daten gezeigt werden, dass die beiden Methoden unterschiedlich robust gegenüber Abweichungen zu den Trainingsdaten sind. So zeigt die komplexere GAN-basierte Methode pix2pix eine geringere Generalisierungsfähigkeit.

Zudem wurde eine Machbarkeitsstudie durchgeführt, um Voraussetzungen und Grenzen der Synthese von Bildmodalitäten zu identifizieren. Hier konnte gezeigt werden, dass die Synthese von Bildmodalitäten auch mit anderen Markern funktioniert. Voraussetzung ist, dass ein eindeutiger Zusammenhang zwischen den

beiden Markern vorliegt. Dieser kann auch, wie bei der Membransynthese gezeigt wurde, komplexer und ohne lokale Übereinstimmung der Signale ausgeprägt sein.

Weiterhin wurde die Leistung der Modelle und die Praxisrelevanz der Markerttransformation anhand eines realen Anwendungsfalls überprüft. Ziel hierbei war die Bestimmung des Zellzyklusstadiums einzelner Zellen. Zur Unterstützung wird hierbei der Proliferationsmarker Ki-67 eingesetzt. In den Experimenten konnte gezeigt werden, dass durch die Synthese des Proliferationsmarkers signifikant bessere Klassifikationsgüten erzielt werden konnten als ohne diesen Marker.

## 6 Zusammenfassung und Ausblick

Fortschrittliche Sensortechnologien und automatisierte Prozesse ermöglichen die Erfassung großer Datenmengen mit wachsender Komplexität. Unter anderem bietet die Auswertung der Daten erhebliches Potenzial für die Entwicklung neuer Materialien und Wirkstoffe sowie die Verbesserung und Sicherung der Produktqualität. In dieser Arbeit wird der Fokus auf die Auswertung von 3D-Bilddaten gelegt, die in der Medizin, Materialforschung und Biologie von großer Bedeutung sind.

Ein besonderer Schwerpunkt liegt auf der Analyse mikroskopischer Bilddaten von 3D-Zellkulturen, wie Sphäroide und Organoide. Im Vergleich zu 2D-Zellkulturen steigt die Komplexität der Bildinhalte deutlich. Zu den wichtigsten Unterschieden zählen kompaktere Zellanordnungen, Überlagerungen von Objekten durch optische Effekte und variierende Signal-Rausch-Verhältnisse durch Lichtbrechungen. Deep-Learning-basierte Algorithmen bieten für solche komplexen Daten ein großes Potenzial, erfordern jedoch ausreichend viele und qualitativ hochwertige Trainingsdaten. Allerdings ist die manuelle Annotation großer Mengen an 3D-Daten aufgrund ihrer hohen Komplexität kaum realisierbar. Eine Alternative stellt die Synthese von Trainingsdaten dar, bei denen die Annotationen bereits bekannt sind. Aktuelle Methoden zur Generierung solcher Trainingsdaten weisen jedoch Limitierungen auf, die in dieser Arbeit adressiert wurden.

Die Fluoreszenzmikroskopie ist zudem aufgrund von Signalüberlagerungen in der maximalen Anzahl gleichzeitig erfassbarer fluoreszierender Marker begrenzt. Dies führt zu einer direkten Einschränkung der extrahierbaren Informationen. In dieser Arbeit wurden daher Methoden entwickelt, um synthetische Markersignale zu generieren und die Limitierungen damit abzuschwächen.

Im Folgenden werden die zentralen Neuerungen und die wichtigsten Ergebnisse dieser Arbeit zusammengefasst:

- Es wurde ein umfassendes Konzept zur Synthese von Bilddaten entwickelt, welches die zu Beginn identifizierten Probleme löst. Auf Basis eines umfangreichen Methodenvergleichs wurde eine Vorgehensweise erstellt, um hinsichtlich gegebener Kriterien die optimale Methode zu wählen.
- Die neu entwickelte Methode SimOptiGAN ermöglicht erstmals die Integration seltener Elemente und Objekte verschiedener Klassen in synthetische 2D- und 3D-Daten sowie die Generierung von zusätzlichen Klassenannotationen. Zudem übertrifft SimOptiGAN die Segmentierungsleistung eines vortrainierten Modells und von händisch annotierten Trainingsdaten.
- Mit SimOptiGAN+, Mem2NucGAN-P und Mem2NucGAN-U wurden drei Methoden eingeführt, welche erstmalig biophysikalische Simulationen zur Verbesserung der Zellanordnung in den Syntheseprozess integrieren. Mit SimOptiGAN+ konnte die Segmentierungsleistung von SimOptiGAN nochmals gesteigert werden.
- SimOptiGAN ist auch in anderen Forschungsbereichen anwendbar. So wurde anhand von REM-Aufnahmen verdeutlicht, dass SimOptiGAN auch in der Materialentwicklung von Interesse ist.
- Mit DualGenGAN wurde eine GAN-Trainingsstruktur eingeführt, welche es erstmalig ermöglicht, neben Bilddaten auch passende Annotationsdaten zu generieren.
- Es wurde zudem eine neue Möglichkeit geschaffen, gepaarte synthetische Membran- und Nukleussignale zu generieren. Wie gezeigt, verbessert deren Kombination die Ergebnisse der Membransegmentierung.
- Mit SimOptiGAN wurde erstmalig der Einfluss seltener Objekte in Trainingsdaten untersucht. Die Einbringung konnte die Segmentierungsleistung dieser deutlich verbessern, ohne die Leistung bei regulären Objekten merklich zu verschlechtern.

- Mit IMS wurde ein neues Framework zur Synthese von Bildmodalitäten entwickelt, das die genannten Probleme der limitierten Anzahl an gleichzeitig aufnehmbaren Markern löst. Anhand eingeführter Bewertungsmethoden konnte eine hohe Korrelation zwischen synthetischen und realen Signalen nachgewiesen werden. Zudem wurde gezeigt, dass mit synthetischen Markern die Klassifikation von Zellzyklusstadien verbessert wird.
- Es wurden zwei Protokolle entwickelt, um zu überprüfen, ob sich die Modelle von IMS auf durchscheinende Signale stützen. Damit konnte nachgewiesen werden, dass dies nicht zutrifft und die Prognosequalität auch bei Anwendungsdaten hoch bleibt.
- Es wurde erstmalig eine Machbarkeitsanalyse durchgeführt, bei der die Voraussetzungen für die Übertragbarkeit von IMS auf weitere Bildmodalitäten aufgezeigt wurden.

Die entwickelten Methoden sind zwar auf die Synthese von Zellkulturen und deren Signale spezialisiert, aber nicht ausschließlich darauf limitiert. Durch die Anpassung der Parametrierung lassen sich auch andere Strukturen synthetisieren, die viele Objektinstanzen aufweisen. Weiterhin ist SimOptiGAN modular aufgebaut, was eine Anpassung für die Generierung anderer Strukturen erlaubt. So können durch die Adaption der virtuellen Probengenerierung andere Strukturen simuliert werden, ohne die restlichen Module anzupassen. IMS ist weiterhin unabhängig von den Bildinhalten und kann auf beliebige Signalmodalitäten angewendet werden, solange die identifizierten Zusammenhänge zwischen ihnen bestehen.

Das Forschungsfeld der synthetischen Datengenerierung ist von schnellem Fortschritt geprägt. Während der Entstehungszeit dieser Arbeit wurden neue Architekturen und Trainingskonzepte veröffentlicht. Diffusionsmodelle gewannen etwa für die Generierung von Bilddaten an Bedeutung. Während die Modelle bislang eine große Menge an Trainingsdaten benötigen, ist es aufgrund ihrer Leistung interessant, diese Modelle in die entwickelten Synthesemethoden zu integrieren. Ein Vorteil der eigens entwickelten Synthesemethoden ist ihr modularer Aufbau,



der sie nicht auf GAN-Modelle begrenzt und eine einfache Integration neuer Transformationsmethoden ermöglicht.

Ein bestehendes Problem bei der Entwicklung von neuen Methoden zur Trainingsdatensynthese ist das Fehlen eines öffentlichen und vollständig annotierten realen Benchmark-Datensatzes. Ein einheitlicher Benchmark-Datensatz erleichtert die Entwicklung neuer Methoden, da die Ergebnisse neuer Ansätze direkt mit den Resultaten bestehender Verfahren verglichen werden können, die bereits auf diesem Datensatz evaluiert wurden. Die vorliegende Arbeit leistet hier einen ersten Beitrag, indem der zum Vergleich der Synthesemethoden genutzte Datensatz inklusive der Annotationen öffentlich zugänglich gemacht wurde. Auch wurden die Implementierungen der Methoden veröffentlicht.

Um den geringen verbleibenden Annotationsaufwand von SimOptiGAN(+) weiter zu reduzieren, können in zukünftigen Arbeiten Methoden zur Synthese von Objektprototypen untersucht werden. Die Herausforderung besteht darin, eine Möglichkeit zu finden, neue Objektprototypen einschließlich ihrer Annotationen ohne zusätzliches manuelles Annotieren zu generieren. Darüber hinaus kann an Methoden geforscht werden, um die Parametrisierung von SimOptiGAN zu automatisieren. Beispielsweise können Parameter zur Rauschverteilung, Objektdichte sowie zum Helligkeitsabfall anhand realer Daten automatisch festgelegt werden. Eine große Schwierigkeit besteht darin, Methoden zu entwickeln, die eine Extrahierung ohne die Unterstützung von Annotationen ermöglichen, das heißt ohne vorheriges Wissen darüber, wo sich die Objekte im Bild befinden.

Eine weitere Forschungslücke besteht in der Entwicklung von Methoden, um die Übereinstimmung zwischen Bild- und Annotationsdaten quantitativ und präzise zu bewerten. Neben der Analyse, ob GAN-Modelle während der Transformation Unstimmigkeiten verursachen, bietet ein solcher Vergleich auch die Möglichkeit, die Qualität händisch erstellter Annotationsdaten zu beurteilen. Eine derartige Bewertungsmethode ist daher für verschiedene Forschungsfelder von großem Interesse.

Weiterhin nimmt die Bedeutung an Live-Cell-Aufnahmen zu, welche die zusätzliche Analyse von temporalen Effekten erlauben. Durch die zeitliche Dimension

steigt die Komplexität und der zeitliche Aufwand der händischen Annotation nochmals stark an. Die in dieser Arbeit durchgeführte Integration von biophysikalischer Simulation in die Bilddatensynthese bietet eine ideale Grundlage für die Erstellung von synthetischen 3D+t-Aufnahmen. So können die biophysikalischen Methoden und die Methoden zur Bildsynthese in zukünftiger Forschung erweitert werden, um realistische temporale Effekte zu synthetisieren.

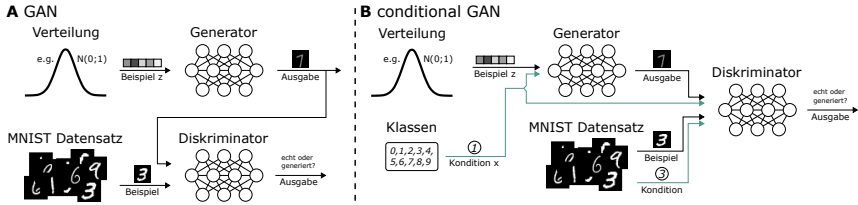
# A Anhang

## A.1 Einführung generativer Netzwerke

### A.1.1 Generative Adversarial Networks

Generative Netzwerke ermöglichen die synthetische Erzeugung beliebiger Daten. Anfangs erforderte das Training generativer Netzwerke allerdings die Entwicklung komplexer und spezialisierter Verlustfunktionen. Goodfellow et al. [199] führen 2014 mit den Generative Adversarial Networks (GANs) eine neue Methode ein, mit der das Training generativer Netzwerke erheblich vereinfacht wird. Der zentrale neue Punkt dieser Methode ist ein Diskriminator, welcher dem generativen Netzwerk als Kontrahent gegenübergestellt wird. Anstatt den Loss für den Generator händisch zu formulieren, übernimmt der Diskriminator die Erstellung der Optimierungsfunktion. Der *Generator* versucht möglichst real aussehende Daten zu erzeugen, um den Diskriminator zu täuschen (vgl. Abbildung A.1a). Der *Diskriminator* wird zeitgleich mit dem Generator trainiert, mit dem Ziel, zwischen realen und synthetischen Daten zu unterscheiden. In einem idealen Trainingsszenario werden beide Modelle immer besser, womit der Generator am Ende des Trainings in der Lage ist, täuschend echte Daten zu generieren.

Der Trainingsprozess stellt ein Minimax-Spiel zwischen dem Generator  $G$  und dem Diskriminator  $D$  dar.  $G$  und  $D$  sind hierbei beliebige differenzierbare Funktionen mit Parametern  $\zeta$  und  $\theta$ . Der Generator bildet Daten  $z \in Z$  der Domäne  $Z$  auf Daten  $y \in Y$  der Domäne  $Y$  ab. Die Daten  $z$  und  $y$  folgen der Wahrscheinlichkeitsverteilung  $p_y(y)$  bzw.  $p_z(z)$ :  $y \sim p_y(y)$  bzw.  $z \sim p_z(z)$ . Als



**Abbildung A.1:** Übersicht der Struktur von Generative Adversarial Networks (GANs). **A** zeigt die Struktur des originalen GANs. Anhand eines *Beispiels*, welches in der Regel aus einer Normalverteilung gezogen wird, generiert der *Generator* eine synthetische *Ausgabe*. Der *Diskriminator* hat dann die Aufgabe, zwischen realen *Beispielen* und synthetischen *Ausgaben* zu unterscheiden. Anhand der *Ausgabe* des *Diskriminators* können die Gewichte des *Generators* optimiert werden. Als Datensatz wird hier beispielhaft der MNIST-Datensatz [200] gezeigt. **B** zeigt die Struktur eines conditional GANs, mit dem Einfluss auf die Ausgaben des *Generators* genommen werden kann. Der *Generator* und *Diskriminator* erhalten als zusätzliche Eingabe eine *Kondition*, welche in dem Beispiel die Klasse bzw. die Zahl des Bildes darstellt. Der *Diskriminator* übernimmt nun zusätzlich die Aufgabe, die Konsistenz zwischen Bild und *Kondition* zu prüfen.

Verteilung  $p_z(z)$  wird in der Regel die Normalverteilung genutzt. Die echte Wahrscheinlichkeitsverteilung der Domäne  $\mathcal{Y}$  ist hingegen unbekannt und wird über die empirische Verteilung der Trainingsdaten abgeschätzt:  $p_{\text{data}}(y)$ . Das gemeinsame Optimierungskriterium von  $G$  und  $D$  ergibt sich mit dem Erwartungswert  $\mathbb{E}$  zu:

$$\min_G \max_D O(G, D) = \mathbb{E}_{y \sim p_{\text{data}}(y)} [\log D(y)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]. \quad (\text{A.1})$$

Die Optimierungsfunktion ist an den Binary-Cross-Entropy-Loss<sup>1</sup> angelehnt<sup>2</sup>. In der Praxis wird die Optimierungsfunktion für eine bessere Trainingsstabilität [201] auf den Least-Squares-Loss<sup>3</sup> umgestellt. Stellen  $a$  und  $b$  die gewünschten

<sup>1</sup>  $\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^N (g_i * \log D(s_i) + (1 - g_i) * \log(1 - D(s_i)))$ , mit dem Datenpunkt  $s_i$ , der wahren Klasse  $g_i$  des Datenpunktes und der Anzahl der Datenpunkte  $N$ .

<sup>2</sup> Im Kontext des maschinellen Lernens bezeichnet der Begriff Loss bzw. Verlustfunktion eine numerische Metrik, welche zur Bewertung der Modellleistung herangezogen wird. Sie bildet die Grundlage für die Optimierung des Modells.

<sup>3</sup>  $\mathcal{L}_{LS} = \frac{1}{N} \sum_{i=1}^N (D(s_i) - g_i)^2$

Ausgaben von  $D$  bei echten bzw. generierten Eingaben dar, ergeben sich für  $G$  und  $D$  die folgenden Optimierungsfunktionen:

$$\min_D O_D(D) = \mathbb{E}_{y \sim p_{\text{data}}(y)} [(D(y) - a)^2] + \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - b)^2], \quad (\text{A.2})$$

$$\min_G O_G(G) = \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - a)^2]. \quad (\text{A.3})$$

Radford et al. [202] erweitern 2015 das klassische GAN zu einem Deep-Convolutional-GAN (DCGAN), welches sich besonders für die Generierung von Bildern eignet.

