Universidade Estadual de Campinas
Instituto de Computação

Guilherme Vieira Leite

Semantic Segmentation of Spheroid Cell Cultures of Cancer

Segmentação Semântica de Culturas Celulares Esferoides de Câncer

CAMPINAS
2025

Guilherme Vieira Leite

Semantic Segmentation of Spheroid Cell Cultures of Cancer

Segmentação Semântica de Culturas Celulares Esferoides de Câncer

Tese apresentada ao Instituto de Computação da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação.

Thesis presented to the Institute of Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Computer Science.

Supervisor/Orientador: Prof. Dr. Hélio Pedrini
Co-supervisor/Coorientadora: Profa. Dra. Carmen Veríssima Ferreira-Halder

Este exemplar corresponde à versão final da Tese defendida por Guilherme Vieira Leite e orientada pelo Prof. Dr. Hélio Pedrini.

CAMPINAS
2025

Ficha catalográfica
Universidade Estadual de Campinas (UNICAMP)
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

L536s Leite, Guilherme Vieira, 1991-
Semantic segmentation of spheroid cell cultures of cancer / Guilherme Vieira Leite. – Campinas, SP : [s.n.], 2025.

Orientador: Hélio Pedrini.
Coorientador: Carmen Veríssima Ferreira-Halder.
Tese (doutorado) – Universidade Estadual de Campinas (UNICAMP), Instituto de Computação.

1. Esferoides celulares. 2. Câncer. 3. Aprendizado profundo. 4. Segmentação semântica. 5. Redes neurais convolucionais. 6. Transformers (Arquitetura de computador). I. Pedrini, Hélio, 1963-. II. Ferreira, Carmen Veríssima, 1969-. III. Universidade Estadual de Campinas (UNICAMP). Instituto de Computação. IV. Título.

Informações complementares

Título em outro idioma: Segmentação semântica de culturas celulares esferoides de câncer
Palavras-chave em inglês:
Spheroids cellular
Cancer
Deep learning
Semantic segmentation
Convolutional neural networks
Transformers (Computer architecture)
Área de concentração: Ciência da Computação
Titulação: Doutor em Ciência da Computação
Banca examinadora:
Hélio Pedrini [Orientador]
Sheila Siqueira Andrade
Edimilson Batista dos Santos
Gabriel Cirac Mendes Souza
José Augusto Saraiva Lustosa Filho
Data de defesa: 10-02-2025
Programa de Pós-Graduação: Ciência da Computação

Objetivos de Desenvolvimento Sustentável (ODS)
ODS: 3. Saúde e bem-estar

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-6871-2069
- Currículo Lattes do autor: http://lattes.cnpq.br/2927269217627994

Universidade Estadual de Campinas
Instituto de Computação

Guilherme Vieira Leite

Semantic Segmentation of Spheroid Cell Cultures of Cancer

Segmentação Semântica de Culturas Celulares Esferoides de Câncer

Banca Examinadora:

- Prof. Dr. Hélio Pedrini
  Instituto de Computação - UNICAMP
- Dra. Sheila Siqueira Andrade
  PlateInnove Biotechnology
- Prof. Dr. Edimilson Batista dos Santos
  Departamento de Computação - UFSJ
- Prof. Dr. Gabriel Cirac Mendes Souza
  FEPI
- Dr. José Augusto Saraiva Lustosa Filho
  Instituto de Computação - UNICAMP

A ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Campinas, 10 de fevereiro de 2025

Agradecimentos

Gostaria de expressar meus sinceros agradecimentos aos meus orientadores. Desde 2018, tenho tido o privilégio de trabalhar sob a orientação do professor Hélio, cuja influência foi fundamental para o meu desenvolvimento como cientista. Sua abordagem permite ampla liberdade aos alunos, garantindo autonomia nos momentos em que era necessário trabalhar de forma independente, ao mesmo tempo em que sempre esteve disponível para orientação e acompanhamento em qualquer dia e horário, quando necessário. À professora Carmen, expresso minha profunda gratidão por aceitar a coorientação e por acreditar neste projeto. Seu compromisso com minha formação foi notável, incentivando minha imersão no objeto de estudo e garantindo não apenas o acesso a materiais e ao laboratório, mas também proporcionando, pessoalmente, minha formação teórica e prática na área de esferoides. Agradeço ainda aos integrantes do laboratório OncoBiomarkers, com destaque aos técnicos Kiko e Cláudia, pelo suporte contínuo e dedicação. O desenvolvimento da ferramenta SAP foi um desafio significativo e, por isso, sou especialmente grato à Beatriz, coautora da ferramenta. Também registro meus agradecimentos aos técnicos do Instituto de Computação, Wilson, William e Daniel, cujo suporte foi essencial para o funcionamento dos equipamentos do laboratório. Sua dedicação ao serviço público, aliada ao profissionalismo e à receptividade, sempre fez com que qualquer solicitação de auxílio fosse atendida com prontidão, acompanhada de uma ótima conversa e um cafezinho. Reconheço, igualmente, a importância dos colegas do laboratório LIV e do Taekwondo, que estiveram presentes ao longo desses anos. Foram muitos momentos de desabafo, mas também discussões que contribuíram significativamente para o avanço deste trabalho, evidenciando a natureza colaborativa da construção do conhecimento. À minha família, expresso minha imensa gratidão. Pelo seu apoio inestimável, tanto para minha formação como indivíduo quanto ao longo de toda a minha trajetória acadêmica. Por fim, e não menos importante, dirijo meus agradecimentos à Jordana, que acompanhou de perto todo esse percurso. Foi por meio dela que conheci a professora Carmen, além de ter recebido inúmeras palestras e esclarecimentos sobre biologia ao longo dos anos. Nossa parceria acadêmica se concretizou em publicações conjuntas, e sua influência em minha trajetória é inegável. Desde o ensino médio, foi ela quem me incentivou a seguir a carreira acadêmica e me apoiou em cada etapa desse processo, nos momentos de tranquilidade e também nos de maior ansiedade. Este estudo foi parcialmente financiado pelo Banco Santander via edital 41/2022 e pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Código de Financiamento 140327/2021-5.

Resumo

Culturas celulares em esferoides são modelos valiosos para o desenvolvimento de fármacos. Entretanto, sua análise ainda demanda muito tempo, tornando essencial a automação para triagens em larga escala. Para enfrentar esse desafio, apresentamos um banco de dados abrangente de imagens de esferoides, acompanhado de protocolos e amostras anotadas, além de um método inovador de segmentação e um pipeline escalável para análises rotineiras de esferoides em larga escala. O banco de dados é composto por imagens em campo claro capturadas em intervalos de 24 horas ao longo do ciclo de vida dos esferoides, geradas no OncoBiomarkers Lab da Universidade Estadual de Campinas (Unicamp). Nossa abordagem de segmentação integra *data augmentation* com arquiteturas de redes neurais convolucionais (CNNs) e Transformers de última geração. Além disso, introduzimos um novo método de *ensemble*, utilizando *late-stage majority vote* de modelos treinados, para melhorar o desempenho na segmentação. A avaliação quantitativa, utilizando o índice Dice como métrica, revelou que arquiteturas baseadas em Transformers se destacam na extração de características essenciais para a segmentação de esferoides, mesmo em cenários com dados limitados. Adicionalmente, o *ensemble* de modelos CNN apresentou desempenho superior em comparação com redes individuais. A análise qualitativa dos resultados de segmentação evidenciou que as métricas quantitativas nem sempre se correlacionam diretamente com a qualidade visual das segmentações. Em conclusão, a segmentação de esferoides cancerígenos continua sendo um desafio e um gargalo em experimentos de alta capacidade. A metodologia proposta demonstrou ser eficaz na geração de máscaras de segmentação confiáveis e, quando integrada ao *Spheroid Analysis Pipeline* (SAP), oferece uma solução de ponta a ponta para testes em larga escala e resultados reprodutíveis. Todos os recursos necessários para a reprodução dos resultados, incluindo a base de dados, os protocolos, os métodos e o manual correspondente, estão publicamente disponíveis para apoiar pesquisas futuras e a extensão deste trabalho.

Abstract

Spheroid cell cultures serve as valuable models for drug development. However, their examination remains time-intensive, necessitating automation for high-throughput screening. To address this challenge, we present a comprehensive dataset of spheroid images, complete with protocols and annotated samples, alongside a novel segmentation method and a scalable pipeline for routine large-scale spheroid analysis. The dataset comprises brightfield images captured at 24-hour intervals throughout the lifespan of spheroids, generated at the Oncobiomarkers Research Laboratory, University of Campinas (Unicamp). Our segmentation approach integrates data augmentation with state-of-the-art convolutional neural networks (CNNs) and Transformer architectures. Furthermore, we introduce a novel ensemble method employing late-stage majority vote fusion of trained models to enhance segmentation performance. Quantitative evaluation, using the Dice index as the benchmark metric, revealed that Transformer-based architectures excel in extracting critical features for spheroid segmentation, even in data-scarce scenarios. Additionally, the ensemble of CNN models demonstrated superior performance compared to individual networks. Qualitative analysis of segmentation outputs highlighted that quantitative metrics alone do not always correlate with visual segmentation quality. In conclusion, cancer spheroid segmentation remains a significant challenge and bottleneck in high-throughput experimentation. Our proposed methodology generates reliable segmentation masks and, when integrated with the Spheroid Analysis Pipeline (SAP), provides an end-to-end solution for large-scale testing and reproducible results. All necessary resources for reproducing the results, including the dataset, protocols, methods, and accompanying manual, are publicly available to support future research and extensions of this work.

List of Figures

1.1 Drug Development Flowchart 14
1.2 2D vs 3D Cell Organization 15
1.3 Spheroid Culture Examples 15

2.1 Classification vs Semantic segmentation 21
2.2 Illustration of Segmentation Masks 22
2.3 Watershed Algorithm Diagram 23
2.4 Illustration of K-means Segmentation 25
2.5 Illustration of 2D Convolution 27
2.6 U-Net Architecture 28
2.7 MultiResUNet Architecture 29
2.8 MultiResUNet Residual Block 30
2.9 MultiResUNet Residual Path 31
2.10 UNet++ Architecture 31
2.11 UNeXt Architecture 32
2.12 Transformer Architecture 34
2.13 TransUNet Architecture 35
2.14 Swin Transformer Architecture 36
2.15 Swin Transformer Block 38
2.16 Swin-UNet Architecture 39
2.17 Dice Score Diagram 41

3.1 Segmentation studies per year 43
3.2 PRISMA diagram 44
3.3 Segmentation method distribution 44

4.1 Our dataset 51
4.2 Spheroid Formation 52
4.3 Annotation software 54
4.4 Lacalle et al. [80]'s dataset samples 55

5.1 Segmentation Method Overview 57

6.1 Protocol 1's Inferences 63
6.2 Protocol 2's Inferences 64
6.3 Protocol 3's Inferences 65
6.4 Protocol 4's Inferences 67
6.5 Protocol 5's Inferences 68
6.6 Protocol 6's Inferences 69
6.7 Ensemble Smoothing Effect 71

7.1 SAP's Datasets Samples 77
7.2 Spheroid Analysis Pipeline (SAP) Overview 77
7.3 SAP Segmentation Overview 78
7.4 SAP Segmentation Illustration 79
7.5 SAP Statistical Analysis Overview 79
7.6 Dataframe boxplot 80

List of Tables

3.1 Spheroid datasets in the literature . . . 45
3.2 Studies using image processing . . . 46
3.3 Studies using software . . . 48
3.4 Studies using deep learning . . . 49

4.1 Our dataset summary . . . 51
4.2 Lumaview Parameters . . . 53
4.3 Lacalle et al. [80]'s dataset . . . 54

6.1 Protocols Summary . . . 62
6.2 Quantitative P1, P2, and P3 . . . 62
6.3 Quantitative P4, P5, and P6 . . . 66
6.4 Ensemble Results . . . 70
6.5 Ablation TransUNet . . . 72
6.6 Ablation Swin-UNet . . . 73
6.7 Ablation U-NetAtt . . . 73

7.1 SAP's Datasets Summary . . . 77

Contents

1 Introduction **13**

1.1 Motivation . . . 13

1.2 Problem Definition . . . 15

1.3 Objectives . . . 16

1.4 Research Questions . . . 16

1.5 Contributions . . . 17

1.6 Publications . . . 18

1.7 Text Organization . . . 18

2 Related Concepts **19**

2.1 Cancer Spheroid Cultures . . . 19

2.2 Bio-Image Capture . . . 19

2.2.1 Phase Contrast . . . 20

2.2.2 Brightfield . . . 20

2.2.3 Fluorescent . . . 20

2.2.4 Confocal . . . 20

2.3 Semantic Segmentation . . . 21

2.4 Machine Learning and Artificial Intelligence . . . 24

2.5 Deep Learning . . . 25

2.6 Convolutional Neural Networks . . . 26

2.7 Convolutions . . . 27

2.7.1 U-Net . . . 27

2.7.2 MultiResUNet . . . 29

2.7.3 U-Net++ . . . 29

2.8 UNeXt . . . 31

2.9 Transformers . . . 33

2.9.1 TransUNet . . . 35

2.9.2 Swin Transformer . . . 36

2.9.3 Swin-UNet . . . 38

2.10 Bias . . . 40

2.11 Evaluation and Metrics . . . 40

2.11.1 Loss Function . . . 40

2.11.2 Metrics . . . 41

3 Related Work **42**

3.1 Literature Sorting . . . 43

3.2 Public Datasets . . . 43

3.3 Found Literature . . . 45

3.3.1 Image Processing 45
3.3.2 Specific Software 47
3.3.3 Deep Learning 47

4 Materials 50
4.1 Our Dataset 50
4.1.1 Cancer Cell Culture 51
4.1.2 Image Capture 52
4.1.3 Mask Annotation 53
4.2 Lacalle's Dataset 53
4.3 Computational Resources 55

5 Cancer Spheroid Segmentation Method 56
5.1 Pre-Processing Phase 58
5.1.1 Random Brightness and Contrast 58
5.1.2 Blur 58
5.1.3 Rotation 58
5.1.4 Perspective 59
5.1.5 Optical Distortion 59
5.1.6 Piecewise Affine 59
5.1.7 Horizontal and Vertical Flip 59
5.1.8 Random Crop 59
5.1.9 Patching 59
5.2 Ensemble 60

6 Results 61
6.1 Protocols 1, 2, and 3 62
6.2 Protocols 4, 5, and 6 66
6.3 Ensemble 70
6.4 Ablation 71
6.4.1 Transformer Parameters 72
6.4.2 U-NetAtt 73
6.5 Discussion 74

7 A Quantitative Spheroid Analysis Pipeline (SAP) 75
7.1 Materials 75
7.1.1 Cancer Cell Culture 75
7.1.2 Datasets 76
7.2 Spheroid Analysis Pipeline (SAP) 77
7.2.1 Segmentation Module 78
7.2.2 Dataframe 78
7.2.3 Statistical Analysis Module 78
7.2.4 Data Visualization Module 80
7.3 Observations 80

8 Conclusions 82
8.1 Final Considerations 82
8.2 Limitations and Directions for Future Work 86

Chapter 1

Introduction

In this chapter, we present an overview of the thesis by outlining its motivation, defining the problem under investigation, and establishing the primary objectives. We then highlight the key contributions made through this research, followed by a summary of related publications that have resulted from this work. Finally, we provide an outline of the organization of this thesis, detailing the structure and content of each chapter to guide the reader through the document.

1.1 Motivation

Cancer continues to pose a significant global health challenge, responsible for nearly 10 million deaths worldwide in 2020 and approximately 19 million new cases reported that year [142]. The development of new drugs, including cancer treatments, is a complex, costly, and time-consuming process. According to Wouters et al. [165] and Hughes et al. [64], between 2014 and 2018, the FDA approved 355 new drugs in the United States, yet cost data was publicly available for only 63 of them. Their analysis estimated that, factoring in failed trials, developing a single drug costs around US$1 billion. They also emphasized that gathering sufficient evidence to justify a screening trial often requires 12 to 15 years of research.

Bringing a drug to market involves four key steps, starting with Research & Development (R&D), which is often the most time-intensive step. This step focuses on understanding the disease and identifying potential drug targets. For example, if an enzyme is found to be overexpressed in cancer and linked to tumor progression, researchers explore strategies to discover new compounds or modify existing ones. Once a target is validated, the compounds move to the next phase for biological testing [43].

The second step, the Preclinical phase, is divided into two sub-phases, illustrated in the top row of Figure 1.1. The first, called *in vitro*, involves testing the compound on human cell cultures to assess effectiveness, optimal dosage, and toxicity. The second, *in vivo*, tests the compounds on animals [64].

Compounds that successfully complete preclinical testing advance to human trials, conducted in three progressive clinical phases. These trials, which expand in scale over time, are critical for assessing the drug's behavior in the human body and ensuring patient safety, as depicted in the bottom row of Figure 1.1. Once sufficient evidence of a drug's safety and efficacy is collected, researchers proceed to the final step: submitting the drug for regulatory review and approval.

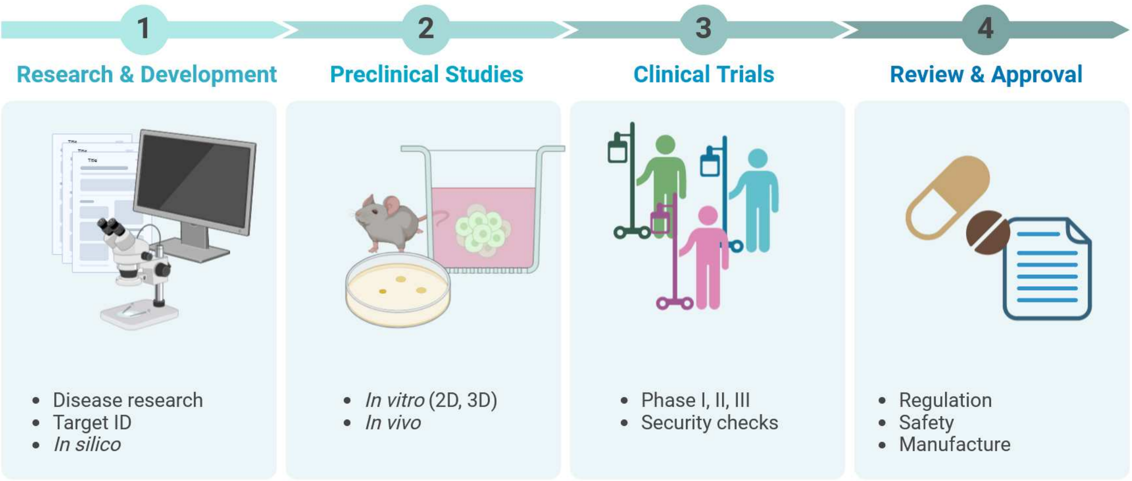


Figure 1.1: A flowchart illustrating the stages of drug development: (1) The *in silico* phase, where computer simulations are used to evaluate the compound. (2) The *in vitro* phase, involving tests on cultured cells. (3) The *in vivo* phase, where the compound is tested in living animals. (4) The four phases of clinical trials, during which the compound is assessed in human participants.

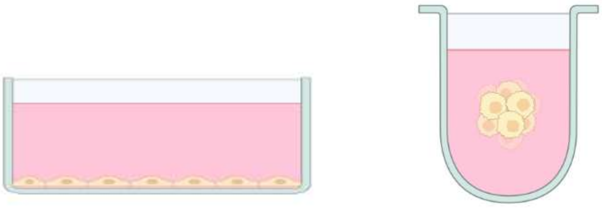
Clinical trials are the most expensive phase of drug development, with only 10% of candidates ultimately reaching the market. This high rejection rate is often linked to limitations in preclinical testing, which may fail to filter out unviable compounds early [41]. To address this, researchers have focused on improving preclinical studies, particularly by advancing *in vitro* models and reducing reliance on animal testing.

A very common *in vitro* approach is the two-dimensional (2D) cell culture, in which cells grow in a monolayer on a plastic or glass surface, as illustrated in Figure 1.2a. This method has several advantages, including reduced ethical concerns compared to animal models, simpler training requirements, and the availability of highly consistent cell batches for reproducibility. Additionally, cells can be frozen for long-term storage, unlike live animals, which age rapidly. Maintaining 2D cultures is also more cost-effective than operating animal facilities [135].

Despite these benefits and their critical role in research [71, 117], 2D cell cultures have notable limitations. Their main disadvantage is the lack of physiological relevance, as they fail to replicate the complex cell-to-cell interactions and microenvironment dynamics found in living tissues. This is particularly important in cancer research, where the tumor microenvironment plays a crucial role in disease progression.

Other limitations of 2D cultures include altered gene expression, increased sensitivity to compounds, absence of systemic effects, and poor predictive accuracy for drug efficacy. Essential factors such as nutrient diffusion and waste removal, which are critical for tumor growth, are also poorly represented in these models [46, 117].

To overcome these challenges, researchers have developed 3D culture models, including spheroids and organoids. In spheroid models, cells are cultured in suspension within

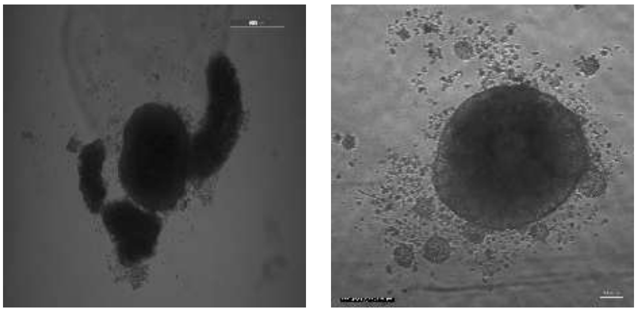


(a) 2D Culture (b) 3D Culture

Figure 1.2: A cross-sectional illustration of cell organization in different culture types: (a) The flat, monolayer structure of a 2D culture, with cells adhering to the bottom of the plate. (b) The spherical arrangement of a three-dimensional (3D) culture, where cells are suspended and form a cohesive spheroid shape.

a nutrient-rich medium, either with or without scaffolds to support cellular organization, as shown in Figures 1.2b and 1.3. These models better mimic the physiological microenvironment by reproducing key *in vivo* phenomena, such as intercellular communication, cell-stroma interactions, and differentiation.

As a result, 3D cultures offer greater accuracy in drug screening compared to 2D cultures, contributing to improving the predictive value of *in vitro* assays. They also provide higher throughput than *in vivo* models, are more automation-friendly, and demonstrate superior reproducibility.



(a) Lacalle et al. [80] (b) Leite et al. [84]

Figure 1.3: Spheroid cultures described in the literature often differ in scale and focus.

1.2 Problem Definition

Drug development is costly [64, 165], and while 3D cultures offer significant advantages over 2D cultures in cancer research [155], high-throughput testing remains a challenge [174]. A key bottleneck is the continuous analysis of spheroids as they grow and respond to tested compounds [121].

One common method involves examining each spheroid under a microscope to assess its various features: whether the cells at the edges are loose or compact, changes in nuclear appearance (darkening or lightening), cell migration from the spheroid, or morphological ones, like changes in the spheroid's area. However, manual analysis of these cultures is time-consuming and prone to errors due to human fatigue and subjectivity.

Automating the process of delineating spheroid boundaries and quantifying their areas could significantly accelerate drug discovery in cancer research. This task, known as semantic segmentation, is a well-established problem in computer vision [148]. In this study, we aim to harness the power of artificial intelligence to meet the needs of laboratory workflows, enabling fast, accurate, and efficient analysis of cancer spheroids.

1.3 Objectives

In this thesis, we seek to lay the material basis for understanding this field and to pave the way for integrating computer vision with cancer spheroid research, above all aiming to automate the spheroid analysis process. To achieve this, we have defined the following objectives:

- O1. Conduct a comprehensive literature review on the application of semantic segmentation to cancer spheroid cultures, explaining terminology, identifying trends, highlighting limitations, and uncovering knowledge gaps.
- O2. Create a publicly accessible dataset of cancer spheroid images, including high-quality annotated masks and relevant metadata, to support future research and promote reproducibility by other researchers.
- O3. To develop a semantic segmentation method based on deep learning architectures that is both versatile and practical for the everyday needs and challenges of cancer spheroid experimentation.
- O4. Implement a practical application of the proposed method within the daily workflow of a biology laboratory focused on cancer spheroid research.

1.4 Research Questions

Here we list the research questions that we came up with to reach our previously stated objectives, these were used to guide our next steps, which in turn generated new questions.

- Q01. Which techniques are mostly used to produce the segmentation results?
- Q02. What is the role of deep learning approaches in these scenarios?
- Q03. How available are the datasets in this field?
- Q04. Are these datasets ready to be used for deep learning training?
- Q05. Which metrics are used for comparison between methods?
- Q06. Are CNNs better suited for this task compared to more complex architectures like Transformers?
- Q07. By combining these two types of architectures, can we improve on segmentation performance?

- Q08. Which architecture type generalized its learning from the few samples we have available?
- Q09. Are the datasets generalizing enough to produce a segmentation on different type of cell culture?
- Q10. What is the minimum necessary on a pipeline to automate spheroid analysis?

1.5 Contributions

This thesis offers three key contributions. First, it introduces a meticulously curated and annotated dataset of cancer spheroid images, serving as a valuable resource for researchers in this field. Second, it proposes a novel deep learning-based method for semantic segmentation, enabling precise and automated analysis of spheroid images. Third, it develops a robust statistical analysis pipeline, facilitating quantitative evaluation and deeper insights into segmented data. Collectively, these contributions advance the state-of-the-art in cancer spheroid image analysis and enables for future research and practical applications.

Our dataset[1] is a critical resource for advancing research of image analysis on cancer spheroid biology, offering high-quality, expertly annotated images of spheroid structures. Each image has undergone careful review to ensure accurate segmentation, supporting reliable and reproducible computational studies. To promote open science and encourage further exploration, the dataset is publicly accessible, enabling researchers worldwide to use and build upon it in their investigations of cancer spheroids and related areas.

Additionally, this work introduces a segmentation method specifically designed to cancer spheroid images, enabling high-throughput, quantitative analysis without human intervention. This method captures the distinct morphology of spheroids, allowing for detailed data extraction across large datasets. Its application has generated significant insights, culminating in scientific publications that validate its effectiveness and impact. By enhancing segmentation techniques, this contribution empowers researchers to perform analyses efficiently, driving progress in cancer research through improved image-based quantification.

The statistical analysis pipeline[2] developed in this thesis provides quantitative insights by integrating the results of spheroid segmentation. It includes tools for conducting t-tests and analysis of variance (ANOVA), enabling comparative analysis across different treatment groups. This ensures robust, reproducible evaluations and helps identify statistically significant differences between experimental conditions. By combining feature extraction with rigorous statistical methods, the pipeline supports in-depth analysis in cancer spheroid studies and related applications.

[1] https://github.com/guilhermevleite/dataset-spheroid-segmentation

[2] https://github.com/guilhermevleite/Spheroid-Analysis-Pipeline

1.6 Publications

Our research has yielded several significant results, which have been submitted to esteemed international journals and conferences in the fields of computing and biology. The following works are presented as either published or currently under review for submission.

- G. V. Leite, J. M. Azevedo-Martins, C. V. Ferreira-Halder, H. Pedrini. *Cancer Spheroid Segmentation Based on Vision Transformer.* IEEE International Conference on Visual Communications and Image Processing, Jeju, South Korea, pp. 1-5, December 2023.

- G. V. Leite, J. M. Azevedo-Martins, C. V. Ferreira-Halder, H. Pedrini. *Semantic Segmentation Techniques for Human-Cancer Spheroid Cell Cultures: A Survey.* Image Analysis and Stereology (Submitted).

- Aires-Lopes, B, G. V. Leite, C. V. Ferreira-Halder, H. Pedrini. *Spheroid Analysis Pipeline (SAP) for Overcoming the Bottleneck of 3D Cell Culture Image Measurement in High Throughput Screening.* Computer Methods and Programs in Biomedicine (Submitted).

- G. V. Leite, J. M. Azevedo-Martins, C. V. Ferreira-Halder, H. Pedrini. *Deep Learning Ensemble Towards Cancer Spheroid Semantic Segmentation.* Computers in Biology and Medicine (Submitted).

1.7 Text Organization

The following chapters are structured to provide a thorough understanding of the research context, methodology, and findings. Chapter 2 introduces foundational background information necessary for interpreting the study's aims and approach. Chapter 3 reviews existing studies and theoretical frameworks relevant to the research, highlighting gaps the thesis aims to address. Chapter 4 describes the specific resources and materials utilized, clarifying the tools and data essential for the proposed method. Chapter 5 outlines the methodology in detail, explaining the procedures and theoretical basis guiding the study's approach. In Chapter 6, the experimental setups and their outcomes are presented, offering insight into the study's findings. Chapter 7 shows a real-world application of the spheroid segmentation towards high-throughput testing in form of a statistical analysis pipeline. Finally, Chapter 8 synthesizes key insights, contributions, and potential directions for future research.

Chapter 2

Related Concepts

In this chapter, we provide essential background information that establishes the foundational knowledge necessary for understanding all the aspects of this thesis. We explain comprehensively the key concepts in image processing, semantic segmentation techniques, and cancer spheroid cultures, enabling readers to fully understand the context and significance of our research. Ultimately, this chapter serves as a starting point for readers to engage with the subsequent discussions on the related literature, materials, method, and experimental results presented later in the work.

2.1 Cancer Spheroid Cultures

Cancer spheroid cultures provide a method for developing cancer models in three-dimensional (3D) space, closely mimicking the morphology and microenvironment of cancer *in vivo*. These 3D models represent a significant improvement over two-dimensional (2D) cultures by replicating key features such as cancer cell-cell interactions, interactions between cancer cells and the surrounding stroma, and cellular density.

One characteristic of spheroid cultures is the differentiation of their layers and the biological processes within them, such as the formation of a hypoxic core. In this layer, cells at the center experience nutrient and oxygen deprivation, while those in the outer layers have abundant access to these resources. This layered structure has a significant impact on the penetration of therapeutic compounds, which can either enhance or hinder their effectiveness.

There are several established methods to form cancer spheroid cultures, each with its advantages and applications depending on the experimental goals, such as liquid overlay, hanging drop, scaffold-based, scaffold-free, spinner flask, microfluidic, and magnetic levitation. For the purposes of this study, the specific method used to generate our spheroids is detailed in Chapter 4.

2.2 Bio-Image Capture

In the context of this thesis, we refer to image capture as the microscopy technique employed to capture the cancer spheroid culture images, and as such, there are various techniques available to do so. Here we describe the ones we came across in the related literature most frequently, while also briefly talking about their functionality, and most importantly for this thesis, their output features.

2.2.1 Phase Contrast

Phase contrast is a microscopy technique designed to enhance the visibility of transparent and colorless samples, such as living cells, by converting subtle differences in light waves into variations in intensity. As light passes through a specimen, differences in refractive index and thickness cause slight phase shifts, which are otherwise undetectable to the human eye. Using specialized optical components, these shifts are transformed into visible contrast, revealing intricate cellular structures without the need for staining [172].

2.2.2 Brightfield

Brightfield is microscopy a technique in which a specimen is illuminated with white light from below, and the image is formed based on the light absorbed, reflected, or refracted by the sample. This method creates contrast between the specimen and its background, making it ideal for observing fixed and stained samples or naturally pigmented specimens. However simple, cost-effective, and suitable for a broad range of applications it is, brightfield microscopy has limited effectiveness for transparent or colorless samples, as they lack contrast. As a result, it is often used with staining techniques to enhance visualization [17].

2.2.3 Fluorescent

Fluorescence microscopy is a powerful imaging technique that uses fluorescent dyes or naturally fluorescent molecules to visualize specific components within a specimen. By illuminating the sample with light of a specific wavelength, fluorescent molecules absorb this energy and emit light at a longer wavelength, producing highly specific and brightly colored signals against a dark background. This method enables detailed visualization of cellular structures, proteins, and molecular processes. Its high sensitivity and specificity make it an essential tool in cell biology, molecular biology, and medical diagnostics [85].

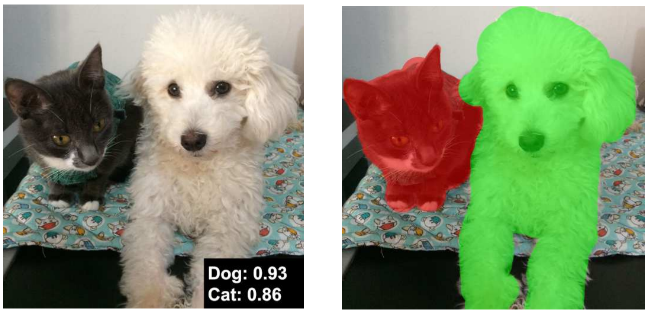
2.2.4 Confocal

Confocal microscopy is an advanced optical imaging technique that enhances resolution and contrast by using a pinhole to eliminate out-of-focus light from specimens, producing sharp, high-resolution images. Unlike traditional fluorescence microscopy, confocal microscopy scans the sample point-by-point with a focused laser beam and collects the light from a specific focal plane, creating optical layers. This allows for the reconstruction of detailed 3D images of thick specimens by compiling multiple focal planes [110].

2.3 Semantic Segmentation

Image segmentation [53] involves dividing an image into multiple cohesive regions, representing one of the oldest and most extensively studied problems in computer vision. However, traditional segmentation methods often overlook the semantic relationships between regions.

Semantic segmentation addresses this limitation by assigning meaningful labels to these segments, achieving a pixel-wise classification where each pixel in the image is categorized. This process, illustrated in Figure 2.1, is inherently more computationally demanding than standard image classification, which assigns a single label to the entire image [98].



(a) Image Classification (b) Semantic Segmentation

Figure 2.1: A comparison between image classification and semantic segmentation. (a) In image classification, the entire image is assigned to a single class, with the output representing the probability of each class. In this case, there are two classes: cat and dog, and the algorithm classifies the image as a dog. (b) In contrast, semantic segmentation involves classifying each individual pixel of the image and assigning it to a specific class. The output of semantic segmentation is a pixel-wise classification. In this figure, there are three classes: cat, dog, and background.

Semantic segmentation methods have evolved significantly over time. Before the advent of deep learning, techniques such as conditional random fields, pixel-based operations, and superpixel analysis dominated the field. The introduction of early deep learning models marked a shift, where classification networks were adapted to classify superpixels. However, the true advancement came with the development of fully convolutional neural networks, which replaced fully connected layers with convolutional ones, enabling pixel-level classification. This not only improved segmentation accuracy but also led to the revival of conditional random fields for refining pixel class probabilities and paved the way for modern encoder-decoder models.

In deep learning-based segmentation, the process is performed on a pixel-by-pixel basis, where each pixel in the image is assigned a classification. These deep architectures typically take an image as input and produce a black-and-white segmentation mask as output, where the white pixels represent the region of interest and the black pixels represent the background, as shown in Figure 2.2. As expected, this pixel-wise classification requires more computational resources compared to standard image classification, where a single class is assigned to the entire image [98].

