Universiteit Leiden
The Netherlands

Computer Science - AI

Identifying cell cycle phase using DNA staining in time-lapse microscopy.

Vicent Descals Carbonell

Supervisors:
Lu Cao & Sylvia Le Dévédec

MASTER THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

June 27, 2024

Abstract

This thesis introduces a novel method for identifying cell cycle stages using deep learning, with a focus on single-channel DNA staining in time-lapse microscopy. Utilizing a UNet architecture with EfficientNet encoders, the research minimizes imaging complexity and data needs while preserving high segmentation precision. Experimental validation confirmed the efficacy of the model, attaining **93.72% average precision**, **92.87% average recall**, and **93.29% average F1 score** in classifying breast cancer cell phases. Although the study employs static images, it underscores the potential for future enhancements by integrating temporal dynamics using other deep learning models with hybrid transformer-CNN architectures. This work contributes to cellular biology analysis by identifying single cells and classifying their cell cycle phases using a single DNA staining method.

Contents

1 Introduction

Understanding the intricacies of the cell cycle is crucial in the fields of biology and medicine, serving as a foundation for deciphering the complex mechanisms of cellular growth, division, and repair. The cell cycle plays a pivotal role in the maintenance of tissue homeostasis and the response to cellular damage. Moreover, aberrations in the cell cycle are often linked to various diseases, particularly cancer, where uncontrolled cell division leads to tumor growth and progression. Therefore, precise analysis of the cell cycle is essential for developing targeted cancer therapies and improving our understanding of oncogenesis.

Traditional methods for analyzing cell cycle phases, while effective, often demand extensive imaging and data processing, presenting significant challenges in terms of efficiency and resource utilization. Recent advances in light microscopy and automated algorithms for cell segmentation have paved the way for more precise and high-throughput analysis of cells [AKZG+18, BBHea17]. In this thesis, an innovative method is introduced that utilizes deep learning to boost the precision and efficiency of identifying cell cycle phases. By employing advanced models commonly used in medical imaging, this research simplifies the imaging process and significantly reduces the data requirements typically associated with multi-channel imaging techniques. This shift not only streamlines the analytical process but also opens new avenues for rapid, scalable, and more accessible cell cycle analysis, positioning deep learning as a transformative tool in cellular biology.

1.1 Biological Background

Accurately identifying **cell cycle** phases is crucial for understanding cellular processes, particularly in cancer research, where cell cycle dysregulation is a hallmark. The **cell cycle** is the series of events that take place in a cell, leading to its division and replication. The main phases of the cell cycle include **G1** (cell growth), **S** (DNA synthesis), **G2** (preparation for mitosis), and **M** (mitosis). Accurate identification of these phases is critical for studies in cell biology and medicine, as it aids in the understanding of cellular processes and disease states. Traditional methods for cell cycle analysis have faced several challenges, primarily due to the complexity and resource-intensive nature of the imaging and data processing required. These challenges include the need for extensive **multi-channel imaging**, which is time-consuming and demands substantial computational resources for image analysis.

In recent years, the development of **fluorescence ubiquitination cell cycle indicators (FUCCI)** has provided a significant advancement in cell cycle analysis. FUCCI uses fluorescent proteins to label different phases of the cell cycle, enabling precise tracking of cellular progression through various stages. Cells express different fluorescent proteins during different phases of the cell cycle, allowing researchers to dynamically track the cell cycle using fluorescence microscopy [FT23]. This method involves imaging through four separate channels, each representing a distinct cell cycle phase, thus providing detailed insights into cell cycle dynamics [SJ21]. However, despite its effectiveness, FUCCI presents challenges such as the need for **multi-channel imaging** and the extra effort required for staining.

The requirement for four-channel imaging significantly increases the complexity of the data acquisition process, making it more challenging to implement in high-throughput settings [AL20]. This complexity often hinders the widespread application of FUCCI, especially in environments with limited resources. Additionally, the multi-channel approach necessitates substantial imaging

and data processing resources, making it less accessible for high-throughput or resource-limited settings [JB17]. This limitation can be a significant barrier in large-scale studies where efficiency and resource management are critical.

To address these limitations, this research introduces an optimized method that leverages **deep learning** to achieve comparable results using **single-channel DNA staining** in time-lapse microscopy. By narrowing down from four channels to just one, the imaging process becomes faster and more cost-effective, alleviating the extensive resource allocation typically required in multi-channel fluorescence imaging. This simplification has profound implications for resource allocation in research settings, making high-throughput analysis more accessible and less invasive to cells. This can facilitate long-term studies and reduce the potential for **phototoxicity** and other cellular stress responses associated with multi-channel imaging, thereby preserving cell viability and integrity.