Das beschriebene GAN ermöglicht zwar die Generierung realistischer Bilder aus zufälligen Vektoren. Allerdings ist eine gezielte Beeinflussung der ausgegebenen Bilder nur schwer erreichbar. So kann, wie im gezeigten Beispiel (siehe Abbildung A.1a), nicht festgelegt werden, welcher Klasse (Zahl) das generierte Bild angehören soll. Damit sind die Klassen der generierten Daten unbekannt, was ein nachträgliches Annotieren erfordert, um sie für das Training von Modellen verwenden zu können. Mit dem conditional GAN (cGAN) stellt Mirza et al. [203] 2014 eine Möglichkeit vor, um die generierten Daten auf direktem Wege steuern zu können. Damit wird ermöglicht, dass anhand von vorgegebenen Klassen bzw. Annotationen, passende Bilder generiert werden. Dazu werden sowohl der *Generator* als auch der *Diskriminator* mit einer zusätzlichen Information  $x$  beaufschlagt, welche als Kondition bezeichnet wird (vgl. Abbildung A.1b). Das Optimierungskriterium aus Gleichung A.1 erweitert sich so zu:

$$\min_G \max_D O(G, D) = \mathbb{E}_{x,y} [\log D(y, x)] + \mathbb{E}_{x,z} [\log(1 - D(G(z, x), x))]. \quad (\text{A.4})$$

Aus Gründen der Übersichtlichkeit wird nun auf die ausführliche Schreibweise der Verteilungsannahmen verzichtet. Neben der Prüfung, ob die Daten echt oder synthetisch sind, erweitert sich die Aufgabe des Diskriminators so implizit um die Prüfung, ob die Daten und die Kondition zusammenpassen.

Isola et al. [191] stellt 2017 auf Basis der cGAN-Trainingsstruktur eine universelle Implementierung vor (pix2pix), um Bilder der Domäne  $\mathcal{X}$  in die Domäne  $\mathcal{Y}$  zu überführen (siehe Abbildung A.2a). Sie bezeichnen diesen Prozess als Bild-zu-Bild-Transformation. In diesem Kontext stellen  $x$  und  $y$  Bilder der Domäne  $\mathcal{X}$  bzw.  $\mathcal{Y}$  dar.

Zwar kann auch ohne Adversarial-Training (Gleichung A.4) eine Bild-zu-Bild-Transformation erlernt werden, indem der Generator über die Differenz zwischen generiertem Bild  $\hat{y}$  und dem dazugehörigen Zielbild  $y$  trainiert wird. Dazu kann beispielsweise die L1-Norm ( $p = 1$ ) oder die L2-Norm ( $p = 2$ ) genutzt werden:

$$\min_G \mathcal{L}_{\text{Diff}}(G) = \mathbb{E}_{x,y,z} [\|y - G(z, x)\|_p]. \quad (\text{A.5})$$

Das als Differenz-Loss bezeichnete Optimierungskriterium führt aber im Vergleich zum Adversarial-Loss (Gleichung A.4) des Diskriminators zu weniger scharfen Bildern [204]. Allerdings erwies sich die Kombination von Adversarial-Loss und Differenz-Loss als vorteilhaft [101]. Das pix2pix-Modell verwendet daher eine Kombination aus beiden Loss-Typen, wobei der Einfluss des Differenz-Loss über einen Parameter  $\lambda$  gesteuert werden kann. Da die L2-Norm im Vergleich zur L1-Norm, zu etwas unschärferen Bildern führt [191, 205], wird die L1-Norm als Differenz-Loss verwendet. Insgesamt ergibt sich der bei pix2pix verwendete Loss zu:

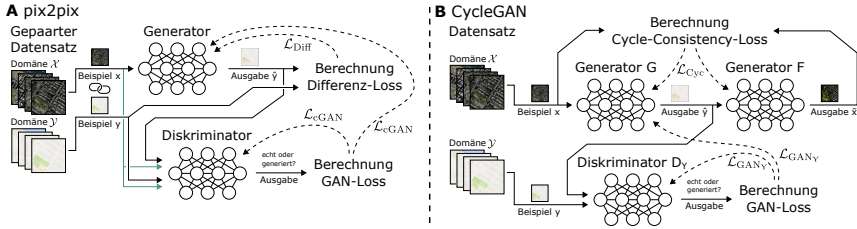
$$\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda_{L1} * \mathcal{L}_{\text{Diff}_{L1}}(G), \text{ mit} \quad (\text{A.6})$$

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y} [\log D(y, x)] + \quad (\text{A.7})$$

$$\mathbb{E}_{x,z} [\log(1 - D(x, G(z, x)))] \text{ und} \quad (\text{A.8})$$

$$\mathcal{L}_{\text{Diff}_{L1}}(G) = \mathbb{E}_{x,y,z} [\|y - G(x, z)\|_1]. \quad (\text{A.9})$$

Mit dem pix2pix-Modell wurde auch eine neue Architektur für den Diskriminator vorgestellt. Da durch den Differenz-Loss  $\mathcal{L}_{L1}$  bereits die Übereinstimmung in



**Abbildung A.2:** Übersicht der Netzwerkstrukturen für die Bild-zu-Bild-Transformation. **A** zeigt die Architektur von pix2pix, die auf der Struktur eines cGANs basiert. Im Unterschied zu einem klassischen cGAN verwendet pix2pix Bilder als Konditionen und kombiniert den GAN-Loss mit einem zusätzlichen Differenz-Loss. Dieser Differenz-Loss wird durch den Vergleich zwischen dem Zielbild und dem vom Generator erzeugten Ausgabe berechnet. **B** zeigt die Struktur des CycleGANs, das im Gegensatz zu pix2pix auch mit ungepaarten Bildern trainiert werden kann. Dies wird durch den sogenannten Cycle-Consistency-Loss ermöglicht. Dabei wird sichergestellt, dass ein Bild nach einer Hin- und Rücktransformation wieder dem ursprünglichen Bild entspricht. Die dargestellte Struktur zeigt lediglich die Transformation von der Domäne  $\mathcal{X}$  nach Domäne  $\mathcal{Y}$ . Die entgegengesetzte Richtung (von Domäne  $\mathcal{Y}$  nach Domäne  $\mathcal{X}$ ) erfolgt analog.

den niedrigen Bildfrequenzen<sup>4</sup> erreicht wird, konzentriert sich der als PatchGAN bezeichnete Diskriminator auf die hohen Frequenzanteile [191]. Im Gegensatz zur klassischen Architektur eines Diskriminators, der das gesamte Bild auf einen einzigen Skalarwert reduziert, bewertet das PatchGAN einzelne, sich überlappende Bildausschnitte (Patches) separat. Dies wird durch den Einsatz eines reinen Convolutional-Networks erreicht, das für jeden Patch eine Einschätzung über dessen Echtheit ausgibt.

Das pix2pix-Modell kann genutzt werden, um beispielsweise auf Basis von Skizzen (Kondition) vollständige Bilder zu generieren. Aufgrund der Konditionierung werden für das Training allerdings gepaarte Bilder benötigt. Das bedeutet, dass für jedes Bild der Domäne  $\mathcal{Y}$  ein dazu passendes Bild der Domäne  $\mathcal{X}$  benötigt wird. Für eine Transformation von beispielsweise Nacht- zu Tagaufnahmen wird so für jedes Motiv eine identische Aufnahme bei Tag als auch bei Nacht benötigt. Dies stellt eine erhebliche Einschränkung dar, da die Aufnahme solcher gepaarten

<sup>4</sup> Die niedrigen Bildfrequenzen stellen die groben Strukturen innerhalb der Bilder dar, während die hohen Bildfrequenzen die Feinheiten wie klare Kanten darstellen.

Daten in der Praxis äußerst aufwendig ist oder in manchen Fällen sogar unmöglich sein kann. Damit eignet sich die Methode nicht für die Generierung von annotierten Trainingsdaten, da für die realen Bilder die Annotationen nicht ohne händisches Annotieren verfügbar sind.

Mit dem 2017 eingeführten CycleGAN [181] (siehe Abbildung A.2b) wird die Voraussetzung der gepaarten Trainingsdaten umgangen. Während beim pix2pix-Modell die Übereinstimmung der generierten Bilder mit den Konditionierungen anhand des Diskriminators und des Reconstruction Losses sichergestellt wird, steht diese Möglichkeit bei ungepaarten Daten aufgrund der für die realen Daten fehlenden Konditionen nicht mehr zur Verfügung. Beim CycleGAN wird das Problem durch die Einforderung der Cycle-Consistency gelöst. Einfach ausgedrückt beschreibt die Cycle-Consistency, dass ein Bild  $x$  aus Domäne  $\mathcal{X}$ , welches in die Domäne  $\mathcal{Y}$  und wieder zurück in Domäne  $\mathcal{X}$  übersetzt wird, zumindest näherungsweise mit dem Ausgangsbild  $x$  übereinstimmen soll. Damit wird gewährleistet, dass durch die Transformation Informationen bezüglich des originalen Bildes erhalten bleiben, die während der Rücktransformation zur Rekonstruktion des Ausgangsbilds genutzt werden können. Auf diese Weise wird auch ohne direkte Konditionierung sichergestellt, dass die Inhalte der generierten Bilder mit denen der Ausgangsbilder übereinstimmen. Durch die Möglichkeit, ungepaarte Daten mittels CycleGAN zu nutzen, eröffnet sich ein breites Anwendungsfeld. Ein bekanntes Beispiel hierfür ist die Transformation von Landschaftsaufnahmen in Gemälde berühmter Künstler. Da diese Künstler bereits verstorben sind und keine direkten Vorlagen existieren, kann eine solche Transformation nur anhand ungepaarter Daten erlernt werden. Eine im Sinne der Trainingsdatensynthese viel wichtigere Anwendungsfeld stellt die Möglichkeit dar, anhand von realen Aufnahmen und unabhängig synthetisch generierten Annotationen realistische synthetische Bilder zu generieren, bei denen die Annotationen bekannt sind.

Im Detail werden beim CycleGAN zur Ausnutzung der Cycle-Consistency zwei Generatoren  $G, F$  und zwei Diskriminatoren  $D_X, D_Y$  trainiert. Ein Bild  $x$  der Domäne  $\mathcal{X}$  wird von dem Generator  $G$  in die Domäne  $\mathcal{Y}$  überführt ( $\hat{y}$ ). Anschließend erfolgt durch den zweiten Generator  $F$  eine Rücktransformation des



generierten Bildes  $\hat{y}$  in die Ausgangsdomäne  $\mathcal{X}$ , woraus sich das rekonstruierte Bild  $\tilde{x}$  ergibt. Dieser Ablauf erfolgt äquivalent auch in die entgegengesetzte Richtung. Die Erwartung ist, dass nach der Hin- und Rücktransformation ein Bild entsteht, das dem Ausgangsbild nahezu identisch ist. Um dies zu erreichen, wird der Cycle-Consistency-Loss verwendet, der die Differenzen zwischen dem Originalbild  $x$  bzw.  $y$  und dem rekonstruierten Bild  $\tilde{x}$  bzw.  $\tilde{y}$  bestraft:

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_x [\|F(G(x)) - x\|_1] + \mathbb{E}_y [\|G(F(y)) - y\|_1]. \quad (\text{A.10})$$

Die Aufgabe der beiden Diskriminatoren besteht weiterhin darin, zwischen realen und generierten Bildern zu unterscheiden, und damit sicherzustellen, dass die generierten Bilder auch zu der entsprechenden Domäne passen. Der Diskriminator  $D_Y$  bewertet die Echtheit der von dem Generator  $G$  generierten Bilder, während der Diskriminator  $D_X$  die Echtheit der vom Generator  $F$  erzeugten Bilder bewertet. Im Unterschied zum pix2pix-Modell erhalten die beiden Diskriminatoren keine zusätzliche Kondition. Der GAN Loss ergibt sich zu:

$$\mathcal{L}_{GAN_Y}(G, D_Y) = \mathbb{E}_y [\log D_Y(y)] + \mathbb{E}_x [\log(1 - D_Y(G(x)))] \quad (\text{A.11})$$

bzw.

$$\mathcal{L}_{GAN_X}(F, D_X) = \mathbb{E}_x [\log D_X(x)] + \mathbb{E}_y [\log(1 - D_X(F(y)))] \quad (\text{A.12})$$

Zusammengefasst ergibt sich für das Training des CycleGANs folgendes Optimierungsproblem:

$$\min_{G, F} \max_{D_X, D_Y} \mathcal{L}_{GAN_Y}(G, D_Y) + \mathcal{L}_{GAN_X}(F, D_X) + \lambda_{cyc} \mathcal{L}_{cyc}(G, F). \quad (\text{A.13})$$

Mit  $\lambda_{cyc}$  kann der Einfluss des Cycle-Consistency-Loss eingestellt werden.

Da es bei RGB-Bildern teilweise zu Unstimmigkeiten in der Farbkomposition zwischen Eingangs- und Ausgangsbild gekommen ist, führen Zhu und Park et al. [181] einen weiteren Loss ein. Die Idee ist, dass der Generator in einem zusätzlichen Schritt bereits ein Bild der Zieldomäne als Eingang erhält. Der Generator soll das Bild nicht verändern, da es bereits der Zieldomäne entspricht. Das als Identity-Loss bezeichnete Kriterium ergibt sich zu:

$$\mathcal{L}_{identity}(G, F) = \mathbb{E}_y [\|G(y) - y\|_1] + \mathbb{E}_x [\|F(x) - x\|_1]. \quad (\text{A.14})$$

Unabhängig vom CycleGAN präsentierte Kim et al. [206] im gleichen Zeitraum eine fast identische Methode für die Bild-zu-Bild-Transformation anhand von ungepaarten Daten. Unterschiede lassen sich nur in den Diskriminatoren finden. Allerdings erlangte das CycleGAN über die Zeit eine deutlich größere Reichweite.

Im selben Jahr stellen auch Liu et al. [207] einen weiteren Ansatz für die Bild-zu-Bild-Transformation anhand von ungepaarten Daten vor. Das UNIT-Framework basiert auf der Annahme, dass Bilder der Domäne  $\mathcal{X}$  und  $\mathcal{Y}$  auf einen gemeinsamen latenten Raum abgebildet werden können. Für ein Bilderpaar  $x$  und  $y$  existiert also ein gemeinsamer Punkt  $z$  in diesem latenten Raum. Über Generatoren  $G_x$  und  $G_y$  können aus dem Punkt  $z$  die Bilder  $x$  bzw.  $y$  rekonstruiert werden:  $x = G_x(z)$  bzw.  $y = G_y(z)$ . Über Encoder  $E_x$  und  $E_y$  können die Bilder der beiden Domänen wiederum auf den gemeinsamen Punkt  $z$  im latenten Raum abgebildet werden:  $z = E_x(x)$  bzw.  $z = E_y(y)$ . Eine Transformation eines Bildes der Domäne  $\mathcal{X}$  in die Domäne  $\mathcal{Y}$  lässt sich über eine Funktion beschreiben:  $F_{x \rightarrow y}(x) = G_y(E_x(x))$ . Äquivalent existiert die Funktion, welche die Transformation in die entgegengesetzte Richtung beschreibt:  $F_{y \rightarrow x}(y) = G_x(E_y(y))$ .