Image segmentation is an older technique that predates the use of deep machine learning approaches [148]. Early methods included data clustering, region growing, edge detection, genetic algorithms, and fuzzy logic, among others. However, with the rise of deep machine learning, segmentation has increasingly been carried out using convolutional neural networks, with specialized architectures designed specifically for this task, as discussed later in this chapter.

Image segmentation is an older technique with approaches that precede those involving deep machine learning [148]. The first methods developed include data clustering, region growing, edge detection, genetic algorithms, fuzzy logic, among others. However, with the advent of deep machine learning, segmentation began to be performed through convolutional neural networks, with architectures created specifically for this topic, as we explain later in this chapter.

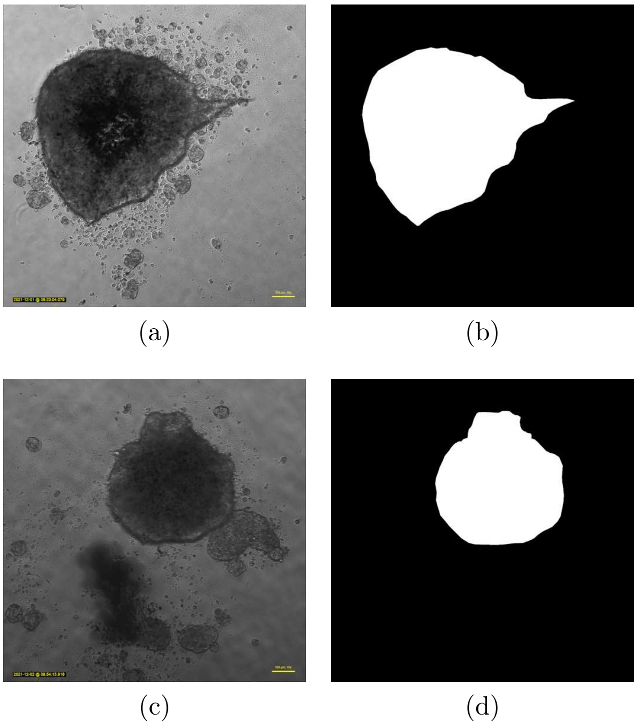


Figure 2.2: Illustration of two cultures and their corresponding segmentation masks. Images (a) and (b) show the input, while images (c) and (d) display the respective outputs. It is important to note that the region of interest does not necessarily encompass the entire cell culture, as demonstrated in (d), where the cell aggregation in the southeast corner of the culture is excluded from the region of interest.

In biological applications, image segmentation is crucial across several fields, including medical imaging, histopathology, and microscopy. It allows researchers to isolate and quantify specific features, such as cells, tissues, and organelles, from complex backgrounds. Accurate segmentation enhances the visualization and interpretation of biological phenomena, facilitating a better understanding of cellular morphology, distribution, and interactions.

The development of advanced algorithms, particularly those utilizing deep learning and convolutional neural networks, has significantly increased segmentation accuracy and efficiency, overcoming challenges related to variations in the size, shape, and appearance of biological structures. By enabling high-throughput analysis and improving result reproducibility, computer-based image segmentation plays an essential role in advancing our knowledge of biological systems and disease mechanisms, ultimately aiding in the improvement of diagnostic and therapeutic approaches.

Otsu

Proposed by Otsu et al. [107], this method determines the optimal threshold for image segmentation automatically by maximizing the variance between classes of pixel intensities. Using histogram analysis, it minimizes intra-class variance to achieve effective separation of foreground and background without the need for manual input or prior knowledge of the image context.

The Otsu algorithm is renowned for its robustness and computational efficiency, making it a favored approach for various object detection and segmentation tasks. Its capability to determine thresholds for binary segmentation automatically has led to its extensive application in fields such as medical and biological imaging.

Watershed

This algorithm is a powerful image segmentation technique, initially developed in the field of topography. It simulates the natural flow of water over a landscape's ridges and valleys to identify regions of interest through a "flooding" process [154], as depicted in Figure 2.3.

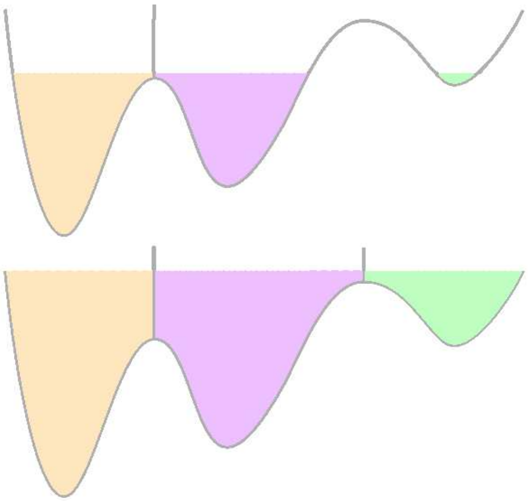


Figure 2.3: Watershed illustration of the seeds being initialized at the bottom of the valleys and the flood algorithm filling the "water" from the seeds, until two regions touch each other, creating a segmentation ridge.

In this method, a grayscale image is interpreted as a topographic surface, where pixel intensities represent peaks (higher values) and valleys (lower values). The algorithm begins by flooding the surface from chosen seed points, gradually filling the valleys. Region boundaries are defined where the floodwaters from different seed points meet, segmenting the image into distinct regions. This approach is particularly effective for separating objects in images with complex intensity variations, where traditional methods may struggle to delineate intricate boundaries.

The watershed algorithm excels in preserving object boundaries, making it highly indicated for applications such as cell and nuclei segmentation in biomedical imaging, as well as geospatial and satellite image analysis [13, 16, 29, 39, 55, 60, 68, 106, 122]. However, it can be sensitive to noise, and the choice of seed points may lead to over-segmentation, resulting in excessive fragmentation of the image. To address these challenges, pre-processing steps like noise reduction and post-processing techniques such as region merging are often employed to enhance segmentation accuracy and better capture the structures of interest.

K-Means Clustering

The K-Means algorithm [140] is a widely used method for image segmentation, dividing an image into regions based on pixel intensity similarities. It clusters pixels into k groups, where k is a user-specified parameter representing the desired number of segmented regions, as illustrated in Figure 2.4.

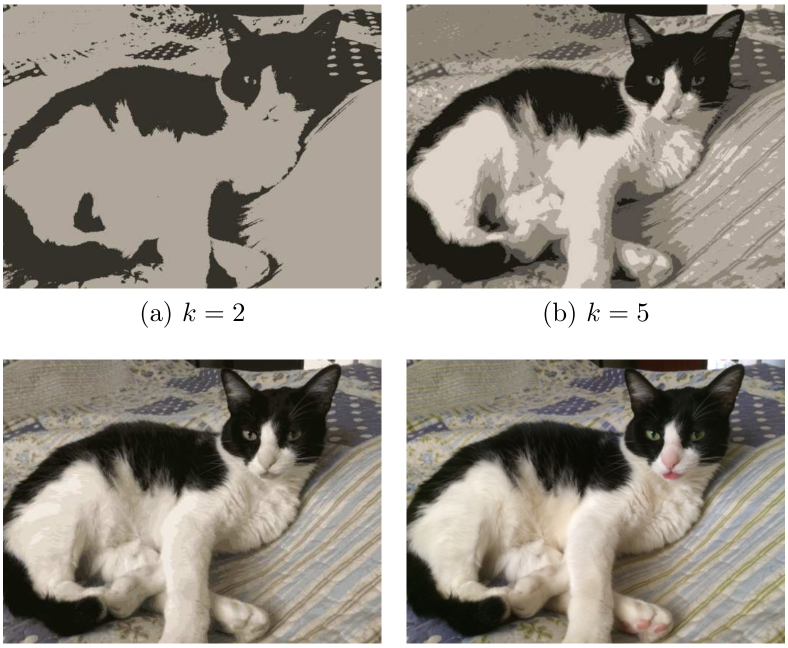
The algorithm begins by randomly initializing k centroids, determined by pixel intensity values. Each pixel is then assigned to the nearest centroid. After assignment, the centroids are recalculated as the mean of the pixel values within each cluster. This process of assignment and recalculation repeats iteratively until the centroids stabilize and no longer change significantly. The result is an image segmented into k regions, where pixels within the same region share similar intensity characteristics [140].

While K-Means is computationally efficient and relatively simple, it may struggle with complex images featuring varying intensity distributions or overlapping objects. Moreover, determining the optimal value for k can be challenging, often requiring domain expertise or experimentation to achieve the best results.

2.4 Machine Learning and Artificial Intelligence

The term Artificial Intelligence (AI) was first coined in 1956 during the Dartmouth Summer Research Project on Artificial Intelligence conference, organized by John McCarthy and Marvin Minsky focused on this topic [21]. AI refers to the use of computers to replicate human intelligence, enabling them to perform tasks that are traditionally better suited to human capabilities. It is a broad field encompassing areas such as image, text, and audio recognition, robotics, classification systems, and the development of expert systems. Machine Learning is referred to as a domain of AI [53].

Machine Learning, which emerged around 1980, focuses on systems that learn from data and improve based on experience. Machine learning algorithms are typically categorized into supervised, unsupervised, and reinforcement learning, depending on the type of supervision provided during training. In supervised learning, data and corresponding labels are provided to the system during training to help it learn to associate the data with the correct labels. Once trained, the model can generalize and correctly label new, unseen data. Unsupervised learning, in contrast, deals with input data without labels, aiming to discover patterns or structures in the data to generate meaningful outputs. Reinforcement



(a) $k = 2$

(b) $k = 5$

(c) $k = 20$

(d) Original

Figure 2.4: Illustration of the K-Means segmentation, defined by the number of chosen centroids. Each image is colored by k different colors, segmenting each region by its color.

learning involves an agent learning through its actions in a controlled environment, where rewards are used to guide the agent's learning towards achieving a desired goal [53].

A neural network mimics the structure of the human brain, using artificial neurons, that usually consists of an input layer, hidden layers, and an output layer. These layers are made up of neurons, or perceptrons, which are responsible for processing information. In a neural network, each input in a layer is associated with a weight, and each neuron has a bias. The perceptron calculates the weighted sum of the input and its corresponding weight, adds the bias, and applies an activation function (such as a step function, sigmoid, or hyperbolic tangent) to determine the neuron's activation or deactivation based on a threshold [53].

During the training process, the network learns by comparing its predicted output with the actual label and performing error backpropagation. This process adjusts the weights within the network to improve its predictions, allowing it to correctly associate inputs with their corresponding outputs.

2.5 Deep Learning

Deep learning technology has become a driving force behind numerous research advancements in various fields, including object identification, speech transcription, user interest analysis, and interdisciplinary applications. Its widespread use is attributed to its ability to extract knowledge from raw data, such as image pixels, and solve complex decision-making problems in real-world scenarios [82].

Proposed by Hinton [59] in 2006, deep learning refers to neural networks with multiple layers, where each hidden layer is pre-trained individually. This technique enables computers to learn from experience and understand the world through a hierarchical structure of concepts, with each concept defined by its relation to simpler ones [53]. Inspired by the neural connections in the human brain, deep learning utilizes Artificial Neural Networks (ANN), which are composed of artificial neurons organized in different layers. The configuration of these layers defines the network's architecture, allowing for complex processing of information.

The complexity of a network is determined by its size, specifically the number of parameters and operations it involves. While more complex architectures tend to offer better adaptability to challenging problems, they also come with the downside of slower and more resource-intensive training, requiring significant computational power. As a result, there is a growing preference for simpler networks that, although smaller, are still sufficiently complex to represent problems effectively and deliver strong performance. However, it's essential to recognize that this involves a trade-off between effectiveness and efficiency, with the optimal choice depending on the specific problem and the available computational resources.

2.6 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are deep neural networks primarily composed of convolutional, pooling, and dense layers. The convolutional layer uses a convolution filter that moves across the image, creating an activation map by weighting neurons in the region where the filter is applied. This layer is crucial because it analyzes regions of the image as a whole, considering not only individual pixels but also the surrounding neighborhood.

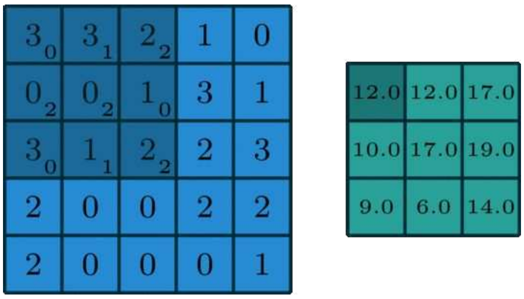
Following the convolutional layers, pooling layers are used to simplify and shrink the output. Pooling is achieved by selecting the maximum, minimum, or average of a region, known as max-pooling, min-pooling, and average pooling, respectively. Finally, the dense (or fully-connected) layer consists of neurons that are fully connected to those in the previous layer.

Within CNNs, certain architectures are specifically designed for image segmentation. Common segmentation techniques focus on pixel grayscale levels, detecting discontinuities or similarities: the first identifies abrupt changes in grayscale, while the second groups similar pixels into the same region [111]. Several classic CNN architectures have significantly advanced network development, including LeNet [81], AlexNet [78], VGGNet [136], GoogLeNet [143], and ResNet [57].

Due to their robustness, CNN-based segmentation methods are widely used across various fields, including remote sensing, geology, medicine, and biology. They are applied to a broad range of tasks, from face detection to traffic monitoring, by simplifying complex problems through image partitioning, which helps separate objects from the background.

2.7 Convolutions

In computer vision, convolutional layers are the building block of deep neural networks. In the context of deep learning, a convolution operation is an element-wise multiplication between matrices, followed by an addition [111]. The input image is multiplied with a kernel, as illustrated in Figure 2.5, after which these values are added together. The resulting sum is the new pixel value of the output image. The next value of the output image is calculated by sliding the kernel matrix one element over.



3_0	3_1	2_2	1	0
0_2	0_2	1_0	3	1
3_0	1_1	2_2	2	3
2	0	0	2	2
2	0	0	0	1

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

3*0 + 3*1 + 2*2 + 0*2 + 0*2 + 1*0 + 3*0 + 1*1 + 2*2 = 12

Figure 2.5: Illustration of a 2D convolution, extracted from Bai [7]. Showcasing the element-wise multiplication of the kernel with the input image, followed by the sum of these values, that led to the pixel value of 12 in the cell (0, 0) of the output.

Usually, a convolution operation has three parameters. (i) The kernel size determines the size of the kernel matrix. Using large kernels requires more computational power. However, they create a larger receptive field, which can help the network to learn better features. (ii) The stride is the step size in which the kernel is slid. For example, a stride value of one moves the kernel one element over, and a stride value of two skips the adjacent element. A value greater than two can be used to downsample images. (iii) Padding defines how the image border is handled. For instance, a padded convolution is when the image borders are filled with a dummy value, such as zero, so that the output of the convolution has the same size as the input. On the other hand, unpadded convolutions operate only over the image values, which creates a smaller image than the input [50]. In summary, a convolution applies a filter to an input image, and the repetition of this process generates a feature map. We will return to the receptive field topic in Section 2.10.

2.7.1 U-Net

The U-Net is a deep learning encoder-decoder architecture originally proposed for medical image segmentation [90], and it has significantly advanced the field due to its ability to achieve high-quality results with fewer training samples. Its architecture is named “U-Net” because of the U-shaped structure formed by its layers, as shown in Figure 2.6. The network consists of convolutional layers, beginning with the encoder and followed by the decoder, allowing it to produce effective segmentation even when trained on small datasets [125].

According to Ronneberger et al. [125], the U-Net architecture consists of two main paths: a contraction path and an expansive path. The contraction path, also known as

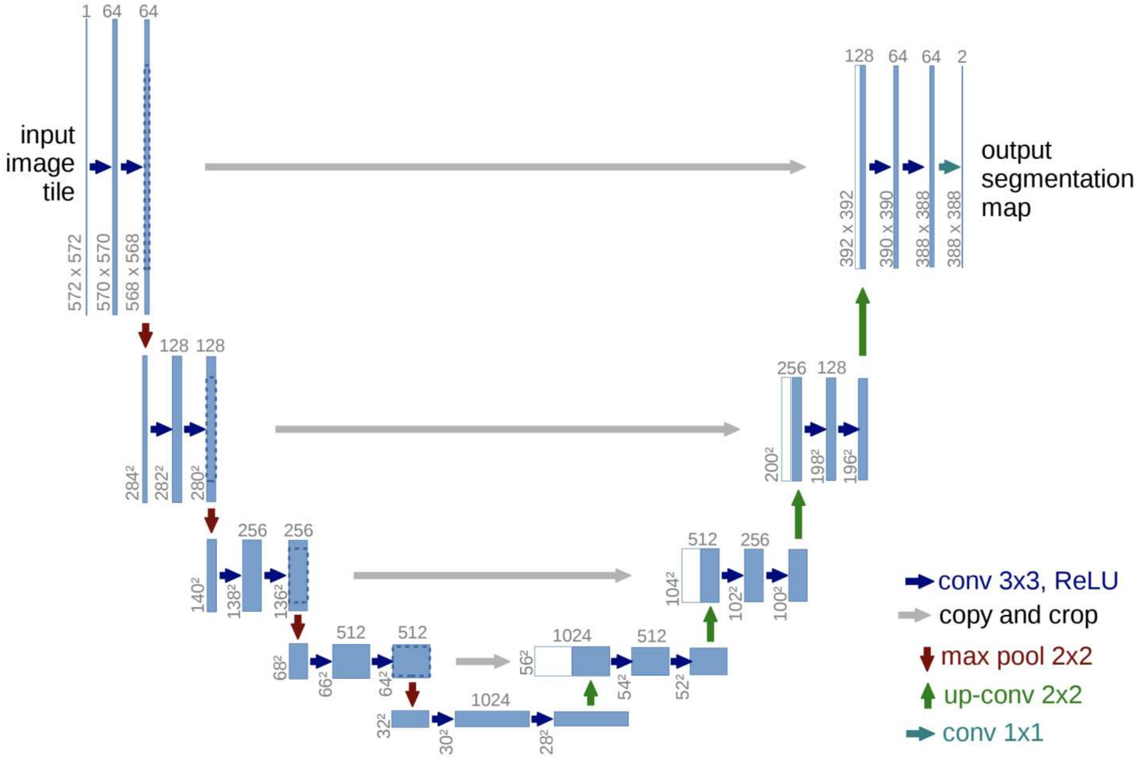


Figure 2.6: The diagram of the U-Net architecture, as presented by Ronneberger et al. [125], illustrates the organization of layers in a U-shape. The encoder is located on the left side, where convolutions are applied to increase the number of feature channels. On the right side, the decoder utilizes transposed convolutions (not to be confused with deconvolutions) to enhance resolution. The connections between the encoder and decoder are represented by skip connections, which facilitate the flow of information between corresponding layers.

the encoder, is represented on the left side of the U-shape in Figure 2.6. This path follows a typical deep learning structure, where the network progressively reduces feature sizes while increasing the number of channels.

The encoder in U-Net applies 3×3 convolutions, followed by a Rectified Linear Unit (ReLU) activation and a downsampling max-pooling operation with a 2×2 filter and stride of 2. This sequence of convolutions reduces the image resolution at each step, which works well for classification tasks but can be problematic in segmentation, where maintaining a high output resolution close to the input size is desired. The output from U-Net is slightly smaller than the input, which is addressed by the decoder, that attempts to restore the feature resolution to the original size.

The decoder works by concatenating the feature map from the encoder with 2×2 transposed convolutions, shown as horizontal lines in Figure 2.6, which reduce the number of feature channels while increasing the resolution. After each concatenation, two 3×3 convolutions are applied with a ReLU layer. Finally, a 1×1 convolutional layer is used at the output to map the features to the classification classes.

2.7.2 MultiResUNet

MultiResUNet [66] is an advanced deep learning architecture proposed for segmenting various modalities of biomedical images, building upon the U-Net architecture [125], as illustrated in Figure 2.7. This architecture enhances the traditional U-Net by incorporating multi-resolution feature extraction, which significantly improves the segmentation of complex structures in images.

Particularly effective in biomedical image analysis, MultiResUNet combines high-resolution and low-resolution features, enabling the model to capture both fine details and contextual information. Through a series of convolutional layers operating at multiple resolutions, the architecture preserves spatial hierarchies and generates richer feature representations. Additionally, the integration of skip connections helps maintain crucial spatial information, resulting in more accurate segmentations.

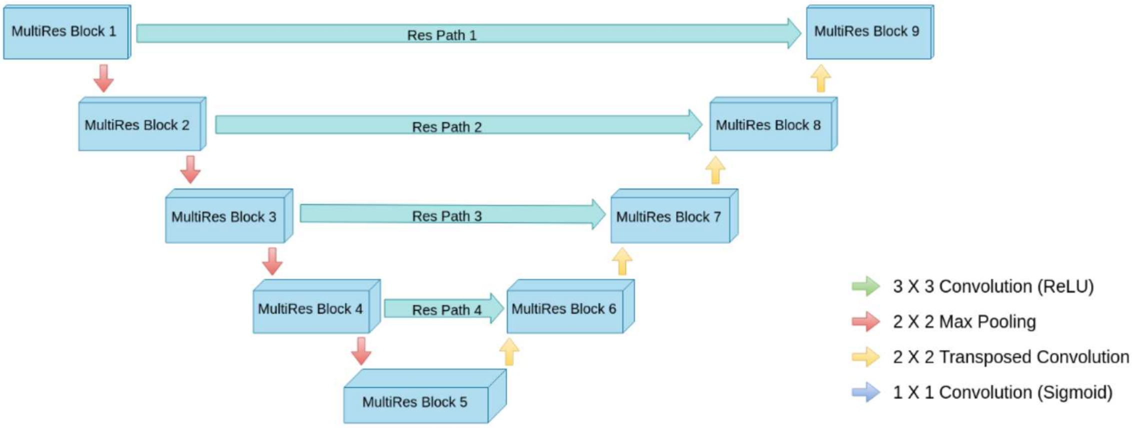


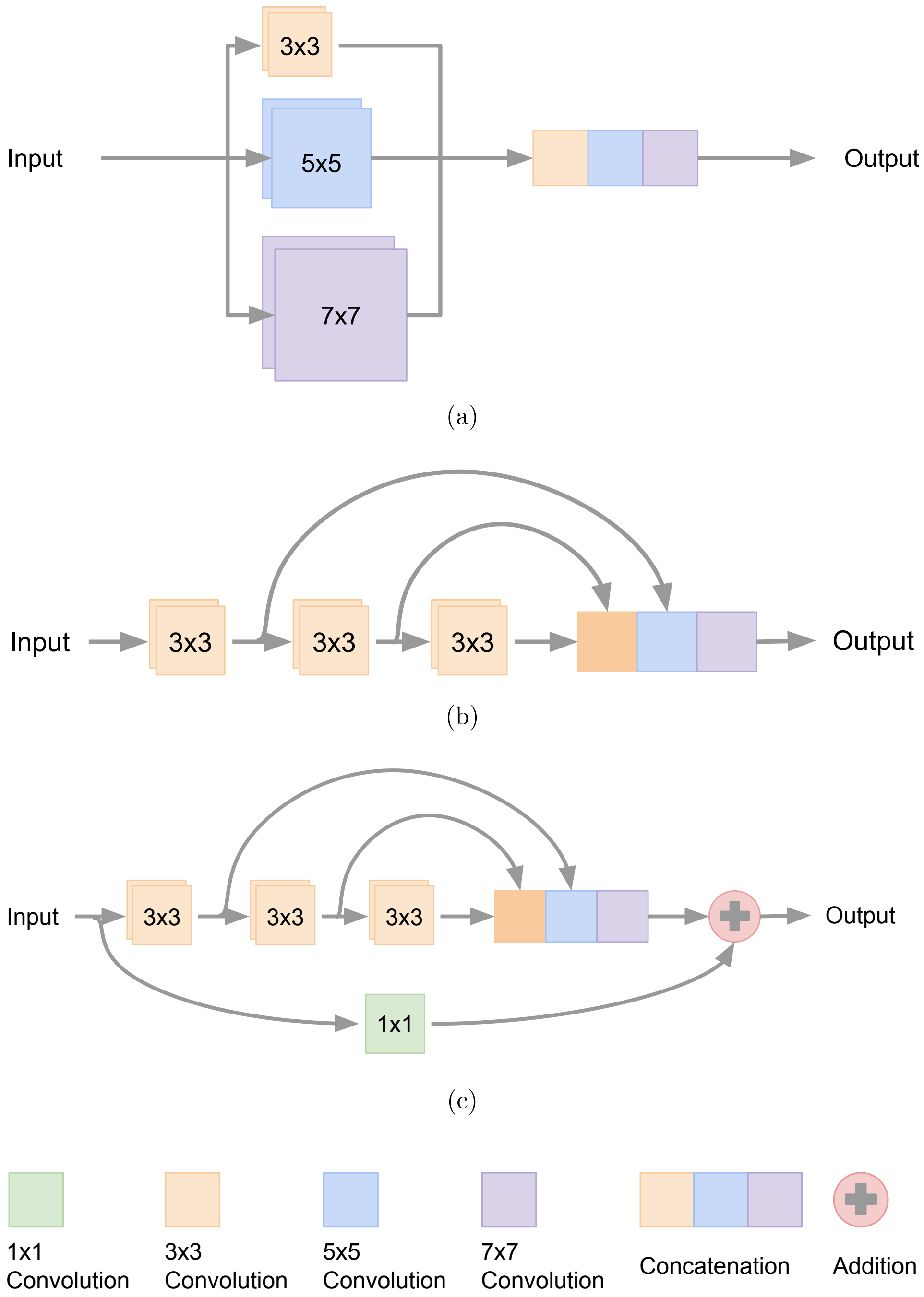
Figure 2.7: MultiResUNet Architecture, extracted from Ibtehaz and Rahman [66].

In MultiResUNet, each pair of convolutional layers in the traditional U-Net architecture is replaced by a MultiResUNet residual block. As shown in Figure 2.8, the MultiResUNet residual block consists of successive 3×3 convolutional layers, with their feature maps concatenated. Additionally, a 1×1 convolutional layer is included to add the input to the concatenated output of the 3×3 convolution layers.

Furthermore, the standard connections in U-Net are replaced by residual paths, as illustrated in Figure 2.9. These residual paths involve the successive addition of feature maps, which are derived from parallel convolution operations using both 3×3 and 1×1 filters.

2.7.3 U-Net++

The U-Net++ [175] architecture builds upon the U-Net architecture by incorporating nested dense convolutional blocks between the encoder and decoder, as illustrated in Figure 2.10. Within the connections between the encoder and decoder, convolution operations are applied, denoted by $X^{i,j}$, where each convolutional layer is preceded by concatenating the output of the previous convolutional layer in the same block with the upsampled output from the convolutional layer of the lower block.



(a)

(b)

(c)

Figure 2.8: MultiResUNet Residual Block. (a) Inception Block and its parallel convolutions with 3×3, 5×5, and 7×7 filters, to preserve spacial features within different scales. (b) Operation in which the 5×5 and 7×7 filters are factored in a succession of 3×3 filters. (c) A 1×1 convolution is added to the concatenated features in form of residual connection.

A notable feature of this architecture is Deep Supervision [4], represented by L in Figure 2.10. This mechanism computes the average output across multiple branches and selects the optimal branch to generate the final segmentation map, minimizing error and improving computational efficiency. The demonstrated effectiveness of U-Net++ in various segmentation tasks highlights its potential to advance image analysis in domains such as medical imaging and cancer research.

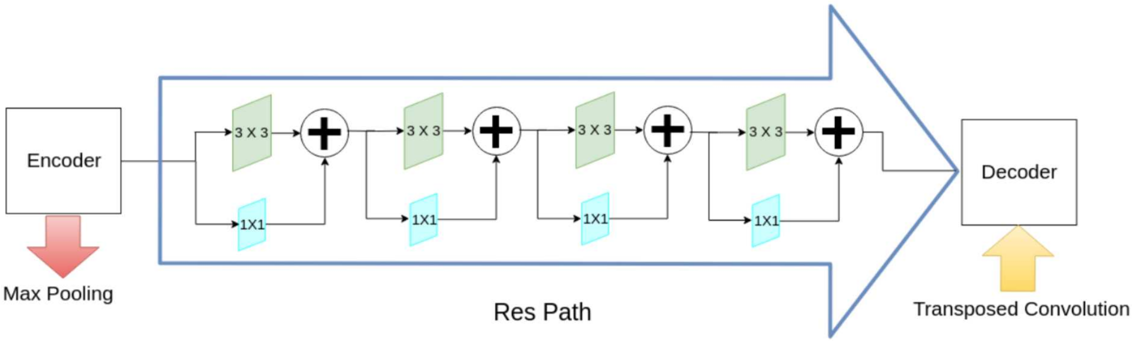


Figure 2.9: MultiResUNet residual path, extracted from Ibtehaz and Rahman [66].

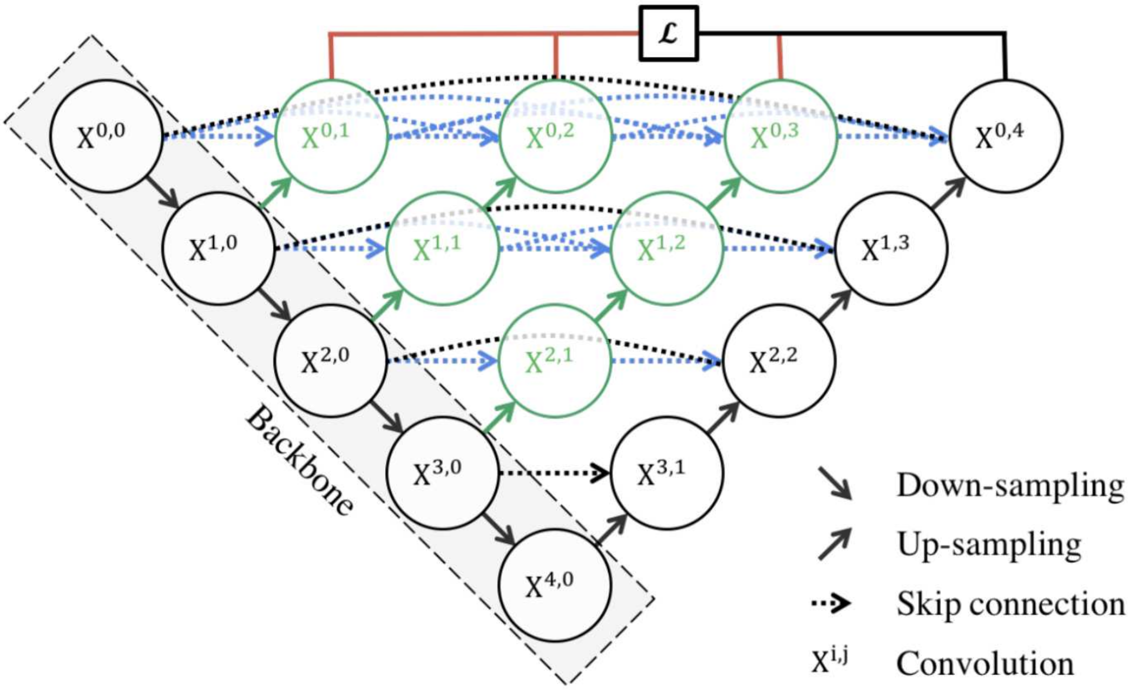


Figure 2.10: UNet++ architecture, extracted from Zhou et al. [175].

2.8 UNeXt

UNeXt, proposed by Valanarasu and Patel [149], is a convolutional multi-layer perceptron (MLP)-based network designed for efficient image segmentation in point-of-care applications. It aims to operate on low-tier hardware while maintaining adequate segmentation performance. To achieve this, the authors kept the 5-layer encoder-decoder structure and skip connections from the U-Net architecture but introduced modifications in two stages: an early convolutional stage and an MLP-based latent stage. They further proposed a tokenized MLP block and incorporated a shift mechanism inspired by the Swin Transformer.

As illustrated in Figure 2.11, the encoder contains three convolutional blocks, followed by two tokenized MLP blocks. The decoder mirrors this structure in reverse, with two tokenized MLP blocks followed by three convolutional blocks. Each encoder block reduces the feature resolution by a factor of 2, which is restored in the decoder. The feature channels in the encoder are set to 32, 64, 128, 160, and 256, with the decoder using

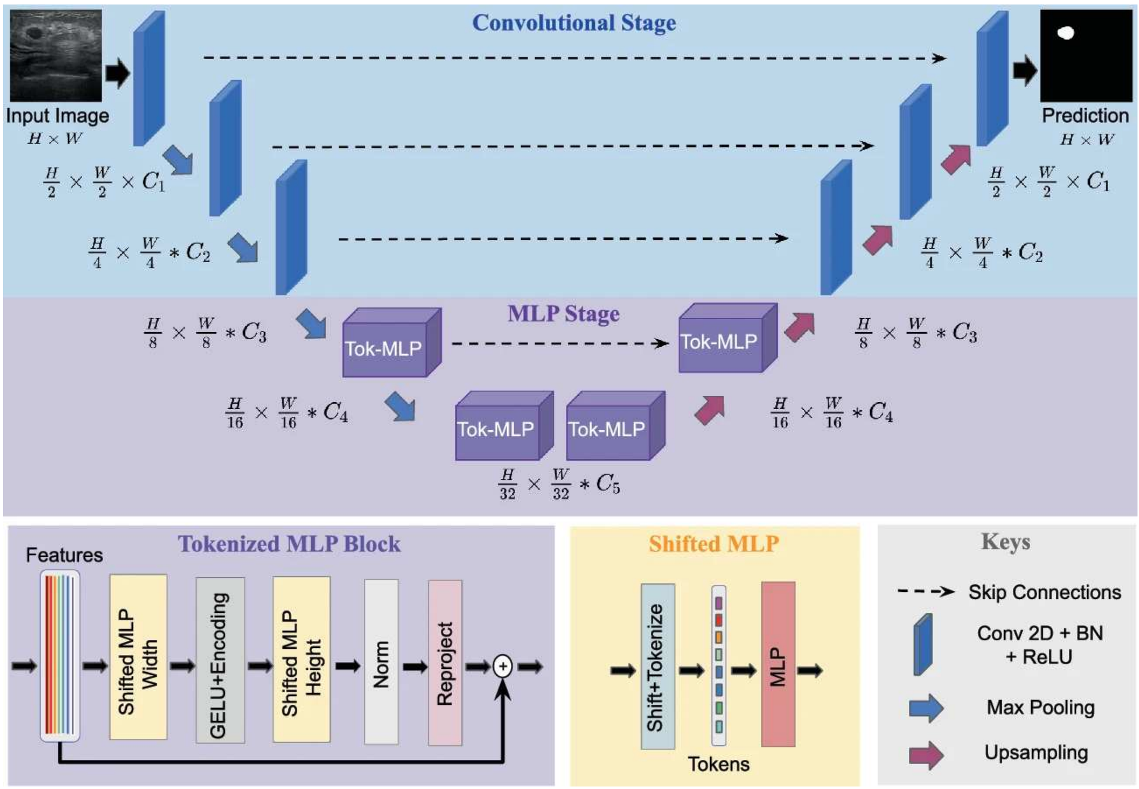

Convolutional Stage
Input Image
$H \times W$
$\frac{H}{2} \times \frac{W}{2} \times C_1$
$\frac{H}{4} \times \frac{W}{4} * C_2$
Prediction
$H \times W$
$\frac{H}{2} \times \frac{W}{2} \times C_1$
$\frac{H}{4} \times \frac{W}{4} * C_2$
MLP Stage
$\frac{H}{8} \times \frac{W}{8} * C_3$
Tok-MLP
Tok-MLP
$\frac{H}{8} \times \frac{W}{8} * C_3$
$\frac{H}{16} \times \frac{W}{16} * C_4$
Tok-MLP
Tok-MLP
$\frac{H}{16} \times \frac{W}{16} * C_4$
$\frac{H}{32} \times \frac{W}{32} * C_5$
Tokenized MLP Block
Features
Shifted MLP Width
GELU+Encoding
Shifted MLP Height
Norm
Reproject
Shifted MLP
Shift+Tokenize
MLP
Tokens
Keys
Skip Connections
Conv 2D + BN + ReLU
Max Pooling
Upsampling

Figure 2.11: UNeXt architecture, extracted from Valanarasu and Patel [149].

the reverse order. Skip connections link each encoder block to its corresponding decoder block.