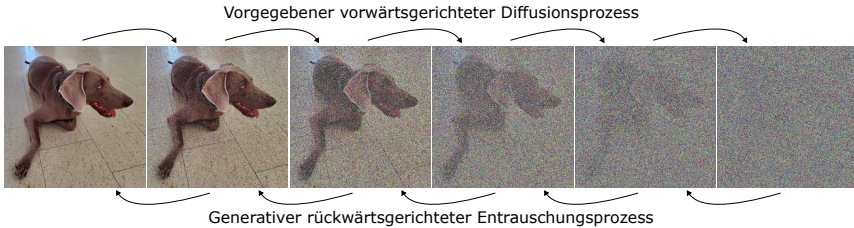
In der Realisierung des UNIT werden für die Funktionen  $F_{x \rightarrow y}(x)$  und  $F_{y \rightarrow x}(y)$  Variational Autoencoder (VAE) [208] verwendet. Für die Realisierung des geteilten latenten Raumes wird eine Zwischenrepräsentation  $h$  eingeführt, womit ein Punkt  $z$  erst in  $h$  und anschließend in das Bild  $x$  bzw.  $y$  überführt wird. Während  $z$  als eine grobe Beschreibung des Bildes gesehen werden kann, stellt  $h$  eine erste

Realisierung dieser Beschreibung dar. Beispielhaft sei  $z$  eine Beschreibung eines Bauteils, wobei  $h$  die technische Zeichnung und  $x$  bzw.  $y$  die Realisierung des Bauteils aus zwei verschiedenen Werkstoffen darstellt. Die Generatoren stellen nun eine Komposition aus  $G_{L,1x}$  bzw.  $G_{L,y}$  und  $G_H$  dar:  $G_x = G_{L,x} \circ G_H$  und  $G_y = G_{L,y} \circ G_H$ . Äquivalent gilt für die Encoder:  $E_x = E_H \circ E_{L,x}$  und  $E_y = E_H \circ E_{L,y}$ . In der Praxis wird der geteilte latente Raum  $z$  und die Zwischenrepräsentation  $h$  über das Teilen von Gewichten der letzten Schichten der beiden Encoder bzw. die Gewichts-Teilung der ersten Schichten der Generatoren forciert. Aufgrund der Komplexität des Ansatzes wird nicht auf die zum Training verwendeten Loss-Funktionen eingegangen und stattdessen auf den originalen Artikel verwiesen [207].

Die beschriebenen Modelle stellen wichtige Meilensteine und verbreitete Ansätze für die Bildsynthese dar. Daneben existieren andere GAN-basierte Methoden bzw. auf spezielle Anwendungsfälle optimierte Modelle. Ein genauerer Überblick für GANs im Allgemeinen und über GANs und deren Verwendung im medizinischen und biologischen Kontext kann in [209, 210] bzw. [211] gefunden werden.

## A.1.2 Denoising-Diffusion-Probabilistic-Models (DDPM)

Neben GAN-basierten Ansätzen zeigen auch Modelle auf Basis von Denoising-Diffusion-Probabilistic-Models (DDPMs) [212] Erfolg bei der Synthese von Bild-daten [213, 214, 215]. Bei DDPMs werden qualitativ hochwertige Bilder durch die iterative Verbesserung von verrauschten Bildern generiert. Der Trainingsprozess wird in einen *vorgegebenen vorwärtsgerichteten Diffusionsprozess* und einen *generativen rückwärtsgerichteten Entrauschungsprozess* aufgeteilt (vgl. Abbildung A.3). Beim vorwärtsgerichteten Prozess wird den Daten über eine Reihe von Zeitschritten Gaußsches Rauschen hinzugefügt, wodurch die ursprüngliche Datenverteilung in eine einfache Gaußsche Verteilung umgewandelt wird. Mathematisch gesehen wird bei jedem Zeitschritt  $t$  Rauschen zu den Daten hinzugefügt, was zu zunehmend verrauschten Versionen der Daten führt.



**Abbildung A.3:** Funktionsweise von Denoising Diffusion Probabilistic Models (DDPMs). DDPMs erzeugen Bilder, indem sie verrauschte Bilder iterativ verbessern. Im vorgegebenen vorwärtsgerichteten Diffusionsprozess wird schrittweise Gaußsches Rauschen hinzugefügt, im generativen rückwärtsgerichteten Entrauschungsprozess entfernt ein neuronales Netzwerk dieses Rauschen schrittweise, um die ursprünglichen Daten wiederherzustellen. Für die Generierung von neuen Daten wird nach dem Training ein zufällig erstelltes Rauschbild schrittweise entrauscht.

Der rückwärtsgerichtete Prozess zielt darauf ab, die verrauschten Daten ausgehend von einem vollständig verrauschten Bild schrittweise zu entrauschen. Das wird anhand eines neuronalen Netzwerks erreicht, welches darauf trainiert wird, das bei jedem Zeitschritt hinzugefügte Rauschen vorherzusagen. Für eine detaillierte mathematische Beschreibung des Prozesses wird auf [216] verwiesen.

Die meist komplexeren DDPMs erfordern allerdings im Vergleich zu GANs längere Trainings- und Inferenzzeiten, benötigen mehr Rechenressourcen und größere Datensätze, da sie bei kleinen Trainingsmengen häufig zum Auswendiglernen tendieren [217, 218]. In einem solchen Fall werden keine neuen Bilder von den Modellen erzeugt, sondern nur die realen Trainingsbilder, nahezu identisch wiedergegeben. In dieser Arbeit wird sich auf die Verwendung von GANs konzentriert. Daher wird auf eine detaillierte Beschreibung verzichtet und stattdessen auf ein Survey von Yang et al. [216] verwiesen, wo ein Überblick über Diffusion Models im Allgemeinen und Möglichkeiten zur Konditionierung gegeben wird. Dennoch wird im späteren Verlauf der Arbeit auf die Übertragbarkeit der eigens entwickelten Methoden auf DDPMs eingegangen.

## A.2 Einführung in die Zellzyklusstadien

Der Zellzyklus einer Zelle lässt sich in zwei Hauptphasen unterteilen: die Interphase, in der die Zelle unter anderem durch Verdopplung der DNA auf die Teilung vorbereitet wird, und die Mitose, in der die eigentliche Zellteilung erfolgt. Die Interphase gliedert sich weiter in die Abschnitte G1, S und G2, während die Mitose in Prophase, Metaphase, Anaphase, Telophase und Zytokinese unterteilt ist. Zellen, die sich nicht im Teilungsstadium befinden, werden der G0-Phase zugeordnet. Weitere Details hierzu finden sich in [219].

## A.3 Beschreibung des Modelltrainings und der Anwendung

Der Code der Modelle ist wie in Abschnitt 3.2 beschrieben öffentlich zugänglich. In einer zugehörigen Readme-Datei finden sich Anweisungen, wie die Analyse reproduziert werden kann. Zusätzlich sind sämtliche Daten, einschließlich der Rohbilddateien, Zwischen- und Endergebnisse sowie der Parametereinstellungen, auf Zenodo verfügbar<sup>5</sup>. Im Folgenden wird eine detaillierte Beschreibung des Trainings der StarDist Modelle sowie des Erstellungsprozesses der synthetischen Daten gegeben.

### A.3.1 StarDist-Modelle

Die vorgestellten Methoden zur Datensynthese werden anhand der Segmentierungsleistung eines 3D-StarDist-Modells bewertet. Für das Training und die Inferenz werden die bereitgestellten Beispielskripte aus dem StarDist-Repository<sup>6</sup> verwendet. In dem Trainingsskript erfolgt eine Normalisierung der Bilder und

---

<sup>5</sup> <https://zenodo.org/doi/10.5281/zenodo.11240361>

<sup>6</sup> <https://github.com/stardist/stardist>

eine Nachbearbeitung, welche kleine Löcher in den Annotationen schließt. Anschließend erfolgt eine zufällige Aufteilung der Trainingsdaten in einen Trainings- und Validierungsdatensatz im Verhältnis von 0,85 zu 0,15, wobei mindestens ein Bild im Validierungsdatensatz enthalten ist. Wichtige Parameter sind wie folgt festgelegt: Der Wert für  $n\_rays$  wird auf den Standardwert von 96 gesetzt, und wenn die Anisotropie größer als 1,5 ist, wird ein unterabgetastetes Raster verwendet. Die Patchgröße für das Training beträgt  $96 \times 256 \times 256 \text{ px}^3 (z, y, x)$ , und die Batchgröße ist auf zwei festgelegt. In Fällen, in denen manuell annotierte Patches für das Training verwendet werden, wird die Patchgröße auf  $32 \times 128 \times 128 \text{ px}^3 (z, y, x)$  reduziert, um der Größe der Ground-Truth-Patches zu entsprechen. Als Augmentierungsmethoden werden die Funktionen aus dem Beispielskript für Spiegelungen, Rotationen und Intensitätsänderungen genutzt. Die Seedings der Zufallsgeneratoren werden aus dem Skript entfernt, um statistische Analysen zu ermöglichen. Alle Modelle werden für 100 Epochen trainiert. Nach dem Training werden die Schwellenwerte (*probability* und *non-maximum suppression*) basierend auf den Validierungsdaten optimiert. Für die Inferenz werden die Gewichte der Epoche mit dem besten Validierungs-Loss ausgewählt, und die Bilddaten werden wie im Beispielskript vor der Prozessierung normalisiert.

### A.3.2 SimOptiGAN

Mit SimOptiGAN werden insgesamt fünf Bilder generiert. Die zugehörigen Parameterdateien sind im öffentlich zugänglichen Datensatz verfügbar<sup>5</sup>. Die Implementierung des Optimierungsmodells basiert auf der offiziellen Implementierung von CycleGAN. Das Optimierungsmodell wird mit fünf simulierten Bildern und einem großen realen Bild trainiert. Während des Trainings werden zufällig Patches der Größe  $32 \times 256 \times 256 \text{ px}^3 (z, y, x)$  aus den Trainingsbildern extrahiert. Das Modell wird 3000 Epochen mit einer konstanten Lernrate von  $2 \cdot 10^{-4}$  trainiert, gefolgt von weiteren 3000 Epochen mit einer linear abnehmenden Lernrate. Die Datensatzgröße wird virtuell verdoppelt, was zu einer Verdoppelung der verarbeiteten Proben pro Epoche führt, um Leistungsprobleme im Zusammenhang mit dem Tracking der Modellleistung zu beheben. Nach dem Training wird die

Epoche ausgewählt, die visuell die besten Ergebnisse liefert (Epoche 2994). Die Inferenz auf den simulierten Trainingsdaten erfolgt mittels eines Patch-basierten Ansatzes, wobei dieselbe Patchgröße wie im Training und eine Überlappung von 50% verwendet wurde. Vor dem Zusammensetzen der Bilder werden überlappende Bereiche von den prozessierten Patches entfernt. Das anschließende Training des Segmentierungsmodells wird mit vier optimierten Bildern durchgeführt.

### A.3.3 SimOptiGAN+

SimOptiGAN+ nutzt die Ergebnisse der biophysikalischen Simulation in Form der BioSim-Instanzen. Für jedes der vier BioSim-Instanz-Bilder wird mithilfe von SimOptiGAN+ ein synthetisches Zellkernbild erzeugt. Die zugehörigen Parameterdateien sind im Daten-Repository<sup>5</sup> verfügbar. Für die nachfolgende Optimierung wird das zuvor trainierte Optimierungsmodell von SimOptiGAN genutzt. Das Training des Segmentierungsmodells erfolgt wiederum mit vier optimierten Bildern.

### A.3.4 Mem2NucGAN-P

Die Methode Mem2NucGAN-P erfordert für das Training des cGAN-basierten Transformationsmodells segmentierte reale Membranbilder. Die Segmentierung dieser Bilder wird mit dem vortrainierten Cellpose *cyto2*-Modell durchgeführt, wobei die Parameter *diameter* und *stitch\_threshold* auf 22 bzw. 0,5 gesetzt werden. Alle weiteren Parameter blieben unverändert. Darüber hinaus benötigt Mem2NucGAN-P für das Training ungepaarte binäre Segmentierungsmasken von Zellkernen. Genutzt wird dafür ein zuvor eingeführtes Segmentierungsmodell [176], welches auf synthetischen Daten trainiert wird. Das Training des cGAN-Transformationsmodells erfolgt mit vier realen Nukleusaufnahmen, dazugehörigen binären Membransegmentierungen und vier ungepaarte binäre Zellkernsegmentierungen. Die Implementierung des cGAN basiert auf der pix2pix-Architektur, nutzt allerdings ein ResNet als Generatormodell. Das cGAN wird

über 1400 Epochen mit einer konstanten Lernrate von  $2 \cdot 10^{-4}$  trainiert, gefolgt von weiteren 1400 Epochen mit einer linear abnehmenden Lernrate. In diesem Fall wird die letzte Epoche (Epoche 2800) für weitere Berechnungen ausgewählt. Ähnlich wie beim Optimierungsmodell von SimOptiGAN wird die Datensatzgröße um das Vierfache virtuell vergrößert, was zu einer ähnlichen Anzahl an verarbeiteten Samples pro Training führt (11200 für Mem2NucGAN-P im Vergleich zu 12000 für SimOptiGAN). Die Implementierung nutzt denselben Patch-basierten Trainings- und Inferenzansatz wie das SimOptiGAN-Optimierungsmodell. Mit vier BioSim-Membranen generiert das trainierte cGAN vier synthetische Bilder von Zellkernsignalen inklusive den entsprechenden Annotationen. In einem nachfolgenden Verarbeitungsschritt werden die Netzwerkausgaben in Instanzannotationen umgewandelt. Das anschließende Training des Segmentierungsmodells wird mit diesen vier generierten Bildern und den entsprechenden Instanzannotationen durchgeführt.

### A.3.5 Mem2NucGAN-U

Mem2NucGAN-U verwendet ein CycleGAN-Modell zur Datengenerierung, was das direkte Training mit BioSim-Membranen in Kombination mit realen Zellkernbildern und ungepaarten binären Zellkernsegmentierungen als Trainingsdaten erlaubt. Für das Training des CycleGANs werden vier BioSim-Membranen, vier reale Zellkernbilder und vier binäre Zellkernsegmentierungen verwendet. Das Modell wird 1400 Epochen mit einer konstanten Lernrate von  $2 \cdot 10^{-4}$  trainiert, gefolgt von weiteren 1400 Epochen mit einer linear abnehmenden Lernrate, wobei die letzte Epoche (2800) für die Inferenz ausgewählt wird. Wie bei den anderen GAN-basierten Modellen wird die Datensatzgröße virtuell um das Vierfache erhöht, um sicherzustellen, dass während des Trainings eine ähnliche Anzahl von Proben verarbeitet wird. Der gleiche patchbasierte Ansatz, wie bei der Optimierung von SimOptiGAN, wird auch hier verwendet. Mit denselben vier BioSim-Membranen, die auch während des Trainings verwendet werden, generiert das trainierte CycleGAN vier synthetische Bilder, die synthetische Zellkernsignale und die entsprechenden synthetischen Annotationen enthalten. Ein



nachfolgender Verarbeitungsschritt wird durchgeführt, um die generierten Annotationen in Instanzannotationen umzuwandeln. Das anschließende Training des Segmentierungsmodells erfolgt mit den vier generierten Zellkernbildern und den entsprechenden Instanzannotationen.

## **A.4 Ablation Study bezüglich des zusätzlichen Segmentierungs-Diskriminators von DualGenGAN**

Mit DualGenGAN wird in dieser Arbeit eine neue GAN-Trainingsstruktur eingeführt, welche es erlaubt, neben den Bilddaten auch zugehörige Annotationsdaten zu generieren (siehe Abschnitt 2.2.3). Um die Wichtigkeit des eingeführten Diskriminators zu untersuchen, der für die Sicherstellung der Generierung realistischer Annotationen verantwortlich ist, wird eine Ablationsstudie für die CycleGAN-basierte Mem2NucGAN-U-Methode durchgeführt. Es ist hervorzuheben, dass dieser zusätzliche Diskriminator die einzige Komponente ist, die für realistisch aussehende Annotationssignale und einen hohen Kontrast zwischen Vorder- und Hintergrundsignalen sorgt. Das Weglassen dieses Diskriminators führt zu generierten Annotationen, die dem Eingangssignal zwar ähnlich sind, jedoch einen deutlich reduzierten Kontrast aufweisen. Zudem verschlechtert sich die Stabilität des Trainingsprozesses merklich. Eine mögliche Erklärung dafür liegt im kleineren Loss-Term des Generatormodells. Da, abgesehen von der finalen Faltungsschicht, der Pfad beider Ausgangskanäle identisch ist, beeinflusst der Loss-Term des zusätzlichen Diskriminators einen Großteil des Netzwerks und seiner Parameter. Weiterhin sei zu erwähnen, dass der Identity-Loss nicht genutzt werden kann, um einen hohen Kontrast der Annotationssignale zu erzielen, da die Anzahl der Eingabe- und Ausgabekanäle zwischen den beiden Generatoren unterschiedlich ist.

Vor der Einführung des zusätzlichen Diskriminators wurde ein weiterer Ansatz entwickelt und getestet. Bei diesem wurde ein Differenz-Loss (L1) zwischen

dem generierten Bild (konstant) und den erzeugten Annotationssignalen genutzt, um eine Übereinstimmung zwischen beiden Signalen zu erzwingen. Dieser Loss wurde zusätzlich mit einem weiteren Loss kombiniert, welcher ein Ausgangssignal durchsetzen soll, bei dem die Werte entweder null oder eins sind:

$$\mathcal{L}_{\text{Seg}}(G_{A2B}) = \mathbb{E}_{xy \sim p_{\text{data}}(x)} [\|G(x)_{\text{img}} - G(x)_{\text{Seg}}\|_1] + \mathbb{E}_{x \sim p_{\text{data}}(x)} [\| \text{bin}(G(x)_{\text{Seg}}) - G(x)_{\text{Seg}} \|_2]. \quad (\text{A.15})$$

Diese Trainingsstruktur erwies sich jedoch als weniger stabil und lieferte im Vergleich zum zusätzlichen Diskriminator schlechtere Ergebnisse.

## A.5 Statistische Analyse der Segmentierung

Für den Vergleich der durchschnittlichen Leistung der getesteten Segmentierungsmodelle in Abschnitt 5.2.4 wird eine statistische Analyse durchgeführt. Jedes StarDist-Segmentierungsmodell wird sechsmal trainiert ( $n = 6$ ) und die Leistung anhand der durchschnittlichen Ergebnisse von drei Ground-Truth-Patches bewertet. Aufgrund der geringen Stichprobengröße ( $n = 6$ ) wird ein einseitiger Welch's t-Test angewendet, da dieser keine gleichen Varianzen oder eine Normalverteilung voraussetzt und somit konservativer ist. Vordefinierte Signifikanzniveaus sind wie folgt festgelegt: \*  $p < 0.05$ , \*\*  $p < 0.01$ , and \*\*\*  $p < 0.001$ .

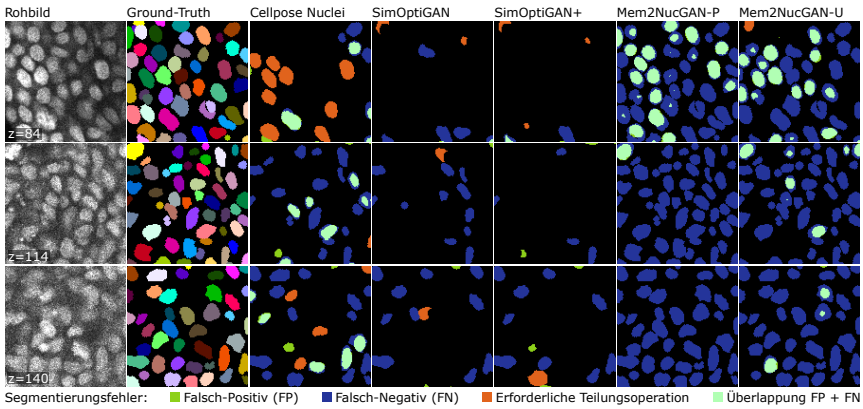
## A.6 Grafikspeicher bei Laufzeitexperiment zur Synthese von Bildmodalitäten

Bei dem Vergleich des Grafikkartenspeicherbedarfs beim Training (vgl. Tabelle 5.2) zeigt sich, dass beide Modelle ähnlich viel Speicher benötigten. Trotz des zusätzlichen Diskriminatornetzwerkes des pix2pix-Modells liegt dessen Speicherbedarf mit 16,8 GB etwas unter dem des ResUnet-Modells (17,2 GB). Ein Grund hierfür ist die verwendete Batch-Größe während des Trainings. Wird diese wie

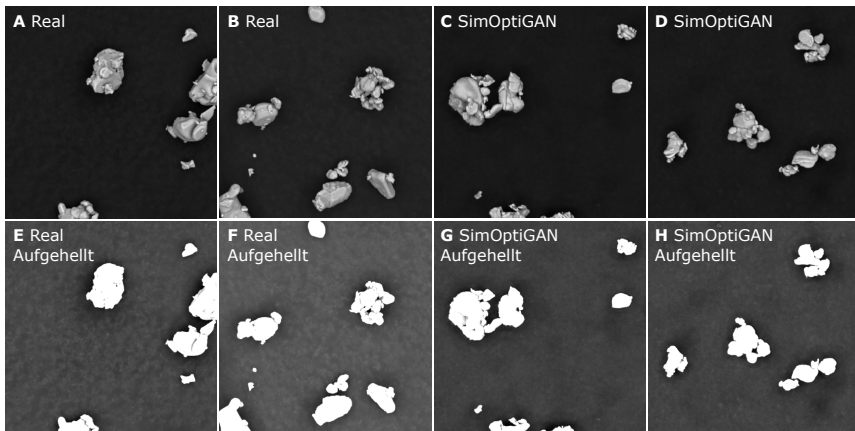
beim pix2pix-Modell von zwei auf eins gesetzt, reduziert sich der benötigte Grafikartenspeicher auf 11,6 GB.

## A.7 Klassifikationsnetzwerk Zell-Zyklus-Klassifikation

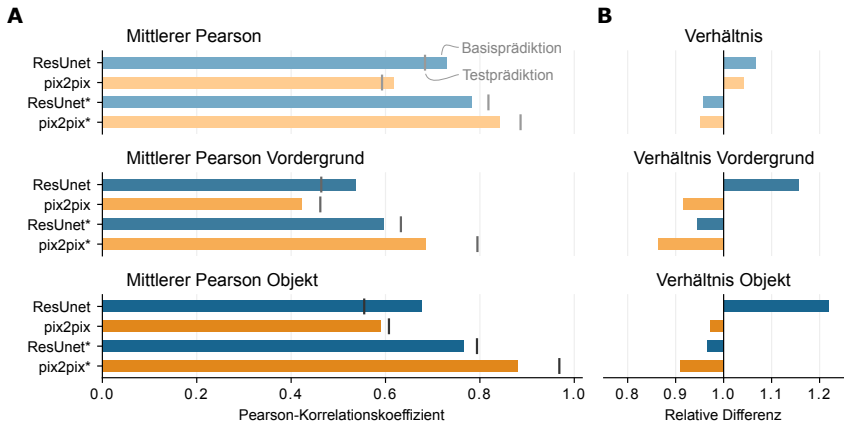
Das für die Klassifikation verwendete Netzwerk nutzt einen Convolutional Encoder, gefolgt von drei Fully-Connected Schichten. Die Aufgabe des Netzwerkes ist es, anhand von  $48 \times 64 \times 64 \text{ px}^3$  großen Bildausschnitten mit zwei oder drei Kanälen, welche auf den Mittelpunkt der zu klassifizierenden Zelle zentriert sind, die Klasse vorherzusagen. Es wurden auch andere Ausschnittgrößen und auch die Skalierung der Bounding Boxen der einzelnen Zellen auf eine einheitliche Größe untersucht. Allerdings konnten damit keine besseren Ergebnisse erzielt werden. Die Struktur des Encoders ist von der des U-Nets abgeleitet, bei der Convolution-Blocks bestehend aus einer 3D-Convolution  $3 \times 3 \times 3$  gefolgt von einer ReLU Aktivierungsfunktion, Batch-Normalisierung und einer  $2 \times 2$  Max-Pooling Operation mit einem Stride von zwei. Beginnend bei 32 verdoppeln sich jeweils die Filterschichten der insgesamt vier Convolution-Blocks bis zu einer Anzahl von 256. Darauf folgen zwei Fully-Connected Schichten mit 120 und 84 Neuronen und ReLU Aktivierungsfunktion. Eine letzte Fully-Connected Ausgangsschicht besitzt fünf Neuronen (eins pro Klasse) mit einer Softmax-Aktivierungsfunktion.



**Abbildung A.4:** Qualitativer Vergleich der Segmentierungsergebnisse der einzelnen Modelle anhand der visualisierten Fehler. Zur besseren Visualisierung wird nur ein repräsentativer optischer Schnitt der drei Ground-Truth-Ausschnitten gezeigt. Die erste und zweite Spalte zeigen das Rohbildsignal und die entsprechende Ground-Truth, während die angrenzenden Spalten die bei der Segmentierung gemachten Fehler der Modelle einschließlich falsch-negativer, falsch-positiver und erforderlicher Teilungsoperationen zeigen. Die originalen Segmentierungsergebnisse finden sich in Abbildung 5.5. Übernommen und übersetzt aus [177].



**Abbildung A.5:** Weitere echte und mittels SimOptiGAN generierte REM-Aufnahmen. A-D zeigen die Aufnahmen mit unveränderter Intensität und E-H gibt eine Darstellung mit erhöhter Helligkeit zur verbesserten Visualisierung des Hintergrundes. A bzw. E und B bzw. F stellen repräsentative reale Aufnahmen dar, während C bzw. G und D bzw. H mit SimOptiGAN generierte synthetische Aufnahmen zeigt.



**Abbildung A.6:** **A** Pearson-Korrelation zwischen den Intensitäten des realen registrierten Ki-67-Signals aus der Testaufnahme und denen des aus den Nukleussignalen aus der Basisaufnahme generierten synthetischen Ki-67-Signalen. Die Korrelation wird zwischen allen Voxel-Intensitäten (Mittlerer Pearson), allen Vordergrund Voxel-Intensitäten (Mittlerer Pearson Vordergrund) und allen Nukleus-Intensitäten (Mittlerer Pearson Objekt) bestimmt. Die Vordergrundbereiche und die Objektinstanzen werden dabei über eine vorherige Segmentierung bestimmt. Verglichen werden die Ergebnisse des ResUnet mit denen des pix2pix basierten Ansatzes. Die mit Sternchen markierten Modelle sind speziell auf der Testaufnahme trainiert worden. Die vertikalen Striche zeigen als Vergleich die Korrelation der aus den Nukleussignalen der Testaufnahme generierten Ki-67-Signale. **B** Verhältnis der Korrelationen zwischen beiden, auf Basis von der Basis- bzw. Testaufnahme, generierten Ki-67-Signalen.

# Abkürzungen und Symbole

## Abkürzungen und Begriffe

3D-ClassAnnotator	In dieser Arbeit eingeführte grafische Benutzerschnittstelle für die effiziente Klassenannotation von Objekten in 3D-Daten.
Alternative Färbung	Entwickelte Methode zur Analyse des Einflusses von Bleed-Trough-Effekten.
AU	AU stellt eine willkürliche Einheit dar (engl.: arbitrary unit).
Basisaufnahme	Erste Aufnahme der zweistufigen Färbung ohne den Zielmarker.
Basisprädiktion	Synthetischer Marker, welcher anhand der Basisaufnahme generiert wurde.
BioSim-Instanzen	Instanzannotationen, welche mittels biophysikalischer Simulation generiert wurden
BioSim-Membranen	Aus BioSim-Instanzen abgeleitete binäre Membransignale
Bleed-Trough	Das Durchscheinen von fluoreszierenden Signalen in einen benachbarten Farbkanal eines Mikroskopes.
Bounding Box	Ein rechteckiger Bereich, welcher ein Objekt umschließt.
Caspase-3	Fluoreszenzmarker für Apoptose

Cellpose	Deep-Learning-basierter Algorithmus für die Segmentierung von Zelldaten.
Cellpose Nuclei	Öffentliches, vortrainiertes Modell von Cellpose zur Segmentierung von Nuklei.
cGAN	conditional Generative Adversarial Network
Cross Talk	Das Durchscheinen von fluoreszierenden Signalen in einen benachbarten Farbkanal eines Mikroskopes.
CT	Computertomografie
CycleGAN	GAN Struktur zur Bild-zu-Bild-Transformation anhand ungepaarter Daten.
DAPI	Fluoreszenzmarker für die Visualisierung von Zellkernen.
DCGAN	Deep Convolutional Generative Adversarial Network
DDPMs	Denoising-Diffusion-Probabilistic-Models
DET	Detektionsmetrik der Cell-Tracking-Challenge
DRAQ5	Fluoreszenzmarker für die Visualisierung von Zellkernen.
DS-Klassifikation	Eingeführter 3D-Datensatz zur Zellzyklus-Klassifikation.
DS-Seltene-Elemente	Entwickelter 3D-Datensatz zur Analyse von seltenen Elementen in Trainingsdaten.
DS-Marker-Syn	Eingeführter 3D-Datensatz zur Synthese von Bildmodalitäten.
DS-Train-Syn	Eingeführter 3D-Datensatz zur Synthese von Trainingsdaten.
DS-BioSim	Eingeführter 3D-Datensatz mit biophysikalisch simulierten Zellgrenzen.

DS-Seg-GT	Eingeführter 3D-Datensatz zur Bewertung der Synthesemethoden auf echten Daten.
DualGenGAN	GAN-Struktur zur dualen Generierung von Bild- und Annotationsdaten.
FID	Fréchet-Inception-Distance
GAN	Generative Adversarial Network
Ground-Truth	Grundwahrheit, meistens in Bezug auf händisch erstellte Annotationen.
GUI	Grafische Benutzerschnittstelle (engl.: graphical user interface)
IMS	Framework zur Synthese von Bildmodalitäten (engl.: image modality synthesis).
IS	Inception-Score
Ki-67	Fluoreszenzmarker für Zellproliferation
KID	Kernel-Inception-Distance
Klassendistanz	In dieser Arbeit eingeführte Metrik zur Bewertung der Klassifikationsleistung.
Kreuzvalidierung	Genutzter Leave-One-Out-Ansatz, um Modelle auf Ground-Truth-Daten zu trainieren und auszuwerten.
Loss	Im Kontext des maschinellen Lernens bezeichnet der Begriff Loss bzw. Verlustfunktion eine numerische Metrik, welche zur Bewertung der Modellleistung herangezogen wird. Sie bildet die Grundlage für die Optimierung des Modells.
Mask R-CNN	Netzwerkstruktur zur Segmentierung von Objekten in Bildern.



Mem2NucGAN-P	In dieser Arbeit eingeführte Methode, um anhand biophysikalisch simulierter Membranstrukturen Nukleusmarker zu generieren. Diese Methode basiert auf GANs, welche mit gepaarten Daten trainiert werden.
Mem2NucGAN-U	In dieser Arbeit eingeführte Methode, um anhand biophysikalisch simulierter Membranstrukturen Nukleusmarker zu generieren. Diese Methode basiert auf GANs, welche mit ungepaarten Daten trainiert werden.
MRT	Magnetresonanztomografie
Multiplexing	Die Nutzung mehrerer fluoreszierender Marker in einer Probe.
Nyquist-Rate	In der konfokalen Mikroskopie bestimmt sie die minimale Abtastrate, um das volle Auflösungsvermögen des Systems zu erfassen.
PCA	Hauptkomponentenanalyse (engl.: principal component analysis)
pix2pix	GAN Struktur zur Bild-zu-Bild-Transformation anhand gepaarter Daten.
PMT	Photoelektronenvervielfacher (engl.: photomultiplier tube)
PSF	Punktspreizfunktion (engl.: point spread function)
REM	Rasterelektronenmikroskop
RMSE	Quadratwurzel des mittleren quadratischen Fehlers (engl.: root mean squared error)
Rohbild	Unverarbeitetes reales Bild
SEG	Segmentierungsmetrik der Cell-Tracking-Challenge

SimOptiGAN	In dieser Arbeit eingeführte Synthesemethode, welche Simulation und Optimierung auf GAN-Basis kombiniert.
SimOptiGAN+	In dieser Arbeit eingeführte Erweiterung von SimOptiGAN, welche Methoden zur biophysikalischen Simulation in den Syntheseprozess integriert.
SpCycleGAN	Spatially Constrained CycleGAN
Spherolyzer	In dieser Arbeit eingeführte grafisch geführte Software für die Analyse von Drug-Screenings.
SSID	Structure Similarity Index
Stardist	Deep-Learning-basierter Algorithmus für die Segmentierung von Zelldaten.
SVM	Support Vector Machine
Testaufnahme	Zweite Aufnahme der zweistufigen Färbung mit dem Zielmarker.
Testprädiktion	Synthetischer Marker, welcher anhand der Testaufnahme generiert wurde.
U-Net	Netzwerkstruktur zur Segmentierung von Objekten in Bildern.
z-Komensation	Erhöhung der Laserintensität eines Fluoreszenzmikroskops in tieferen Schichten, um einer abfallenden Helligkeit entgegenzuwirken.
Zweistufige Färbung	Entwickelte Methode zur Analyse des Einflusses von Bleed-Trough-Effekten.