The convolutional stage, depicted at the top of Figure 2.11, consists of a convolution layer, a batch normalization layer, and a ReLU activation function. In the encoder, convolutional blocks incorporate max-pooling layers, while the decoder employs bilinear interpolation layers as a computationally lighter alternative to transpose convolutions.

The shifted MLP stage, shown at the bottom of Figure 2.11, introduces a shift mechanism that operates on channel axes before tokenization. Inspired by the Swin Transformer [88], this technique forces a locality bias within the MLP block, as opposed to the global processing used in the ViT model [166]. Each tokenized MLP block contains two MLPs: one shifts features along the width and the other along the height.

The tokenized MLP stage, depicted in the middle of Figure 2.11, tokenizes features before feeding them into the shifted MLP blocks. First, the features are shifted and projected into tokens. These tokens are processed by a shifted MLP block (operating across the width) and then passed through a depthwise convolutional layer (DW-Conv), which encodes positional information more efficiently than standard positional encodings. The use of DW-Conv, combined with a GELU activation function, ensures lower computational cost.

Subsequently, the features are passed through another shifted MLP (operating across the height) and combined with the original tokens as residual information. Finally, layer normalization is applied to normalize across tokens, and the features are re-projected.

In comparison to the original U-Net architecture, UNeXt reduces the number of parameters by a factor of 72, decreases computational complexity by a factor of 68, and improves inference speed by a factor of 10.

2.9 Transformers

To optimize the encoding of input sequences and the decoding of output values, Vaswani et al. [152] introduced the Transformer network in 2017, which replaces the concept of recurrence with an attention mechanism, designed for natural language tasks. Attention-based networks leverage mechanisms that aggregate and process input information dynamically based on the data itself, primarily utilizing self-attention and multi-head self-attention techniques.

The self-attention mechanism processes a sequence of input elements by calculating the relationships between each element, referred to as a key, and all other elements in the sequence, including itself. This process determines how much each element incorporates information from others. Practically, this mechanism is implemented through the inner product of vector representations of the input sequence, where the resulting values quantify the degree of relationship between elements [55].

In the Transformer architecture, multi-head attention employs eight attention heads in parallel, enabling the model to learn and combine multiple perspectives of the input sequence while reducing training time through parallelized operations. The architecture has six encoder blocks and six decoder blocks, as depicted in Figure 2.12 as N×, each consisting of attention layers. These layers are essential in capturing relationships within the data, allowing the model to incorporate information from the global context effectively.

The initial step of the Transformer network involves tokenizing the input words (or image patches), converting them into token vectors typically represented by integer values. These tokens are then mapped to word embeddings, which represent each word as a 512-dimensional vector. A common technique for generating embeddings is word2vec [14], introduced by Google in 2013, which employs continuous bag-of-words (CBOW) and continuous skip-gram models to produce distributed word representations. Usually, the input size across all layers of the Transformer network is fixed, set to 512 in the original implementation. This consistent layer size reduces computational complexity and simplifies the control of data flow throughout the model.

Following the embedding process, positional encodings are added to the word embeddings to capture the order of words within the sequence. These encodings can be computed using various methods, such as sine and cosine functions of different frequencies. The positional values are combined with the embeddings through addition or other operations, including multiplication, subtraction, division, or square root functions.

During training, the weights of the query, key, and value matrices are optimized. The input vectors are multiplied by these matrices to produce the query (Q), key (K), and value (V) vectors, which are subsequently used to compute attention. The attention mechanism employed in the Transformer is known as Scaled Dot-Product Attention, characterized by the inclusion of a scaling factor, $\sqrt{d_k}$, where d_k represents the dimensionality of the key vector, set to 64. This process happens in each attention head, is mathematically defined by Equation 2.1, using Q, K, and V to capture relationships between elements in

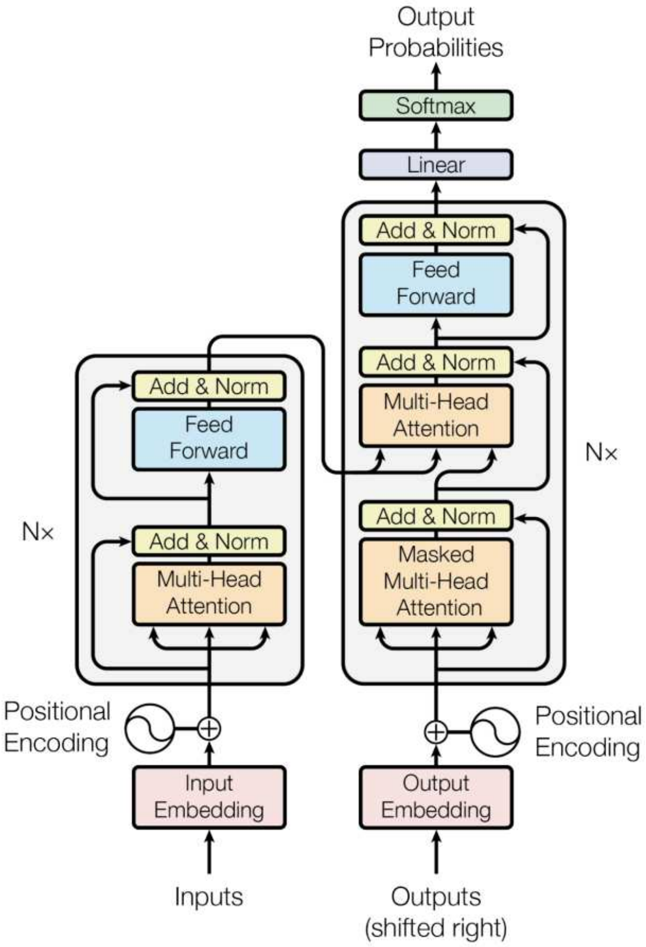


Figure 2.12: Transformer architecture, encoder (right) and decoder (left), extracted from Smyrek and Stelzer [138].

the input sequence effectively.

$$\mathrm{Attention(Q, K, V)} = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \cdot V. \tag{2.1}$$

Once the eight attention heads compute their outputs, the results are concatenated and passed to the Add and Norm layer, which joins residual connections. These residual connections brings information from earlier layers, combining it with the current layer's output before applying a normalization operation. This mechanism ensures the preservation of essential data, such as positional encodings, throughout the model.

After the Add and Norm layer, feedforward blocks are used, comprising of two fully connected layers with ReLU as the activation function. Following the feedforward block, another Add and Norm layer is applied, leading into the decoder block. The decoder organization mirrors the encoder but includes an additional Masked Multi-Head Attention Mechanism layer. This layer prevents the model from accessing future terms in the sequence during training, enabling the network to learn to predict the next elements. Keep in mind that this architecture was firstly proposed to work on text inputs.

The final stage of the Transformer network consists of a linear layer, which generates the output sequence by applying a linear transformation to determine the next elements.

A softmax function then refines this prediction by identifying the single most likely element, allowing the model to produce the output sequence one token at a time.

The Transformer architecture represents the state of the art in natural language processing and has demonstrated significant promise in fields such as computer vision and audio processing [13, 24]. However, limitations such as inefficiency in handling long sequences, high computational demands of self-attention, and difficulty in generalizing from limited data have prompted the development of Transformer-based models to address these downsides.

Despite their transformative impact on Natural Language Processing (NLP), applying Transformers to computer vision tasks poses challenges due to the fundamental differences between text and image data. High-resolution images contain significantly more information, and spatial relationships between pixels are far more critical than sequential dependencies.

2.9.1 TransUNet

Proposed by Chen et al. [28], TransUNet is a medical image segmentation architecture designed to use the strengths of Transformer and convolutional neural network (CNN) models. By employing the Transformer architecture and a CNN as the encoder, the model effectively leverages the global context captured during encoding alongside the spatial features during decoding. The encoder of this hybrid design incorporates a self-attention mechanism, functioning in a sequence-to-sequence prediction strategy. The self-attention features generated during encoding are upsampled and combined with higher-resolution features from the CNN decoder via skip connections, enabling precise localization of features.

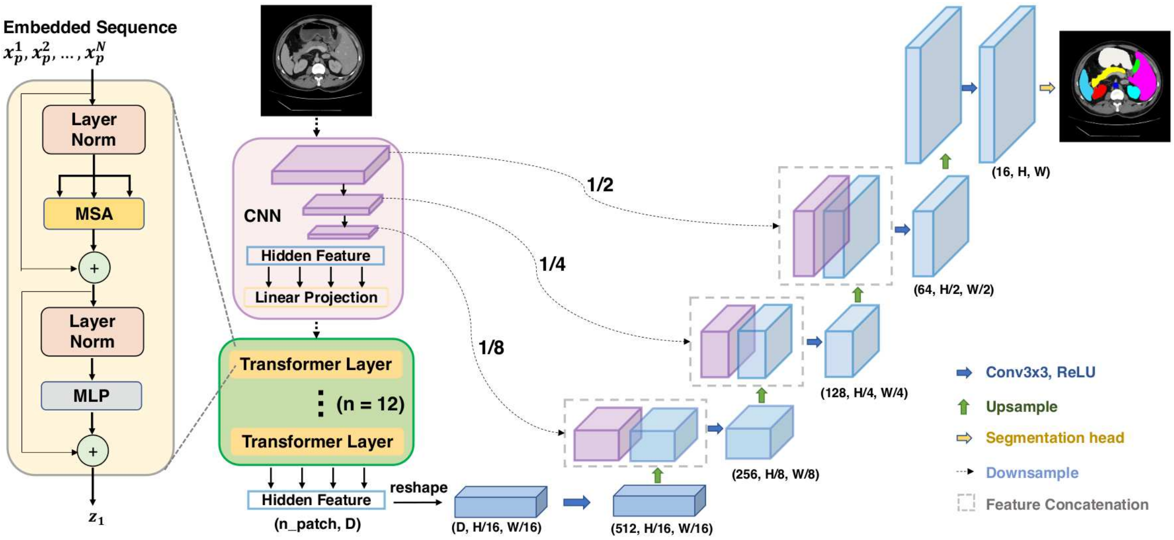


Figure 2.13: TransUNet architecture, extracted from Chen et al. [28].

As illustrated in Figure 2.13, the architecture adopts a CNN-Transformer hybrid as its encoder. Initially, the CNN operates as a feature extractor while simultaneously providing skip connections from the high-resolution features. Before the input proceeds to the Transformer, it is tokenized by reshaping into a sequence of flattened 2D patches. These patches are mapped into a latent D-dimensional embedding space through a trainable linear projection. Positional embeddings are learned and encoded to preserve spatial information, which is subsequently added to the patch embeddings to retain positional context.

The Transformer encoder, depicted on the left side of Figure 2.13, consists of L layers of Multihead Self-Attention (MSA) and Multi-Layer Perceptron (MLP) blocks. Both mechanisms are preceded by layer normalization. This design, in along with the CNN, utilizes the high-resolution features in the decoder, achieving better performance compared to architectures that solely rely on the Transformer for encoding.

On the decoder path, shown on the right side of Figure 2.13, the authors propose a cascaded upsampler approach. This approach entails multiple upsampling steps to decode the learned features. The features are reshaped to match the high-resolution features from the skip connections and concatenated together. Following this concatenation, the resulting features are processed through two upsampling operations, a convolutional layer, and a ReLU activation layer. This sequence reconstructs the segmentation mask at the input resolution.

2.9.2 Swin Transformer

The Swin Transformer [88] introduces a hierarchical representation strategy that enables the processing of visual information in a multi-scale way, similar to convolutional neural networks (CNNs). Starting with small, fixed-size image patches, it progressively merges neighboring patches to reduce spatial resolution and join features, forming a type of pyramid structure, illustrated in Figure 2.14.

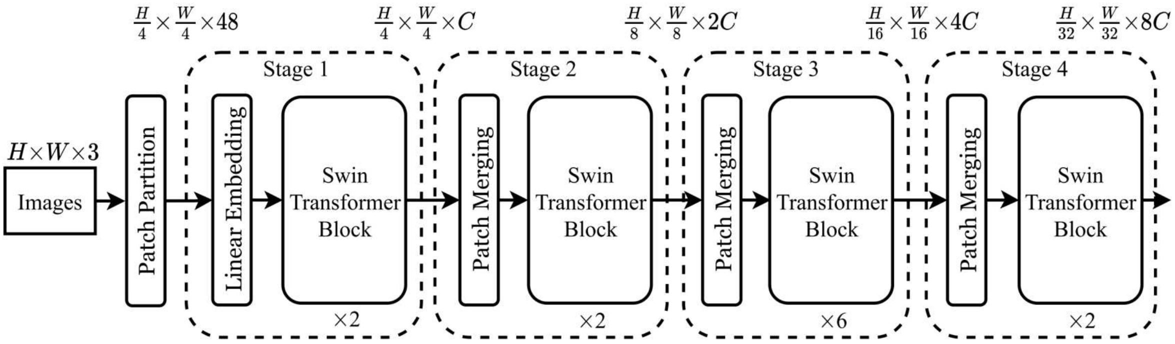


Figure 2.14: Swin Transformer architecture, extracted from Sharma et al. [132].

This approach captures fine-grained details at lower levels while encoding contextual information at higher levels, making it adequate for tasks requiring both local and global understanding. By combining this hierarchical design with efficient attention mechanisms, the Swin Transformer achieves a balance between computational efficiency and rich feature representation, making it adept at handling high-resolution images.

A key innovation of the Swin Transformer is its shifted window partitioning mechanism, which facilitates efficient local-global interaction during attention computation. Initially, images are divided into non-overlapping windows, in which self-attention is computed independently to constrain computational complexity. To overcome the limitations of isolated attention regions, the model introduces a shift in window alignment between consecutive layers, ensuring that patches from adjacent windows are included in the attention field without increasing computational overhead. Alternating between regular and shifted windows enables the Swin Transformer to capture both localized and contextual features effectively, addressing the challenges of strictly localized attention.

The Swin Transformer processes images using an integrated approach of patch partitioning, window-based multi-head self-attention (W-MSA), and shifted window multi-head self-attention (SW-MSA) to capture features across scales efficiently. Patch partitioning divides the input image into fixed-size patches, which are embedded into feature vectors for subsequent layers. W-MSA computes self-attention within these non-overlapping patches, significantly reducing the complexity compared to global attention. SW-MSA further enhances this design by shifting window alignments between layers, enabling cross-window feature interactions and improving the model's ability to aggregate features effectively. This innovative combination ensures computational efficiency while maintaining powerful feature extraction capabilities, making the Swin Transformer scalable and effective for a wide range of computer vision tasks.

The Swin Transformer addresses key challenges in computer vision by employing a novel architecture that strikes a balance between expressiveness and efficiency, thereby enhancing performance across diverse applications. Its advantage over other Transformer models in computer vision lies in two innovative aspects: its hierarchical architecture and the sliding window-based self-attention mechanism.

The hierarchical architecture of the Swin Transformer is implemented through successive stages that group patches from the previous layer, reducing the number of patches by a factor of four and doubling the number of feature channels. To achieve this, each patch is transformed into a linear embedding of dimension C. After that, each 2×2 block of patches is selected and concatenated into a $4C$-dimensional vector, which is then mapped into a $2C$ linear embedding. This process reduces the image dimensions by a factor of four and doubles the embedding size at each stage, which represents larger and larger areas of the original image. This design allows the network to extract information at multiple scales, with each patch becoming more representative of the image as the grouping process advances.

The second key innovation of the Swin Transformer lies in its use of Swin Transformer blocks, which apply local self-attention within fixed-size windows, each consisting of a number of patches. This contrasts with the global self-attention used in traditional vision Transformers, which makes the computational cost that scales quadratically with image resolution. By applying self-attention only within fixed-size windows, the Swin Transformer reduces the number of self-attention operations, resulting in proportional scaling with the input resolution. However, simply applying self-attention within these windows would limit the representational power of the network, as it would ignore the global context provided by traditional global self-attention. To overcome this limitation, each Swin Transformer block incorporates both a Multi-head Self Attention (MSA) layer with windows (W-MSA) and an MSA layer with sliding windows (SW-MSA), where the window alignment is shifted to enable cross-window interactions. This approach enhances the network's representational power, allowing it to capture similar features as a model with global self-attention, while offering superior computational efficiency.

Figure 2.15 illustrates the layers within each Swin Transformer block, highlighting the W-MSA and SW-MSA self-attention layers, along with linear normalization (LN) and fully connected (MLP) layers.

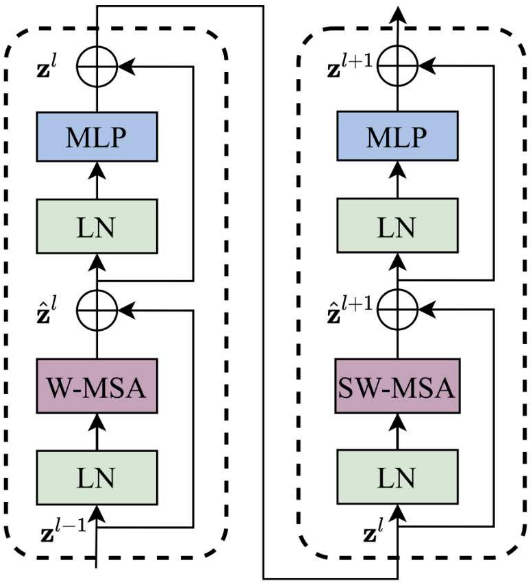


Figure 2.15: Swin Transformer block, extracted from Sharma et al. [132], showing two consecutive blocks.

The Swin Transformer V2 is an optimized version of the original Swin Transformer architecture, focusing on addressing three key limitations: unstable training, discrepancies in input image sizes between pretraining and fine-tuning, and high GPU consumption. To improve training stability, the authors relocated the normalization layer from the beginning of the Swin Transformer block to the end, thereby stabilizing the information flow from the residual connections. Additionally, they introduced a cosine-based self-attention mechanism, which makes the attention mechanism insensitive to magnitude. To minimize the disparity between pretraining and fine-tuning inputs, the authors implemented log-spaced coordinates, enabling input conversion into a consistent space and reducing the gap between stages. Finally, they adopted several strategies to optimize GPU usage, including zero optimizer, activation checkpointing, and sequential self-attention.

2.9.3 Swin-UNet

Proposed by Cao et al. [25], the Swin-UNet model is a pure Transformer-based architecture inspired by the U-shaped encoder-decoder structure of U-Net. By using Swin Transformer blocks, the authors designed an encoder that projects features into an arbitrary dimensional space C, which are subsequently passed through multiple Swin Transformer blocks.

Patch merging layers are employed in the encoder stage for down-sampling the inputs, while patch expanding layers are utilized in the decoder stage for up-sampling. The decoder mirrors the encoder in a symmetric fashion, with both components connected via skip connections, as is customary in U-Net-based networks, as illustrated in Figure 2.16.

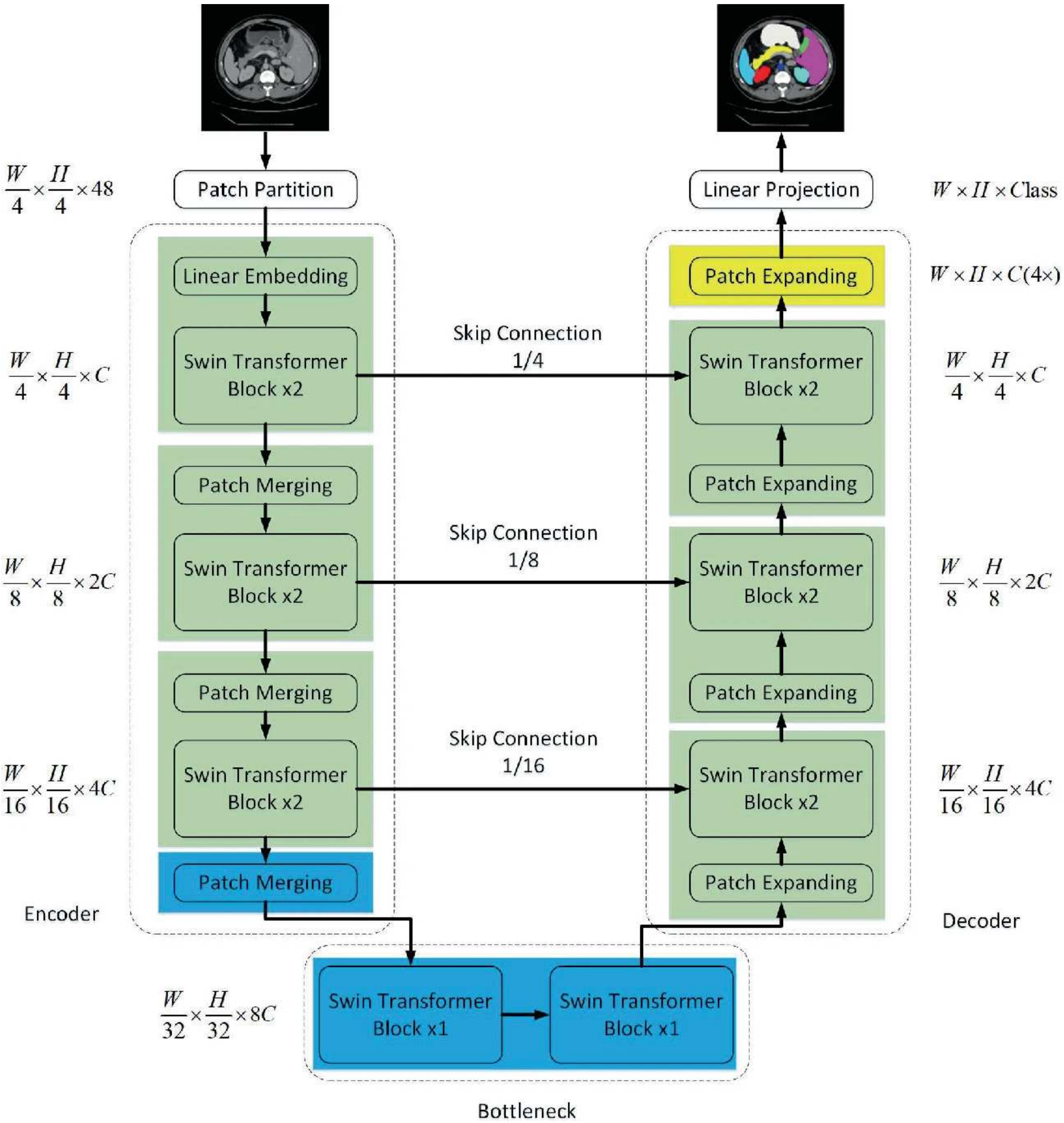


Figure 2.16: Swin-UNet architecture, extracted from Cao et al. [25].

The encoder begins by feeding the C-dimensional tokenized inputs into two consecutive Swin Transformer blocks to learn feature representations without altering the feature dimension or resolution. Following this, a patch merging layer reduces the number of tokens by a factor of 2, down-sampling the input while doubling the feature dimension. This process is repeated three times, as shown in Figure 2.16.

The patch merging layer operates by dividing the input into four regions and concatenating them, thereby reducing the spatial resolution by a factor of 2. However, as this concatenation increases the feature dimension by a factor of 4, a linear layer is then applied to adjust the feature dimension to double the original, rather than quadrupling it.

At the bottleneck of this U-shaped architecture, two Swin Transformer blocks are employed to learn deep feature representations while maintaining both their resolution and dimensionality. The authors explain that this choice addresses the challenge of Transformers being difficult to converge when too deep. Additionally, skip connections are utilized to integrate multi-scale features from the encoder with the up-sampled features in the decoder. These skip connections are followed by a linear layer to ensure the concatenated feature dimension matches the up-sampled features.

For the decoder, the authors adopt a symmetric design based on Swin Transformer blocks, incorporating patch expanding layers for up-sampling the deep features. These layers reshape the feature maps to double the resolution while halving the feature dimension.

Specifically, the patch expanding layer applies a linear layer to the input features, doubling their dimensionality. This is followed by a rearrange operation that expands the resolution by a factor of 2 and reduces the feature dimension by a factor of 4, achieving the desired spatial up-sampling.

2.10 Bias

The receptive field bias in Convolutional Neural Networks (CNNs) stems from their reliance on local convolutional filters and pooling operations, which inherently focus on capturing spatially local patterns and relationships within a fixed neighborhood, as shown in Figure 2.5. This bias ensures that CNNs are efficient for tasks with strong spatial structures, such as image processing, where local dependencies and translation invariance are important.

In contrast, Transformers operate with minimal inductive bias, relying instead on a self-attention mechanism that dynamically weighs the relevance of all elements in the input sequence, regardless of their spatial or sequential proximity. This approach enables Transformers to model global context and long-range dependencies effectively but comes at the cost of requiring larger datasets and computational resources to learn these relationships from zero. While CNNs are naturally tailored for local feature extraction, Transformers offer greater flexibility, making them suitable for tasks with complex, non-local patterns.

2.11 Evaluation and Metrics

During the training process, the network weights are iteratively updated to reflect patterns present in the input data. After each epoch, the model's performance is evaluated using a validation dataset to assess progress and identify the need for adjustments.

This evaluation is conducted by monitoring a validation metric and minimizing the loss function, ensuring that the model's learning remains aligned with the desired objectives.

2.11.1 Loss Function

The loss function utilized for evaluating the segmentation models in this dissertation is Binary Cross-Entropy (BCE), we opted for this metric since our preliminary results showed that it was adequate to train the networks, a widely used approach for binary classification tasks. This method is particularly suitable when distinguishing between two classes, such as classifying a pixel as either part of a spheroid or the background.

A single neuron suffices to represent this binary information, as it can output a value of 1 (indicating spheroid) or 0 (indicating background). Here, y_i denotes the true label for data point i, and p_i represents the predicted probability for the same data point.

The BCE is computed using Equation 2.2, where y refers to the ground truth and p corresponds to the network's prediction.

$$\mathrm{L} = -y\ \log(p) - (1-y)\ \log(1-p). \tag{2.2}$$

2.11.2 Metrics

The Dice coefficient [35], also known as the F1-score, is a statistical metric commonly used to evaluate the performance of segmentation models. It measures the overlap between the predicted segmentation and the ground truth, providing a value between 0 and 1, where 1 indicates perfect agreement and 0 signifies no overlap, illustrated in Figure 2.17.

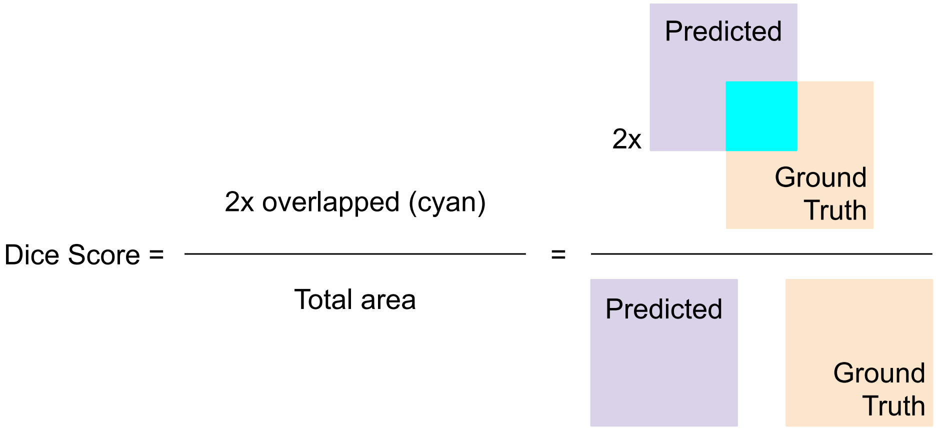


Figure 2.17: Diagram illustrating the Dice score, in which it doubles the overlap between prediction and groundtruth, and divides it by the sum of their areas.

The Dice coefficient is particularly useful in tasks involving imbalanced data, as it emphasizes the correct classification of smaller regions. Mathematically, it is defined as twice the intersection of the predicted and true regions divided by the sum of their individual areas, ensuring a balance between precision and recall as shown in Equation 2.3. This makes it a robust and interpretable metric for assessing segmentation of spheroid, specially its borders, since it emphasizes correctness over small regions [35].

$$\mathrm{Dice} = \frac{2|X \cap Y|}{|X| + |Y|}. \tag{2.3}$$

Chapter 3

Related Work

In this chapter, we present a curated review of the scholarly works and prior research that enabled and support this thesis, and served as the material for the survey we created to explore our study field. This chapter highlights the significant contributions from the field of image segmentation, cancer spheroid analysis, and the application of deep learning techniques, establishing a comprehensive background for our study. This review serves to validate our approach, illustrating how our work builds upon and extends existing knowledge, while also addressing critical gaps that needs further exploration.

To investigate the scope of our field, we searched for studies that utilized any form of image segmentation to analyze cancer spheroid cultures. This search was conducted using relevant databases such as PubMed, IEEE Xplore, Scopus, and Web of Science. We limited our review to works published after 2015 and written in either Portuguese or English.

The selection consisted of multiple screening stages. Initially, studies were excluded based on their titles alone, eliminating irrelevant topics. In the second stage, both titles and abstracts were reviewed to identify studies likely to have employed segmentation in their analysis or methodology, as some biology-focused papers did not explicitly mention this processing. Finally, a full-text reading was conducted on the articles, as illustrated in Figure 3.2.

The studies included differed in the level of detail provided about their segmentation approaches. From each selected study, we extracted these variables to guide our analysis.

- **Segmentation Approach**: Which technique was employed to segment the cell culture.
- **Cell Line**: The specific cancer type or the name of the cell line used by the authors.
- **Imaging Method**: The modality used to capture the input images.
- **Dataset Accessibility**: Whether the dataset used in the study is publicly available.
- **Study Objective**: The paper's goal.

To collect this information, each paper was thoroughly examined to extract statements from the authors about the key variables. Subsequently, the papers were categorized according to their segmentation methods. We used the open-source software Zotero [47] for collecting, annotating, and organizing the literature.

3.1 Literature Sorting

A total of 3, 219 papers were initially identified over the selected databases. Figure 3.1 highlights an increasing publication trend over the years, reflecting a rising interest within the scientific community in the interdisciplinary field of 3D cell cultures and image segmentation.

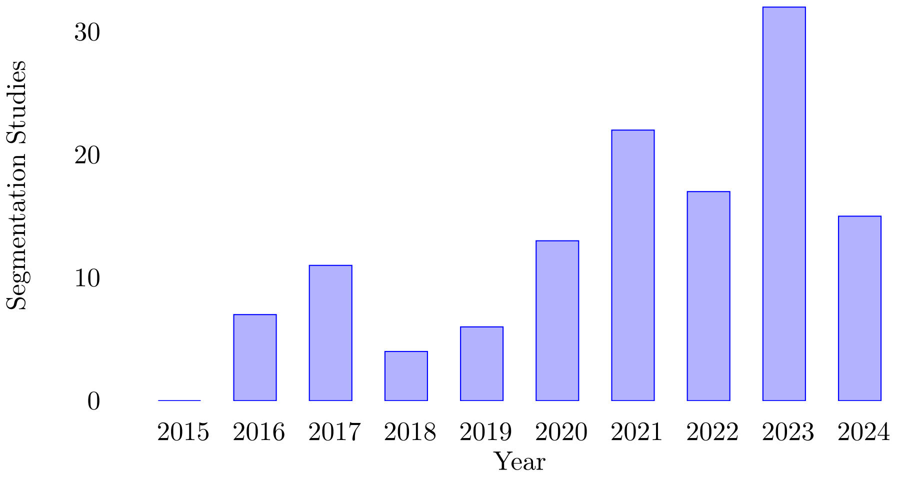


Figure 3.1: Distribution of studies per year, illustrating an increasing interest of cancer-spheroid segmentation papers.

After removing duplicates across databases, the number of unique papers dropped to 2, 938. We then applied a filter to include only studies published after 2015, resulting in 2, 382 papers. Two rounds of preliminary screening based on titles and abstracts further reduced the pool, as shown in the PRISMA diagram (Figure 3.2). Ultimately, 118 papers advanced to the full-text review phase.

The studies were classified by segmentation method, yielding the following results: 6 papers utilized manual segmentation, 48 employed image processing pipelines, 30 used specific software, and 36 relied on deep learning models. Figure 3.3 illustrates the distribution of these methods. Details of each method and the associated studies are provided in the next sections, starting with the identified datasets.

3.2 Public Datasets

Publicly accessible datasets are essential for reproducibility and the development of new models. However, fully open datasets remain scarce. Table 3.1 provides a summary of the datasets identified, including their sample sizes and availability status. The “Availability” column includes the following categories: “Yes” indicates that both images and annotations are freely downloadable; “Yes*” means the dataset is reportedly available,

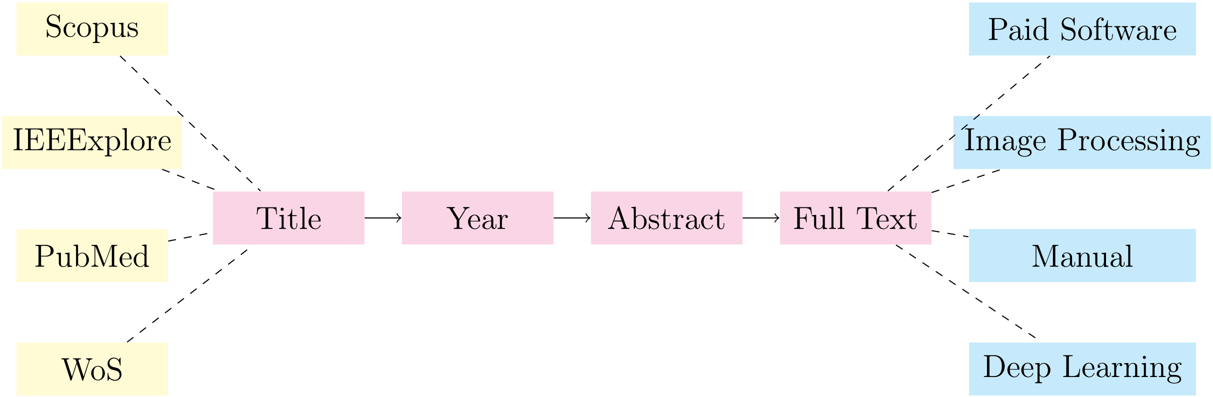


Figure 3.2: PRISMA diagram outlining the data extraction process. The initial selection was based on titles, followed by the exclusion of studies published before 2015. This was followed by abstract screening and a thorough full-text review. In the final stage, the papers were classified according to their segmentation method (manual, image processing, specific software, or deep learning).

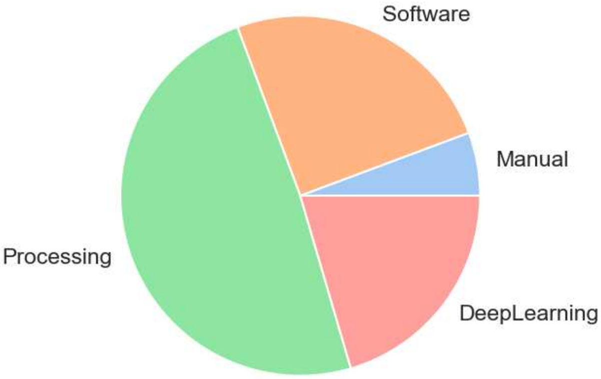


Figure 3.3: Segmentation methods found in the related literature.

but access could not be confirmed; and “Request” denotes datasets that can be obtained by contacting the authors. Datasets that were entirely inaccessible are excluded from the list.

In the rare cases where datasets were openly available with functional download links, we found minimal information about the data in the corresponding publications, although there were a few notable exceptions. Additionally, the number of available samples is relatively small compared to other deep learning datasets. This scarcity is largely due to the significant effort required to maintain cell cultures, achieve consistency across plate wells under controlled conditions, and perform the labor-intensive task of annotating each sample for segmentation.

Table 3.1: Datasets in the literature pertaining cancer spheroid images for segmentation.

Authors	**# Images**	**Available**	**Year**
Liu et al. [86]	-	Yes*	2024
Lee et al. [83]	-	Request	2024
Abd El-Sadek et al. [1]	-	Request	2024
Leite et al. [84]	294	Yes	2023
Mukashyaka et al. [101]	-	Yes*	2023
Tebon et al. [146]	100	Yes*	2023
Kaseva et al. [74]	12	Yes*	2022
Matthews et al. [92]	66	Yes*	2022
Winkelmaier and Parvin [164]	-	Yes*	2021
Spiller et al. [139]	-	Yes*	2021
Lacalle et al. [80]	-	Yes	2021
Diosdi et al. [40]	-	Yes*	2021
Karabag et al. [72]	517	Yes*	2019
Ahonen et al. [3]	-	Request	2017

3.3 Found Literature

We classified the segmentation methods into four primary categories: image processing, specific software, deep learning, and manual segmentation. Manual segmentation, in this context, involves human annotation of spheroid surroundings, typically using image editors or similar tools.

As noted by Deckers et al. [37], annotating and analyzing each sample manually takes approximately 80 to 100 seconds. While this may initially seem efficient, the time required accumulates rapidly in small-scale experiments and becomes infeasible for high-throughput testing. Consequently, only a handful of studies employed this method [14, 38, 73, 94, 118, 132]. More scalable alternatives, such as image processing pipelines, provide faster solutions for segmentation.

3.3.1 Image Processing

Image processing uses algorithms to perform segmentation, but these algorithms often fall short when used alone, requiring additional pre- or post-processing steps. Together, these elements form what is known as a segmentation pipeline.

Most of the reviewed studies (48 papers) utilized image processing for segmentation, as detailed in Table 3.2. This table lists the authors, the specific algorithms used, image types, cell-culture lines, and publication year, arranged in descending order by year. While various image types were mentioned, the primary focus was on brightfield and fluorescent images.

Among the algorithms, Otsu's method was the most prevalent, followed by the Watershed algorithm. Fluorescent images were frequently used due to their effectiveness in segmentation, as they produce high-contrast images where regions of interest stand out as bright spots against a dark background. In these images, pixels exceeding a low threshold are typically identified as part of the target region.

Table 3.2: Studies that applied image processing algorithms for segmentation. The table columns list the authors, the algorithm used, the input image type, the cancer culture studied, and the publication year. An asterisk (*) indicates methods applied without additional pre- or post-processing. Abbreviations include: Brightfield (BF), Fluorescent (FL), Phase-Contrast (PC), Mass Spectrometry (MS), and Differential Interference Contrast (DIC).

Authors	**Algorithm**	**Image Type**	**Cancer Culture**	**Year**
Chang et al. [27]	K-Means	Fluorescent	Bone	2024
Van Hemelryk et al. [150]	-	Confocal	Prostate	2023
Mendonca et al. [95]	BG Subtraction	Light-Sheet	Breast	2023
L. Fillioux et al. [79]	Canny Edge	Brightfield	Colon	2023
Dimitriou et al. [39]	Watershed	Confocal	Breast	2023
Chen et al. [29]	Watershed	Confocal	Liver	2023
Akbaba et al. [5]	Simple Threshold	MiniOpto Tomography	Pancreas	2023
Petrovic et al. [113]	Otsu	Darkfield	Lung	2022
Deckers et al. [37]	Otsu*	Brightfield	hPDCs	2022
Chambost et al. [26]	Otsu	Brightfield	Brain	2022
Schurr et al. [130]	Otsu*	Fluorescent	-	2021
Ndyabawe et al. [104]	Hysteresis	Confocal	Brain	2021
Hof et al. [60]	Watershed*	Light-Sheet, BF	Pancreas, Bile Duct	2021
Grosser et al. [55]	Watershed*	Confocal	Breast, Cervical	2021
Gillette et al. [52]	Otsu	Fluorescent	Pancreas	2021
Gil et al. [51]	Otsu	One-Photon Redox	Colorectal	2021
Diosdi et al. [40]	Otsu	Fluorescent	Breast	2021
Alsehli et al. [6]	Subtraction	Fluorescent	hiPSC Stem	2021
Wardwell-Swanson et al. [162]	Color Segmentation	Confocal	Lung, Gastric	2020
Schuster et al. [131]	-	Phase Contrast	Pancreas	2020
Nürnberg et al. [106]	Watershed	Confocal	Colon and others	2020
Henser-Brownhill et al. [58]	Fuzzy C-Means	Ptychographic	Melanoma	2020
Favreau et al. [45]	Fuzzy C-Means	Selective Plane	Colorectal	2020
Edwards et al. [42]	Simple Threshold	Light-Sheet	Renal	2020
Tobias et al. [147]	Bisecting K-Means	Mass Spectrometry	Colon	2019
Murali et al. [103]	Simple Threshold	Fluorescent	Melanoma	2019
Michálek et al. [97]	Triangle Threshold	Confocal, MS	-	2019
Kovac et al. [77]	Ellipsoid Fit	Synthetic Images	-	2019
Keller et al. [75]	Simple Threshold	Brightfield, Confocal	Breast	2019
Karabag et al. [72]	Canny Edge	Face Scanning Electron	Cervical	2019
Hou et al. [61]	Graph Cut	Confocal	Lung	2018
Boutin et al. [16]	Watershed	Fluorescent	Breast, Glioblastoma	2018
Borten et al. [15]	Adaptive Threshold	DIC	Breast and others	2018
Wan et al. [158]	Entropy Threshold	Brightfield, FL	Endothelial, Ovarian	2017
Smyrek and Stelzer [138]	Simple Threshold	Light-Sheet	Glioblastoma	2017
Schmitz et al. [128]	Otsu	Light-Sheet	Breast	2017
Reijonen et al. [122]	Watershed	Confocal	Liver	2017
Piccinini et al. [114]	Otsu	Light-Sheet	Lung	2017
Moriconi et al. [100]	Frangi Filtering	Brightfield	Glioblastoma	2017
Cannon et al. [24]	Otsu	Fluorescent	Breast	2017
Bulin et al. [22]	Otsu	Confocal	Pancreas	2017
Ahonen et al. [3]	Local Entropy Filter	Confocal	Lung	2017
Walsh et al. [157]	Simple Threshold	Fluorescent	Pancreas	2016
Jagiella et al. [68]	Watershed	Brightfield, FL	Lung	2016
Cisneros Castillo et al. [34]	-	Brightfield	Glioblastoma	2016
Cheng et al. [32]	Simple Threshold, K-Mean	Brightfield, FL	Breast	2016
Bilgin et al. [13]	Watershed*	Fluorescent	Mammary	2016

Some methods in Table 3.2 are marked with an asterisk (*), indicating that the studies used only the core algorithm without any additional processing steps, which may compromise segmentation quality. Notably, few studies, such as Lacalle et al. [80], published both datasets and segmentation metrics, enabling meaningful performance comparisons.

For building segmentation pipelines, several tools are available, with FIJI software being particularly popular. FIJI [127], a distribution of the open-source platform ImageJ [129], allows users to design and automate image processing workflows. It supports batch processing and includes useful plugins for enhancing functionality, along with basic image editing features.

The widespread use of image processing methods likely stems from their accessibility and ease of use. Tools such as FIJI are free and come with extensive online tutorials, making them approachable for users with varying skill levels. Additionally, these methods offer a practical balance between automation and manual adjustments, significantly reducing the time needed for image analysis even when results are imperfect.

3.3.2 Specific Software

Specific software methods can be considered a subset of image processing, as they also involve applying pipelines to images. However, these pipelines are often pre-configured within laboratory equipment, such as microscopes, or locked behind paywalls. Due to their limited customizability, restricted accessibility, and black-box nature, we have categorized them separately.

We identified 30 studies employing this approach, summarized in Table 3.3. The results are diverse, with no single software emerging as dominant. Noteworthy examples include Ibrahim et al. [65], which used the Fiji WEKA plugin. Although tools such as WEKA are technically plugins, their design as black-box solutions justifies their separate classification.

3.3.3 Deep Learning

Deep learning models have achieved significant success in biological image segmentation tasks, including cancer spheroid segmentation. As a result, many studies have explored the use of deep neural networks for this purpose. For this category, we include any study that utilizes a deep learning architecture for segmentation, whether by training from scratch or using pre-trained models.

Notably, one study proposed both an image processing pipeline and a deep learning method. However, since the primary focus was on the deep learning approach, we categorized it accordingly. Table 3.4 summarizes the 36 studies that employed deep learning models for cancer spheroid segmentation, along with their chosen architectures.

U-Net and its variations were the most commonly used architectures, given their effectiveness in biological segmentation tasks [125]. Additionally, several studies implemented post-processing steps to improve results, addressing issues such as removing small artifacts or refining boundaries identified by the model.

Despite the growing success of Transformer-based models across artificial intelligence

Table 3.3: Summary of studies utilizing specific software for segmentation, including commercial software, embedded tools, and plugins for other platforms. The table columns provide the authors, the software used, the input image type, the cancer culture examined, and the publication year. Abbreviations: Brightfield (BF), Fluorescent (FL), and Phase-Contrast (PC).

Authors	**Software**	**Image Type**	**Cancer Culture**	**Year**
Sun et al. [141]	Fiji WEKA	Confocal	Bladder	2024
Lee et al. [83]	AnaSP	Brightfield	Breast	2024
Abd El-Sadek et al. [1]	Fiji FCR	Confocal	Gastric	2024
Wang and Hummon [161]	SCiLS Lab	Mass Spectrometry	Colon	2023
Wang et al. [159]	Imaris	Confocal	Breast	2023
Sinenko et al. [137]	CeelPose	Fluorescent	Breast	2023
Ramm et al. [121]	CellProfiler	Brightfield	Breast, Prostate	2023
Hu et al. [63]	CellProfiler	Fluorescent	Breast	2023
Hu et al. [62]	CellProfiler	Fluorescent	Breast	2023
Xie et al. [169]	SCiLS Lab	Mass Spectrometry	Colon	2022
Tanaka et al. [144]	Neurolucida	Confocal	Liver	2022
Powell et al. [116]	Pipeline Pilot	Brightfield	Colorectal	2022
Perini et al. [112]	INSIDIA	Brightfield, FL	Brain, Pancreas	2022
Lotsberg et al. [91]	Ilastik	-	Lung	2022
Koch et al. [76]	MorphoLibJ	Brightfield	Liver	2022
Kang et al. [70]	Imaris	Confocal	Pancreas	2022
Ibrahim et al. [65]	Fiji WEKA	Brightfield	-	2022
Spiller et al. [139]	Harmony PerkinElmer	Brightfield	Colorectal	2021
Sargenti et al. [126]	Nikon NIS Elements AR	Confocal	Colon	2021
Choo et al. [33]	CellProfiler	Brightfield, FL	Prostate	2021
Berg et al. [11]	MINS (MatLab)	Confocal	Liver	2021
Aguilar Cosme et al. [2]	AnaSP	Inverted Light	Melanoma	2021
Zanotelli et al. [171]	CellProfiler	Mass Cytometric	Colorectal and others	2020
Tasnadi et al. [145]	3D Active Surface	-	-	2020
Shirai et al. [133]	CL-Quant	Phase Contrast	Renal	2020
Liu et al. [87]	SCiLS Lab	Mass Spectrometry	Colorectal, Liver	2018
Veelken et al. [153]	Automated Cellular Analysis	Confocal	Melanoma	2017
Mittler et al. [99]	HCS Studio	Brightfield	Prostate	2017
Garvey et al. [49]	CellTracker	Confocal	Lung	2016
Barbier et al. [9]	DIPimage (MatLab)	Confocal	Prostate	2016

applications, only a few studies applied them to cancer spheroid segmentation. This suggests that while Transformers are popular in fields such as natural language processing and image classification, their use in biological image segmentation is still rising.

Many studies treated deep learning models as black-box systems, which can limit the benefits of fine-tuning. Optimizing architectures or tailoring loss functions could significantly enhance performance. By not fully customizing these models, researchers may miss opportunities to boost accuracy and efficiency in segmentation tasks, especially in the biological domain.

Table 3.4: List of studies that applied deep learning for segmentation, including authors, deep learning architectures, input image types, cancer cultures, and publication years. Studies using multiple methods are included here, with an emphasis on deep learning approaches. Authors who made minor modifications to U-Net are also categorized under U-Net. Abbreviations: Brightfield (BF), Fluorescent (FL), and Phase-Contrast (PC).

Authors	**Architecture**	**Image Type**	**Cancer Culture**	**Year**
Liu et al. [86]	3D U-Net + Watershed	Confocal	Gastric	2024
García-Domínguez et al. [48]	U-Nets, GANs	Brightfield, FL	Brain, Colon	2024
Zhang et al. [174]	AU2Net	Brightfield	Bladder	2023
Zhang et al. [173]	EGO-Net	OCT	Colon	2023
X. Deng et al. [168]	MacrOrga	Brightfield	Pancreas and others	2023
X. Deng et al. [167]	CAMPEOD	Brightfield, PC	Pancreas	2023
Vong et al. [156]	U-Net	-	Brain	2023
Tebon et al. [146]	U-Net	Interferometry	Breast	2023
Shuyun et al. [134]	-	Spectroscopy	Liver	2023
Rieken Münke et al. [123]	U-Net	Brightfield	Cervix	2023
Qin et al. [119]	TransOrga	Brightfield, PC	Pancreas and others	2023
Park et al. [108]	U-Net	Brightfield, FL	Colon	2023
Ngo et al. [105]	EfficientNet	-	Breast	2023
Mukashyaka et al. [102]	Stardist-3D	Confocal	Breast	2023
Mukashyaka et al. [101]	Stardist-3D and others	Confocal	Breast	2023
Maylaa et al. [93]	U-Net, YOLOv5	Brightfield	Cervical	2023
Jadav et al. [67]	3D U-Net	Volume Electron	Lung	2023
Leite et al. [84]	U-Net and others	Brightfield	Stomach and others	2023
Erdem et al. [44]	U-Net	Phase-Contrast	Breast	2023
Beydag-Tasöz et al. [12]	Stardist-3D	Light-Sheet	Pancreas	2023
Bao et al. [8]	U-Net, EGO-Net	OCT	Stomach and others	2023
Wang et al. [160]	RDAU-Net	Brightfield	Bladder	2022
Merivaara et al. [96]	U-Net + Watershed	Confocal	Prostate, Liver	2022
Matthews et al. [92]	U-Net	Brightfield, PC	Pancreas and others	2022
Kaseva et al. [74]	U-Net, 3D U-Net	Confocal	Liver	2022
Bruch et al. [19]	U-Net, 3DResU-Net	Confocal	Lung, Pancreas	2022
Beghin et al. [10]	U-Net	Light-Sheet	Pancreas	2022
Yao et al. [170]	AD-GAN	Confocal	Lung, Cervical	2021
Winkelmaier and Parvin [164]	U-Net	Confocal	Breast	2021
Wen et al. [163]	U-Net + Watershed	Fluorescent	Cervical	2021
Lacalle et al. [80]	U-Net and others	Brightfield, FL	Brain, Colon	2021
Grexa et al. [54]	Mask R-CNN and others	Brightfield, FL	Breast, Liver	2021
Chen et al. [30]	PSP-U-Net	Brightfield	Breast, Colon, Lung	2021
Chen et al. [31]	U-Net	EIT Sensor	Breast	2021
Rahkonen et al. [120]	U-Net	Phase Contrast	Prostate	2020
Bruningk et al. [20]	U-Net	Brightfield, FL	Intestine, Tongue	2020

Chapter 4

Materials

In this chapter, we provide a detailed description of the various materials and resources utilized in our methodology. This chapter outlines the specific datasets, imaging techniques, and computational tools that support our segmentation method and analysis.

A key component of this chapter is the introduction of our proposed dataset, which includes a collection of annotated cancer spheroid images that have been meticulously curated to support high-throughput quantitative analysis. We elaborate on the characteristics of the dataset, including the diverse range of spheroid sizes, shapes, and imaging conditions, as well as the annotation process used to generate accurate segmentation masks.

As discussed in Chapter 3, a significant challenge in this field is the scarcity of publicly available as well as the high-quality annotated datasets. Many studies rely on small, internal datasets and often do not make these datasets accessible to the broader research community. A notable exception is the work by Lacalle et al. [80], which graciously responded to our inquiries and provided access to their dataset, representing one of the few instances of data sharing in this area.

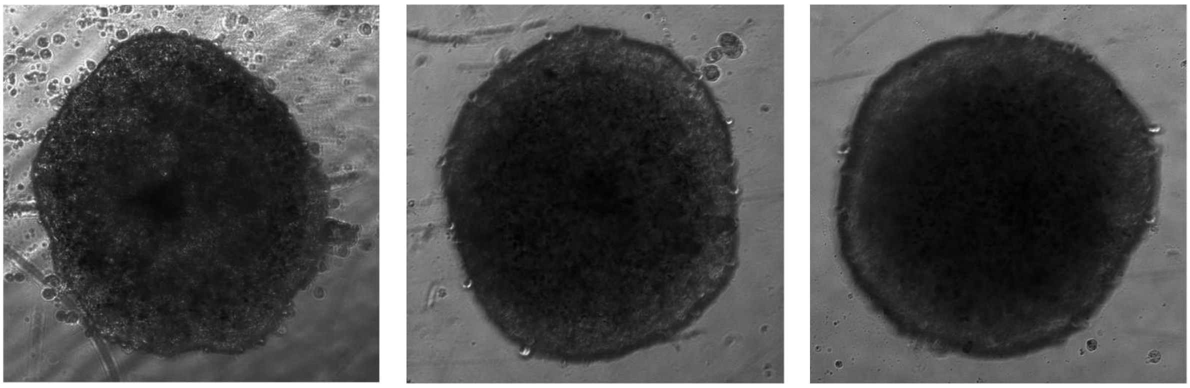
Given this landscape, we committed to creating a publicly available dataset of cancer spheroid images accompanied by high-quality segmentation masks, ensuring minimal barriers for other researchers to utilize the data for training their models. To facilitate accessibility, our dataset is well-organized and freely available on our GitHub page, with no restrictions on access or usage.

4.1 Our Dataset

Our dataset[1] comprises brightfield images of gastric cancer spheroid cultures, the segmentation mask annotations that we created, and an accompanying Comma-Separated Values (CSV) file detailing metadata. The CSV file includes information such as the sample file name, the spheroid's lifetime at the time of capture, the cell line used, the microscope's optical lens magnification, the well coordinates where the culture was formed, the scientist responsible for capturing the image, the date the image was taken, and a difficulty score indicating the level of challenge in segmenting the sample.

[1] `https://github.com/guilhermevleite/dataset-spheroid-segmentation`

Initially, our dataset consisted of 294 images and corresponding masks, provided without any data augmentation. To optimize the dataset for deep learning applications, we incorporated data augmentation and established a train/test protocol. This protocol enhances reproducibility and mitigates the risk of data contamination across different training sessions. We allocated 80% of the samples for training and 20% for testing, ensuring a balanced distribution between the two sets. A summary of the dataset is presented in Table 4.1, and Figure 4.1 illustrates some of our samples.



(a) 48 hours (b) 72 hours (c) 96 hours

Figure 4.1: Examples of brightfield images from our dataset illustrate the same culture captured at 48, 72, and 96 hours after treatment application. These images highlight the compaction process that occurs as the spheroid develops. As cells aggregate more densely, the central region of the spheroid becomes darker, while the borders remain lighter, reflecting the structural changes over time.

Table 4.1: Features from our dataset.

Samples	**Masks**	**Lens**	**Cell Line**	**Capturing Time**	**Resolution**	**Colorspace**
294	294	10×	Kato-III	48h, 72h, 96h	1600×	Grayscale

4.1.1 Cancer Cell Culture

Our dataset consists of gastric cancer cell cultures, specifically the Kato-III cell line. Initially, the cells were cultivated in a monolayer using DMEM medium supplemented with 10% fetal bovine serum, 100U/mL penicillin, and 100μg/mL streptomycin. The cultures were maintained in a humidified atmosphere at 37°C with 5% CO_2. To ensure the integrity of the experiments, the cells were routinely tested for mycoplasma contamination.

We chose to form the spheroids using a magnetization-based approach. For this, we utilized the Bio-Assembler n3D assay from Greiner (Biosciences, Houston/TX) [36], following the manufacturer's protocol. After culturing the Kato-III cells in a monolayer (2D model) for 24 hours to reach approximately 80% confluence, we added Nanoshuttles (NS) at a ratio of 1.2μL per 10,000 cells. The cells were then incubated overnight at 37°C. Following incubation, the cells were detached and transferred into a 96-well cell-repellent plate, with 10.000 cells per well. The plate was placed on a magnetic drive and incubated for an additional 24 hours at 37°C with 5% CO_2 to promote spheroid formation. Finally, 48 hours post-spheroid formation, the cultures were either treated with 50μM of iRF or left untreated as controls. The entire process was documented and is illustrated in Figure 4.2.

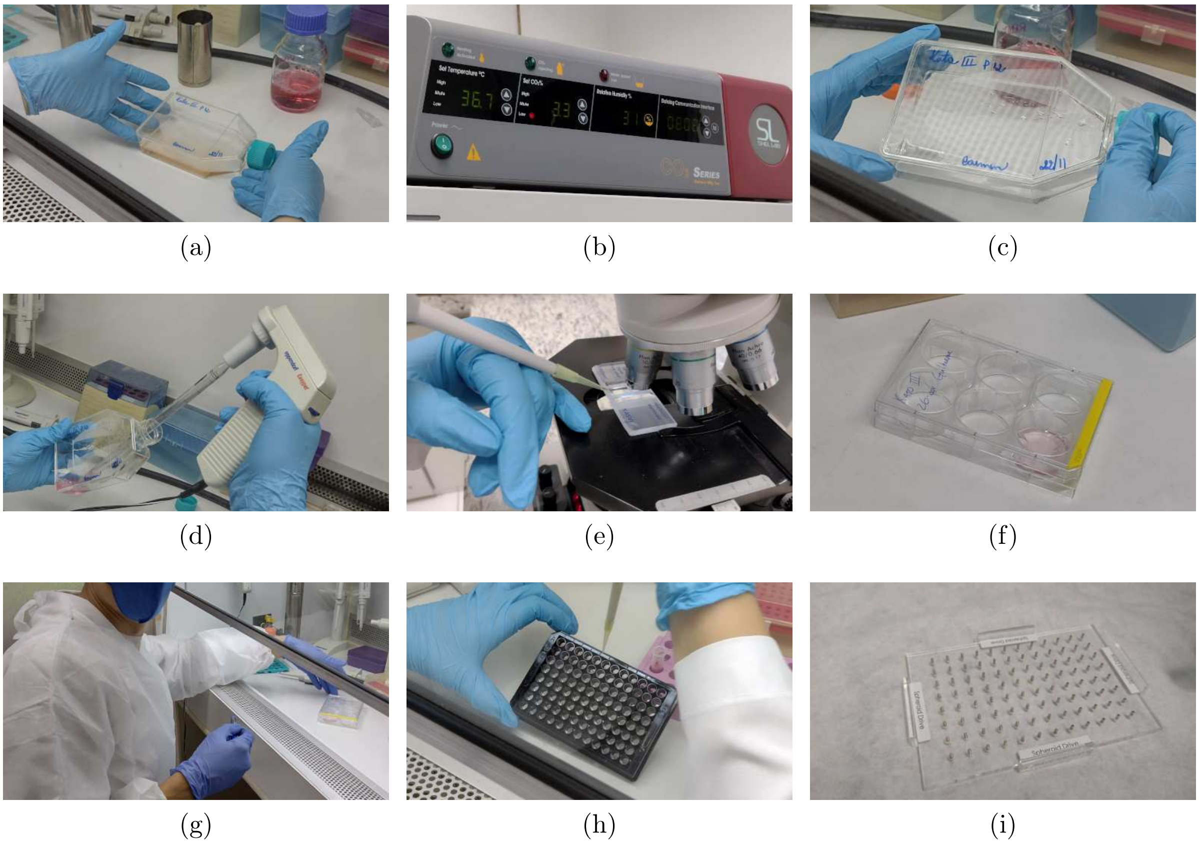

(a) (b) (c) (d) (e) (f) (g) (h) (i)

Figure 4.2: Spheroid formation. (a) The KATO-III cells were cultured in a bottle with DMEM medium supplemented with 10% fetal bovine serum, 1% penicillin-streptomycin (100 U/mL and 100μg/mL, respectively), and (b) maintained at 37°C in a humidified incubator with 5% CO_2. (c) When the bottle was 80% confluent, the cells were gently detached from the bottom using trypsin by (d) pipetting to achieve a single-cell suspension. (e) Cell concentration were determined using a hemocytometer. (f) 10,000 of cells were plated in a 6-well plate. (g) After 24h, metallic nanoparticles were applied. The magnetized cells were detached by tripsin after 24h, and (h) 10,000 were applied in a 96-well cell-repellent plate and placed atop on a (i) magnetic drive. After 24h, the drive is removed, and after more 48h, the spheroid is then formed.

The cell culture reagent DMEM (Dulbecco's Modified Eagle Medium), along with streptomycin sulfate and penicillin, were obtained from Nutricell (Campinas, SP, Brazil), while the fetal bovine serum was sourced from Gibco (Invitrogen, NY, USA).

4.1.2 Image Capture

To capture the spheroid images, we utilized a Lumascope light microscope, placing the 96-well plate containing the cultures described in Section 4.1.1 onto the microscope stage, using 10× optical lens. The Lumascope allows for the acquisition of both brightfield and fluorescent images; however, for the purposes of this thesis, only the brightfield imaging capability was employed. Image acquisition and digitization were facilitated using the Lumaview software, provided by the Lumascope manufacturer, to communicate with the microscope's camera.

Spheroids were sampled every 24 hours following treatment application. Due to the natural shape and size of the spheroids, it was not possible to simultaneously focus on both the nucleus and the borders. To address this, two images were captured per sample: one focused on the spheroid nucleus and the other on its borders. To minimize stress on the cultures, the plate was returned to the greenhouse every 30 minutes of capturing, maintaining conditions of 30°C and 5% CO_2. The images were digitized into grayscale Tag Image File (TIF) format at a resolution of 1600 × 1600 pixels using the Lumaview software. Table 4.2 shows the parameters used in the Lumaview software.

Table 4.2: Parameters used on the Lumeview software to capture the brightfield images of the spheroid cultures.

Brightness	**Gain**	**Exposure**
14.9	1.750	178.8

4.1.3 Mask Annotation

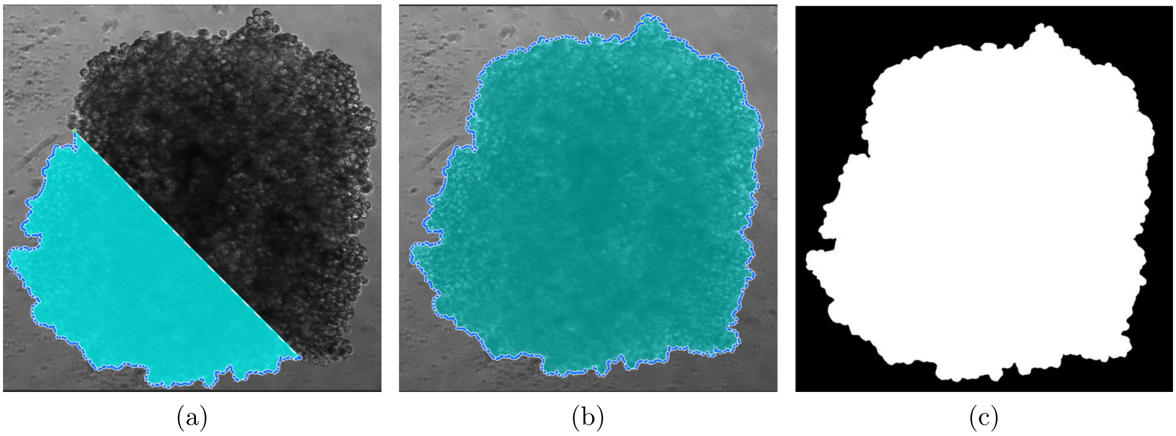
One significant contribution of our work is the creation of annotated cancer spheroid images, a resource that is both scarce and invaluable for training deep neural networks. To achieve our objective of quantifying the 2D area of spheroids, we manually annotated each sample to accurately delineate the regions corresponding to the cultures. This process involved using online tools to create segmentation lines directly on the original images.

As illustrated in Figure 4.3, we outlined the spheroids by placing vertices to form polygons that closely approximate their shapes. These polygons were then exported as binary mask images, where the background was represented in black and the regions of interest in white. To ensure accuracy, all initial annotations were subsequently reviewed by a specialist and corrected based on their recommendations. This whole process generated black and white images, with 1600 × 1600 pixels, which were saved in PNG format.

4.2 Lacalle's Dataset

Lacalle et al. [80] developed and published six datasets, three of which consist of brightfield images and the other three of fluorescence images, all accompanied by annotated segmentation masks. However, the associated publication did not provide a train/test protocol to facilitate reproducibility of their results. Key features of these datasets are summarized in Table 4.3.

The authors of the dataset annotated their images using ImageJ software, which produces annotation files containing only the vertices of the segmentation masks. To convert these annotations into usable image masks, we developed a Python script that reconstructs the segmentation masks from the vertices and saves them as image files. Given



(a) (b) (c)

Figure 4.3: Illustration of the annotation process, where a polygon was created to outline the spheroid shape and subsequently exported as a binary mask image. (a) Displays a partially annotated spheroid in cyan. (b) Shows the fully segmented spheroid. (c) Presents the final binary mask image.

Table 4.3: Summary of the datasets presented by Lacalle et al. [80], categorized by imaging method: brightfield and fluorescence. The second row provides the dataset names, while the third row, Samples, indicates the number of images available per dataset. The fourth row, Resolution, specifies the image resolution. The Microscope row identifies the brand of the microscope used for image acquisition, and the Lens row details the magnification of the lenses employed. The Format row describes the file format in which the images are stored, and the Type row specifies the colorspace of the images. Finally, the Culture row highlights the culture medium used to grow the spheroids (Suspended and Collagen).

Method	Brightfield			Fluorescence		
Name	**BL5S**	**BN2S**	**BN10S**	**FL5C**	**FL5S**	**FN2S**
Samples	50	154	105	19	50	34
Resolution	1296 × 966	1002 × 1004	1002 × 1004	1296 × 966	1296 × 966	1002 × 1004
Microscope	Leica	Nikon	Nikon	Leica	Leica	Nikon
Lens	5×	2×	10×	5×	5×	2×
Format	TIFF	ND2	ND2	TIFF	TIFF	ND2
Type	RGB	Grayscale	Grayscale	RGB	RGB	Grayscale
Culture	Susp.	Susp.	Susp.	Colla.	Susp.	Susp.

the binary nature of these datasets, the spheroid pixels were assigned a white color, while the background pixels were painted black. Figure 4.4 shows examples of the spheroid images alongside their reconstructed segmentation masks.

While working with the dataset, we discovered differences between many of the segmentation masks and their corresponding images. Upon contacting the authors, they provided access to a repository containing all their brightfield images, but none of the fluorescence data. Additionally, the organization described in Table 4.3 was absent from this repository. Instead, the data was divided into manually annotated samples and those annotated using a baseline algorithm developed by the authors. Consequently, rather than focusing on generating comparable metrics, our tests prioritize reproducing the method-

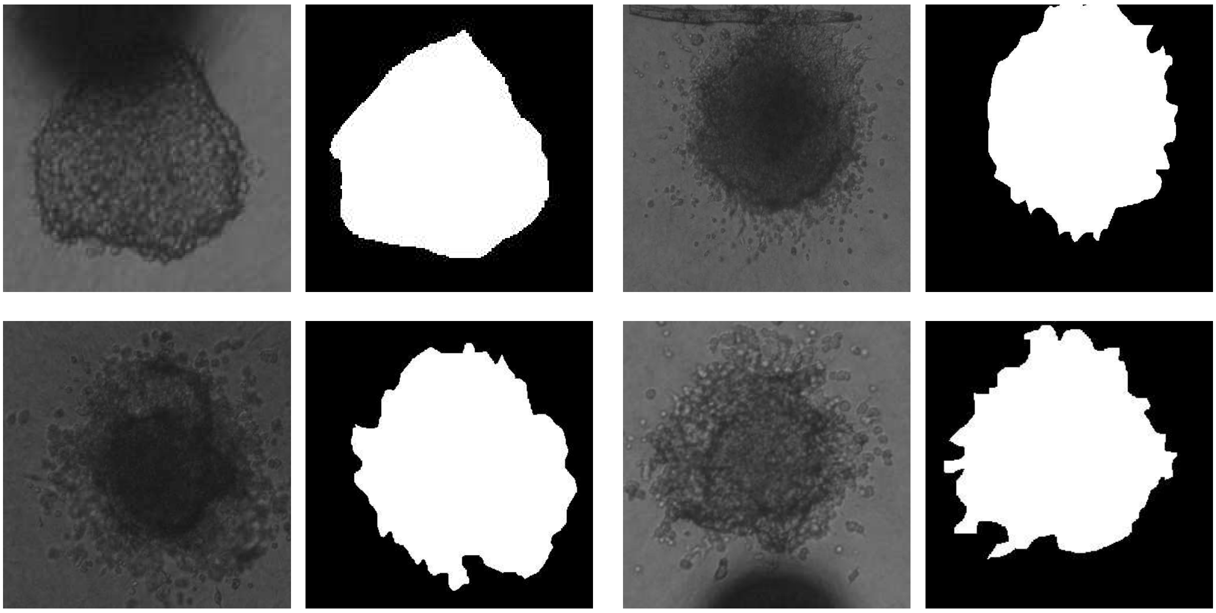


Figure 4.4: Examples of brightfield images from the dataset by Lacalle et al. [80]. The images have been cropped to enhance visualization, and the corresponding segmentation masks were generated using our Python script.

ology used in other studies.