## Symbole und Variablen

$a$	Gewünschte Ausgabe des Diskriminators bei echten Daten
$b$	Gewünschte Ausgabe des Diskriminators bei unechten Daten
$d_{\text{edt}}$	Euklidische Distanz eines Punktes zum Rand eines Objektes
$k_{\text{norm}}$	Normierungsfaktor für die Distanz $d_{\text{edt}}$
$r$	Euklidische Distanz eines Punktes zum Zentrum eines Objektes
$d_{\text{norm}}$	Normierte Distanz eines Punktes zum Rand eines Objektes
$p_{\text{position}}$	Wahrscheinlichkeit für die Auswahl einer bestimmten Position
$k_{\text{vol}}$	Schwellenwert zur Überprüfung der Überlappung von Zelle und Nukleus
$N_{\text{pos}}$	Maximale Anzahl an Versuchen, ein Objekt an verschiedenen Positionen zu platzieren
$N_{\text{obj}}$	Maximale Anzahl an geprüften Objekten für einen Platzierungsversuch
$n_{\text{aug}}$	Anzahl an Augmentierungen für ein extrahiertes Objekt
$D$	Diskriminator
$D_X$	Diskriminator zur Bewertung der Daten der Domäne X
$D_Y$	Diskriminator zur Bewertung der Daten der Domäne Y
$D_{\text{Seg}}$	Diskriminator zur Bewertung der Annotation
$\mathbb{E}$	Erwartungswert

$F$	Generatormodell
$G$	Generatormodell
$I$	Menge aller Voxel eines Bildes
$I_{VG}$	Menge aller Voxel eines Bildes, welche sich im Vordergrund befinden
$KD$	Neu eingeführte Klassendistanz
$\mathcal{L}$	Lossfunktion
$\mathcal{L}_{BCE}$	Binary-Cross-Entropy-Loss
$\mathcal{L}_{cyc}$	Cycle-Consistency-Loss
$\mathcal{L}_{Diff}$	Differenz-Loss
$\mathcal{L}_{GAN}$	Least-Squares-Loss
$\mathcal{L}_{GAN_S}$	Eingeführter GAN-Loss zur Verbesserung der generierten Annotationen
$\mathcal{L}_{identity}$	Identity-Loss
$\mathcal{L}_{LS}$	Least-Squares-Loss
$\lambda$	Parameter zur Steuerung des Einflusses einer bestimmten Lossfunktion
$\max$	Maximumfunktion
$\min$	Minimumfunktion
$N_k$	Anzahl an Klassen
$N_s$	Anzahl an Samples
$o$	Mittlere Intensität eines Objekts
$O$	Optimierungskriterium

$p(\cdot)$	Wahrscheinlichkeitsverteilung
$p_r$	Verhältnis der Korrelationskoeffizienten zwischen der Basis- und Testprädiktion
$r_{z,\hat{z}}$	Pearson-Korrelationskoeffizient
$r_{z,\hat{z}; \text{VG}}$	Pearson-Korrelationskoeffizient zwischen Vordergrundvo- xel
$r_{o,\hat{o}; \text{Obj}}$	Pearson-Korrelationskoeffizient zwischen mittleren Ob- jektintensitäten
$\theta$	Gewichte eines neuronalen Netzwerks
$v$	Annotationen
$x$	Daten der Domäne X; Bedingung
$\mathcal{X}$	Domäne X
$y$	Daten der Domäne Y; Klasse
$\hat{y}_z$	Prognostizierter Zielmarker der Basisaufnahme
$y_b$	Basismarker der Basisaufnahme
$\mathcal{Y}$	Domäne Y
$z$	Rauschterm
$z_i$	Intensität eines Voxels $i$
$z_b$	Basismarker der Testaufnahme
$z_z$	Zielmarker der Testaufnahme
$\hat{z}_z$	Prognostizierter Zielmarker der Testaufnahme
$\mathcal{Z}$	Domäne Z
$\zeta$	Gewichte eines neuronalen Netzwerks

$\bar{\cdot}$	Mittelwert
$\hat{\cdot}$	Synthetische Signale generiert durch einfache Transformation
$\tilde{\cdot}$	Synthetische Signale generiert durch Hin- und Rücktransformation
$\rightarrow$	Abbildung
$\circ$	Komposition
$\sim$	Verteilt wie

# Abbildungsverzeichnis

1.1	Vereinfachte Darstellung eines konfokalen Fluoreszenzmikroskops. . .	5
1.2	Visualisierung einer Domänenlücke zwischen realen und syn- thetischen Daten. . . . .	15
2.1	Übersicht der Datenauswertung mittels Deep-Learning-Algorithmen mit markierten Zielstellungen der Arbeit. . . . .	30
2.2	Konzept von SimOptiGAN zur Synthese von Aufnahmen. . . . .	37
2.3	Detailansichten der in Abbildung 2.2 dargestellten Blöcke Pro- bengenerierung, Aufnahmesimulation und Optimierung. . . . .	40
2.4	Veranschaulichung der Wirkung verschiedener Parametereinstel- lungen auf die mit SimOptiGAN generierten 3D-Aufnahmen. . . . .	43
2.5	Konzept zur Einbringung von biophysikalischen Simulationen in den Syntheseprozess. . . . .	46
2.6	Entwickelte Möglichkeiten zur Generierung von Nukleusaufnah- men auf Basis biophysikalisch simulierter binärer Membransi- gnale. . . . .	47
2.7	Detailansicht der weiter untersuchten Methoden SimOptiGAN+, Mem2NucGAN-P und Mem2NucGAN-U zur Generierung von Nukleusaufnahmen, basierend auf den biophysikalischen Simulationen.	50
2.8	Detailansicht der adaptierten Probengenerierung von SimOptiGAN+. .	50
2.9	Überblick der eingeführten Trainingsstruktur <i>DualGenGAN</i> für die zusätzliche Generierung von Annotationsdaten. . . . .	57
2.10	Generierungsprozess von synthetischen Nukleusaufnahmen mit zugehörigen Instanzannotationen. . . . .	60
2.11	Berechnung der KID-Metrik zwischen realen und synthetischen 3D-Aufnahmen anhand von Bildausschnitten orthogonaler Ebenen. .	65
2.12	Ablauf der Bewertung synthetischer Trainingsdaten auf Basis der Segmentierungsgüte auf realen Daten. . . . .	66

2.13	Problemstellung inkompatibler Marker und Übersicht des Frameworks zur Synthese von Bildmodalitäten. . . . .	68
2.14	Überblick über das Training der Modelle zur Synthese von Bildmodalitäten und zwei Evaluierungsmethoden zur Prüfung, ob sich Transformationsmodelle an eventuell vorhandenen durchscheinenden Signalen orientieren. . . . .	75
2.15	Darstellung verschiedener Zellzyklusstadien. . . . .	78
2.16	Konzept zur Klassifikation von Zellzyklusstadien. . . . .	80
2.17	Darstellung der drei Versuchsanordnungen zum Testen der Methoden zur Synthese von Bildmodalitäten anhand der Klassifikation von Zellzyklusstadien. . . . .	81
3.1	Übersicht der erstellten Softwarepakete zur Synthese von Trainingsdaten und deren Zusammenspiel. . . . .	86
3.2	Veranschaulichung des Ablaufs zur Auswertung von 3D-Sphäroid-Daten mittels Spherolyzer. . . . .	87
3.3	Grafische Benutzeroberfläche für die Parametrierung der Software zur Auswertung von 3D-Zellkulturaufnahmen. . . . .	90
3.4	3D-ClassAnnotator: Entwickelte grafische Oberfläche zur Erstellung der Klassenannotationen. . . . .	91
4.1	Synthetische 3D-Objektprototypen zweier Klassen: seltene und reguläre Objekte. . . . .	95
4.2	2D-Schnitte von realen 3D-Aufnahmen für die Ground-Truth-Erstellung und das Training der Transformationsmodelle und 2D-Schnitte eines Ergebnisses der biophysikalischen Simulation mit daraus generierten synthetischen Membranen. . . . .	99
4.3	2D-Schnitte von beispielhaft ausgewählten 3D-Aufnahmen, welche für die Erstellung des Datensatzes <i>DS-Marker-Syn</i> genutzt wurden. . . . .	103
4.4	Visualisierung des Sphäroids, welcher für die Erstellung des Trainingsdatensatzes ( <i>DS-Klassifikation</i> ) genutzt wurde und zusätzlich Beispiele der daraus extrahierten Bildausschnitte. . . . .	105
5.1	Visualisierung der Domänenlücke zwischen synthetischen und realen Daten. . . . .	109



5.2	Exemplarische 2D-Bildausschnitte des Datensatzes <i>DS-Seltene-Elemente</i> mit zugehörigen Segmentierungen für die Analyse von seltenen Elementen in den Trainingsdaten. . . . .	112
5.3	Segmentierungsergebnisse zur Analyse von seltenen Elementen in den Trainingsdaten. . . . .	112
5.4	Vergleich von realen und synthetisch generierten Bildern mit Angabe des KID zwischen synthetischen und realen Daten. . . . .	117
5.5	Segmentierungsqualität der Nukleus-Segmentierungsmodelle, welche auf verschiedenen synthetischen und realen Datensätzen trainiert wurden. . . . .	120
5.6	Qualitativer Vergleich der Segmentierungsergebnisse von Modellen, welche auf verschiedenen synthetischen und realen Daten trainiert wurden. . . . .	122
5.7	2D-Visualisierung der mittels Mem2NucGAN-P generierten Aufnahme und Annotationen . . . . .	125
5.8	Vergleich der Konsistenz zwischen generierten Bildern und zugehörigen generierten Annotationen. . . . .	126
5.9	Vergleich des Einflusses des zusätzlichen Nukleusmarkers auf die Segmentierungsleistung von Membransegmentierungsmodellen. .	129
5.10	Vergleich zwischen realen und mittels SimOptiGAN generierten Rasterelektronenmikroskopaufnahmen. . . . .	132
5.11	Vergleich zwischen realen und mit SimOptiGAN generierten 3D-Aufnahmen unterschiedlicher Zellkulturen. . . . .	135
5.12	Ergebnisse der Bild-Optimierung mittels des trainierten CycleGANs von SimOptiGAN. . . . .	136
5.13	Bildausschnitte zweier 3D-Aufnahmen von Nukleus und dazugehörigen Ki-67-Signalen und den auf Basis der realen Nukleus-signale generierten Ki-67-Prädiktionen. . . . .	140
5.14	Pearson-Korrelation zwischen den Intensitäten des realen Ki-67-Signals und des synthetischen Ki-67-Signals. . . . .	141
5.15	Beispielhafte Bildausschnitte der Datensätze <i>DS-Klassifikation-Real</i> und <i>DS-Klassifikation-Syn</i> . . . . .	144
5.16	Vergleich der Güte der Klassifikationsmodelle für die Klassifikation von Zellzyklusstadien. . . . .	146

5.17	Konfusionsmatrizen der Klassifikationsmodelle für die Klassifikation von Zellzyklusstadien. . . . .	146
5.18	Repräsentative Bildausschnitte der zweistufigen 3D-Aufnahme des Bleed-Through-Tests. . . . .	149
5.19	Ergebnisse des Bleed-Trough-Tests für die Methoden zur Synthese von Bildmodalitäten. . . . .	150
5.20	Pearson-Korrelation zwischen den realen Signalen des Zielmarkers und denen anhand des Basismarkers generierten synthetischen Signalen für verschiedene Markerkombinationen. . . . .	152
5.21	Repräsentative Bildausschnitte von realen und synthetisch generierten 3D-Signalen von verschiedenen Markerkombinationen. . . . .	154
A.1	Übersicht der Struktur von Generative Adversarial Networks (GANs). . . . .	166
A.2	Übersicht der Netzwerkstrukturen für die Bild-zu-Bild-Transformation. . . . .	169
A.3	Funktionsweise von Denoising Diffusion Probabilistic Models (DDPMs) . . . . .	174
A.4	Qualitativer Vergleich der Segmentierungsergebnisse der einzelnen Modelle anhand der visualisierten Fehler. . . . .	182
A.5	Weitere realen und mittels SimOptiGAN generierten Rasterelektronenmikroskopaufnahmen. . . . .	182
A.6	Zusätzliche Ergebnisse des Bleed-Trough-Tests für die Methoden zur Synthese von Bildmodalitäten. . . . .	183

# Tabellenverzeichnis

- 2.1 Vergleich der Vor- und Nachteile des ResUnet im Vergleich zu pix2pix. 71
- 5.1 Vergleich der zur Generierung von 3D-Trainingsdaten benötigten Zeit. 114
- 5.2 Vergleich der zum Training und zur Anwendung benötigten Re-  
chenzeit und Speicherbedarf von zwei Methoden zur Synthese  
von Bildmodalitäten. . . . . 138

## Eigene Veröffentlichungen

- [21] F. Keller, **R. Bruch**, R. Schneider, J. Meier-Hubberten, M. Hafner, and R. Rudolf, “A scaffold-free 3-D co-culture mimics the major features of the reverse Warburg effect in vitro,” *Cells*, vol. 9, no. 8, p. 1900, 2020.
- [22] F. Keller, **R. Bruch**, F. Clauder, M. Hafner, and R. Rudolf, “Extracellular matrix components regulate bone sialoprotein expression in MDA-MB-231 breast cancer cells,” *Cells*, vol. 10, no. 6, p. 1304, 2021.
- [23] S. J. Hörner, N. Couturier, **R. Bruch**, P. Koch, M. Hafner, and R. Rudolf, “hiPSC-derived schwann cells influence myogenic differentiation in neuromuscular cocultures,” *Cells*, vol. 10, no. 12, p. 3292, 2021.
- [24] E. Nuernberg, **R. Bruch**, M. Hafner, R. Rudolf, and M. Vitacolonna, “Quantitative analysis of whole-mount fluorescence-stained tumor spheroids in phenotypic drug screens,” in *3D Cell Culture: Methods and Protocols*, Z. Sumbalova Koledova, Ed. New York, NY: Springer US, 2024, pp. 311–334.
- [29] E. Nürnberg, M. Vitacolonna, J. Klicks, E. von Molitor, T. Cesetti, F. Keller, **R. Bruch**, T. Ertongur-Fauth, K. Riedel, P. Scholz *et al.*, “Routine optical clearing of 3D-cell cultures: simplicity forward,” *Frontiers in Molecular Biosciences*, vol. 7, 2020.
- [31] **R. Bruch**, P. M. Scheickl, R. Mikut, F. Loosli, and M. Reischl, “epiTracker: a framework for highly reliable particle tracking for the quantitative analysis of fish movements in tanks,” *SLAS Technology*, vol. 26, no. 4, pp. 367–376, 2021.

- [56] **R. Bruch**, R. Rudolf, R. Mikut, and M. Reischl, “Evaluation of semi-supervised learning using sparse labeling to segment cell nuclei,” *Current Directions in Biomedical Engineering*, vol. 6, no. 3, pp. 398–401, 2020.
- [57] M. Vitacolonna, **R. Bruch**, A. Agaçi, E. Nürnberg, T. Cesetti, F. Keller, F. Padovani, S. Sauer, K. M. Schmoller, M. Reischl *et al.*, “A multiparametric analysis including single-cell and subcellular feature assessment reveals differential behavior of spheroid cultures on distinct ultra-low attachment plate types,” *Frontiers in Bioengineering and Biotechnology*, vol. 12, 2024.
- [59] M. Vitacolonna<sup>1</sup>, **R. Bruch**<sup>1</sup>, R. Schneider, J. Jabs, M. Hafner, M. Reischl, and R. Rudolf, “A spheroid whole mount drug testing pipeline with machine-learning based image analysis identifies cell-type specific differences in drug efficacy on a single-cell level,” *BMC Cancer*, vol. 24, p. 1542, 2024.
- [68] L. Rettenberger, F. R. Münke, **R. Bruch**, and M. Reischl, “Mask R-CNN outperforms U-Net in instance segmentation for overlapping cells,” *Current Directions in Biomedical Engineering*, vol. 9, no. 1, pp. 335–338, 2023.
- [149] M. Böhländ, **R. Bruch**, K. Löffler, and M. Reischl, “Unsupervised GAN epoch selection for biomedical data synthesis,” *Current Directions in Biomedical Engineering*, vol. 9, no. 1, pp. 467–470, 2023.
- [176] **R. Bruch**, F. Keller, M. Böhländ, M. Vitacolonna, L. Klinger, R. Rudolf, and M. Reischl, “Synthesis of large scale 3D microscopic images of 3D cell cultures for training and benchmarking,” *PLOS ONE*, vol. 18, no. 3, p. e0283828, 2023.