4.3 Computational Resources

All tests were conducted using an NVIDIA GTX 1080 Ti GPU with Python [151] code incorporating may of its libraries. To minimize variability, we maintained as much as possible hyperparameters across all models, including the loss function, learning rate, learning rate scheduler, and early stopping criteria. Given the high computational demands of deep learning algorithms, which exceed the capabilities of standard laptops, some experiments were carried out in the cloud using Google Colab.

The cloud environment provided a system equipped with an NVIDIA Tesla P100 GPU (16 GB, CUDA 11.2) and an Intel Xeon CPU (2 GHz). Python's extensive adoption in scientific research has led to the development of numerous ready-to-use libraries and tools for deep learning. Consequently, our solution relied on Python [151] and its ecosystem of libraries, including SciPy [69], NumPy [56], PyTorch [109], Einops [124], and OpenCV [18].

Chapter 5

Cancer Spheroid Segmentation Method

In this chapter, we present a thorough description of our methodology developed for cancer spheroid segmentation, detailing each critical step involved in the process[1]. We begin by outlining the necessary pre-processing techniques applied to the images. Following this, we describe our training and testing setup, including the architecture of the segmentation model, the dataset splits, and the training protocols employed to optimize performance.

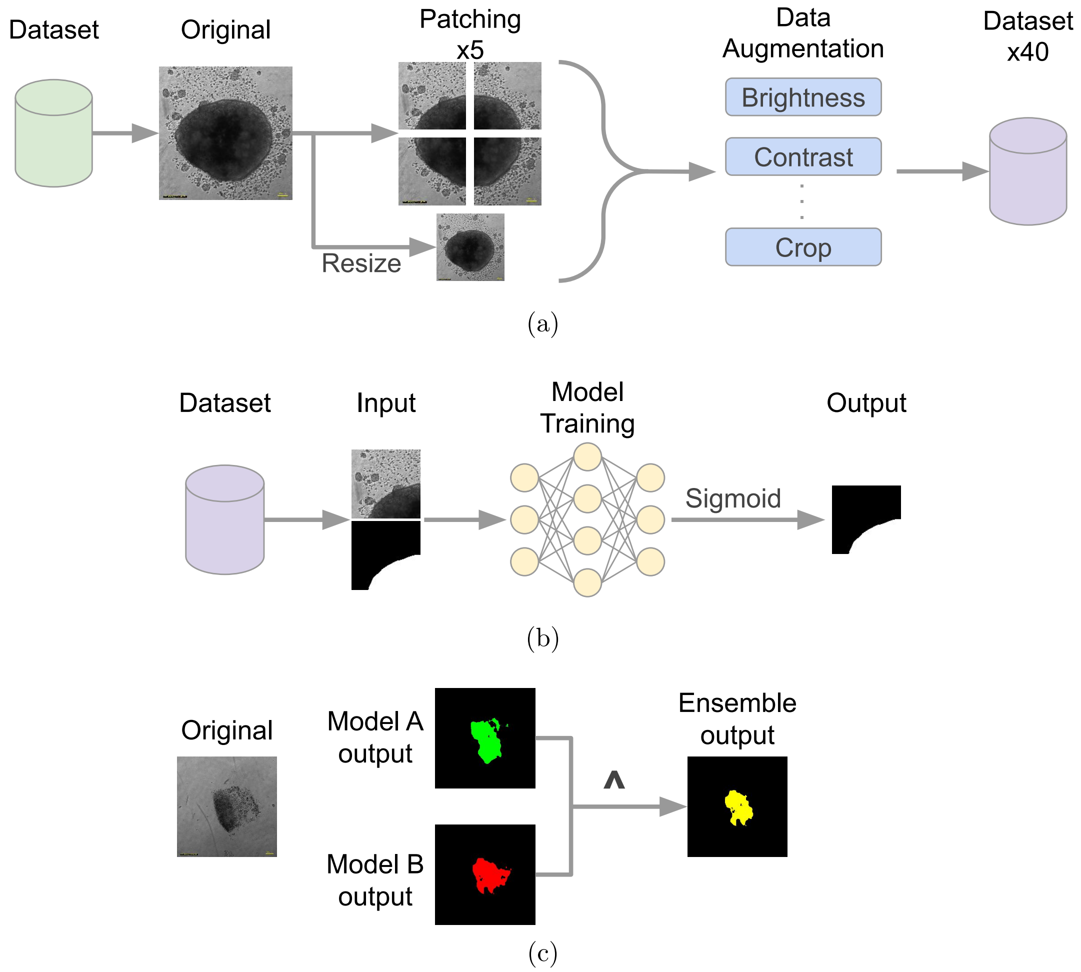
Figure 5.1 presents a diagram overviewing the proposed methodology for performing semantic segmentation on spheroid cell cultures. The process begins with an input spheroid image paired with its corresponding segmentation mask. Both the image and mask are resized and normalized to a standardized format before undergoing a series of image augmentation steps, detailed bellow. These augmented inputs are then utilized for model training, ensuring the network learns robust features across varying conditions.

To determine which architectures best fit the cancer spheroid task, we scanned the literature in search of adequate neural networks, in doing so we found two major paradigms to split our methodology: the CNN-based and the Transformer-based networks. We also selected on MLP-based architecture, because of its lightweight design. In conclusion, to perform the cancer spheroid segmentation process, we will apply the following architectures, sorted by their number of parameters in millions: starting with the MLP-based, UNeXt [149] (1.5m), followed by the CNN-based, MultiResUNet [66] (7.2m), U-Net [90] (7.7m), U-Net++ [175] (9m), and finally the Transformer-based ones, Swin-UNet [25] (41.35m), TransUNet [28] (105.32m), and SwinV2 [89] (197m). We detailed these architectures in Chapter 2.

As illustrated in Figure 5.1 (c), we introduce a novel approach for cancer spheroid segmentation, which is specifically designed to improve both accuracy and reliability. This method uses the power of an ensemble of networks to generate the segmentation mask. By utilizing multiple models within the ensemble, the method aims to make use of the strengths and unique characteristics of each network. To further optimize the process, our approach incorporates the output masks of pre-trained models, thereby reducing the need for retraining and maximizing the utility of existing learned features. By joining the outputs of these models, the proposed method combines different perspectives and learned features, producing a more robust and consistent segmentation outcome. This

[1] `https://github.com/guilhermevleite/spheroid_keras`

not only improves the precision of the segmentation process but also contributes to its trustworthiness, making it particularly well-suited for the detailed analysis required in cancer spheroid studies



(a)

(b)

(c)

Figure 5.1: Segmentation method overview. To save time on training time, we opted to save in disk our pre-processed dataset, as shown in (a) in which every input image was patched into four quadrants, and along side with the original image were resized to a more manageable resolution, quadrupling the size of our dataset. Afterwards, we executed eight strategies of data augmentation in those five images, generating a dataset 40 times bigger. This new dataset was saved to disk and used to train our models. (b) The training paradigm, in which a input image and its segmentation ground truth are loaded from disk and used to train the model. The model's output goes through a sigmoid function to generate a segmentation map, called output. (c) Our ensemble strategy, the late fusion majority vote, in which we join the output from different trained models to generate a new and better segmentation map.

The proposed method generates a segmentation mask that undergoes both quantitative and qualitative evaluation. Quantitatively, the Dice coefficient is used to measure the overlap between predicted and ground truth masks, providing an objective evaluation of segmentation accuracy. Additionally, the segmentation results are qualitatively reviewed by a specialist to ensure practical relevance and validate the model's performance, as detailed in Chapter 6. This combined evaluation approach ensures a comprehensive assessment of the method's effectiveness, which is proven applicable as we detail in Chapter 7.

5.1 Pre-Processing Phase

Deep learning models usually achieve optimal performance when trained on large datasets comprising thousands of samples. However, our dataset described in Chapter 4 consists of a limited number of highly detailed images. To address this limitation, pre-processing techniques can be employed to augment the dataset by generating additional data from existing samples, providing a cost-effective way to enhance model performance. Furthermore, resizing the input images to more manageable dimensions is a common practice to optimize hardware utilization without compromising efficiency.

Data augmentation encompasses techniques designed to increase the number of samples without incurring the cost of acquiring new ones. However, the selected augmentations must align with the specific context and scope of the dataset. For example, rotation is a commonly used augmentation that is suitable for cancer spheroid images but may not be applicable to datasets featuring car images. Considering these factors, we determined that the following operations were appropriate for cancer spheroid cultures. All augmentations were implemented using the Python library Albumentations [23].

To accelerate the training process and effectively increase the number of samples, all augmentations were pre-generated and saved to disk. These augmented samples were subsequently loaded by the model's data feeder during training, ensuring efficiency and consistency in data handling.

5.1.1 Random Brightness and Contrast

The random brightness and contrast augmentation adjusts the intensity and contrast levels of an image within a specified range, introducing variability that enhances the model's robustness to lighting conditions. This technique is particularly beneficial for datasets like cancer spheroids, where slight variations in illumination can simulate real-world scenarios, helping the model generalize better to unseen data [23].

5.1.2 Blur

The blur augmentation applies a slight blurring effect to images, mimicking out-of-focus conditions or noise that may occur during image acquisition. This technique enhances the model's ability to handle imperfections in the data [23].

5.1.3 Rotation

The rotation augmentation involves rotating images by a random angle within a specified range, introducing variability in orientation. This technique is particularly useful for datasets like cancer spheroids, where objects can appear in different orientations, helping the model become invariant to rotational changes and improving robustness [23].

5.1.4 Perspective

The perspective augmentation applies a transformation that simulates changes in the viewing angle, altering the spatial perspective of the image. This technique is useful for enhancing the model's robustness to variations in viewpoint [23].

5.1.5 Optical Distortion

The optical distortion augmentation simulates irregular deformations that mimic lens imperfections or other optical anomalies. This technique helps the model become more resilient to such distortions [23].

5.1.6 Piecewise Affine

The piec-wise affine augmentation applies localized geometric transformations by spreading grid points on an image, and randomly moves the neighborhood of these points around. This technique introduces subtle, realistic deformations [23].

5.1.7 Horizontal and Vertical Flip

The horizontal and vertical flip augmentations invert images along their respective axes, introducing variations in orientation. These augmentations are particularly effective for datasets where the subject's orientation is not fixed, helping the model learn invariance to flipping and improving its generalization [23].

5.1.8 Random Crop

Random crop augmentation involves randomly extracting smaller regions from the original image, allowing for the model to learn from diverse perspectives and focus on different parts of the spheroid. This approach is particularly effective in combating overfitting, as it encourages the model to recognize spheroid boundaries and textures regardless of their position within the image [23].

5.1.9 Patching

Given the large size of the original images, simply resizing them to 256 pixels would result in significant information loss. To address this, we partitioned each image into four quadrants of 800 × 800 pixels, ensuring that critical details were preserved. This approach not only increased the effective size of the dataset by a factor of four but also allowed us to retain the original full-size images as input. Consequently, the model was exposed to multiple scales of the same sample, promoting scale-invariant learning. This strategy significantly expanded the dataset and yielded notable improvements in model performance, particularly in tasks requiring detailed feature extraction.

5.2 Ensemble

Another innovative aspect of our methodology lies in leveraging the outputs of multiple models to construct an improved segmentation model through ensemble techniques. This approach, widely recognized in the literature, can be implemented in various ways. We opted for late fusion majority voting.

This ensemble method integrates the predictions of multiple segmentation models at the decision level by assigning each pixel in the segmentation output the class label that receives the majority vote among all models. Each model independently processes the input microscopy image, generating a pixel-wise segmentation map, and the final segmentation is formed by aggregating these outputs through majority voting.

As illustrated in Figure 5.1(c), the segmentation maps generated by two distinct models are shown in red and green. Subsequently, a pixel-wise majority voting operation, represented by the $\wedge$ symbol, classifies a pixel as part of the spheroid only if both models agree on its class, with the resulting output shown in yellow in Figure 5.1(c). This approach mitigates individual model errors and enhances segmentation robustness.

In our methodology, we expect this technique to enhance spheroid segmentation at its borders, as segmentation models often face challenges in correctly delineating the boundaries of cancer cultures, as we will show in the next chapter.

Chapter 6

Results

In this chapter, we provide a detailed description of our testing setup, including the results and conclusions derived from our experiments. We outline the testing protocols to facilitate future result reproduction and describe the specific ablation studies conducted to evaluate the impact of various components on performance. Each modification's effect on the outcome is analyzed. Finally, we discuss the results in depth, highlighting key discoveries and their implications towards spheroid segmentation.

In the following sessions, we will outline the testing protocols employed in this study, providing a overview of both the quantitative and qualitative outcomes. This analysis will detail the methods used to assess performance and day-to-day usability, ensuring a robust understanding of the data. Quantitative results will include the dice score , while qualitative results will offer insights into the specialist point of view of the data, adding depth to our understanding of the methodology.

We utilized three datasets for training our models: Our Dataset, Lacalle Manual, and Lacalle SpheroidJ. Additionally, two datasets were used for testing: Our Dataset Test and Lacalle Test. Regarding the models tested, we employed a selection of deep learning architectures introduced in Chapter 2. These included well-known convolutional neural networks (CNNs), such as U-Net, U-Net++, and MultiResUNet. We also evaluated transformer-based models, including SwinV2 Transformer, TransUNet, and Swin-UNet. Finally, we tested a multilayer perceptron (MLP)-based model, UNeXt.

To gain a deeper understanding of our data, we approached testing from multiple perspectives. Consequently, we developed six protocols, described in the following sections, to analyze what the models learned and how effectively they generalize this learning.

We assumed that every dataset has an intrinsic scientist bias, in which the samples were generally cultured, captured, and sometimes annotated by the same professional, creating such bias. This bias is not a problem on its own, however, it causes samples from the same dataset to share some features, such as capture hardware, lightning conditions, cell line culture, and well-to-well similarities. In an attempt to minimize the loss function, these similarities may lead the models to learn some features that are specific to this dataset, and face a drop in performance once presented with new data.

6.1 Protocols 1, 2, and 3

Protocols 1, 2, and 3, summarized in Table 6.1, involve training and testing the networks within a single, consistent dataset, thereby keeping the intrinsic bias present in this experiment. This approach establishes a baseline performance by checking each network's ability to learn and generalize within the same data environment.

By focusing on a controlled dataset, it becomes possible to analyze the base capabilities of the networks, providing a reference point before introducing more complex variables or testing conditions.

Table 6.1: Protocols summary, stating wherein each model was trained and later tested.

	Protocol 1	**Protocol 2**	**Protocol 3**	**Protocol 4**	**Protocol 5**	**Protocol 6**
Train	Our	SpheroidJ	Manual	Our	SpheroidJ	Manual
Test	Our	Lacalle	Lacalle	Lacalle	Our	Our

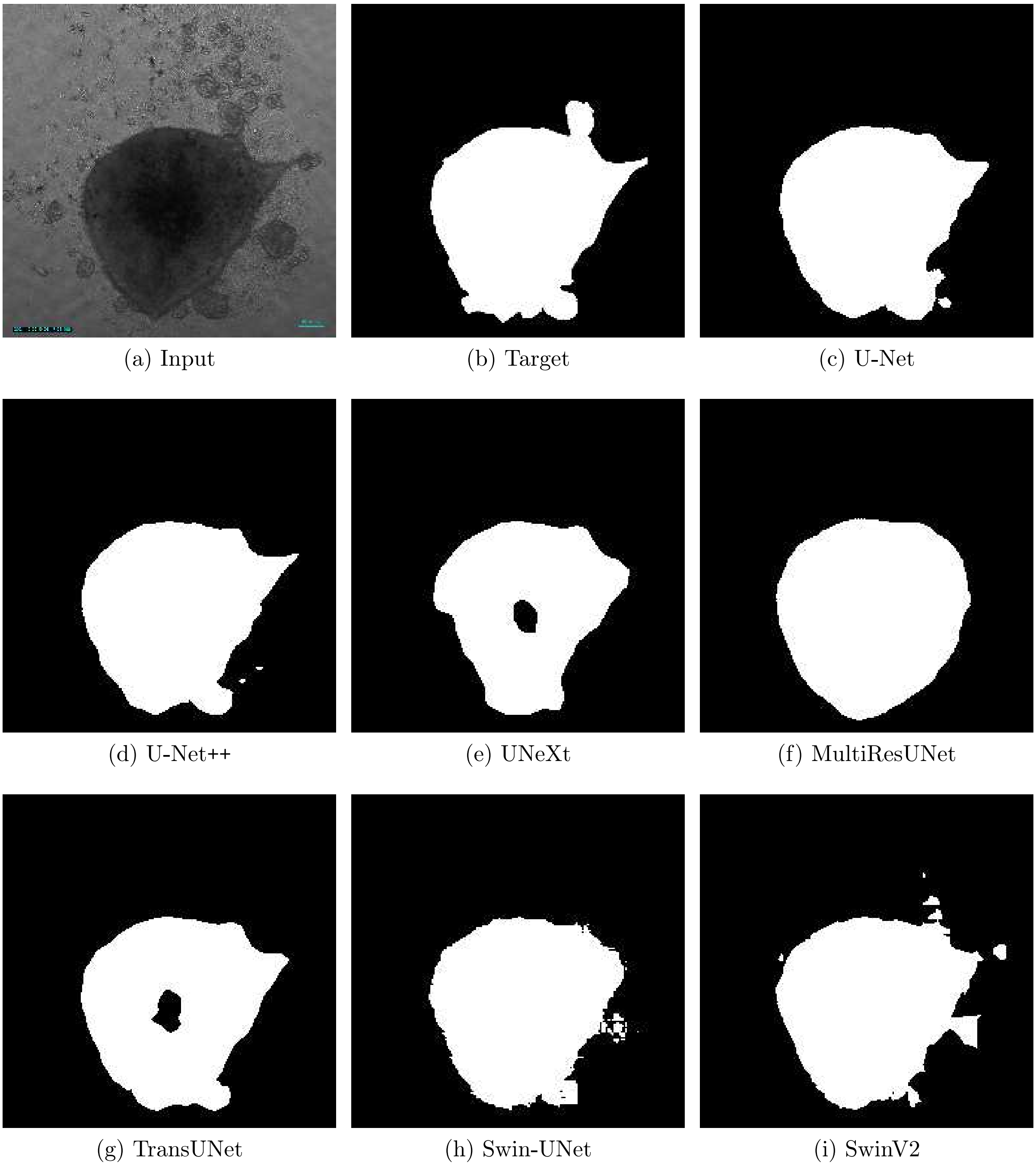
Table 6.2 presents the results of the protocols 1, 2, and 3, with each row corresponding to a protocol and each column displaying the Dice scores of the respective models. Protocol 1 revealed a competitive performance between the CNNs and Transformer models. The top three models achieved scores above 95%, with SwinV2 obtaining the highest score (95.84), followed by U-Net++ (95.74) and U-Net (95.06). Notably, the smallest model, UNeXt, which was expected to have the lowest performance, did not rank as the worst-performing model. SwinV2 is the most optimized and largest model in our list, making its top performance expected.

Table 6.2: Results of protocols 1, 2, and 3. Each row represents a protocol, in ascending order, and every column shows the dice score percent obtained by a given model. The best results are highlighted in bold.

Protocol	**U-Net**	**U-Net++**	**UNeXt**	**MultiResUNet**	**TransUNet**	**Swin-UNet**	**SwinV2**
P1	95.06	95.74	88.65	87.10	91.66	91.35	**95.84**
P2	89.81	90.51	76.16	77.10	90.10	77.23	**95.07**
P3	79.78	66.47	69.60	68.74	83.50	70.52	**96.44**

Figure 6.1 presents a qualitative comparison between the different models under Protocol 1 and their corresponding target segmentation masks. While the SwinV2 model achieved the highest score, it exhibited a tendency to produce a ragged segmentation at the spheroid borders. Additionally, the models demonstrated a consistent tendency to oversmooth the segmentation boundaries, consequently failing to capture crevices in the delineation.

In Protocol 2, which involved training on the SpheroidJ dataset and testing on the Lacalle Test dataset, we found a different scenario compared to Protocol 1. In this case, the majority of results produced a Dice score below 90. In particular, SwinV2 achieved the highest score (95.07), followed by U-Net++ (90.51) and TransUNet (90.10). The SpheroidJ dataset, being the largest in this study, may have contributed to the improved performance of the Transformer-based TransUNet model. However, despite also being a



(a) Input (b) Target (c) U-Net

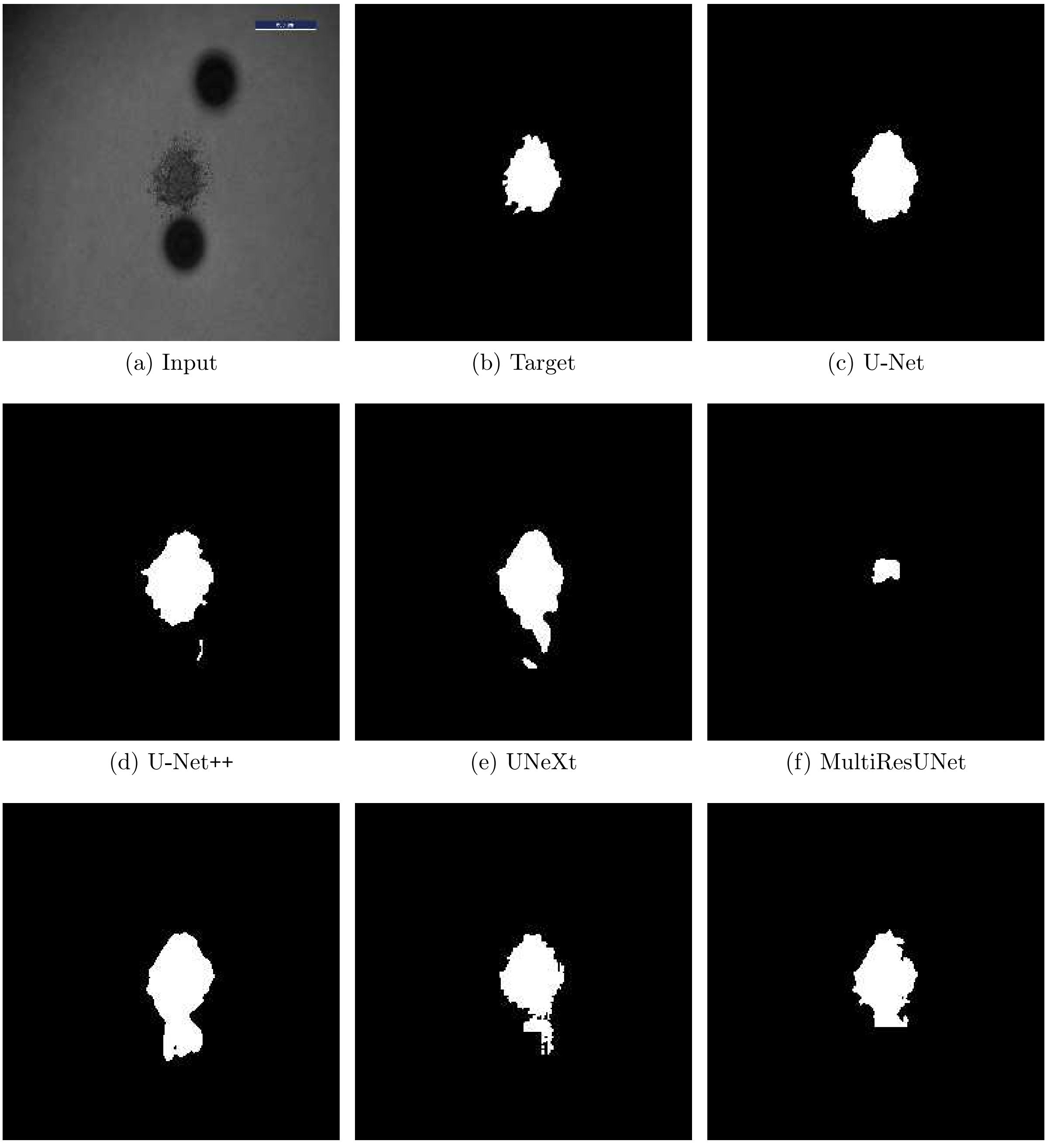
(d) U-Net++ (e) UNeXt (f) MultiResUNet

(g) TransUNet (h) Swin-UNet (i) SwinV2

Figure 6.1: Obtained from Protocol 1 by the models.

Transformer model, Swin-UNet did not achieve a Dice score of 90 or higher. Moreover, this protocol generated a balanced ranking between the CNNs and the Transformers.

In Protocol 2, the models were tested on the Lacalle Test dataset, which presents additional challenges compared to our test dataset. For example, as shown in Figure 6.2a, the target spheroid is located at the center of the image, with two rounded shadows close by. This configuration prevents the segmentation models from relying solely on contrast to delineate the culture. Consequently, the models UNeXt, TransUNet, Swin-UNet, and SwinV2 incorrectly classified the lower shadow as part of the spheroid. Additionally, we observe that Swin-UNet and SwinV2 produced ragged segmentation outcomes, whereas the CNN-based models U-Net and U-Net++ generated smoother borders.



(a) Input (b) Target (c) U-Net

(d) U-Net++ (e) UNeXt (f) MultiResUNet

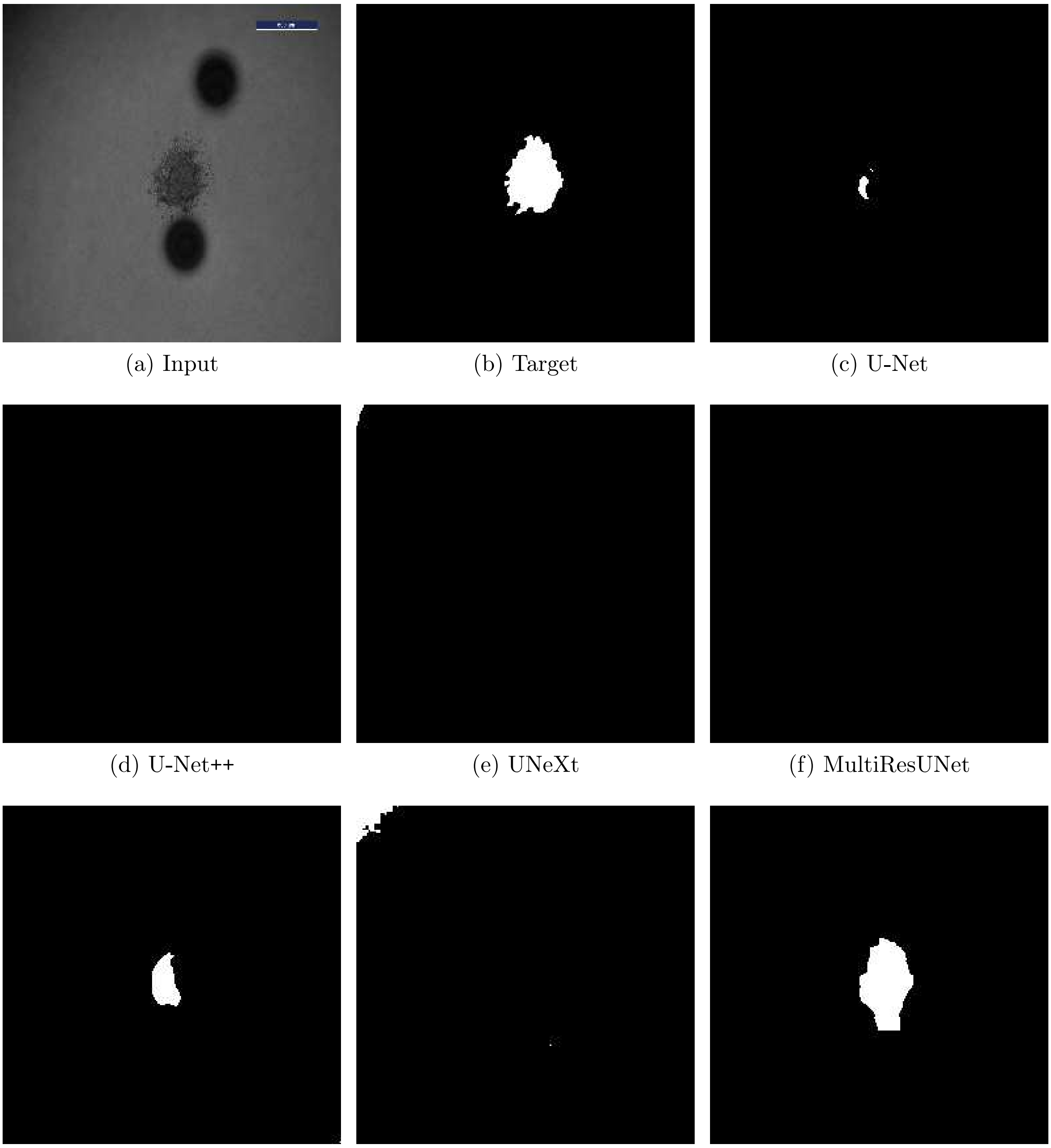
(g) TransUNet (h) Swin-UNet (i) SwinV2

Figure 6.2: Obtained from Protocol 2 by the models.

The last row in Table 6.2 corresponds to Protocol 3, in which the models were trained on the Lacalle Manual dataset and tested on the Lacalle Test dataset. Although the Lacalle Manual dataset is the second largest in this study, it is considerably smaller than the SpheroidJ. This protocol resulted in a significant drop in performance, particularly for the CNN models. Unexpectedly, the U-Net++ model, which had consistently ranked among the top three performers in previous tests, was ranked last in this scenario. Moreover, the only model achieving a Dice score above 90 was the SwinV2.

These results suggest that the contextual differences between the Manual and Test datasets are substantial, leading to reduced model performance. In contrast, the stronger performance observed in other protocols may indicate greater contextual similarities between the SpheroidJ and the Test dataset, and also our training and testing datasets.

Figure 6.3 was generated using the same input as Figure 6.2 to maintain a consistent basis for comparison. However, in this instance, all models produced unsatisfactory segmentations, including the SwinV2. Notably, both U-Net++ and MultiResUNet classified the entire image as background, indicating that these models failed to generalize their learning from the Manual dataset.



(a) Input

(b) Target

(c) U-Net

(d) U-Net++

(e) UNeXt

(f) MultiResUNet

(g) TransUNet

(h) Swin-UNet

(i) SwinV2

Figure 6.3: Obtained from Protocol 3 by the models.

The results from Protocols 1 and 3 raised questions about whether the models effectively learned the necessary features to segment cancer spheroid cultures or merely adapted to specific characteristics of each dataset to produce the segmentation results. To address these uncertainties, we designed the following protocols to provide further insight into these doubts.

6.2 Protocols 4, 5, and 6

In Table 6.3 we summarized the results of our cross-testing protocols, in which the models were tested on a dataset with a different creation bias, in contrast with the previous protocols where the models were tested in the same context as their training. This methodology allows us to evaluate the model's generalization capabilities by exposing it to variations in data characteristics, such as differences in noise levels, resolution, or features. Cross-testing is particularly useful for identifying overfitting to the training dataset, as well as for assessing the models' ability to learn transferable features for new data.

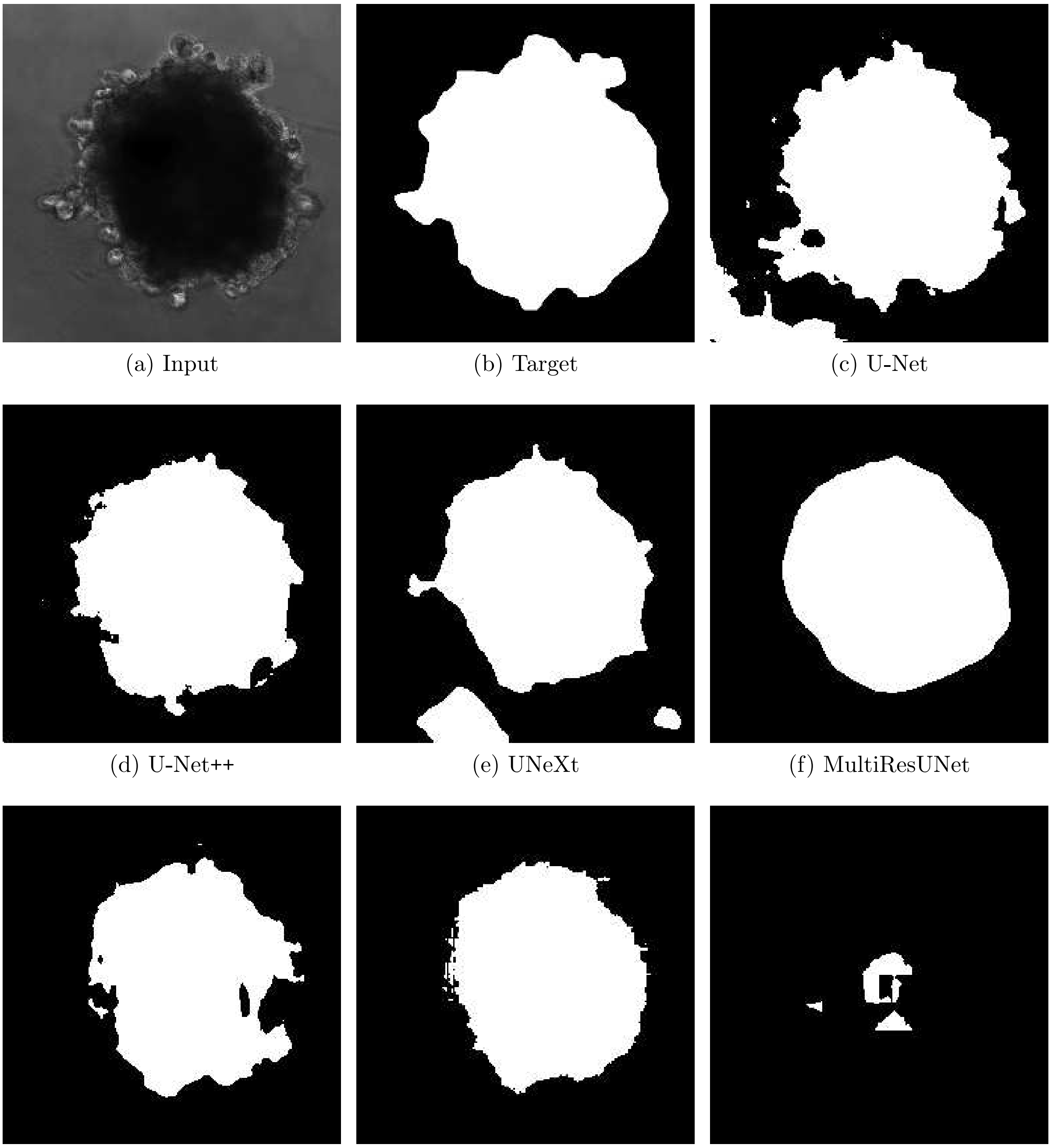
The results for Protocol 4, presented in the first row of Table 6.3, correspond to training on our dataset and testing on the Lacalle Test dataset. A noticeable drop in performance was observed for all models compared to Protocol 1. Although SwinV2 achieved a Dice score of 83.00, the second-best model scored as low as 62.34. This outcome is somewhat expected, given the smaller size of our dataset, which may limit the models' ability to generalize. The MultiResUNet was the most affected by this dataset discrepancy, obtaining a Dice score of only 29.43.

Table 6.3: Results of protocols 4, 5, and 6. Each row represents a protocol, in ascending order, and every column shows the dice score percent obtained by a given model. The best results are highlighted in bold.

Protocol	**U-Net**	**U-Net++**	**UNeXt**	**MultiResUNet**	**TransUNet**	**Swin-UNet**	**SwinV2**
P4	50.69	52.99	34.55	29.43	62.34	42.19	**83.00**
P5	77.05	77.32	37.87	70.43	73.86	66.24	**89.84**
P6	30.78	25.80	60.09	22.84	73.90	72.23	**86.77**

Figure 6.4 shows the obtained inferences for Protocol 4. In this instance it is possible to notice the discrepancy between quantitative metrics and qualitative analysis. For instance the TransUNet inference obtained 62.34 Dice score, which could be considered a good metric. However, we can notice that the qualitative analysis of such segmentation shows a segmentation map that lacks in resemblance to the spheroid region.

Protocol 5 involved training on the SpheroidJ dataset and testing on our test dataset. This setup ensured that the models were trained on a significantly larger and more heterogeneous dataset compared to the testing one. The results align with the expectations for this protocol. Although there was a decline in performance compared to Protocol 2, the top three models maintained Dice scores above 77.00, with SwinV2 achieving the highest score (89.84). The increased amount of training data benefited all models, however, the improvement was more noticeable for the CNN-based models than for the Transformer-based models when compared to Protocol 4.



(a) Input (b) Target (c) U-Net

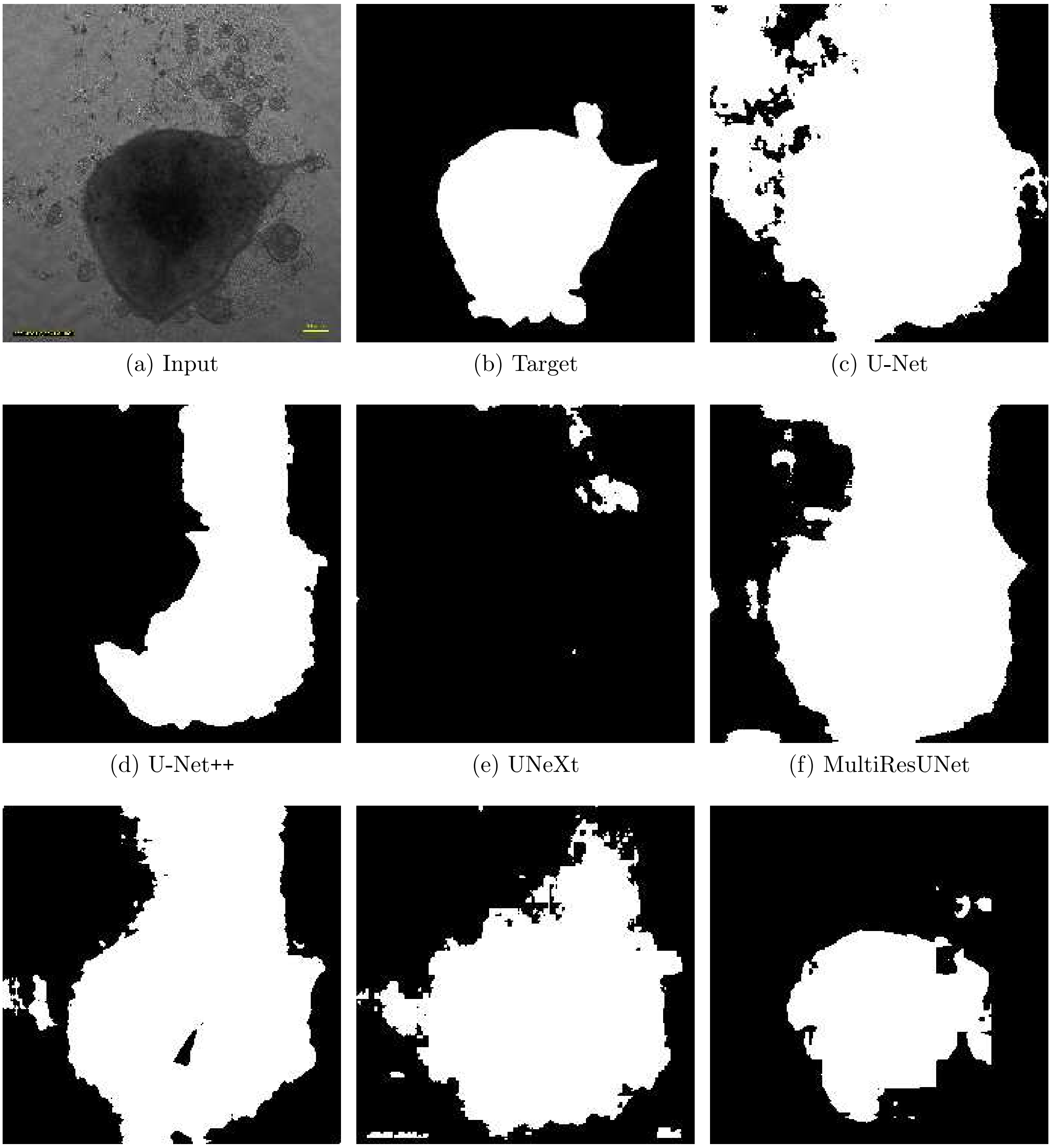
(d) U-Net++ (e) UNeXt (f) MultiResUNet

(g) TransUNet (h) Swin-UNet (i) SwinV2

Figure 6.4: Inferences obtained from Protocol 4.

Protocol 5 generated the outputs in Figure 6.5, in which we can notice that in most cases the model over segmented the cell culture, extrapolating its region and producing unusable masks. In other instances, such as SwinV2 and Swin-UNet the models produced a ragged segmentation that seldom resembles the target mask.

We also noticed that when training the MultiResUNet model on any dataset from Lacalle, in which a majority of the images do include the scale-bar legend, and tested on our dataset, in which every image has the scale-bar, the model will frequently segment the scale-bar as part of the spheroid culture. This behavior also emerged from TransUNet, and Swin-UNet, although in a lesser degree.



(a) Input (b) Target (c) U-Net

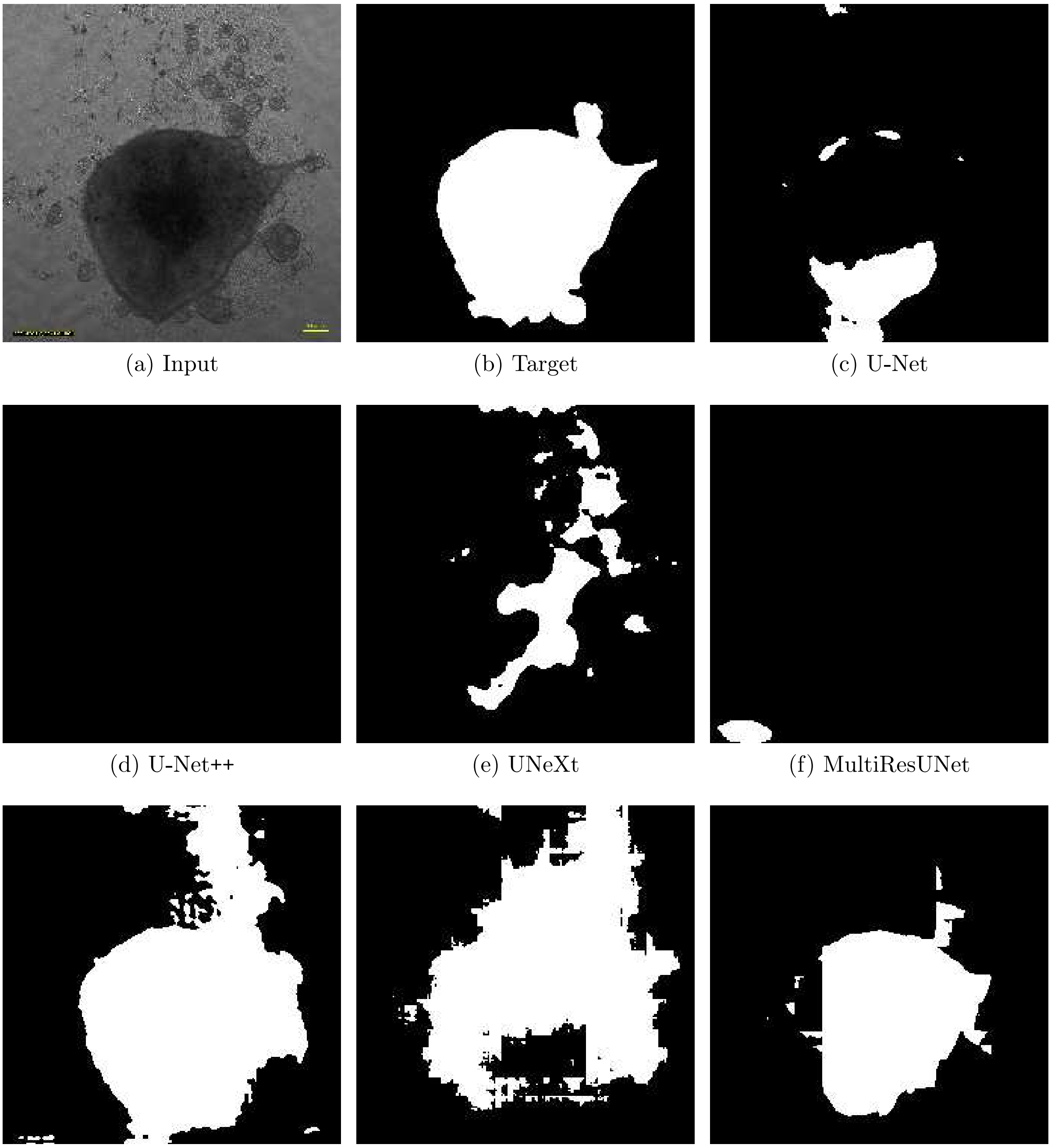
(d) U-Net++ (e) UNeXt (f) MultiResUNet

(g) TransUNet (h) Swin-UNet (i) SwinV2

Figure 6.5: Obtained from Protocol 5 by the models.

In Protocol 6, we trained on the Lacalle Manual dataset and tested on our test dataset. The results align with expectations, falling between the outcomes of Protocols 4 and 5. However, an interesting development emerged: the UNeXt model (60.09) outperformed all CNN models. Although UNeXt is the smallest model, and thus expected to perform the worst, it was able to extract significant features from the Manual dataset and apply them effectively on our test dataset. Additionally, the Transformer models ranked in the top three, led by the SwinV2 (86.77), while the CNN models ranked lower, with MultiResUNet (22.84) scoring the lowest. This result for the transformer models can be attributed to their ability to capture long-range dependencies, in contrast to the short-range limitations of CNNs.

In Protocol 6's inferences we can notice more tame segmentations compared to Protocol 5. However, the models still produced unusable segmentation maps. The Transformers came close to follow the spheroid region, with the SwinV2 coming the closest, but with a very ragged segmentation. Lastly, as we mentioned previously, this protocol shows an instance in which the MultiResUNet segmented the scale-bar as part of the cell culture.


(a) Input (b) Target (c) U-Net

(d) U-Net++ (e) UNeXt (f) MultiResUNet

(g) TransUNet (h) Swin-UNet (i) SwinV2

Figure 6.6: Obtained from Protocol 6 by the models.

6.3 Ensemble

In this section, we present the results obtained from our ensemble methodology. To evaluate which combinations of models provided the most valuable contributions to cancer spheroid semantic segmentation, and which combinations could further improve it. The following combinations were tested: UNeXt and TransUNet, UNeXt and U-Net++, U-Net and TransUNet, and U-Net and U-Net++.

The UNeXt model, being the smallest among the selected models, ranked between the worst performances within the group. Consequently, we decided to investigate whether pairing this model with either a CNN or a Transformer could enhance its performance to achieve competitive results. Furthermore, this approach aimed to identify which type of information, convolutional or Transformer-based, is best suited to complement the UNeXt model's capabilities.

The first and second rows in Table 6.4 demonstrate that pairing the UNeXt model with other models did not result in any significant performance improvements. In instances where the ensemble outperformed the UNeXt model alone, the gains were minimal. However, in Protocol 6, the ensemble caused a substantial drop of 40 points in the score. Overall, these findings suggest that incorporating the UNeXt model into an ensemble is more likely to decrease the performance of other models rather than provide any substantial advantage.

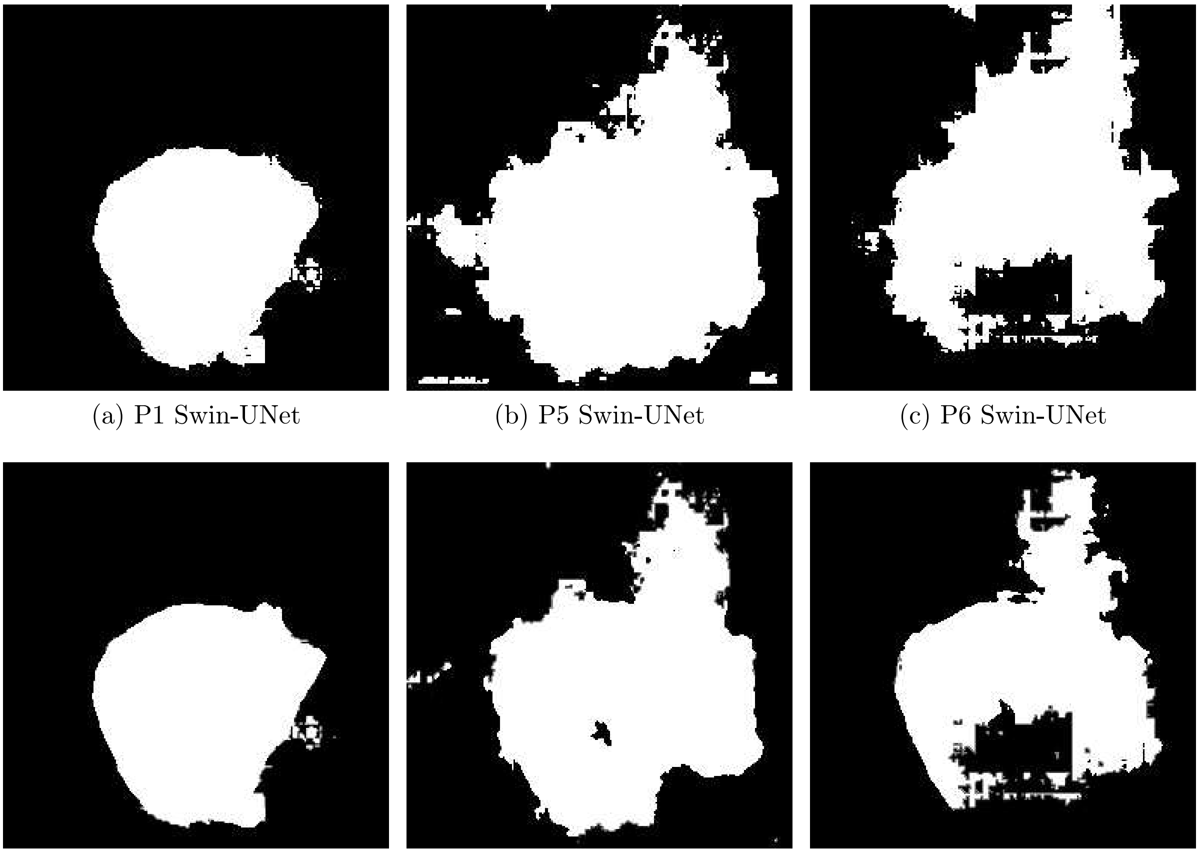
Table 6.4: Each row corresponds to a specific combination of models, while each column displays the Dice score achieved for each protocol. Results that outperformed the individual models are highlighted in bold.

Ensemble	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**
UNeXt $\wedge$ TransUNet	89.27	76.15	68.88	31.69	38.71	62.70
UNeXt $\wedge$ U-Net++	89.81	76.29	58.21	30.78	39.29	18.97
U-Net $\wedge$ TransUNet	93.30	88.99	78.12	44.25	**81.73**	30.78
U-Net $\wedge$ U-Net++	**95.98**	**90.80**	**80.07**	**54.98**	75.34	**35.83**

For our next ensemble, we opted to combine contrasting paradigms, CNN and Transformer, by pairing the U-Net with the TransUNet model, third row in Table 6.4. Despite this attempt, the results once again failed to surpass the performance of the individual models. However, in Protocol 5, the ensemble approach achieved its first victory over the individual models by improving the Dice score. This outcome, shows that mixing paradigms was not particularly effective for cancer spheroid segmentation.

Given the satisfactory results from U-Net and U-Net++ in previous experiments, we decided to investigate whether combining two CNNs would yield better results. The ensemble of these two models produced a score increase in most cases, with particular emphasis on Protocol 6, where the score improved by 5 points. This positive outcome suggests that both models effectively learned the essential features needed for spheroid culture segmentation, highlighting the potential benefits of combining CNNs in this context.

Interestingly, we identified an emerging consequence of the ensembling mechanism: the smoothing of ragged segmentations. This behavior suggests that the ensemble may play a supporting role, particularly benefiting the SwinV2 model. Although SwinV2 achieved the best Dice score, it produced overly ragged segmentations. The ensemble, by contrast, helped smooth these segmentations, potentially enhancing the overall quality. In the second row of Figure 6.7, we present the smoothed segmentation obtained from the ensemble.



(a) P1 Swin-UNet (b) P5 Swin-UNet (c) P6 Swin-UNet

(d) P1 Swin-UNet ∧ TransUNet (e) P5 Swin-UNet ∧ TransUNet (f) P6 Swin-UNet ∧ TransUNet

Figure 6.7: Ensembling Swin-UNet and TransUNet improved the ragged segmentations. The top row displays the ragged segmentations produced by the Swin-UNet model, while the bottom row presents the segmentation for the same spheroid using an ensemble of the Swin-UNet with the TransUNet model. It is evident that the ragged effect is mitigated in the bottom row, highlighting the smoothing effect introduced by the ensemble approach.

It is important to note that the ensemble approach we employed, late-stage majority voting, requires agreement between both models for the pixel to be included. This causes expected performance drops, as some correct pixels may be omitted during the process. However, the remaining pixels undergo a double-checking mechanism, making them more reliable and trustworthy in the segmentation outcome.

6.4 Ablation

In our ablation studies, we focused on examining the effects of network variations, and the attention block to understand their impacts on model performance. The exploration of network variations allowed us to evaluate the effectiveness of specific features in the Transformer models.

Additionally, we added an attention block into the U-Net model bottleneck to understand the impact of such technique towards spheroid segmentation. This approach helped us gain a nuanced perspective on the elements most crucial to optimizing our model.

6.4.1 Transformer Parameters

Given the promise of the Transformer models in other fields we decided to further investigate their effectiveness by altering the parameters of the TransUNet and the Swin-UNet models. Initially we considered changing the patch size parameter, however, due to their architecture the patch size has to remain fixed as 1 × 1.

Regarding the TransUNet model, we altered its B parameter. Figure 2.13 shows that in this architecture the Transformer layer is repeated several times (green box); This repetition is controlled by the B parameters. In other words, it controls how many times the features are passed through the Transformer layers to be learned from.

The results of such variation are shown in Table 6.5. We opted to vary the parameters between 4, 8 , and 16, and re-run the six protocols explained above. All previous test with this model were executed with the value at eight. In the results, we can notice a shy variation on the Dice score of a few points, specially when it comes to improve the metric.

In Protocol 4 the variation resulted in a drop of 5 points in the score when decreasing and an even larger drop of 14 points when increasing. As we explained from Table 6.3, Protocol 4 demands a lot from the network by being trained on a smaller dataset, and altering the network parameters showed no effect.

Overall, this variation has little effect and shows that passing the features through the Transformer block more times is counter productive, but doing it less time could benefit the model, since it saves on resources and has little effect on the outcome.

Table 6.5: TransUNet ablation parameter B variation. Every row refers to a protocol, and each column shows the Dice score obtained for every parameter variation. The best result of a protocol is bolded.

Protocol	B 04	B 08	B 16
P1	91.09	**91.66**	91.58
P2	87.49	**90.10**	89.37
P3	83.30	**83.50**	82.04
P4	57.35	**62.34**	47.84
P5	71.65	73.86	**75.99**
P6	**75.37**	73.90	73.81

As for the Swin-UNet model we altered the C parameter. This parameter controls the depth of the features when moving deeper into the architecture, shown in Figure 2.16 by the letter C. We assumed that by altering this parameter the model carries more information into the encoder, and by doing so is able to produce better inferences from the input data. We tested with the following values: 16, 32 (standard), 64, and 128.

What Table 6.6 shows is that we were correct, and in general the model obtained better Dice score by increasing the parameter. Initially, we tested only as far as 64, but the results were so promising that we increased it to 128 (maximum we could fit into our GPU's memory), and the model once again improved its performance. Opposed to the previous experiment, decreasing this value to save on resources was not advantageous, with an exception in P1, which lost just a little of the Dice score. It is also worth noting that the Protocol 4 benefited very little from the increase of C, as if the network had reached its peak performance from that learning.

Table 6.6: Swin-UNet ablation parameter C variation. Every row refers to a protocol, and each column shows the dice score obtained for every parameter variation. The best result of a protocol is bolded.

Protocol	**C 16**	**C 32**	**C 64**	**C 128**
P1	88.88	**91.35**	88.25	87.74
P2	66.87	77.23	81.08	**88.08**
P3	55.71	70.52	70.97	**82.52**
P4	29.01	42.19	29.48	**42.21**
P5	69.24	66.24	**76.28**	75.13
P6	61.50	72.23	78.68	**80.20**

6.4.2 U-NetAtt

Given the success of the Transformer models and their reliance on attention blocks [152], we decided the investigate the effects of the attention mechanism towards segmentation when applied to a CNN, in this case, the U-Net architecture. To do so, we added the multi-head attention layer, implemented by PyTorch, right before the bottleneck shown in Figure 2.6. In doing so, we expect the attention mechanism to execute long-range dependencies, which should allow for better segmentation maps.

In Table 6.7 it can be seen that with respect to the initial protocols 1, 2, and 3, the attention mechanism produced little effect, being most significant in P2 and P4, increasing the Dice score by a few points. Its greater effect was observed in the cross-testing protocols (5 and 6), in which it dropped the score from P5 by more than 20 points while also increasing the P6 score by 50 points.

In general, the attention mechanism alone was too ineffective to produce better results. As we have seen before with the SwinV2 results, this mechanism in combination with other factors can produce much better segmentation. However, as a standalone solution it increases the complexity of the models without assuring any improvement.

Table 6.7: U-NetAtt. Every row refers to a model, and each column shows the dice score obtained on such protocol. The best model of a protocol is bolded.

Model	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**
U-Net	95.06	89.81	**79.78**	50.69	**77.05**	30.78
U-NetAtt	**95.99**	**92.66**	79.21	**53.59**	51.97	**81.38**

6.5 Discussion

On average, the more complex models (Transformers) surpassed the lighter ones (CNNs). This result is primarily attributed to the SwinV2 model, which scored best in every protocol we proposed. Although we anticipated that the Transformers would achieve top performance, we also expected some challenges, as they typically require significantly larger datasets for training, which we did not have available. This limitation prompted the use of our data augmentation method.

When comparing between protocols, Protocol 1 got consistently higher scores across all models. However, when the training was transferred to a different dataset, as seen in Protocol 4, there was a significant drop in performance. This decline suggests that while the models were able to learn the features from our dataset, the data may have lacked the necessary variability or quantity to enable effective extrapolation to other contexts.

SwinV2 performance in the cross-testing protocols (4, 5, and 6) was particularly surprising, as it was able to maintain a significantly better result. In these protocols, the model demonstrated a wide gap in performance between itself and the second place, showing its robustness and effectiveness in producing cancer spheroid segmentation maps.

In contrast, we selected the UNeXT model due to the authors' emphasis on creating a small, efficient model that could be useful in point-of-care settings, such as being executed during a doctor's appointment. We anticipated that this model could potentially be adapted for use in laboratory bench applications. It was therefore surprising to observe that, in several instances, UNeXT outperformed other methods. However, despite these instances of improved performance, the small model struggled to produce acceptable segmentations, as evidenced by both quantitative and qualitative analyses.

Although quantitative metrics serve as valuable indicators and are necessary for result reproducibility, our analysis demonstrated that they may not always align with a qualitative evaluation. Through the qualitative analysis, we observed instances where segmentation maps, despite obtaining high Dice scores, failed to capture important details, such as the core and the borders of the spheroid. These discrepancies highlight the need for further post-processing steps in certain cases to ensure more accurate and complete segmentations.

The qualitative results showed that ragged segmentation was a common issue in most of the Transformer inferences. We believe this was caused by the patching mechanism in these models, where the input is divided into smaller patches that are processed individually, as detailed in Section 2.9. However, the ensemble approach was able to mitigate this effect by requiring agreement between two models before an image was segmented. This mechanism effectively smoothed the segmentation borders, resulting in more reliable and refined segmentations, as illustrated in Figure 6.7.

Ensembles such as U-Net $\wedge$ U-Net++ consistently outperformed both CNN and Transformer averages in protocols P1, P2, P3, and P4, demonstrating their effectiveness in these scenarios. However, their performance declined in protocols P5 and P6, where Transformer-based models showed superior results. This suggests that the U-Net $\wedge$ U-Net++ ensemble may struggle to extend its learning to more complex scenarios, where Transformer models excel.

Chapter 7

A Quantitative Spheroid Analysis Pipeline (SAP)

This chapter introduces a configurable, user-friendly pipeline for spheroid analysis that facilitates semantic segmentation into day-to-day bench applications [1]. Unlike other tools, which often lack integrated statistical analysis and require careful data transfers prone to errors and inefficiencies, our solution eliminates these issues through integrated automation. Designed to tackle the challenges of microscope imaging, which demands high sensitivity, the pipeline not only generates usable segmentation masks but also facilitates statistical analysis of compound effects and produces data visualization plots. Its modular and customizable design makes it accessible to a diverse range of users, ensuring rapid processing of large sample datasets. The general workflow includes three key stages: segmentation, analysis, and visualization, which will be detailed as follows.

7.1 Materials

For this application, three datasets were created to demonstrate the efficacy of the proposed pipeline. It is important to note that these datasets differ from those presented in Chapter 4. Additionally, they are not publicly available due to ongoing experimental results associated with their respective cultures.

7.1.1 Cancer Cell Culture

Dulbecco's Modified Eagle Medium (DMEM), streptomycin sulfate, and penicillin were utilized as reagents, all of which were sourced from Nutricell (Campinas, SP, Brazil). Additionally, Fetal Bovine Serum (FBS) was obtained from Gibco (Invitrogen, NY, USA).

The SKMEL-28 human melanoma cell line was kindly provided by Prof. Silvya Stuchi Maria-Engler from the University of São Paulo (USP, Brazil). The cells were cultured in DMEM supplemented with 10% FBS, 100 U/mL penicillin, and 100μg/mL streptomycin. Incubation was conducted under a humidified atmosphere at 37°C with 5% CO_2. Regular testing was performed to ensure the absence of mycoplasma contamination.

[1] `https://github.com/guilhermevleite/Spheroid-Analysis-Pipeline`

Spheroids were generated using the Bio-Assembler n3D assay (Biosciences, Houston, TX), following the manufacturer's protocol, similar to Chapter 4. Cells were initially cultured for 24 hours as a monolayer (2D) model. Nanoshuttles™were added at a concentration of 1.6μL per 10e4 cells, and the culture was continued. After 24 hours, the cells were detached, transferred to a cell-repellent plate (10e4) cells per well), and incubated over a magnet for an additional 24 hours. Subsequently, the magnet was removed, and the plate was incubated for another 24 hours before treating the spheroids.

We treated the SKMEL-28 spheroids with 0.7μM, 1μM, 2.5μM, and 5μM of solution 1 over ten days, while doing drug holiday periods. Solution 1 was applied to the plates and incubated in a humidified atmosphere at 37°C with 5% CO_2 for 48 hours. Subsequently, the treatment medium was completely replaced with DMEM with 10% FBS and 1% penicillin/streptomycin added to it, followed by incubation under the same conditions for an additional 48 hours. The treatment with solution 1 was repeated twice, with drug holiday periods between applications. Similarly, SKMEL-103 spheroids were treated with 10μM of solution 2 over ten days, following the same drug holiday protocol.

7.1.2 Datasets

Three image datasets were utilized in this study, all of which were generated in Oncobiomarkers Laboratory using a LumaScope microscope equipped with a 10× objective lens, illustrated in Figure 7.1. Imaging was performed under standardized conditions for light intensity, gain, and exposure settings. Spheroids were captured every 48 hours over the ten-day experiment, with each image having a resolution of 1600 × 1600 pixels.

The output images were systematically organized into daily folders and labeled by the following name conversion: "control_x" or "test_y_x", in which, x refers to the experiment replicate ID, and y indicates the corresponding treatment. This organization was essential to enable the pipeline to identify the samples associated with each treatment or control condition.

Dataset 1 comprises 77 images of SKMEL-28 spheroids. Each day includes two experimental conditions: control and treatment with solution 1, at a concentration of 0.7μM. The dataset contains five images of the control samples and six images of the treated spheroid per day.

Dataset 2 is composed of 78 images of SKMEL-103 spheroids. Each day features two experimental conditions: control and treatment with solution 2 at a concentration of 10μM. The dataset includes six images of spheroids for both the control and treated conditions, totallying 12, per day.

Dataset 3 comprises 125 images of SKMEL-28 spheroids. Each day includes four experimental conditions: control and treatment with solution 1 at concentrations of 1μM, 2.5μM, and 5μM. The dataset contains six images of control spheroids and five images per concentration of solution 1 per day.

Table 7.1 provides an overview of the datasets. It is important to note that certain samples were excluded due to cases where a replicate failed to survive between captures. However, this discrepancy does not impact the performance of the segmentation algorithm, and we propose a solution in the statistical analysis methodology.

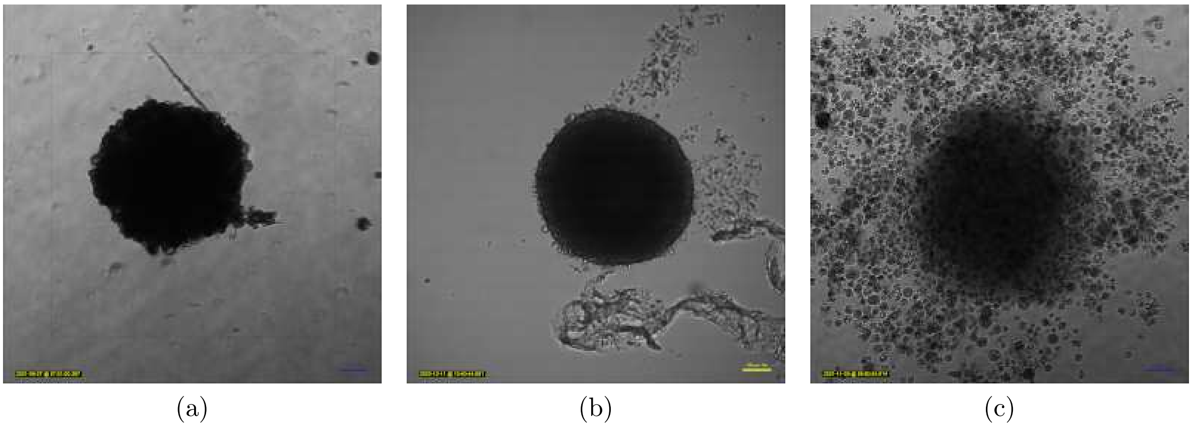


Figure 7.1: Samples from each dataset. (a) Dataset 1, SKMEL-28, treated with 0.7μL of solution 1. (b) Dataset 2, SKMEL-103, treated with 10μL of solution 2. (c) Dataset 3, SKMEL-28, treated with 5μL of solution 1.

Table 7.1: Summary of each dataset. Column Treatment refers to which treatment was employed. Column concentration indicates the concentrations that each treatment was used. Column Control refers to the number replicates samples in the control group per day, and the same to the Treatment column. Finally, Total column indicates the total number of samples per dataset.

Dataset	**Treatment**	**Concentration**	**# Control**	**# Treatment**	**# Total**
1	1	0.7μM	5	6	77
2	2	10μM	6	6	78
3	1	1, 2.5, 5μM	6	5	125

7.2 Spheroid Analysis Pipeline (SAP)

Our pipeline was designed with modularity as a key feature and is divided into three main components: segmentation, statistical analysis, and data visualization, as illustrated in Figure 7.2. To ensure flexibility, the output data of each module is structured to allow integration and substitution of modules depending on to the user's preferences. Additionally, ease of use was a main consideration, and the pipeline requires minimal or no coding skills to set up and operate.

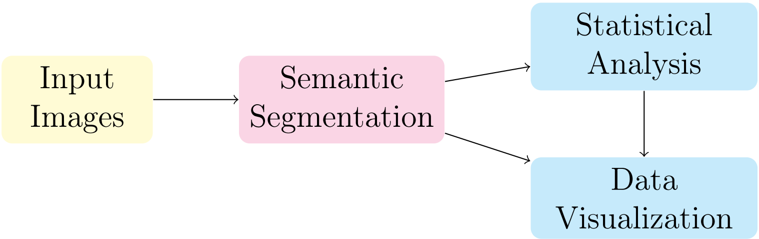


Figure 7.2: Spheroid Analysis Pipeline (SAP) overview. Images are input in temporal batches (daily batches in our examples), then fed to the segmentation stage. The areas found in the segmentation are used to produce p-values from the statistical stage, and finally a data visualization is produced from both stages.

7.2.1 Segmentation Module

To make our pipeline accessible for public use, we chose a straightforward segmentation approach, employing the widely recognized Otsu algorithm alongside simple pre- and post-processing steps, as illustrated in Figure 7.3.