---

<sup>1</sup> Geteilte Erstautorschaft. Die Autoren M. Vitacolonna und R. Bruch haben gleichwertig zur Arbeit beigetragen.

- [177] **R. Bruch**, M. Vitacolonna, E. Nürnberg, S. Sauer, R. Rudolf, and M. Reischl, “Improving 3D deep learning segmentation with biophysically motivated cell synthesis,” *Communications Biology*, vol. 8, no. 1, 2025.
- [178] M. Böhlend, **R. Bruch**, S. Bäuerle, L. Rettenberger, and M. Reischl, “Improving generative adversarial networks for patch-based unpaired image-to-image translation,” *IEEE Access*, vol. 11, pp. 127 895–127 906, 2023.
- [180] E. Nürnberg, M. Vitacolonna, **R. Bruch**, M. Reischl, R. Rudolf, and S. Sauer, “From in vitro to in silico: a pipeline for generating virtual tissue simulations from real image data,” *Frontiers in Molecular Biosciences*, vol. 11, 2024.
- [188] **R. Bruch**, M. Vitacolonna, R. Rudolf, and M. Reischl, “Prediction of fluorescent Ki67 staining in 3D tumor spheroids,” *Current Directions in Biomedical Engineering*, vol. 8, no. 2, pp. 305–308, 2022.

# Literaturverzeichnis

- [1] F. Tao, B. Xiao, Q. Qi, J. Cheng, and P. Ji, “Digital twin modeling,” *Journal of Manufacturing Systems*, vol. 64, pp. 372–389, 2022.
- [2] J. Kocic, N. Jovicic, and V. Drndarevic, “Sensors and sensor fusion in autonomous vehicles,” in *2018 26th Telecommunications Forum (TELFOR)*. IEEE, 2018, pp. 420–425.
- [3] H. S. Stein and J. M. Gregoire, “Progress and prospects for accelerating materials science with automated and autonomous workflows,” *Chemical Science*, vol. 10, no. 42, pp. 9640–9649, 2019.
- [4] J. Schuetzke, S. Schweidler, F. R. Muenke, A. Orth, A. D. Khandelwal, B. Breitung, J. Aghassi-Hagmann, and M. Reischl, “Accelerating materials discovery: Automated identification of prospects from x-ray diffraction data in fast screening experiments,” *Advanced Intelligent Systems*, vol. 6, no. 3, 2023.
- [5] I. Despotović, B. Goossens, and W. Philips, “Mri segmentation of the human brain: Challenges, methods, and applications,” *Computational and Mathematical Methods in Medicine*, vol. 2015, pp. 1–23, 2015.
- [6] A. Wadhwa, A. Bhardwaj, and V. Singh Verma, “A review on brain tumor segmentation of mri images,” *Magnetic Resonance Imaging*, vol. 61, pp. 247–259, 2019.
- [7] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, “Deep learning for cardiac image segmentation: A review,” *Frontiers in Cardiovascular Medicine*, vol. 7, 2020.

- [8] W. Wu, Z. Qi, and L. Fuxin, “Pointconv: Deep convolutional networks on 3D point clouds,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9621–9630.
- [9] T. Cortinhal, F. Kurnaz, and E. E. Aksoy, “Semantics-aware multi-modal domain translation: From lidar point clouds to panoramic color images,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 2021, pp. 3032–3048.
- [10] T. Cortinhal and E. E. Aksoy, “Depth- and semantics-aware multi-modal domain translation: Generating 3D panoramic color images from lidar point clouds,” *Robotics and Autonomous Systems*, vol. 171, p. 104583, 2024.
- [11] S. Milz, M. Simon, K. Fischer, M. Pöpperl, and H.-M. Gross, *Points2Pix: 3D Point-Cloud to Image Translation Using Conditional GANs*. Springer International Publishing, 2019, pp. 387–400.
- [12] Z. Song, W. Chen, D. Campbell, and H. Li, “Deep novel view synthesis from colored 3D point clouds,” in *Lecture Notes in Computer Science*. Springer International Publishing, 2020, pp. 1–17.
- [13] P. J. Withers, C. Bouman, S. Carmignato, V. Cnudde, D. Grimaldi, C. K. Hagen, E. Maire, M. Manley, A. Du Plessis, and S. R. Stock, “X-ray computed tomography,” *Nature Reviews Methods Primers*, vol. 1, no. 1, 2021.
- [14] I. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken, “Computer analysis of computed tomography scans of the lung: a survey,” *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 385–405, 2006.
- [15] U. Kubitscheck, Ed., *Fluorescence Microscopy: From Principles to Biological Applications*, 2nd ed. Weinheim, Germany: Wiley-VCH, 2013.
- [16] D. Murphy and M. Davidson, *Fundamentals of Light Microscopy and Electronic Imaging*, 2nd ed., ser. Fundamentals of Light Microscopy and Electronic Imaging. Hoboken, NJ: Wiley-Blackwell, 2012.



- [17] P. P. Mondal and A. Diaspro, *Fundamentals of Fluorescence Microscopy: Exploring Life with Light*, 1st ed. Dordrecht, Netherlands: Springer, 2014.
- [18] P. I. H. Bastiaens and A. Squire, “Fluorescence lifetime imaging microscopy: spatial resolution of biochemical processes in the cell,” *Trends in Cell Biology*, vol. 9, no. 2, pp. 48–52, 1999.
- [19] J. Pawley, *Handbook of biological confocal microscopy*, 3rd ed. New York, NY: Springer, 2006.
- [20] G. Herberich, R. Windoffer, R. E. Leube, and T. Aach, “Signal and noise modeling in confocal laser scanning fluorescence microscopy,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*. Springer, 2012, pp. 381–388.
- [21] F. Keller, **R. Bruch**, R. Schneider, J. Meier-Hubberten, M. Hafner, and R. Rudolf, “A scaffold-free 3-D co-culture mimics the major features of the reverse Warburg effect in vitro,” *Cells*, vol. 9, no. 8, p. 1900, 2020.
- [22] F. Keller, **R. Bruch**, F. Clauder, M. Hafner, and R. Rudolf, “Extracellular matrix components regulate bone sialoprotein expression in MDA-MB-231 breast cancer cells,” *Cells*, vol. 10, no. 6, p. 1304, 2021.
- [23] S. J. Hörner, N. Couturier, **R. Bruch**, P. Koch, M. Hafner, and R. Rudolf, “hiPSC-derived schwann cells influence myogenic differentiation in neuromuscular cocultures,” *Cells*, vol. 10, no. 12, p. 3292, 2021.
- [24] E. Nuernberg, **R. Bruch**, M. Hafner, R. Rudolf, and M. Vitacolonna, “Quantitative analysis of whole-mount fluorescence-stained tumor spheroids in phenotypic drug screens,” in *3D Cell Culture: Methods and Protocols*, Z. Sumbalova Koledova, Ed. New York, NY: Springer US, 2024, pp. 311–334.
- [25] N. Alépée, “State-of-the-art of 3D cultures (organs-on-a-chip) in safety testing and pathophysiology,” *ALTEX - Alternatives to animal experimentation*, vol. 31, no. 4, pp. 441–477, 2014.

- [26] A. P. Browning, J. A. Sharp, R. J. Murphy, G. Gunasingh, B. Lawson, K. Burrage, N. K. Haass, and M. Simpson, “Quantitative analysis of tumour spheroid structure,” *eLife*, vol. 10, p. e73020, 2021.
- [27] J. Drost and H. Clevers, “Organoids in cancer research,” *Nature Reviews Cancer*, vol. 18, no. 7, pp. 407–418, 2018.
- [28] J. Tang, J. Shi, and J. Liu, “Editorial: Advances in 3D cell culture for drug screening and toxicology evaluation,” *Frontiers in Bioengineering and Biotechnology*, vol. 11, 2023.
- [29] E. Nürnberg, M. Vitacolonna, J. Klicks, E. von Molitor, T. Cesetti, F. Keller, **R. Bruch**, T. Ertongur-Fauth, K. Riedel, P. Scholz *et al.*, “Routine optical clearing of 3D-cell cultures: simplicity forward,” *Frontiers in Molecular Biosciences*, vol. 7, 2020.
- [30] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [31] **R. Bruch**, P. M. Scheickl, R. Mikut, F. Loosli, and M. Reischl, “epiTracker: a framework for highly reliable particle tracking for the quantitative analysis of fish movements in tanks,” *SLAS Technology*, vol. 26, no. 4, pp. 367–376, 2021.
- [32] L. Gole, K. H. Ong, T. Boudier, W. Yu, and S. Ahmed, “OpenSegSPIM: a user-friendly segmentation tool for SPIM data,” *Bioinformatics*, vol. 32, no. 13, pp. 2075–2077, 2016.
- [33] X. Lou, M. Kang, P. Xenopoulos, S. Muñoz-Descalzo, and A.-K. Hadjantonakis, “A rapid and efficient 2D/3D nuclear segmentation method for analysis of early mouse embryo and stem cell image data,” *Stem Cell Reports*, vol. 2, no. 3, pp. 382–397, 2014.
- [34] B. Mathew, A. Schmitz, S. Muñoz-Descalzo, N. Ansari, F. Pampaloni, E. Stelzer, and S. Fischer, “Robust and automated three-dimensional segmentation of densely packed cell nuclei in different biological specimens

- with lines-of-sight decomposition,” *BMC Bioinformatics*, vol. 16, no. 1, 2015.
- [35] C. McQuin, A. Goodman, V. Chernyshev, L. Kametsky, B. A. Cimini, K. W. Karhohs, M. Doan, L. Ding, S. M. Rafelski, D. Thirstrup *et al.*, “CellProfiler 3.0: Next-generation image processing for biology,” *PLOS Biology*, vol. 16, no. 7, p. e2005970, 2018.
- [36] J. Ollion, J. Cochenne, F. Loll, C. Escudé, and T. Boudier, “TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization,” *Bioinformatics*, vol. 29, no. 14, pp. 1840–1841, 2013.
- [37] A. Schmitz, S. C. Fischer, C. Mattheyer, F. Pampaloni, and E. H. K. Stelzer, “Multiscale image analysis reveals structural heterogeneity of the cell microenvironment in homotypic spheroids,” *Scientific Reports*, vol. 7, no. 1, p. 43693, 2017.
- [38] J. Stegmaier, J. C. Otte, A. Kobitski, A. Bartschat, A. Garcia, G. U. Nienhaus, U. Strähle, and R. Mikut, “Fast segmentation of stained nuclei in terabyte-scale, time resolved 3D microscopy image stacks,” *PLoS ONE*, vol. 9, no. 2, p. e90036, 2014.
- [39] J. Stegmaier, F. Amat, W. Lemon, K. McDole, Y. Wan, G. Teodoro, R. Mikut, and P. Keller, “Real-time three-dimensional cell segmentation in large-scale microscopy data of developing embryos,” *Developmental Cell*, vol. 36, no. 2, pp. 225–240, 2016.
- [40] A. Bartschat, E. Hübner, M. Reischl, R. Mikut, and J. Stegmaier, “XPIWIT—an XML pipeline wrapper for the insight toolkit,” *Bioinformatics*, vol. 32, no. 2, pp. 315–317, 2016.
- [41] V. Ulman, M. Maška, K. E. G. Magnusson, O. Ronneberger, C. Haubold, N. Harder, P. Matula, P. Matula, D. Svoboda, M. Radojevic *et al.*, “An objective comparison of cell-tracking algorithms,” *Nature methods*, vol. 14, no. 12, p. 1141, 2017.

- [42] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki *et al.*, “Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BraTS challenge,” *arXiv preprint arXiv:1811.02629*, 2018.
- [43] J. C. Caicedo, A. Goodman, K. W. Karhohs, B. A. Cimini, J. Ackerman, M. Haghighi, C. Heng, T. Becker, M. Doan, C. McQuin *et al.*, “Nucleus segmentation across imaging experiments: the 2018 data science bowl,” *Nature Methods*, vol. 16, no. 12, pp. 1247–1253, 2019.
- [44] S. Graham, Q. D. Vu, M. Jahanifar, M. Weigert, U. Schmidt, W. Zhang, J. Zhang, S. Yang, J. Xiang, X. Wang *et al.*, “Conic challenge: Pushing the frontiers of nuclear detection, segmentation, classification and counting,” *Medical Image Analysis*, vol. 92, p. 103047, 2024.
- [45] Z. Jiang, C. Ding, M. Liu, and D. Tao, “Two-stage cascaded u-net: 1st place solution to BraTS challenge 2019 segmentation task,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Cham, Switzerland: Springer, 2020, pp. 231–241.
- [46] A. E. Kavur, N. S. Gezer, M. Barış, S. Aslan, P.-H. Conze, V. Groza, D. D. Pham, S. Chatterjee, P. Ernst, S. Özkan *et al.*, “CHAOS challenge - combined (CT-MR) healthy abdominal organ segmentation,” *Medical Image Analysis*, vol. 69, p. 101950, 2021.
- [47] J. Ma, R. Xie, S. Ayyadhury, C. Ge, A. Gupta, R. Gupta, S. Gu, Y. Zhang, G. Lee, J. Kim *et al.*, “The multimodality cell segmentation challenge: toward universal solutions,” *Nature Methods*, vol. 21, no. 6, pp. 1103–1113, 2024.
- [48] M. Maška, V. Ulman, P. Delgado-Rodriguez, E. Gómez-de Mariscal, T. Nečasová, F. A. Guerrero Peña, T. I. Ren, E. M. Meyerowitz, T. Scherr, K. Löffler *et al.*, “The Cell Tracking Challenge: 10 years of objective benchmarking,” *Nature Methods*, vol. 20, no. 7, pp. 1010–1020, 2023.

- [49] A. M. Mendrik, K. L. Vincken, H. J. Kuijf, M. Breeuwer, W. H. Bouvy, J. de Bresser, A. Alansary, M. de Bruijne, A. Carass, A. El-Baz *et al.*, “MR-BrainS challenge: Online evaluation framework for brain image segmentation in 3T MRI scans,” *Computational Intelligence and Neuroscience*, vol. 2015, pp. 1–16, 2015.
- [50] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, “The multimodal brain tumor image segmentation benchmark (BraTS),” *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.
- [51] E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, and D. V. Valen, “Deep learning for cellular image analysis,” *Nature Methods*, vol. 16, no. 12, pp. 1233–1246, 2019.
- [52] N. U. Din and J. Yu, “Training a deep learning model for single-cell segmentation without manual annotation,” *Scientific Reports*, vol. 11, p. 23995, 2021.
- [53] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, “A guide to deep learning in healthcare,” *Nature Medicine*, vol. 25, pp. 24–29, 2019.
- [54] Y. Roh, G. Heo, and S. E. Whang, “A survey on data collection for machine learning: A big data - AI integration perspective,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1328–1347, 2021.
- [55] S. Wang, C. Li, R. Wang, Z. Liu, M. Wang, H. Tan, Y. Wu, X. Liu, H. Sun, R. Yang *et al.*, “Annotation-efficient deep learning for automatic medical image segmentation,” *Nature Communications*, vol. 12, 2021.
- [56] **R. Bruch**, R. Rudolf, R. Mikut, and M. Reischl, “Evaluation of semi-supervised learning using sparse labeling to segment cell nuclei,” *Current Directions in Biomedical Engineering*, vol. 6, no. 3, pp. 398–401, 2020.

- [57] M. Vitacolonna, **R. Bruch**, A. Agaçi, E. Nürnberg, T. Cesetti, F. Keller, F. Padovani, S. Sauer, K. M. Schmoller, M. Reischl *et al.*, “A multiparametric analysis including single-cell and subcellular feature assessment reveals differential behavior of spheroid cultures on distinct ultra-low attachment plate types,” *Frontiers in Bioengineering and Biotechnology*, vol. 12, 2024.
- [58] M. Bohland, O. Neumann, M. P. Schilling, M. Reischl, R. Mikut, K. Löffler, and T. Scherr, “ciscNet - a single-branch cell nucleus instance segmentation and classification network,” in *2022 IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)*. IEEE, 2022, pp. 1–5.
- [59] M. Vitacolonna<sup>1</sup>, **R. Bruch**<sup>1</sup>, R. Schneider, J. Jabs, M. Hafner, M. Reischl, and R. Rudolf, “A spheroid whole mount drug testing pipeline with machine-learning based image analysis identifies cell-type specific differences in drug efficacy on a single-cell level,” *BMC Cancer*, vol. 24, p. 1542, 2024.
- [60] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI*. Cham, Switzerland: Springer, 2015, pp. 234–241.
- [61] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald *et al.*, “U-Net: deep learning for cell counting, detection, and morphometry,” *Nature Methods*, vol. 16, pp. 67–70, 2018.
- [62] X. Li, Y. Wang, Q. Tang, Z. Fan, and J. Yu, “Dual U-Net for the segmentation of overlapping glioma nuclei,” *IEEE Access*, vol. 7, pp. 84 040–84 052, 2019.
- [63] T. Scherr, K. Löffler, M. Böhlend, and R. Mikut, “Cell segmentation and tracking using distance transform predictions and movement estimation with graph-based matching,” *arXiv preprint arXiv:2004.01486*, 2020.