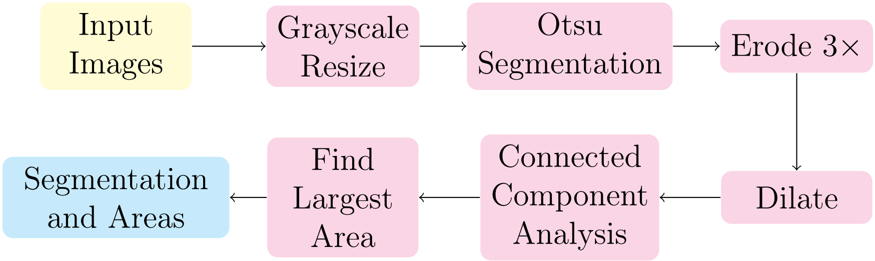


Figure 7.3: SAP segmentation steps overview.

The first step in our pipeline involves loading the image into a grayscale color space and resizing it to a resolution of 512 × 512 pixels, as shown in Figure 7.4a. Subsequently, the Otsu segmentation algorithm, provided by the OpenCV library, is applied (Figure 7.4b). To reduce noise generated during segmentation and to separate regions that are barely touching, we perform three erosion operations followed by one dilation operation, as illustrated in Figure 7.4c. Once the regions have been cleared, we extract all connected components from the image (Figure 7.4d). Finally, a post-processing step is applied to identify the spheroid region by assuming that the largest connected component represents the spheroid (Figure 7.4e). In our implementation, each step is modularly separated, allowing users to customize the pipeline by substituting techniques or adjusting parameters as desired.

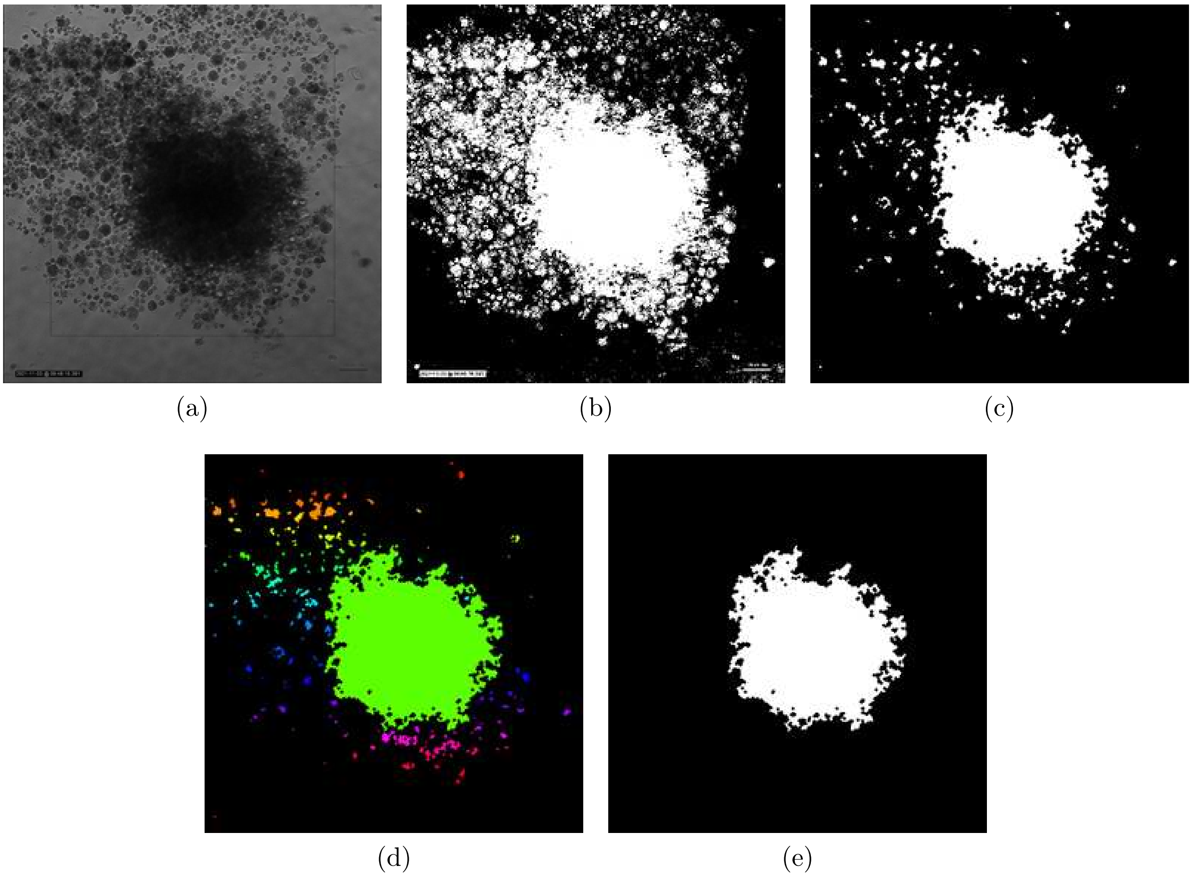
7.2.2 Dataframe

The connected component analysis also quantifies the size of each region, or its area, and stores this information in a Pandas dataframe, which is subsequently exported to the file system as a CSV file, via Pandas library.

The dataframe contains columns regarding image capture interval (time), the sample ID, the segmented area, and which experiment it is part of, as in control group or *treatment_x*, for instance. This step is crucial for ensuring modularity, as the statistical analysis that follows relies solely on the dataframe as its input. Consequently, the segmentation pipeline can be fully replaced without impacting code integration, provided it generates a compatible dataframe.

7.2.3 Statistical Analysis Module

Our pipeline is designed to work with two experimental conditions: (i) a single compound or treatment group tested against a control group, analyzed using the t-test to evaluate statistical differences between distributions, and (ii) multiple compounds or treatment groups tested against a control group, analyzed using the ANOVA test.



(a) (b) (c)

(d) (e)

Figure 7.4: SAP segmentation illustration. (a) The input image is converted to grayscale and resized; while resizing is not mandatory, it impacts the pipeline's throughput. (b) The Otsu algorithm produces an initial segmentation that includes noise around the borders and incorrectly segments the bottom region, which is not part of the spheroid. (c) Three erosion operations remove most of the noise and disconnect the central region from the bottom region, while a dilation restores some of the original shape. (d) Each detected connected component is assigned a unique hue. (e) The largest connected component is identified and segmented as the spheroid region.

Additionally, our ANOVA results include a Tukey's test to identify significance in pairwise comparisons. It is important to note that while the pipeline supports any number of samples in each group, the statistical tests require a minimum sample size to produce meaningful and reliable results.

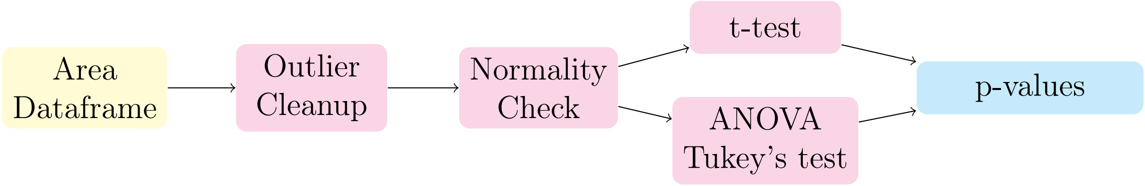


Figure 7.5: SAP statistical analysis overview.

Figure 7.5 illustrates the steps of our statistical analysis, using the dataframe generated by the segmentation process. To ensure that nonviable samples do not interfere with the data, we first remove any outliers through interquartile range (IQR) filtering. This method divides the data into quartiles and calculates the difference between the 75th percentile (Q3) and the 25th percentile (Q1). Values exceeding Q3 × 1.5 or falling below Q1 × 1.5 are considered outliers. This step is optional and can be omitted.

Both statistical tests we selected require the data to follow a normal distribution. Therefore, we assess normality using the Shapiro-Wilk test before proceeding. If the data fails to meet the normality assumption, the pipeline generates a warning to inform the user not to proceed with the analysis.

Finally, we perform either a t-test or ANOVA followed by Tukey's post-hoc test to compute the desired p-values. These p-values are reported in a temporal fashion (e.g., daily or hourly).

7.2.4 Data Visualization Module

Finally, the dataframe is visualized using an interactive boxplot graph, implemented by the Plotly [115] library, which depicts the area distributions divided by time intervals (e.g., days in our example). The graph is fully configurable, allowing users to customize legends and labels to ensure clarity and relevance, as illustrated in Figure 7.6.

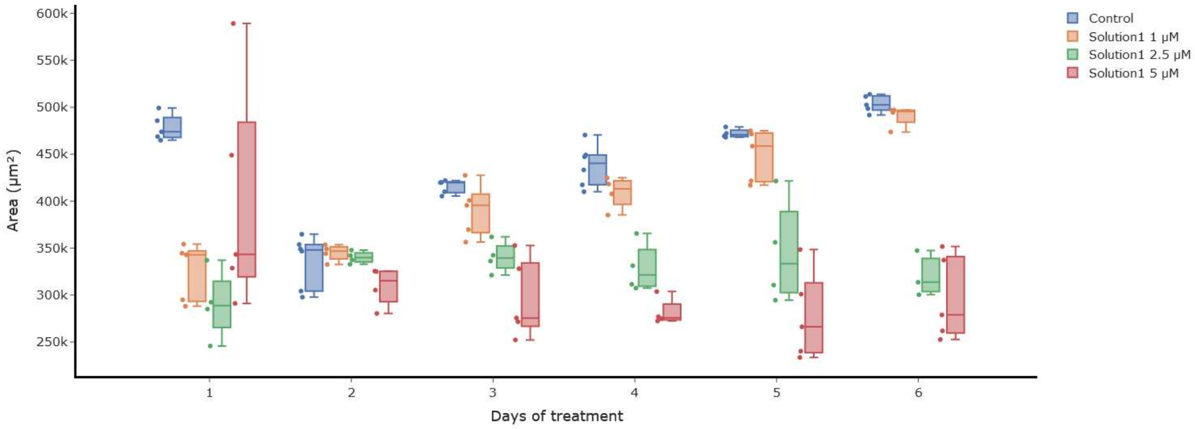


Figure 7.6: Dataframe boxplot rendered via Plotly [115] library.

7.3 Observations

The three datasets utilized in this chapter have not been made publicly available, as our collaborators are still in the process of publishing results derived from these data. Nonetheless, this underscores the immediate usability, utility, and demand for our pipeline, which has already been seamlessly integrated into the routine analysis workflows of the researchers involved.

Our pipeline has been made publicly available on a dedicated GitHub page, where users can download the notebook file (.ipynb). The notebook script is designed for seamless integration with Google Colab, where users can upload it and grant the necessary authorization to access files via Google Drive upon the first execution. To ensure proper functionality, the pipeline requires that images are organized into specific folders and adhere to a predefined naming convention. Detailed instructions for these requirements are provided in a user manual, which is also publicly accessible on the GitHub page.

Chapter 8

Conclusions

This chapter presents the main conclusions drawn from the research conducted during this doctoral study. It highlights the key findings that contribute to the understanding and advancement of the field, while also addressing the limitations encountered throughout the work. Furthermore, potential directions for future research are outlined, aiming to build upon the foundations established and to explore new opportunities for further investigation.

8.1 Final Considerations

In conclusion, three-dimensional (3D) cell culture spheroids are widely used in various applications, such as studying biological processes, diseases, and drug development, while also improving the reliability of *in vitro* studies. However, large-scale spheroid image analysis remains a significant challenge. This thesis addresses this challenge by presenting four key contributions to the field: a comprehensive literature survey on cancer spheroid segmentation, a new publicly available dataset of cancer spheroid samples with accurate annotations, a novel neural network ensemble method for semantic segmentation of these cultures, and an analytical pipeline that uses these segmentation outcomes to provide statistical insights into experimental development. To the best of our knowledge, our pipeline is the first to fully integrate automated cancer spheroid segmentation with statistical analysis for area estimation.

Motivated by an empirical observation, we initiated our survey to identify commonly used methods for analyzing the morphology of 3D cancer cultures, which later became our objective (O1). Our findings highlighted existing knowledge gaps in cancer spheroid segmentation, revealing a considerable use of image processing techniques for analysis, followed by proprietary software and, to a lesser extent, deep learning models. Furthermore, our survey underscored the scarcity of available datasets suitable for training deep learning models.

Inspired by our survey and anticipating the direction of our methodology, we began the development of Objective O2: a cancer spheroid dataset. This dataset was designed with deep learning training in mind, ensuring that we (i) manually annotated all ground truth information, (ii) split the data into training/validation and testing sets, (iii) established clear protocols for result reproduction and comparison, and (iv) made the dataset publicly available for long-term access, allowing it to be downloaded without the need for permission or special access.

Objective O3 was achieved through our methodology, in which we explained our data augmentation strategy, followed by a training and testing phase. We selected a set of segmentation models from the literature that could potentially be effective for cancer spheroid segmentation. In our protocols, we conducted a comparative analysis between different paradigms, evaluating the models' performance and their ability to learn the necessary patterns, as well as their capacity to generalize to new data. Additionally, we introduced a novel approach to cancer spheroid segmentation by ensembling two models to leverage their strengths for improved segmentation results. Finally, we fine-tuned various aspects of the models to identify which modifications had the greatest impact on cancer segmentation.

Our results revealed a clear gap between CNNs and Transformers, with the latter demonstrating a notable advantage. Transformers outperformed CNNs not only in Protocols 1, 2, and 3 but also in the cross-testing Protocols 4, 5, and 6, which evaluated the models' ability to generalize to new data. In our ensemble testing, combining two CNN models proved most effective, with the U-Net and TransUNet ensemble being the only exception. Qualitative analysis highlighted certain limitations of the Transformer models, particularly their tendency to produce ragged segmentations. Similarly, the CNN model MultiResUNet encountered significant challenges, especially during cross-testing. The UNeXt model did not meet our expectations, producing consistently low scores.

The quantitative spheroid analysis pipeline (SAP) presented in Chapter 7 was the outcome of Objective O4. This tool was developed for everyday use, not only for spheroid analysis but also for broader experimental analysis. Our pipeline successfully integrated the temporal aspect of experiments, provided spheroid segmentation, and delivered coherent analyses of the samples in relation to both time and the compounds under investigation. Designed with non-programmer users in mind, the tool was released in an open-source and modular format, ensuring adaptability to different user needs. We made the pipeline publicly available, along with detailed instructions on how to customize and utilize its various components.

In the course of producing this thesis, we developed several scientific articles that have either been published in prominent international conferences or are currently under review, as referenced in Chapter 1. This chapter also outlines the research questions we formulated for this work, and here, we are finally able to provide answers to them.

Q01. Which techniques are mostly used to produce the segmentation results?

In our literature review in Chapter 3, we identified three main approaches used to produce cancer spheroid segmentation results, ranked from most common to least. The first approach involves traditional image processing algorithms, which utilize various techniques such as thresholding (Otsu), edge detection, and region-growing (Watershed) to segment spheroids. These methods are widely used due to their simplicity and efficiency but may struggle with complex or noisy data. The second approach, specific software, includes commercially available tools that often integrate both image processing and machine learning techniques for segmentation. While these tools are user-friendly and optimized for specific applications, they may not be as adaptable or accessible as open-source alternatives. The third and increasingly prevalent method is deep learning, which leverages neural networks, particularly convolutional neural networks (CNNs), to learn and predict spheroid boundaries from datasets. This method has shown superior accuracy and robustness in other contexts, especially for complex segmentation tasks, though it requires substantial computational resources and large annotated datasets. Finally, manual segmentation is still used in certain cases, although it is time-consuming and prone to human error, making it less viable for large-scale studies.

Q02. What is the role of deep learning approaches in these scenarios?

We found that when applied, deep learning models were primarily used as a "black-box" solution, where the models were trained to produce segmentations without significant adaptation to the specific requirements of the cancer spheroid domain. The most commonly used architecture in this area was U-Net, as seen in Table 3.4, known for its effectiveness in medical image segmentation. While deep learning has the potential to improve segmentation accuracy, its application has been somewhat limited by the lack of adequate training data and the use of models that were not adapted to cancer spheroid segmentation.

Q03. How available are the datasets in this field?

In the field of cancer spheroid segmentation, the availability of datasets was a significant challenge. In Table 3.1 we showed that most of the datasets related to cancer spheroid cultures were either dead links or completely inaccessible. We observed a clear trend among authors to protect their data, either by withholding it or requiring direct contact for data release. However, despite these efforts, we found only a single dataset available. This lack of accessible data in the field is one of the key reasons we decided to create our own dataset, aiming to contribute to the limited resources available for training and evaluating deep learning models in cancer spheroid segmentation.

Q04. Are these datasets ready to be used for deep learning training?

No, the available datasets were not ready to be used for deep learning training. As discussed in Section 4.2, the unique dataset we found was published with incorrect annotation masks, rendering it unusable for training purposes. After requesting the correct data, we were given access to files that did not correspond to the protocols described in the paper, further complicating the situation. As a result, we had to create new protocols for future experimentation and to ensure the replicability of our results. This experience highlighted the challenges of working with existing datasets in this field and shows the need for more reliable and properly annotated datasets of cancer spheroid cultures.

Q05. Which metrics are used for comparison between methods?

As discussed in Chapter 3, the primary metric used for comparison between methods in the few works that released their metrics is the Dice score, which is well-suited for evaluating segmentation tasks. However, we also identified instances where less appropriate metrics, such as simple accuracy, were reported. These metrics are less effective for segmentation evaluation as they fail to account for the overlap and spatial correspondence required in segmentation.

Q06. Are CNNs better suited for this task compared to more complex

architectures like Transformers?

No, CNNs (such as U-Net, U-Net++, and MultiResUNet) are not a better fit for segmenting cancer spheroid cultures compared to more complex architectures like Transformers (SwinV2, Swin-UNet, and TransUNet). In our experiments, which included testing with three protocols (P1, P2, and P3), Transformers consistently outperformed CNN-based models in all cases, as showed in Table 6.2. Even when we combined two CNNs (U-Net $\wedge$ U-Net++) in an ensemble, the Dice score did not surpass the results achieved by the SwinV2 model. These findings suggest that Transformers, particularly SwinV2, are more effective at learning the necessary features for accurate cancer spheroid segmentation.

Q07. By combining these two types of architectures, can we improve on segmentation performance?

Combining these architectures (U-Net $\wedge$ TransUNet) did lead to an improvement in the Dice score on Protocol 5 compared to using the models independently, as shown in Table 6.4, similar to the performance boost observed when combining two CNNs. However, this improvement was still insufficient to outperform the score achieved by the SwinV2 Transformer model, which is still the top performer.

Q08. Which architecture type generalized its learning from the few samples we have available?

Table 6.3, shows that Transformers demonstrated superior generalization capability when learning from the limited dataset available, as evidenced by Protocol 4, where models were tested on their ability to apply learned features in a different context. Considering the Transformers, the SwinV2 model was the best performer, achieving a significantly higher performance than any other model, highlighting the Transformer's ability to generalize from small datasets.

Q09. Are the datasets generalizing enough to produce a segmentation on different type of cell culture?

Protocols 4, 5, and 6 also evaluated the ability of models to generalize by training on one type of cell culture and testing on another. These experiments, shown in Tabled 6.3, showed a drop in performance, indicating challenges in generalization. For CNN-based models, the datasets were not sufficient to do such cross learning, as their performance declined too much. In contrast, Transformers demonstrated better generalization capabilities, with the SwinV2 model being the most robust. Showing that the datasets provided enough data to allow for some models to generalize for other cell-line context.

Q10. What is the minimum necessary on a pipeline to automate spheroid analysis?

In Chapter 7 we identified the necessity for the pipeline to account for the timing and conditions of the samples. For example, if samples are captured every 24 hours over three days, the pipeline must provide information about the spheroid progression between these intervals. Additionally, the pipeline should support compound analysis for any number of compounds, such as comparing control groups with experimental groups exposed to specific compounds. Regarding segmentation, we determined that a combination of an Otsu algorithm, cleaning operations (e.g., erosion and dilation) to filter out noise, and connected component analysis to isolate the most promising region were the minimum requirements for producing usable segmentations. Finally, to generate meaningful insights, it was essential to include statistical tools, such as t-tests or ANOVA, alongside data visualization methods to facilitate interpretation of the results.

8.2 Limitations and Directions for Future Work

Throughout this work, we encountered several limitations and challenges. Initially, we found that the majority of datasets referenced in the literature were unavailable for download. As we collected our own images, it became evident that creating an extensive dataset was both costly and labor-intensive, resulting in a smaller dataset limited to samples from a single cell line. Another significant challenge was the lack of segmentation-focused research in which authors detailed their testing protocols and provided data for result reproduction and comparison. Consequently, we were unable to directly compare our results with those studies. Lastly, our generated segmentation masks are relatively small compared to the original image sizes. This limitation arises from the nature of the architectures used. However, some techniques exist that could potentially mitigate this issue.

Certain aspects of cancer spheroid analysis did not fit within the timeframe of the doctorate or would have significantly increased the scope of the thesis, these were therefore assigned as future work. One important area is the explainability of neural networks, which are often regarded as black-box models. Interpretability involves extracting information about the model's internal workings to help humans understand its outputs, increasing trust and enabling validation of the inferences. Additionally, there is a need to enhance segmentation accuracy, particularly at the borders of the cell cultures. This could be achieved by assigning weight to border errors during training or by employing alternative techniques.

From a biological perspective, incorporating biochemical data into the analysis would be a valuable addition. Spheroid cultures show morphological variations due to nutrient and oxygen diffusion gradients, which decrease from the outer layers to the nucleus. These differences create distinct regions classified as hypoxic, quiescent, and proliferative. Quantifying the growth and morphology of these regions is crucial for evaluating the effects of compounds. By utilizing biochemical markers and fluorescent microscopy, it is possible to visualize these regions. Developing a neural network capable of identifying these regions without the need for costly biochemical analyses would be a significant advancement, contributing to both automation and cost-effectiveness in the field.

In conclusion, upon comparing paradigms, the convolutions in a CNN introduce an inductive bias, where only the pixels within the kernel window are considered during each convolution. In contrast, Transformers utilize a multi-head attention mechanism that takes into account the relationships between all tokens, thereby producing a global bias for the information. This attention mechanism considers a significantly larger amount of information, requiring substantially more data to capture patterns, making the training of Transformers a more resource-intensive task. The Swin Transformer offers a middle ground between the local and global biases by introducing an attention window, which allows for the consideration of more information without being entirely global. As a result, the Swin Transformer utilizes the attention mechanism while requiring less data for training. This approach has proven to be highly effective for cancer spheroid segmentation, outperforming all other methods we tested.

Lastly, the automation of segmentation tasks, combined with integrated statistical analysis, represents a transformative step toward more efficient, reliable, and reproducible cancer research. By replacing manual segmentation, which typically requires approximately two minutes per sample, our pipeline has demonstrated the ability to process an entire week's worth of samples in under five minutes, a significant reduction in analysis time. This advancement not only accelerates research workflows but also minimizes human error, enhancing the consistency of results. Currently in use for daily analyses, the pipeline has received positive feedback, showing its practicality and potential despite existing limitations.

Bibliography

[1] I. Abd El-Sadek, R. Morishita, T. Mori, S. Makita, P. Mukherjee, S. Matsusaka, and Y. Yasuno. Label-Free Visualization and Quantification of the Drug-Type-Dependent Response of Tumor Spheroids by Dynamic Optical Coherence Tomography. *Scientific Reports*, 14:3366, 2024. 45, 48

[2] J. Aguilar Cosme, D. Gagui, H. Bryant, and F. Claeyssens. Morphological Response in Cancer Spheroids for Screening Photodynamic Therapy Parameters. *Frontiers in Molecular Biosciences*, 8, 2021. 48

[3] I. Ahonen, M. Å kerfelt, M. Toriseva, E. Oswald, J. Schüler, and M. Nees. A High-Content Image Analysis Approach for Quantitative Measurements of Chemosensitivity in Patient-Derived Tumor Microtissues. *Scientific Reports*, 7(1):1–18, 2017. 45, 46

[4] I. Ahonen, M. Åkerfelt, M. Toriseva, E. Oswald, J. Schüler, and M. Nees. A high-content image analysis approach for quantitative measurements of chemosensitivity in patient-derived tumor microtissues. *Scientific Reports*, 7(1):6600, 2017. 30

[5] C. Akbaba, A. Polat, and D. Göktürk. Tracing 2D Growth of Pancreatic Tumoroids Using the Combination of Image Processing Techniques and Mini-Opto Tomography Imaging System. *Technology in Cancer Research and Treatment*, 22, 2023. 46

[6] H. Alsehli, F. Mosis, C. Thompson, E. Hamrud, E. Wiseman, E. Gentleman, and D. Danovi. An Integrated Pipeline for High-Throughput Screening and Profiling of Spheroids Using Simple Live Image Analysis of Frame to Frame Variations. *Methods*, 190:33–43, 2021. 46

[7] K. Bai. A Comprehensive Introduction to Different Types of Convolutions in Deep Learning, 2019. `https://towardsdatascience.com/a-comprehensive-introduction-to-different-types-of-convolutions-in-deep-learning-669281e58215`. 27

[8] D. Bao, L. Wang, X. Zhou, S. Yang, K. He, and M. Xu. Automated Detection and Growth Tracking of 3D Bio-Printed Organoid Clusters Using Optical Coherence Tomography With Deep Convolutional Neural Networks. *Frontiers in bioengineering and biotechnology*, 11:1133090, 2023. 49

[9] M. Barbier, S. Jaensch, F. Cornelissen, S. Vidic, K. Gjerde, R. De Hoogt, R. Graeser, E. Gustin, Y. Chong, J. Hickman, M. Burbridge, E. Verschuren, O. Kallioniemi,

J. Klefström, C. Brisken, V. Rotter, M. Oren, C. Brito, J. Schalken, G. Jenster, W. Van Weerden, J. Vilo, J. Schueler, O. Monni, S. Barry, S. Grünewald, P. Kallio, H.-J. Mueller, A. Nopora, W. Sommergruber, E. Anderson, H. Van Der Kuip, M. Smalley, and E. Boghaert. Ellipsoid Segmentation Model for Analyzing Light-Attenuated 3D Confocal Image Stacks of Fluorescent Multi-Cellular Spheroids. *PLoS One*, 11(6), 2016. 48

[10] A. Beghin, G. Grenci, G. Sahni, S. Guo, H. Rajendiran, T. Delaire, S. B. Mohamad Raffi, D. Blanc, R. de Mets, and H. T. Ong. Automated High-Speed 3D Imaging of Organoid Cultures with Multi-Scale Phenotypic Quantification. *Nature Methods*, 19(7):881–892, 2022. 49

[11] I. Berg, E. Mohagheghian, K. Habing, N. Wang, and G. Underhill. Microtissue Geometry and Cell-Generated Forces Drive Patterning of Liver Progenitor Cell Differentiation in 3D. *Advanced Healthcare Materials*, 10(12), 2021. 48

[12] B. Beydag-Tasöz, J. D'Costa, L. Hersemann, B. Lee, F. Luppino, Y. Kim, C. Zechner, and A. Grapin-Botton. Integrating Single-Cell Imaging and Rna Sequencing Datasets Links Differentiation and Morphogenetic Dynamics of Human Pancreatic Endocrine Progenitors. *Developmental Cell*, 58(21):2292–2308.e6, 2023. 49

[13] C. C. Bilgin, G. Fontenay, Q. Cheng, H. Chang, J. Han, and B. Parvin. BioSig3D: High Content Screening of Three-Dimensional Cell Culture Models. *PLoS One*, 11 (3):e0148379, 2016. 24, 35, 46

[14] T. Booij, M. Klop, K. Yan, C. Szántai-Kis, B. Szokol, L. Orfi, B. Van De Water, G. Keri, and L. Price. Development of a 3D Tissue Culture-Based High-Content Screening Platform That Uses Phenotypic Profiling to Discriminate Selective Inhibitors of Receptor Tyrosine Kinases. *Journal of Biomolecular Screening*, 21(9): 912–922, 2016. 33, 45

[15] M. Borten, S. Bajikar, N. Sasaki, H. Clevers, and K. Janes. Automated Brightfield Morphometry of 3D Organoid Populations by OrganoSeg. *Scientific Reports*, 8(1), 2018. 46

[16] M. E. Boutin, T. C. Voss, S. A. Titus, K. Cruz-Gutierrez, S. Michael, and M. Ferrer. A High-Throughput Imaging and Nuclear Segmentation Analysis Protocol for Cleared 3D Culture Models. *Scientific Reports*, 8(1):11135, 2018. 24, 46

[17] S. Bradbury and P. J. Evennett. *Contrast Techniques in Light Microscopy.* Garland Science, 2020. 20

[18] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. 55

[19] R. Bruch, M. Vitacolonna, R. Rudolf, and M. Reischl. Prediction of Fluorescent Ki67 Staining in 3D Tumor Spheroids. In *Current Directions in Biomedical Engineering*, volume 8, pages 305–308, 2022. 49

[20] S. Bruningk, I. Rivens, C. Box, U. Oelfke, and G. ter Haar. 3D Tumour Spheroids for the Prediction of the Effects of Radiation and Hyperthermia Treatments. *Scientific Reports*, 10(1), 2020. 49

[21] B. G. Buchanan. A (very) brief history of artificial intelligence. *AI Magazine*, 26 (4):53–53, 2005. 24

[22] A.-L. Bulin, M. Broekgaarden, and T. Hasan. Comprehensive High-Throughput Image Analysis for Therapeutic Efficacy of Architecturally Complex Heterotypic Organoids. *Scientific Reports*, 7(1), 2017. 46

[23] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin. Albumentations: Fast and Flexible Image Augmentations. *Information*, 11(2):125, 2020. 58, 59

[24] T. Cannon, A. Shah, and M. Skala. Autofluorescence Imaging Captures Heterogeneous Drug Response Differences between 2D and 3D Breast Cancer Cultures. *Biomedical Optics Express*, 8(3):1911–1925, 2017. 35, 46

[25] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. In *European Conference on Computer Vision*, pages 205–218. Springer, 2022. 38, 39, 56

[26] A. Chambost, N. Berabez, O. Cochet-Escartin, F. Ducray, M. Gabut, C. Isaac, S. Martel, A. Idbaih, D. Rousseau, D. Meyronet, and S. Monnier. Machine Learning-Based Detection of Label-Free Cancer Stem-like Cell Fate. *Scientific Reports*, 12 (1), 2022. 46

[27] D.-M. Chang, H.-H. Hsu, P.-L. Ko, W.-J. Chang, T.-H. Hsieh, H.-M. Wu, and Y.-C. Tung. Rapid Time-Lapse 3D Oxygen Tension Measurements within Hydrogels Using Widefield Frequency-Domain Fluorescence Lifetime Imaging Microscopy (FD-FLIM) and Image Segmentation. *The Analyst*, 149:1727–1737, 2024. 46

[28] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv preprint arXiv:2102.04306*, 2021. 35, 56

[29] X. Chen, M. E. Kandel, S. He, C. Hu, Y. J. Lee, K. Sullivan, G. Tracy, H. J. Chung, H. J. Kong, M. Anastasio, and G. Popescu. Artificial Confocal Microscopy for Deep Label-Free Imaging. *Nature Photonics*, 17(3):250–258, 2023. 24, 46

[30] Z. Chen, N. Ma, X. Sun, Q. Li, Y. Zeng, F. Chen, S. Sun, J. Xu, J. Zhang, H. Ye, J. Ge, Z. Zhang, X. Cui, K. Leong, Y. Chen, and Z. Gu. Automated Evaluation of Tumor Spheroid Behavior in 3D Culture Using Deep Learning-Based Recognition. *Biomaterials*, 272:120770, 2021. 49

[31] Z. Chen, Y. Yang, and P.-O. Bagnaninchi. Hybrid Learning-Based Cell Aggregate Imaging with Miniature Electrical Impedance Tomography. *IEEE Transactions on Instrumentation and Measurement*, 70, 2021. 49

[32] Y.-H. Cheng, Y.-C. Chen, R. Brien, and E. Yoon. Scaling and Automation of a High-Throughput Single-Cell-Derived Tumor Sphere Assay Chip. *Lab on a Chip*, 16(19):3708–3717, 2016. 46

[33] N. Choo, S. Ramm, J. Luu, J. M. Winter, L. A. Selth, A. R. Dwyer, M. Frydenberg, J. Grummet, S. Sandhu, T. E. Hickey, W. D. Tilley, R. A. Taylor, G. P. Risbridger, M. G. Lawrence, and K. J. Simpson. High-Throughput Imaging Assay for Drug Screening of 3D Prostate Cancer Organoids. *SLAS Discovery*, 26(9):1107–1124, 2021. 48

[34] L. Cisneros Castillo, A.-D. Oancea, C. Stüllein, and A. Régnier-Vigouroux. A Novel Computer-Assisted Approach to Evaluate Multicellular Tumor Spheroid Invasion Assay. *Scientific Reports*, 6, 2016. 46

[35] W. R. Crum, O. Camara, and D. L. Hill. Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis. *IEEE Transactions on Medical Imaging*, 25(11):1451–1461, 2006. 41

[36] P. F. de Souza Oliveira, A. V. Faria, S. P. Clerici, E. M. Akagi, H. F. Carvalho, G. Z. Justo, N. Duran, and C. V. Ferreira-Halder. Violacein negatively modulates the colorectal cancer survival and epithelial–mesenchymal transition. *Journal of Cellular Biochemistry*, 123(7):1247–1258, 2022. 51

[37] T. Deckers, G. Hall, I. Papantoniou, J.-M. Aerts, and V. Bloemen. A Platform for Automated and Label-Free Monitoring of Morphological Features and Kinetics of Spheroid Fusion. *Frontiers in Bioengineering and Biotechnology*, 10, 2022. 45, 46

[38] A. Deloria, S. Haider, B. Dietrich, V. Kunihs, S. Oberhofer, M. Knofler, R. Leitgeb, M. Liu, W. Drexler, and R. Haindl. Ultra-High-Resolution 3D Optical Coherence Tomography Reveals Inner Structures of Human Placenta-Derived Trophoblast Organoids. *IEEE Transactions on Biomedical Engineering*, 68(8):2368–2376, 2021. 45

[39] N. Dimitriou, S. Flores-Torres, J. Kinsella, and G. Mitsis. Detection and Spatiotemporal Analysis of In-vitro 3D Migratory Triple-Negative Breast Cancer Cells. *Annals of Biomedical Engineering*, 51(2):318–328, 2023. 24, 46

[40] A. Diosdi, D. Hirling, M. Kovacs, T. Toth, M. Harmati, K. Koos, K. Buzas, F. Piccinini, and P. Horvath. A Quantitative Metric for the Comparative Evaluation of Optical Clearing Protocols for 3D Multicellular Spheroids. *Computational and Structural Biotechnology Journal*, 19:1233–1243, 2021. 45, 46

[41] R. Edmondson, J. J. Broglie, A. F. Adcock, and L. Yang. Three-Dimensional Cell Culture Systems and their Applications in Drug Discovery and Cell-Based Biosensors. *Assay and Drug Development Technologies*, 12(4):207–218, 2014. 14

[42] S. Edwards, V. Carannante, K. Kuhnigk, H. Ring, T. Tararuk, F. Hallböök, H. Blom, B. Önfelt, and H. Brismar. High-Resolution Imaging of Tumor Spheroids

and Organoids Enabled by Expansion Microscopy. *Frontiers in Molecular Biosciences*, 7, 2020. 46