- 
- [64] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning dense volumetric segmentation from sparse annotation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Cham, Switzerland: Springer, 2016, pp. 424–432.
- [65] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu, “Cellpose: a generalist algorithm for cellular segmentation,” *Nature Methods*, vol. 18, pp. 100–106, 2021.
- [66] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, “Cell detection with star-convex polygons,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Cham, Switzerland: Springer, 2018, pp. 265–273.
- [67] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [68] L. Rettenberger, F. R. Münke, **R. Bruch**, and M. Reischl, “Mask R-CNN outperforms U-Net in instance segmentation for overlapping cells,” *Current Directions in Biomedical Engineering*, vol. 9, no. 1, pp. 335–338, 2023.
- [69] M. Y. Lee, J. S. Bedia, S. S. Bhate, G. L. Barlow, D. Phillips, W. J. Fantl, G. P. Nolan, and C. M. Schürch, “CellSeg: a robust, pre-trained nucleus segmentation and pixel quantification software for highly multiplexed fluorescence images,” *BMC Bioinformatics*, vol. 23, p. 46, 2022.
- [70] D. Eschweiler, R. S. Smith, and J. Stegmaier, “Robust 3D cell segmentation: Extending the view of cellpose,” in *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 191–195.
- [71] M. Weigert, U. Schmidt, R. Haase, K. Sugawara, and G. Myers, “Star-convex polyhedra for 3D object detection and segmentation in microscopy,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2020.

- [72] H. Chen and R. F. Murphy, “3DCellComposer - a versatile pipeline utilizing 2D cell segmentation methods for 3D cell segmentation,” *bioRxiv* 2024.03.08.584082, 2024.
- [73] I. Wolf, M. Vetter, I. Wegner, M. Nolden, T. Bottger, M. Hastenteufel, M. Schobinger, T. Kunert, and H.-P. Meinzer, “The medical imaging interaction toolkit (MITK): a toolkit facilitating the creation of interactive software by extending VTK and ITK,” in *Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display*. SPIE, 2004.
- [74] R. Kikinis, S. D. Pieper, and K. G. Vosburgh, “3D Slicer: A platform for subject-specific image analysis, visualization, and clinical support,” in *Intraoperative Imaging and Image-Guided Therapy*, F. A. Jolesz, Ed. New York, NY: Springer, 2013, pp. 277–289.
- [75] J. Ahlers, D. Althviz Moré, O. Amsalem, A. Anderson, G. Bokota, P. Boone, J. Bragantini, G. Buckley, A. Burt, M. Bussonnier *et al.*, “napari: a multi-dimensional image viewer for Python,” Zenodo, 2023.
- [76] M. Amrehn, S. Gaube, M. Unberath, F. Schebesch, T. Horz, M. Strumia, S. Steidl, M. Kowarschik, and A. Maier, “UI-Net: Interactive artificial neural networks for iterative image segmentation based on a user model,” *arXiv preprint arXiv:1709.03450*, 2017.
- [77] J. Chen, L. Ding, M. P. Viana, M. C. Hendershott, R. Yang, I. A. Mueller, and S. M. Rafelski, “The allen cell structure segmenter: a new open source toolkit for segmenting 3D intracellular structures in fluorescence microscopy images,” *bioRxiv*, p. 491035, 2018.
- [78] D. F. Pace, A. V. Dalca, T. Brosch, T. Geva, A. J. Powell, J. Weese, M. H. Moghari, and P. Golland, “Iterative segmentation from limited training data: Applications to congenital heart disease,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer International Publishing, 2018, pp. 334–342.



- [79] D. P. Papadopoulos, A. D. F. Clarke, F. Keller, and V. Ferrari, “Training object class detectors from eye tracking data,” in *Computer Vision – ECCV 2014*. Cham, Switzerland: Springer, 2014, pp. 361–376.
- [80] T. Sakinis, F. Milletari, H. Roth, P. Korfiatis, P. Kostandy, K. Philbrick, Z. Akkus, Z. Xu, D. Xu, and B. J. Erickson, “Interactive segmentation of medical images through fully convolutional neural networks,” *arXiv preprint arXiv:1903.08205*, 2019.
- [81] M. P. Schilling, S. Schmelzer, L. Klinger, and M. Reischl, “KaIDA: a modular tool for assisting image annotation in deep learning,” *Journal of Integrative Bioinformatics*, vol. 19, no. 4, p. 20220018, 2022.
- [82] M. P. Schilling, L. Klinger, U. Schumacher, S. Schmelzer, M. B. López, B. Nestler, and M. Reischl, “AI2Seg: A method and tool for ai-based annotation inspection of biomedical instance segmentation datasets,” in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2023.
- [83] B. Wang, K. W. Liu, K. M. Prastawa, A. Irima, P. M. Vespa, J. D. Van Horn, P. T. Fletcher, and G. Gerig, “4D active cut: An interactive tool for pathological anatomy modeling,” in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2014, pp. 529–532.
- [84] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, “Segment anything,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 4015–4026.
- [85] S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy *et al.*, “ilastik: interactive machine learning for (bio)image analysis,” *Nature Methods*, vol. 16, no. 12, pp. 1226–1232, 2019.
- [86] B. Lutnick, B. Ginley, D. Govind, S. D. McGarry, P. S. LaViolette, R. Yacoub, S. Jain, J. E. Tomaszewski, K.-Y. Jen, and P. Sarder, “An integrated

- iterative annotation technique for easing neural network training in medical image analysis,” *Nature Machine Intelligence*, vol. 1, no. 2, pp. 112–119, 2019.
- [87] A. Top, G. Hamarneh, and R. Abugharbieh, “Active learning for interactive 3D image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*. Berlin, Heidelberg, Germany: Springer Berlin Heidelberg, 2011, pp. 603–610.
- [88] T. Zhou, L. Li, G. Bredell, J. Li, J. Unkelbach, and E. Konukoglu, “Volumetric memory network for interactive medical image segmentation,” *Medical Image Analysis*, vol. 83, p. 102599, 2023.
- [89] K.-K. Maninis, S. Caelles, J. Pont-Tuset, and L. Van Gool, “Deep extreme cut: From extreme points to object segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 616–625.
- [90] C. Rother, V. Kolmogorov, and A. Blake, ““GrabCut”: interactive foreground extraction using iterated graph cuts,” *ACM Transactions on Graphics*, vol. 23, no. 3, p. 309, 2004.
- [91] E. A. Tasnadi, T. Toth, M. Kovacs, A. Diosdi, F. Pampaloni, J. Molnar, F. Piccinini, and P. Horvath, “3D-Cell-Annotator: an open-source active surface tool for single cell segmentation in 3D microscopy images,” *Bioinformatics*, vol. 36, no. 9, pp. 2948–2949, 2020.
- [92] T. Wang, B. Han, and J. Collomosse, “TouchCut: Fast image and video segmentation using single-touch interaction,” *Computer Vision and Image Understanding*, vol. 120, pp. 14–30, 2014.
- [93] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, “User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability,” *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

- [94] M. Koziński, A. Mosinska, M. Salzmann, and P. Fua, “Tracing in 2D to reduce the annotation effort for 3D deep delineation of linear structures,” *Medical Image Analysis*, vol. 60, p. 101590, 2020.
- [95] K. Sugawara, “Training deep learning models for cell image segmentation with sparse annotations,” *bioRxiv*, 2023.
- [96] A. Bearman, O. Russakovsky, V. Ferrari, and L. Fei-Fei, “What’s the point: Semantic segmentation with point supervision,” in *Computer Vision – ECCV 2016*. Cham, Switzerland: Springer, 2016, pp. 549–565.
- [97] A. Khoreva, R. Benenson, J. Hosang, M. Hein, and B. Schiele, “Simple does it: Weakly supervised instance and semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 876–885.
- [98] D. Lin, J. Dai, J. Jia, K. He, and J. Sun, “Scribblesup: Scribble-supervised convolutional networks for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3159–3167.
- [99] I. Yoo, D. Yoo, and K. Paeng, “Pseudoedgenet: Nuclei segmentation only with point annotations,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Cham, Switzerland: Springer, 2019, pp. 731–739.
- [100] T. Zhao and Z. Yin, “Weakly supervised cell segmentation by point annotation,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 10, pp. 2736–2747, 2021.
- [101] D. Pathak, P. Krahenbuhl, and T. Darrell, “Constrained convolutional neural networks for weakly supervised segmentation,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1796–1804.
- [102] S. Mittal, M. Tatarchenko, and T. Brox, “Semi-supervised semantic segmentation with high- and low-level consistency,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2019.

- [103] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le, “Self-training with noisy student improves imagenet classification,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [104] Y. Zhu, Z. Zhang, C. Wu, Z. Zhang, T. He, H. Zhang, R. Manmatha, M. Li, and A. Smola, “Improving semantic segmentation via efficient self-training,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 3, pp. 1589–1602, 2024.
- [105] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in *Proceedings of the 37th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, I. I. I. Hal Daumé and A. Singh, Eds., vol. 119. PMLR, 2020, pp. 1597–1607.
- [106] L. Rettenberger, M. Schilling, S. Elser, M. Böhlend, and M. Reischl, “Self-supervised learning for annotation efficient biomedical image segmentation,” *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 9, pp. 2519–2528, 2023.
- [107] J. E. van Engelen and H. H. Hoos, “A survey on semi-supervised learning,” *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2019.
- [108] L. Jing and Y. Tian, “Self-supervised visual feature learning with deep neural networks: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 11, pp. 4037–4058, 2019.
- [109] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, “A comprehensive survey on transfer learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2019.
- [110] M. P. Schilling, T. Scherr, F. R. Munke, O. Neumann, M. Schutera, R. Mikut, and M. Reischl, “Automated annotator variability inspection for biomedical image segmentation,” *IEEE Access*, vol. 10, pp. 2753–2765, 2022.

- [111] J. Schuetzke, N. J. Szymanski, and M. Reischl, “Validating neural networks for spectroscopic classification on a universal synthetic dataset,” *npj Computational Materials*, vol. 9, no. 1, 2023.
- [112] H. H. Höfer, A. Orth, S. Schweidler, B. Breitung, J. Aghassi-Hagmann, and M. Reischl, “Quantitative convolutional neural network based multi-phase XRD pattern analysis,” *Current Directions in Biomedical Engineering*, vol. 10, no. 4, pp. 307–310, 2024.
- [113] A. Orth, H. Höfer, A. Nefedov, M. Jalali, C. Wöll, and M. Reischl, “ML-based XPS quantification supported by synthetic dataset generation,” *Current Directions in Biomedical Engineering*, vol. 10, no. 4, pp. 482–485, 2024.
- [114] Z. Gong, P. Zhong, and W. Hu, “Diversity in machine learning,” *IEEE Access*, vol. 7, pp. 64 323–64 350, 2019.
- [115] F. R. Münke, J. Schützke, F. Berens, and M. Reischl, “A review of adaptable conventional image processing pipelines and deep learning on limited datasets,” *Machine Vision and Applications*, vol. 35, no. 2, 2024.
- [116] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, “Domain randomization for transferring deep neural networks from simulation to the real world,” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 23–30.
- [117] H. He, Y. Bai, E. A. Garcia, and S. Li, “ADASYN: Adaptive synthetic sampling approach for imbalanced learning,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 1322–1328.
- [118] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Boochoon, and S. Birchfield, “Training deep networks with synthetic data: Bridging the reality gap by domain randomization,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018.

- [119] S. Bäuerle, M. Böhlend, J. Barth, M. Reischl, A. Steimer, and R. Mikut, “CAD-to-real: enabling deep neural networks for 3D pose estimation of electronic control units: A transferable and automated approach for industrial use cases,” *at - Automatisierungstechnik*, vol. 69, no. 10, pp. 880–891, 2021.
- [120] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, “Domain generalization: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–20, 2022.
- [121] A. Westerski and F. W. Teck, “Synthetic data for object detection with neural networks: State of the art survey of domain randomisation techniques,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 21, no. 1, 2023.
- [122] X. Li, X. Wang, F. Deng, and F.-Y. Wang, “Scenarios engineering for trustworthy ai: Domain adaptation approach for reidentification with synthetic data,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 54, no. 11, pp. 6901–6910, 2024.
- [123] S. Dakshit and B. Prabhakaran, “CVAE-based generator for variable length synthetic ECG,” in *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI)*, 2023, pp. 235–244.
- [124] B. Ninon and S. David, *Biomedical Image Synthesis and Simulation : Methods and Applications.*, ser. The MICCAI Society Book Series. London, UK: Academic Press, 2022.
- [125] S. I. Nikolenko, *Synthetic Data for Deep Learning*, ser. Springer Optimization and Its Applications. Cham, Switzerland: Springer, 2021.
- [126] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, “Playing for data: Ground truth from computer games,” in *Lecture Notes in Computer Science*. Cham, Switzerland: Springer, 2016, pp. 102–118.

- [127] D. Svoboda, M. Kozubek, and S. Stejskal, “Generation of digital phantoms of cell nuclei and simulation of image formation in 3D image cytometry,” *Cytometry Part A*, vol. 75A, no. 6, pp. 494–509, 2009.
- [128] D. Wiesner, D. Svoboda, M. Maška, and M. Kozubek, “CytoPacq: a web-interface for simulating multi-dimensional cell imaging,” *Bioinformatics*, vol. 35, no. 21, pp. 4531–4533, 2019.
- [129] D. V. Sorokin, I. Peterlik, V. Ulman, D. Svoboda, T. Necasova, K. Morgaenko, L. Eiselleova, L. Tesarova, and M. Maska, “FiloGen: A model-based generator of synthetic 3-D time-lapse sequences of single motile cells with growing and branching filopodia,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2630–2641, 2018.
- [130] D. Svoboda and V. Ulman, “MitoGen: A framework for generating 3D synthetic time-lapse sequences of cell populations in fluorescence microscopy,” *IEEE Transactions on Medical Imaging*, vol. 36, no. 1, pp. 310–321, 2017.
- [131] J. Stegmaier, J. Arz, B. Schott, J. C. Otte, A. Kobitski, G. U. Nienhaus, U. Strahle, P. Sanders, and R. Mikut, “Generating semi-synthetic validation benchmarks for embryomics,” in *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016.
- [132] M. Weigert, K. Subramanian, S. T. Bundschuh, E. W. Myers, and M. Kreyling, “Biobeam—multiplexed wave-optical simulations of light-sheet microscopy,” *PLOS Computational Biology*, vol. 14, no. 4, p. e1006079, 2018.
- [133] S. Dmitrieff and F. Nédélec, “ConfocalGN: A minimalistic confocal image generator,” *SoftwareX*, vol. 6, pp. 243–247, 2017.
- [134] A. Lehmussola, P. Ruusuvaori, J. Selinummi, H. Huttunen, and O. Yli-Harja, “Computational framework for simulating fluorescence microscope images with cell populations,” *IEEE Transactions on Medical Imaging*, vol. 26, no. 7, pp. 1010–1016, 2007.