[43] C. H. Emmerich, L. M. Gamboa, M. C. Hofmann, M. Bonin-Andresen, O. Arbach, P. Schendel, B. Gerlach, K. Hempel, A. Bespalov, U. Dirnagl, et al. Improving Target Assessment in Biomedical Research: The GOT-IT Recommendations. *Nature Reviews Drug Discovery*, 20(1):64–81, 2021. 13

[44] Y. S. Erdem, A. Ayanzadeh, B. Mayalı, M. Balıkçi, Ö. N. Belli, M. Uçar, Ö. Y. Özyusal, D. P. Okvur, S. Önal, K. Morani, et al. Automated Analysis of Phase-Contrast Optical Microscopy Time-Lapse Images: Application to Wound Healing and Cell Motility Assays of Breast Cancer. In *Diagnostic Biomedical Signal and Image Processing Applications with Deep Learning Methods*, pages 137–154. Elsevier, 2023. 49

[45] P. Favreau, J. He, D. Gil, D. Deming, J. Huisken, and M. Skala. Label-Free Redox Imaging of Patient-Derived Organoids Using Selective Plane Illumination Microscopy. *Biomedical Optics Express*, 11(5):2591–2606, 2020. 46

[46] K. A. Fitzgerald, M. Malhotra, C. M. Curtin, F. J. O'Brien, and C. M. O'Driscoll. Life in 3D is Never Flat: 3D Models to Optimise Drug Delivery. *Journal of Controlled Release*, 215:39–54, 2015. 14

[47] R. R. C. for History and N. Media. Zotero, 2024. URL `https://www.zotero.org`. 43

[48] M. García-Domínguez, C. Domínguez, J. Heras, E. Mata, and V. Pascual. Deep Style Transfer to Deal with the Domain Shift Problem on Spheroid Segmentation. *Neurocomputing*, 569, 2024. 49

[49] C. M. Garvey, E. Spiller, D. Lindsay, C.-T. Chiang, N. C. Choi, D. B. Agus, P. Mallick, J. Foo, and S. M. Mumenthaler. A High-Content Image-Based Method for Quantitatively Studying Context-Dependent Cell Population Dynamics. *Scientific Reports*, 6(1):1–12, 2016. 48

[50] A. Géron. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly, 2022. 27

[51] D. A. Gil, D. Deming, and M. C. Skala. Patient-Derived Cancer Organoid Tracking with Wide-Field One-Photon Redox Imaging to Assess Treatment Response. *Journal of Biomedical Optics*, 26(3), 2021. 46

[52] A. Gillette, C. Babiarz, A. Vandommelen, C. Pasch, L. Clipson, K. Matkowskyj, D. Deming, and M. Skala. Autofluorescence Imaging of Treatment Response in Neuroendocrine Tumor Organoids. *Cancers*, 13(8), 2021. 46

[53] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. 21, 24, 25, 26

[54] I. Grexa, A. Diosdi, M. Harmati, A. Kriston, N. Moshkov, K. Buzas, V. Pietiäinen, K. Koos, and P. Horvath. SpheroidPicker for Automated 3D Cell Culture Manipulation Using Deep Learning. *Scientific Reports*, 11(1):14813, 2021. 49

[55] S. Grosser, J. Lippoldt, L. Oswald, M. Merkel, D. Sussman, F. Renner, P. Gottheil, E. Morawetz, T. Fuhs, X. Xie, S. Pawlizak, A. Fritsch, B. Wolf, L.-C. Horn, S. Briest, B. Aktas, M. Manning, and J. Kas. Cell and Nucleus Shape as an Indicator of Tissue Fluidity in Carcinoma. *Physical Review X*, 11(1), 2021. 24, 33, 46

[56] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. Fernández del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array Programming with NumPy. *Nature*, 585: 357–362, 2020. 55

[57] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 26

[58] T. Henser-Brownhill, R. Ju, N. Haass, S. Stehbens, C. Ballestrem, and T. Cootes. Estimation of Cell Cycle States of Human Melanoma Cells with Quantitative Phase Imaging and Deep Learning. In *International Symposium on Biomedical Imaging*, 2020. 46

[59] G. E. Hinton. Learning multiple layers of representation. *Trends in Cognitive Sciences*, 11(10):428–434, 2007. 25

[60] L. Hof, T. Moreth, M. Koch, T. Liebisch, M. Kurtz, J. Tarnick, S. Lissek, M. Verstegen, L. van der Laan, M. Huch, F. Matthäus, E. Stelzer, and F. Pampaloni. Long-Term Live Imaging and Multiscale Analysis Identify Heterogeneity and Core Principles of Epithelial Organoid Morphogenesis. *BMC Biology*, 19(1), 2021. 24, 46

[61] Y. Hou, J. Konen, D. Brat, A. Marcus, and L. Cooper. TASI: A Software Tool for Spatial-Temporal Quantification of Tumor Spheroid Dynamics. *Scientific Reports*, 8(1), 2018. 46

[62] L. Hu, B. Ter Hofstede, D. Sharma, F. Zhao, and A. Walsh. Comparison of Phasor Analysis and Biexponential Decay Curve Fitting of Autofluorescence Lifetime Imaging Data for Machine Learning Prediction of Cellular Phenotypes. *Frontiers in Bioinformatics*, 3, 2023. 48

[63] L. Hu, N. Wang, J. Bryant, L. Liu, L. Xie, A. West, and A. Walsh. Label-Free Spatially Maintained Measurements of Metabolic Phenotypes in Cells. *Frontiers in Bioengineering and Biotechnology*, 11, 2023. 48

[64] J. P. Hughes, S. Rees, S. B. Kalindjian, and K. L. Philpott. Principles of Early Drug Discovery. *British Journal of Pharmacology*, 162(6):1239–1249, 2011. 13, 15

[65] H. Ibrahim, S. D. Thorpe, M. Paukshto, T. S. Zaitseva, W. Moritz, and B. J. Rodriguez. A Biomimetic High Throughput Model of Cancer Cell Spheroid Dissemination onto Aligned Fibrillar Collagen. *SLAS Technology*, 27(4):267–275, 2022. 47, 48

[66] N. Ibtehaz and M. S. Rahman. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121:74–87, 2020. 29, 31, 56

[67] N. Jadav, S. Velamoor, D. Huang, L. Cassin, N. Hazelton, A.-R. Eruera, L. Burga, and M. Bostina. Beyond the Surface: Investigation of Tumorsphere Morphology Using Volume Electron Microscopy. *Journal of Structural Biology*, 215(4), 2023. 49

[68] N. Jagiella, B. Müller, M. Müller, I. Vignon-Clementel, and D. Drasdo. Inferring Growth Control Mechanisms in Growing Multi-cellular Spheroids of NSCLC Cells from Spatial-Temporal Image Data. *PLoS Computational Biology*, 12(2), 2016. 24, 46

[69] E. Jones, T. Oliphant, and P. Peterson. SciPy: Open Source Scientific Tools for Python, 2001. `http://www.scipy.org`. 55

[70] Y. Kang, J. Deng, J. Ling, X. Li, Y.-J. Chiang, E. Koay, H. Wang, J. Burks, P. Chiao, M. Hurd, M. Bhutani, J. Lee, B. Weston, A. Maitra, N. Ikoma, C.-W. Tzeng, J. Lee, R. DePinho, R. Wolff, S. Pant, F. McAllister, M. Katz, J. Fleming, and M. Kim. 3D Imaging Analysis on an Organoid-Based Platform Guides Personalized Treatment in Pancreatic Ductal Adenocarcinoma. *Journal of Clinical Investigation*, 132(24), 2022. 48

[71] M. Kapalczynska, T. Kolenda, W. Przybyla, M. Zajaczkowska, A. Teresiak, V. Filas, M. Ibbs, R. Blizniak, Ł. Łuczewski, and K. Lamperska. 2D and 3D Cell Cultures - A Comparison of Different Types of Cancer Cell Cultures. *Archives of Medical Science*, 14(4):910, 2018. 14

[72] C. Karabag, M. Jones, C. Peddie, A. Weston, L. Collinson, and C. Reyes-Aldasoro. Segmentation and Modelling of the Nuclear Envelope of HeLa Cells Imaged with Serial Block Face Scanning Electron Microscopy. *Journal of Imaging*, 5(9), 2019. 45, 46

[73] H. Karimi, B. Leszczyński, T. Koł odziej, E. Kubicz, M. Przybył o, and E. Stepień. X-Ray Microtomography as a New Approach for Imaging and Analysis of Tumor Spheroids. *Micron*, 137, 2020. 45

[74] T. Kaseva, B. Omidali, E. Hippeläinen, T. Mäkelä, U. Wilppu, A. Sofiev, A. Merivaara, M. Yliperttula, S. Savolainen, and E. Salli. Marker-Controlled Watershed with Deep Edge Emphasis and Optimized H-minima Transform for Automatic Segmentation of Densely Cultivated 3D Cell Nuclei. *BMC Bioinformatics*, 23(1): 289, 2022. 45, 49

[75] F. Keller, R. Rudolf, and M. Hafner. Towards Optimized Breast Cancer 3D Spheroid Mono-and Co-Culture Models for Pharmacological Research and Screening. *Journal of Cellular Biotechnology*, 5(2):89–101, 2019. 46

[76] M. Koch, S. Nickel, R. Lieshout, S. Lissek, M. Leskova, L. van der Laan, M. Verstegen, B. Christ, and F. Pampaloni. Label-Free Imaging Analysis of Patient-Derived Cholangiocarcinoma Organoids after Sorafenib Treatment. *Cells*, 11(22), 2022. 48

[77] B. Kovac, J. Fehrenbach, L. Guillaume, and P. Weiss. FitEllipsoid: A Fast Supervised Ellipsoid Segmentation Plugin. *BMC Bioinformatics*, 20(1), 2019. 46

[78] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 2012. 26

[79] L. Fillioux, E. Gontran, J. Cartry, J. R. Mathieu, S. Bedja, A. Boilève, P. -H. Cournède, F. Jaulin, S. Christodoulidis, and M. Vakalopoulou. Spatio-Temporal Analysis of Patient-Derived Organoid Videos Using Deep Learning for the Prediction of Drug Efficacy. In *International Conference on Computer Vision Workshops*, pages 3932–3941, 2023. 46

[80] D. Lacalle, H. A. Castro-Abril, T. Randelovic, C. Domínguez, J. Heras, E. Mata, G. Mata, Y. Méndez, V. Pascual, and I. Ochoa. SpheroidJ: An Open-Source Set of Tools for Spheroid Segmentation. *Computer Methods and Programs in Biomedicine*, 200:105837, 2021. 8, 10, 15, 45, 47, 49, 50, 53, 54, 55

[81] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989. 26

[82] Y. LeCun, Y. Bengio, and G. Hinton. Deep Learning. *Nature*, 521:436–444, 2015. 25

[83] J. Lee, Y. Kim, J. Lim, H.-I. Jung, G. Castellani, F. Piccinini, and B. Kwak. Optimization of Tumor Spheroid Preparation and Morphological Analysis for Drug Evaluation. *Biochip Journal*, 18:160–169, 2024. 45, 48

[84] G. V. Leite, J. M. Azevedo-Martins, C. V. Ferreira-Halder, and H. Pedrini. Cancer Spheroid Segmentation Based on Vision Transformer. In *IEEE International Conference on Visual Communications and Image Processing*, pages 1–5, 2023. 15, 45, 49

[85] J. W. Lichtman and J.-A. Conchello. Fluorescence Microscopy. *Nature methods*, 2 (12):910–919, 2005. 20

[86] B. Liu, Y. Zhu, Z. Yang, H. H. N. Yan, S. Y. Leung, and J. Shi. Deep Learning-Based 3D Single-Cell Imaging Analysis Pipeline Enables Quantification of Cell-Cell Interaction Dynamics in the Tumor Microenvironment. *Cancer Research*, 84(4): 517–526, 2024. 45, 49

[87] X. Liu, C. Flinders, S. Mumenthaler, and A. Hummon. MALDI Mass Spectrometry Imaging for Evaluation of Therapeutics in Colorectal Tumor Organoids. *Journal of the American Society for Mass Spectrometry*, 29(3):516–526, 2018. 48

[88] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In *IEEE International Conference on Computer Vision*, pages 10012–10022, 2021. 32, 36

[89] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, et al. Swin Transformer v2: Scaling up Capacity and Resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 12009–12019, 2022. 56

[90] J. Long, E. Shelhamer, and T. Darrell. Fully Convolutional Networks for Semantic Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015. 27, 56

[91] M. Lotsberg, G. Rø sland, A. Rayford, S. Dyrstad, C. Ekanger, N. Lu, K. Frantz, L. Stuhr, H. Ditzel, J. Thiery, L. Akslen, J. Lorens, and A. Engelsen. Intrinsic Differences in Spatiotemporal Organization and Stromal Cell Interactions Between Isogenic Lung Cancer Cells of Epithelial and Mesenchymal Phenotypes Revealed by High-Dimensional Single-Cell Analysis of Heterotypic 3D Spheroid Models. *Frontiers in Oncology*, 12, 2022. 48

[92] J. M. Matthews, B. Schuster, S. S. Kashaf, P. Liu, R. Ben-Yishay, D. Ishay-Ronen, E. Izumchenko, L. Shen, C. R. Weber, and M. Bielski. OrganoID: A Versatile Deep Learning Platform for Tracking and Analysis of Single-Organoid Dynamics. *PLOS Computational Biology*, 18(11):e1010584, 2022. 45, 49

[93] T. Maylaa, F. Windal, H. Benhabiles, G. Maubon, N. Maubon, E. Vandenhaute, and D. Collard. A Hierarchical Deep Learning Framework for Nuclei 3D Reconstruction from Microscopic Stack-Images of 3D Cancer Cell Culture. In *Lecture Notes in Networks and Systems*, volume 579, pages 225–235, 2023. 49

[94] J. McIntosh, L. Yang, T. Wang, H. Zhou, M. Lockett, and A. Oldenburg. Tracking the Invasion of Breast Cancer Cells in Paper-Based 3D Cultures by OCT Motility Analysis. *Biomedical Optics Express*, 11(6):3181–3194, 2020. 45

[95] T. Mendonca, K. Lis-Slimak, A. Matheson, M. Smith, A. Anane-Adjei, J. Ashworth, R. Cavanagh, L. Paterson, P. Dalgarno, C. Alexander, M. Tassieri, C. Merry, and A. Wright. OptoRheo: Simultaneous in situ micro-mechanical sensing and imaging of live 3D biological systems. *Communications Biology*, 6(1), 2023. 46

[96] A. Merivaara, E. Koivunotko, K. Manninen, T. Kaseva, J. Monola, E. Salli, R. Koivuniemi, S. Savolainen, S. Valkonen, and M. Yliperttula. Stiffness-Controlled Hydrogels for 3D Cell Culture Models. *Polymers*, 14(24):5530, 2022. 49

[97] J. Michálek, K. ˇ Stˇ epka, M. Kozubek, J. Navrátilová, B. Pavlatovská, M. MacHálková, J. Preisler, and A. Pruˇ ska. Quantitative Assessment of Anti-Cancer Drug Efficacy from Coregistered Mass Spectrometry and Fluorescence Microscopy Images of Multicellular Tumor Spheroids. *Microscopy and Microanalysis*, 25(6):1311–1322, 2019. 46

[98] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos. Image Segmentation Using Deep Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 21

[99] F. Mittler, P. Obeïd, A. Rulina, V. Haguet, X. Gidrol, and M. Balakirev. High-Content Monitoring of Drug Effects in a 3D Spheroid Model. *Frontiers in Oncology*, 7, 2017. 48

[100] C. Moriconi, V. Palmieri, R. Di Santo, G. Tornillo, M. Papi, G. Pilkington, M. De Spirito, and M. Gumbleton. INSIDIA: A FIJI Macro Delivering High-Throughput and High-Content Spheroid Invasion Analysis. *Biotechnology Journal*, 12(10), 2017. 46

[101] P. Mukashyaka, P. Kumar, D. J. Mellert, S. Nicholas, J. Noorbakhsh, M. Brugiolo, O. Anczukow, E. T. Liu, and J. H. Chuang. Cellos : High-throughput Deconvolution of 3D Organoid Dynamics at Cellular Resolution for Cancer Pharmacology. *bioRxiv: The Preprint Server for Biology*, page 2023.03.03.531019, 2023. 45, 49

[102] P. Mukashyaka, P. Kumar, D. J. Mellert, S. Nicholas, J. Noorbakhsh, M. Brugiolo, E. T. Courtois, O. Anczukow, E. T. Liu, and J. H. Chuang. High-throughput deconvolution of 3D organoid dynamics at cellular resolution for cancer pharmacology with Cellos. *Nature Communications*, 14(1):8406, 2023. 49

[103] V. Murali, B.-J. Chang, R. Fiolka, G. Danuser, M. Cobanoglu, and E. Welf. An Image-Based Assay to Quantify Changes in Proliferation and Viability upon Drug Treatment in 3D Microenvironments. *BMC Cancer*, 19(1), 2019. 46

[104] K. Ndyabawe, M. Haidekker, A. Asthana, and W. Kisaalita. Spheroid Trapping and Calcium Spike Estimation Techniques toward Automation of 3D Culture. *SLAS Technology*, 26(3):265–273, 2021. 46

[105] T. K. N. Ngo, S. J. Yang, B.-H. Mao, T. K. M. Nguyen, Q. D. Ng, Y.-L. Kuo, J.-H. Tsai, S. N. Saw, and T.-Y. Tu. A Deep Learning-Based Pipeline for Analyzing the Influences of Interfacial Mechanochemical Microenvironments on Spheroid Invasion Using Differential Interference Contrast Microscopic Images. *Materials Today. Bio*, 23:100820, 2023. 49

[106] E. Nürnberg, M. Vitacolonna, J. Klicks, E. von Molitor, T. Cesetti, F. Keller, R. Bruch, T. Ertongur-Fauth, K. Riedel, P. Scholz, T. Lau, R. Schneider, J. Meier, M. Hafner, and R. Rudolf. Routine Optical Clearing of 3D-Cell Cultures: Simplicity Forward. *Frontiers in Molecular Biosciences*, 7, 2020. 24, 46

[107] N. Otsu et al. A Threshold Selection Method From Gray-Level Histograms. *Automatica*, 11(285-296):23–27, 1975. 23

[108] T. Park, T. K. Kim, Y. D. Han, K.-A. Kim, H. Kim, and H. S. Kim. Development of a Deep Learning Based Image Processing Tool for Enhanced Organoid Analysis. *Scientific Reports*, 13(1):19841, 2023. 49

[109] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 55

[110] J. Pawley. *Handbook of Biological Confocal Microscopy*, volume 236. Springer Science & Business Media, 2006. 20

[111] H. Pedrini and W. R. Schwartz. *Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações*. Cengage Learning, 2007. 26, 27

[112] G. Perini, E. Rosa, G. Friggeri, L. Di Pietro, M. Barba, O. Parolini, G. Ciasca, C. Moriconi, M. Papi, M. De Spirito, and V. Palmieri. INSIDIA 2.0 High-Throughput Analysis of 3D Cancer Models: Multiparametric Quantification of Graphene Quantum Dots Photothermal Therapy for Glioblastoma and Pancreatic Cancer. *International Journal of Molecular Sciences*, 23(6), 2022. 48

[113] L. Petrovic, M. Oumano, J. Hanlon, M. Arnoldussen, I. Koruga, S. Yasmin-Karim, W. Ngwa, and J. Celli. Image-Based Quantification of Gold Nanoparticle Uptake and Localization in 3D Tumor Models to Inform Radiosensitization Schedule. *Pharmaceutics*, 14(3), 2022. 46

[114] F. Piccinini, A. Tesei, M. Zanoni, and A. Bevilacqua. ReViMS: Software Tool for Estimating the Volumes of 3-D Multicellular Spheroids Imaged Using a Light Sheet Fluorescence Microscope. *BioTechniques*, 63(5):227–229, 2017. 46

[115] Plotly. Plotly, 2024. URL `https://github.com/plotly/plotly.py`. 80

[116] R. T. Powell, M. J. Moussalli, L. Guo, G. Bae, P. Singh, C. Stephan, I. Shureiqi, and P. J. Davies. deepOrganoid: A Brightfield Cell Viability Model for Screening Matrix-Embedded Organoids. *SLAS Discovery*, 27(3):175–184, 2022. 48

[117] S. Przyborski. *Technology Platforms for 3D Cell Culture: A User's Guide*. John Wiley & Sons, 2017. 14

[118] Y. Qi, Y. Liu, Y. Huang, M. Xiong, S. You, B. Wang, and M. Gu. A Three-Dimensional Technique for The Visualization of Mitochondrial Ultrastructural Changes in Pancreatic Cancer Cells. *Journal of Visualized Experiments*, 2023(196), 2023. 45

[119] Y. Qin, J. Li, Y. Chen, Z. Wang, Y.-A. Huang, Z. You, L. Hu, P. Hu, and F. Tan. TransOrga: End-To-End Multi-modal Transformer-Based Organoid Segmentation. In *International Conference on Intelligent Computing*, pages 460–472. Springer, 2023. 49

[120] S. Rahkonen, E. Koskinen, I. Pölönen, T. Heinonen, T. Ylikomi, S. Äyrämö, and M. Eskelinen. Multilabel Segmentation of Cancer Cell Culture on Vascular Structures with Deep Neural Networks. *Journal of Medical Imaging*, 7(2), 2020. 49

[121] S. Ramm, R. Vary, T. Gulati, J. Luu, K. J. Cowley, M. S. Janes, N. Radio, and K. J. Simpson. High-Throughput Live and Fixed Cell Imaging Method to Screen Matrigel-Embedded Organoids. *Organoids*, 2(1):1–19, 2023. 15, 48

[122] V. Reijonen, L. Kanninen, E. Hippeläinen, Y.-R. Lou, E. Salli, A. Sofiev, M. Malinen, T. Paasonen, M. Yliperttula, A. Kuronen, and S. Savolainen. Multicellular Dosimetric Chain for Molecular Radiotherapy Exemplified with Dose Simulations on 3D Cell Spheroids. *Physica Medica*, 40:72–78, 2017. 24, 46

[123] F. Rieken Münke, L. Rettenberger, A. Popova, and M. Reischl. A Lightweight Framework for Semantic Segmentation of Biomedical Images. In *Current Directions in Biomedical Engineering*, volume 9, pages 190–193. De Gruyter, 2023. 49

[124] A. Rogozhnikov. Einops: Clear and Reliable Tensor Manipulations with Einstein-like Notation. In *International Conference on Learning Representations*, 2022. 55

[125] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015. 27, 28, 29, 47

[126] A. Sargenti, F. Musmeci, C. Cavallo, M. Mazzeschi, S. Bonetti, S. Pasqua, F. Bacchi, G. Filardo, D. Gazzola, M. Lauriola, and S. Santi. A New Method for the Study of Biophysical and Morphological Parameters in 3D Cell Cultures: Evaluation in LoVo Spheroids Treated with Crizotinib. *PLoS One*, 16, 2021. 48

[127] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, et al. Fiji: An Open-Source Platform for Biological-Image Analysis. *Nature Methods*, 9(7):676–682, 2012. 47

[128] A. Schmitz, S. Fischer, C. Mattheyer, F. Pampaloni, and E. Stelzer. Multiscale Image Analysis Reveals Structural Heterogeneity of the Cell Microenvironment in Homotypic Spheroids. *Scientific Reports*, 7, 2017. 46

[129] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri. NIH Image to ImageJ: 25 Years of Image Analysis. *Nature Methods*, 9(7):671–675, 2012. 47

[130] J. Schurr, C. Eilenberger, P. Ertl, J. Scharinger, and S. Winkler. Automated Evaluation of Cell Viability in Microfluidic Spheroid Arrays. In *10th International Workshop on Innovative Simulation for Health Care*, pages 27–35, 2021. 46

[131] B. Schuster, M. Junkin, S. Kashaf, I. Romero-Calvo, K. Kirby, J. Matthews, C. Weber, A. Rzhetsky, K. White, and S. Tay. Automated Microfluidic Platform for Dynamic and Combinatorial Drug Screening of Tumor Organoids. *Nature Communications*, 11(1), 2020. 46

[132] M. Sharma, V. Goudar, M. Koduri, F. Tseng, and M. Bhattacharya. Quantitative and Qualitative Image Analysis of in Vitro Co-Culture 3D Tumor Spheroid Model by Employing Image-Processing Techniques. *Applied Sciences*, 11(10), 2021. 36, 38, 45

[133] K. Shirai, H. Kato, Y. Imai, M. Shibuta, K. Kanie, and R. Kato. The Importance of Scoring Recognition Fitness in Spheroid Morphological Analysis for Robust Label-Free Quality Evaluation. *Regenerative Therapy*, 14:205–214, 2020. 48

[134] W. Shuyun, F. Lin, C. Pan, Q. Zhang, H. Tao, M. Fan, L. Xu, K. Kong, Y. Chen, D. Lin, and S. Feng. Laser Tweezer Raman Spectroscopy Combined With Deep Neural Networks for Identification of Liver Cancer Cells. *Talanta*, 264, 2023. 49

[135] S. Silvani, M. Figliuzzi, and A. Remuzzi. Toxicological Evaluation of Airborne Particulate Matter. Are Cell Culture Technologies Ready to Replace Animal Testing? *Journal of Applied Toxicology*, 39(11):1484–1491, 2019. 14

[136] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 26

[137] I. Sinenko, F. Kuttler, V. Simeonov, A. Moulin, P. Aouad, C. Stathopoulos, F. Munier, A. Berger, and P. Dyson. Translational Screening Platform to Evaluate Chemotherapy in Combination With Focal Therapy for Retinoblastoma. *Cancer Science*, 114(9):3728–3739, 2023. 48

[138] I. Smyrek and E. Stelzer. Quantitative Three-Dimensional Evaluation of Immunofluorescence Staining for Large Whole Mount Spheroids with Light Sheet Microscopy. *Biomedical Optics Express*, 8(2):484–499, 2017. 34, 46

[139] E. R. Spiller, N. Ung, S. Kim, K. Patsch, R. Lau, C. Strelez, C. Doshi, S. Choung, B. Choi, and E. F. Juarez Rosales. Imaging-Based Machine Learning Analysis of Patient-Derived Tumor Organoid Drug Response. *Frontiers in Oncology*, 11:771173, 2021. 45, 48

[140] H. Steinhaus. Sur la Division Des Corps Matériels en Parties. *Bulletin of the Polish Academy of Sciences Technical Sciences*, 1(804):801, 1956. 24

[141] Y. Sun, Z. Lu, J. A. Taylor, and J. L. S. Au. Quantitative Image Analysis of Intracellular Protein Translocation in 3-Dimensional Tissues for Pharmacodynamic Studies of Immunogenic Cell Death. *Journal of Controlled Release : Official Journal of the Controlled Release Society*, 365:89–100, 2024. 48

[142] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3):209–249, 2021. 13

[143] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015. 26

[144] S. Tanaka, K. Takizawa, and F. Nakamura. One-Step Visualization of Natural Cell Activities in Non-Labeled Living Spheroids. *Scientific Reports*, 12(1), 2022. 48

[145] E. 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*, 36(9):2948–2949, 2020. 48

[146] P. J. Tebon, B. Wang, A. L. Markowitz, A. Davarifar, B. L. Tsai, P. Krawczuk, A. E. Gonzalez, S. Sartini, G. F. Murray, H. T. L. Nguyen, N. Tavanaie, T. L. Nguyen, P. C. Boutros, M. A. Teitell, and A. Soragni. Drug Screening at Single-Organoid Resolution via Bioprinting and Interferometry. *Nature Communications*, 14(1):3168, 2023. 45, 49

[147] F. Tobias, J. McIntosh, G. Labonia, M. Boyce, M. Lockett, and A. Hummon. Developing a Drug Screening Platform: MALDI-Mass Spectrometry Imaging of Paper-Based Cultures. *Analytical Chemistry*, 91(24):15370–15376, 2019. 46

[148] I. Ulku and E. Akagündüz. A survey on deep learning-based architectures for semantic segmentation on 2D images. *Applied Artificial Intelligence*, 36(1):2032924, 2022. 16, 22

[149] J. M. J. Valanarasu and V. M. Patel. UneXt: MLP-Based Rapid Medical Image Segmentation Network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 23–33. Springer, 2022. 31, 32, 56

[150] A. Van Hemelryk, S. Erkens-Schulze, L. Lim, C. M. A. de Ridder, D. C. Stuurman, G. W. Jenster, M. E. van Royen, and W. M. van Weerden. Viability Analysis and High-Content Live-Cell Imaging for Drug Testing in Prostate Cancer Xenograft-Derived Organoids. *Cells*, 12(10), 2023. 46

[151] G. van Rossum and F. L. Drake Jr. *Python Reference Manual*. Centrum voor Wiskunde en Informatica, Amsterdam, 1995. 55

[152] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is All You Need. In *Advances in Neural Information Processing Systems*, pages 5998–6008, 2017. 33, 73

[153] C. Veelken, G.-J. Bakker, D. Drell, and P. Friedl. Single Cell-Based Automated Quantification of Therapy Responses of Invasive Cancer Spheroids in Organotypic 3D Culture. *Methods*, 128:139–149, 2017. 48

[154] L. Vincent and P. Soille. Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(06):583–598, 1991. 23

[155] M. Vinci, C. Box, and S. A. Eccles. Three-Dimensional (3D) Tumor Spheroid Invasion Assay. *Journal of Visualized Experiments*, 99:e52686, 2015. 15

[156] C. K. Vong, A. Wang, M. Dragunow, T. I.-H. Park, and V. Shim. Quantification of Tumoursphere Migration with a Physics-based Machine Learning Method. *Cytometry Part A*, 2023. 49

[157] A. Walsh, J. Castellanos, N. Nagathihalli, N. Merchant, and M. Skala. Optical Imaging of Drug-Induced Metabolism Changes in Murine and Human Pancreatic Cancer Organoids Reveals Heterogeneous Drug Response. *Pancreas*, 45(6):863–869, 2016. 46

[158] X. Wan, P. Bovornchutichai, Z. Cui, E. O'Neill, and H. Ye. Morphological Analysis of Human Umbilical Vein Endothelial Cells Co-Cultured with Ovarian Cancer Cells in 3D: An Oncogenic Angiogenesis Assay. *PLoS One*, 12(7), 2017. 46

[159] L. Wang, J. Goldwag, M. Bouyea, J. Barra, K. Matteson, N. Maharjan, A. Eladdadi, M. Embrechts, X. Intes, U. Kruger, and M. Barroso. Spatial Topology of Organelle Is a New Breast Cancer Cell Classifier. *Iscience*, 26(7), 2023. 48

[160] X. Wang, C. Wu, S. Zhang, P. Yu, L. Li, C. Guo, and R. Li. A Novel Deep Learning Segmentation Model for Organoid-Based Drug Screening. *Frontiers in Pharmacology*, 13, 2022. 49

[161] Y. Wang and A. B. Hummon. Quantification of Irinotecan in Single Spheroids Using Internal Standards by MALDI Mass Spectrometry Imaging. *Analytical Chemistry*, 95(24):9227–9236, 2023. 48

[162] J. Wardwell-Swanson, M. Suzuki, K. Dowell, M. Bieri, E. Thoma, I. Agarkova, F. Chiovaro, S. Strebel, N. Buschmann, F. Greve, and O. Frey. A Framework for Optimizing High-Content Imaging of 3D Models for Drug Discovery. *SLAS Discovery*, 25(7):709–722, 2020. 46

[163] C. Wen, T. Miura, V. Voleti, K. Yamaguchi, M. Tsutsumi, K. Yamamoto, K. Otomo, Y. Fujie, T. Teramoto, T. Ishihara, K. Aoki, T. Nemoto, E. Hillman, and K. Kimura. 3DeeCellTracker, a Deep Learning-Based Pipeline for Segmenting and Tracking Cells in 3D Time Lapse Images. *eLife*, 10, 2021. 49

[164] G. Winkelmaier and B. Parvin. An Enhanced Loss Function Simplifies the Deep Learning Model for Characterizing the 3D Organoid Models. *Bioinformatics*, 37 (18):3084–3085, 2021. 45, 49

[165] O. J. Wouters, M. McKee, and J. Luyten. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *Jama*, 323(9): 844–853, 2020. 13, 15

[166] B. Wu, C. Xu, X. Dai, A. Wan, P. Zhang, Z. Yan, M. Tomizuka, J. Gonzalez, K. Keutzer, and P. Vajda. Visual Transformers: Token-based Image Representation and Processing for Computer Vision. *arXiv preprint arXiv:2006.03677*, 2020. 32

[167] X. Deng, L. Hu, Z. -H. You, and P. -W. Hu. CAMPEOD: A Cross Attention-Based Multi-Scale Patch Embedding Organoid Detection Model. In *International Conference on Bioinformatics and Biomedicine*, pages 1068–1073, 2023. 49

[168] X. Deng, L. Hu, Z. Jiang, L. Liu, and P. -W. Hu. A Contactless Automated Dynamic Monitoring Method for Organoid Morphology on the Time Axis. In *IEEE Conference on Data Mining Workshops*, pages 412–417, 2023. 49

[169] P. Xie, H. Zhang, P. Wu, Y. Chen, and Z. Cai. Three-Dimensional Mass Spectrometry Imaging Reveals Distributions of Lipids and the Drug Metabolite Associated with the Enhanced Growth of Colon Cancer Cell Spheroids Treated with Triclosan. *Analytical Chemistry*, 94(40):13667–13675, 2022. 48

[170] K. Yao, J. Sun, K. Huang, L. Jing, H. Liu, D. Huang, and C. Jude. Analyzing Cell-Scaffold Interaction through Unsupervised 3D Nuclei Segmentation. *International Journal of Bioprinting*, 8(1):495, 2021-12-30. 49

[171] V. Zanotelli, M. Leutenegger, X.-K. Lun, F. Georgi, N. de Souza, and B. Bodenmiller. A Quantitative Analysis of the Interplay of Environment, Neighborhood, and Cell State in 3D Spheroids. *Molecular Systems Biology*, 16(12), 2020. 48

[172] F. Zernike. Phase Contrast, A New Method for the Microscopic Observation of Transparent Objects. *Physica*, 9(7):686–698, 1942. 20

[173] L. Zhang, L. Wang, S. Yang, K. He, D. Bao, and M. Xu. Quantifying the Drug Response of Patient-Derived Organoid Clusters by Aggregated Morphological Indicators With Multi-Parameters Based on Optical Coherence Tomography. *Biomedical Optics Express*, 14(4):1703–1717, 2023. 49

[174] S. Zhang, L. Li, P. Yu, C. Wu, X. Wang, M. Liu, S. Deng, C. Guo, and R. Tan. A Deep Learning Model for Drug Screening and Evaluation in Bladder Cancer Organoids. *Frontiers in Oncology*, 13:1064548, 2023. 15, 49

[175] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang. UNet++: A Nested U-Net architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, and 8th International Workshop, held in conjunction with MICCAI 2018*, pages 3–11, Granada, Spain, Sept. 2018. Springer. 29, 31, 56