- [135] M. Scalbert, F. Couzinie-Devy, and R. Fezzani, “Generic isolated cell image generator,” *Cytometry Part A*, vol. 95, no. 11, pp. 1198–1206, 2019.
- [136] D. Svoboda and V. Ulman, “Generation of synthetic image datasets for time-lapse fluorescence microscopy,” in *Image Analysis and Recognition*. Berlin, Heidelberg, Germany: Springer, 2012, pp. 473–482.
- [137] T. Ursell and K. C. Huang, “BlurLab - 3D microscopy simulation package,” SimTK, 2016, <https://simtk.org/projects/blurlab> (Last Visited: 02/2025).
- [138] V. Ulman, D. Svoboda, M. Nykter, M. Kozubek, and P. Ruusuvuori, “Virtual cell imaging: A review on simulation methods employed in image cytometry,” *Cytometry Part A*, vol. 89, no. 12, pp. 1057–1072, 2016.
- [139] C. Fu, S. Lee, D. Joon Ho, S. Han, P. Salama, K. W. Dunn, and E. J. Delp, “Three dimensional fluorescence microscopy image synthesis and segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018, pp. 2221–2229.
- [140] K. W. Dunn, C. Fu, D. J. Ho, S. Lee, S. Han, P. Salama, and E. J. Delp, “DeepSynth: Three-dimensional nuclear segmentation of biological images using neural networks trained with synthetic data,” *Scientific Reports*, vol. 9, no. 1, 2019.
- [141] S. Han, S. Lee, A. Chen, C. Yang, P. Salama, K. W. Dunn, and E. J. Delp, “Three dimensional nuclei segmentation and classification of fluorescence microscopy images,” in *Proceedings of the IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020.
- [142] A. Bansal, S. Ma, D. Ramanan, and Y. Sheikh, “Recycle-GAN: Unsupervised video retargeting,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 119–135.
- [143] D. Eschweiler, T. Klose, F. N. Müller-Fouarge, M. Kopaczka, and J. Stegmaier, “Towards annotation-free segmentation of fluorescently labeled cell membranes in confocal microscopy images,” in *Simulation and Synthesis*



- in *Medical Imaging*, N. Burgos, A. Gooya, and D. Svoboda, Eds. Cham, Switzerland: Springer, 2019, pp. 81–89.
- [144] D. Eschweiler, M. Rethwisch, M. Jarchow, S. Koppers, and J. Stegmaier, “3D fluorescence microscopy data synthesis for segmentation and benchmarking,” *PLOS ONE*, vol. 16, no. 12, p. e0260509, 2021.
  - [145] D. Bahr, D. Eschweiler, A. Bhattacharyya, D. Moreno-Andres, W. Antonin, and J. Stegmaier, “CellCycleGAN: Spatiotemporal microscopy image synthesis of cell populations using statistical shape models and conditional gans,” in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021.
  - [146] P. Goldsborough, N. Pawlowski, J. C. Caicedo, S. Singh, and A. E. Carpenter, “CytoGAN: Generative modeling of cell images,” *bioRxiv*, 2017.
  - [147] A. Osokin, A. Chessel, R. E. Carazo Salas, and F. Vaggi, “GANs for biological image synthesis,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2233–2242.
  - [148] K. Yao, K. Huang, J. Sun, and C. Jude, “AD-GAN: End-to-end unsupervised nuclei segmentation with aligned disentangling training,” *arXiv preprint arXiv:2107.11022*, 2021.
  - [149] M. Böhlend, **R. Bruch**, K. Löffler, and M. Reischl, “Unsupervised GAN epoch selection for biomedical data synthesis,” *Current Directions in Biomedical Engineering*, vol. 9, no. 1, pp. 467–470, 2023.
  - [150] M. Böhlend, T. Scherr, A. Bartschat, R. Mikut, and M. Reischl, “Influence of synthetic label image object properties on GAN supported segmentation pipelines,” in *Proceedings 29th Workshop Computational Intelligence*, 2019, pp. 289–305.
  - [151] D. Wiesner, T. Nečasová, and D. Svoboda, “On generative modeling of cell shape using 3D GANs,” in *Image Analysis and Processing – ICIAP 2019*, E. Ricci, S. Rota Bulò, C. Snoek, O. Lanz, S. Messelodi, and N. Sebe, Eds. Cham, Switzerland: Springer, 2019, pp. 672–682.

- [152] A. Chen, L. Wu, S. Han, P. Salama, K. W. Dunn, and E. J. Delp, “Three dimensional synthetic non-ellipsoidal nuclei volume generation using Bézier curves,” in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021.
- [153] W. Wu, Y. Zhao, M. Z. Shou, H. Zhou, and C. Shen, “Diffumask: Synthesizing images with pixel-level annotations for semantic segmentation using diffusion models,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 1206–1217.
- [154] D. Eschweiler, R. Yilmaz, M. Baumann, I. Laube, R. Roy, A. Jose, D. Brückner, and J. Stegmaier, “Denoising diffusion probabilistic models for generation of realistic fully-annotated microscopy image datasets,” *PLOS Computational Biology*, vol. 20, no. 2, p. e1011890, 2024.
- [155] M. U. Akbar, W. Wang, and A. Eklund, “Beware of diffusion models for synthesizing medical images - a comparison with GANs in terms of memorizing brain MRI and chest x-ray images,” *Machine Learning: Science and Technology*, vol. 6, no. 1, p. 015022, 2024.
- [156] H. Sahak, D. Watson, C. Saharia, and D. Fleet, “Denoising diffusion probabilistic models for robust image super-resolution in the wild,” *arXiv preprint arXiv:2302.07864*, 2023.
- [157] A. Tsirikoglou, G. Eilertsen, and J. Unger, “A survey of image synthesis methods for visual machine learning,” *Computer Graphics Forum*, vol. 39, no. 6, pp. 426–451, 2020.
- [158] M. Chen, H. Xue, and D. Cai, “Domain adaptation for semantic segmentation with maximum squares loss,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.
- [159] J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. Efros, and T. Darrell, “CyCADA: Cycle-consistent adversarial domain adaptation,” in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 2018, pp. 1989–1998.

- [160] L. Hoyer, D. Dai, and L. Van Gool, “Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 9924–9935.
- [161] S. R. Richter, H. A. Alhaija, and V. Koltun, “Enhancing photorealism enhancement,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 1700–1715, 2023.
- [162] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [163] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, “Learning from simulated and unsupervised images through adversarial training,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [164] M. Sela, P. Xu, J. He, V. Navalpakkam, and D. Lagun, “GazeGAN - unpaired adversarial image generation for gaze estimation,” *arXiv preprint arXiv:1711.09767*, 2017.
- [165] J. R. Astley, A. M. Biancardi, H. Marshall, P. J. C. Hughes, G. J. Collier, M. Q. Hatton, J. M. Wild, and B. A. Tahir, “A hybrid model- and deep learning-based framework for functional lung image synthesis from multi-inflation ct and hyperpolarized gas mri,” *Medical Physics*, vol. 50, no. 9, pp. 5657–5670, 2023.
- [166] W. Li, Y. Li, W. Qin, X. Liang, J. Xu, J. Xiong, and Y. Xie, “Magnetic resonance image (MRI) synthesis from brain computed tomography (CT) images based on deep learning methods for magnetic resonance (MR)-guided radiotherapy,” *Quantitative Imaging in Medicine and Surgery*, vol. 10, no. 6, pp. 1223–1236, 2020.

- [167] J. M. Wolterink, A. M. Dinkla, M. H. F. Savenije, P. R. Seevinck, C. A. T. van den Berg, and I. Išgum, “Deep mr to ct synthesis using unpaired data,” in *Lecture Notes in Computer Science*. Cham, Switzerland: Springer, 2017, pp. 14–23.
- [168] Z. Zhang, L. Yang, and Y. Zheng, “Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [169] X. Zhang, A. Sisniega, W. B. Zbijewski, J. Lee, C. K. Jones, P. Wu, R. Han, A. Uneri, P. Vagdargi, P. A. Helm *et al.*, “Combining physics-based models with deep learning image synthesis and uncertainty in intraoperative cone-beam CT of the brain,” *Medical Physics*, vol. 50, no. 5, pp. 2607–2624, 2023.
- [170] B. Bai, X. Yang, Y. Li, Y. Zhang, N. Pillar, and A. Ozcan, “Deep learning-enabled virtual histological staining of biological samples,” *Light: Science & Applications*, vol. 12, no. 1, 2023.
- [171] E. M. Christiansen, S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, S. Lipnick, E. Mount, A. O’Neil, K. Shah, A. K. Lee *et al.*, “In silico labeling: Predicting fluorescent labels in unlabeled images,” *Cell*, vol. 173, no. 3, pp. 792–803.e19, 2018.
- [172] Y. Liu, H. Yuan, Z. Wang, and S. Ji, “Global pixel transformers for virtual staining of microscopy images,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 2256–2266, 2020.
- [173] C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson, “Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy,” *Nature Methods*, vol. 15, no. 11, pp. 917–920, 2018.
- [174] X. Chen, M. E. Kandel, S. He, C. Hu, Y. J. Lee, K. Sullivan, G. Tracy, H. J. Chung, H. J. Kong, M. Anastasio *et al.*, “Artificial confocal microscopy for

- deep label-free imaging,” *Nature Photonics*, vol. 17, no. 3, pp. 250–258, 2023.
- [175] S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli, “Scientific machine learning through physics–informed neural networks: Where we are and what’s next,” *Journal of Scientific Computing*, vol. 92, no. 3, 2022.
- [176] **R. Bruch**, F. Keller, M. Böhländ, M. Vitacolonna, L. Klinger, R. Rudolf, and M. Reischl, “Synthesis of large scale 3D microscopic images of 3D cell cultures for training and benchmarking,” *PLOS ONE*, vol. 18, no. 3, p. e0283828, 2023.
- [177] **R. Bruch**, M. Vitacolonna, E. Nürnberg, S. Sauer, R. Rudolf, and M. Reischl, “Improving 3D deep learning segmentation with biophysically motivated cell synthesis,” *Communications Biology*, vol. 8, no. 1, 2025.
- [178] M. Böhländ, **R. Bruch**, S. Bäuerle, L. Rettenberger, and M. Reischl, “Improving generative adversarial networks for patch-based unpaired image-to-image translation,” *IEEE Access*, vol. 11, pp. 127 895–127 906, 2023.
- [179] A. Desmaison, L. Guillaume, S. Triclin, P. Weiss, B. Ducommun, and V. Lobjois, “Impact of physical confinement on nuclei geometry and cell division dynamics in 3D spheroids,” *Scientific Reports*, vol. 8, no. 1, 2018.
- [180] E. Nürnberg, M. Vitacolonna, **R. Bruch**, M. Reischl, R. Rudolf, and S. Sauer, “From in vitro to in silico: a pipeline for generating virtual tissue simulations from real image data,” *Frontiers in Molecular Biosciences*, vol. 11, 2024.
- [181] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2242–2251.
- [182] L. Wu, A. Chen, P. Salama, S. Winfree, K. W. Dunn, and E. J. Delp, “NIS-Net3D: three-dimensional nuclear synthesis and instance segmentation for fluorescence microscopy images,” *Scientific Reports*, vol. 13, no. 1, 2023.

- [183] A. Borji, “Pros and cons of GAN evaluation measures,” *Computer Vision and Image Understanding*, vol. 179, pp. 41–65, 2019.
- [184] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training GANs,” in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Curran Associates, Inc., 2016.
- [185] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “GANs trained by a two time-scale update rule converge to a local nash equilibrium,” in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.
- [186] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, “Demystifying MMD GANs,” *arXiv preprint arXiv:1801.01401*, 2018.
- [187] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [188] **R. Bruch**, M. Vitacolonna, R. Rudolf, and M. Reischl, “Prediction of fluorescent Ki67 staining in 3D tumor spheroids,” *Current Directions in Biomedical Engineering*, vol. 8, no. 2, pp. 305–308, 2022.
- [189] Z. Zhang, Q. Liu, and Y. Wang, “Road extraction by deep residual U-Net,” *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.
- [190] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [191] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

- [192] N. I. George, T.-P. Lu, and C.-W. Chang, “Cost-sensitive performance metric for comparing multiple ordinal classifiers,” *Artificial Intelligence Research*, vol. 5, no. 1, pp. 135–143, 2016.
- [193] W. Nagl, “Die Kernstruktur während des mitotischen und endomitotischen Zellzyklus,” *Berichte der Deutschen Botanischen Gesellschaft*, vol. 81, no. 7, pp. 320–324, 1968.
- [194] R. Wang, A. Kamgoue, C. Normand, I. Léger-Silvestre, T. Mangeat, and O. Gadai, “High resolution microscopy reveals the nuclear shape of budding yeast during cell cycle and in various biological states,” *Journal of Cell Science*, vol. 129, no. 24, pp. 4480–4495, 2016.
- [195] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, “PyTorch: An imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019.
- [196] D. Borland, C. M. McCormick, N. K. Patel, O. Krupa, J. T. Mory, A. A. Beltran, T. M. Farah, C. F. Escobar-Tomlienovich, S. S. Olson, M. Kim *et al.*, “Segmentor: a tool for manual refinement of 3D microscopy annotations,” *BMC Bioinformatics*, vol. 22, no. 1, 2021.
- [197] L. Rettenberger, N. J. Szymanski, Y. Zeng, J. Schuetzke, S. Wang, G. Ceder, and M. Reischl, “Uncertainty-aware particle segmentation for electron microscopy at varied length scales,” *npj Computational Materials*, vol. 10, no. 1, 2024.
- [198] M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley *et al.*, “Content-aware image restoration: pushing the limits of fluorescence microscopy,” *Nature Methods*, vol. 15, no. 12, pp. 1090–1097, 2018.

- [199] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds., vol. 27. Curran Associates, Inc., 2014.
- [200] L. Deng, “The MNIST database of handwritten digit images for machine learning research [best of the web],” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.
- [201] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, “Least squares generative adversarial networks,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [202] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” *arXiv preprint arXiv:1511.06434*, 2015.
- [203] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” *arXiv preprint arXiv:1411.1784*, 2014.
- [204] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, “Autoencoding beyond pixels using a learned similarity metric,” in *Proceedings of The 33rd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. F. Balcan and K. Q. Weinberger, Eds., vol. 48. New York, NY: PMLR, 2016, pp. 1558–1566.
- [205] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, “Loss functions for image restoration with neural networks,” *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.
- [206] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, “Learning to discover cross-domain relations with generative adversarial networks,” in *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. PMLR, 2017, pp. 1857–1865.



- [207] M.-Y. Liu, T. Breuel, and J. Kautz, “Unsupervised image-to-image translation networks,” *arXiv preprint arXiv:1703.00848*, 2017.
- [208] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.
- [209] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, “A review on generative adversarial networks: Algorithms, theory, and applications,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3313–3332, 2023.
- [210] Z. Wang, Q. She, and T. E. Ward, “Generative adversarial networks in computer vision: A survey and taxonomy,” *ACM Computing Surveys*, vol. 54, no. 2, pp. 1–38, 2021.
- [211] X. Yi, E. Walia, and P. Babyn, “Generative adversarial network in medical imaging: A review,” *Medical image analysis*, vol. 58, p. 101552, 2019.
- [212] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 6840–6851.
- [213] P. Dhariwal and A. Nichol, “Diffusion models beat GANs on image synthesis,” in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 8780–8794.
- [214] F. Khader, G. Müller-Franzes, S. Tayebi Arasteh, T. Han, C. Haarburger, M. Schulze-Hagen, P. Schad, S. Engelhardt, B. Baeßler, S. Foersch *et al.*, “Denoising diffusion probabilistic models for 3D medical image generation,” *Scientific Reports*, vol. 13, no. 1, 2023.
- [215] M. Usman Akbar, M. Larsson, I. Blystad, and A. Eklund, “Brain tumor segmentation using synthetic mr images - a comparison of GANs and diffusion models,” *Scientific Data*, vol. 11, no. 1, 2024.

- [216] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang, “Diffusion models: A comprehensive survey of methods and applications,” *ACM Computing Surveys*, vol. 56, no. 4, pp. 1–39, 2023.
- [217] N. Carlini, J. Hayes, M. Nasr, M. Jagielski, V. Sehwag, F. Tramèr, B. Balke, D. Ippolito, and E. Wallace, “Extracting training data from diffusion models,” in *32nd USENIX Security Symposium (USENIX Security 23)*. Anaheim, CA: USENIX Association, 2023, pp. 5253–5270.
- [218] G. Somepalli, V. Singla, M. Goldblum, J. Geiping, and T. Goldstein, “Diffusion art or digital forgery? investigating data replication in diffusion models,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 6048–6058.
- [219] J. Markl, D. Sadava, D. M. Hillis, H. C. Heller, and S. D. Hacker, *Purves Biologie*. Berlin, Heidelberg, Germany: Springer Spektrum, 2019, ch. Zellzyklus und Zellteilung, pp. 307–344.