The proposed method focuses on **breast cancer cells**, a critical research area due to their dynamic nature and the importance of monitoring their evolution. By concentrating on breast cancer cells, we can explore fundamental aspects of cell cycle dynamics and the mutations and behaviors characteristic of cancer progression. This provides a unique opportunity to observe the onset and development of cancerous traits at a cellular level, which is vital for developing targeted therapies. Integrating **deep learning** into this process enhances the practical benefits. **Convolutional neural networks** excel in extracting meaningful patterns from complex datasets, even those limited to single-channel inputs. This capability ensures that, despite the reduced data complexity, the accuracy and reliability of cell phase identification are maintained. The automated nature of deep learning analysis also facilitates faster, more consistent results, enabling researchers to focus on higher-level interpretations and applications of their data [LD21].

A key aspect of this research is the use of knowledge gained from **FUCCI staining** for ground truth annotation. By leveraging the detailed cell cycle phase information provided by FUCCI, we can train the deep learning models to accurately identify cell cycle phases using single-channel DNA staining [MG18]. This hybrid approach combines the precision of FUCCI with the simplicity and efficiency of single-channel imaging, providing a robust solution for cell cycle analysis.

The **cell cycle phases** are further explained as follows: The **G1 phase (Gap 1)** is the first phase where the cell grows and synthesizes proteins necessary for DNA replication. The **S phase (Synthesis)** involves DNA replication, resulting in the duplication of the cell's genetic material. The **G2 phase (Gap 2)** is the second gap phase where the cell continues to grow and prepares for mitosis. Finally, the **M phase (Mitosis)** is where cell division occurs, divided into several stages (prophase, metaphase, anaphase, and telophase) followed by cytokinesis, resulting in two daughter cells. This understanding is visually represented in Figure 1, which illustrates the cell cycle phases using a color-coded diagram.

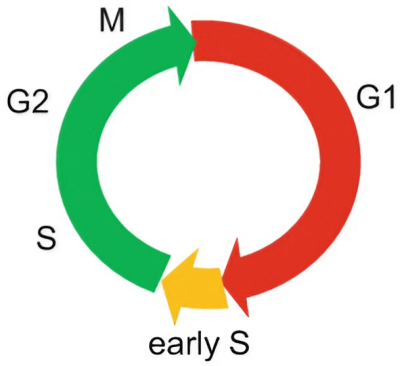

Figure 1: Illustration of the cell cycle phases using a color-coded diagram. The cycle is divided into four main phases: G1 (red), S (green), G2 (green), and M (green), along with a transition point marked as 'early S' (yellow). This visual representation helps in understanding the progression of a cell through its cycle, emphasizing the phases where specific cellular activities occur, such as DNA replication in the S phase and cell division in the M phase.

This research not only streamlines the analytical process but also opens up new avenues for rapid, scalable, and accessible cell cycle analysis. This innovative approach holds the potential to revolutionize how we study cellular functions, offering a more efficient yet equally effective alternative to traditional multi-channel methods. The **single-channel vs. multi-channel imaging** contrast underscores the benefits of the former, which is less data-intensive and simpler in terms of both acquisition and analysis, making it preferable for long-term studies or studies requiring reduced computational resources. In contrast, multi-channel imaging, as used in FUCCI, provides detailed insights by capturing different fluorescent signals that correspond to specific cell cycle phases, but it is more complex and resource-intensive due to the need for sequential imaging and data processing across multiple channels.

1.2 Challenges and Limitations

One of the principal challenges encountered in this research was the creation and preparation of a suitable dataset for training the deep learning model, given the absence of readily available ground truth data. This thesis will detail the customized workflow designed to annotate the dataset automatically, leveraging existing biological knowledge and computational techniques to generate reliable training data.
The challenge of training with a single-channel approach is particularly daunting due to the limited information available compared to multi-channel methods. Single-channel images primarily provide basic structural details, lacking the rich context multi-channel setups offer through additional markers for different cell components or phases. This makes it significantly harder to distinguish subtle features that are crucial for accurate cell cycle analysis, requiring sophisticated computational techniques to extrapolate meaningful insights from less data.
Additional challenges include:

- **Managing the Computational Demands:** Training a sophisticated deep learning model is computationally intensive. The research tackled this by optimizing the neural network

architecture to ensure it remains lightweight and efficient, thus maintaining scalability and accessibility for various research contexts. Techniques such as transfer learning and network pruning were explored to balance computational efficiency without compromising the model's performance.

- **Addressing Model Biases and Variability:** Ensuring the model performs consistently across different imaging setups and experimental conditions is critical. This research implemented robust validation strategies, using a diverse set of images from varied experimental setups to tune the model. This approach helps mitigate potential biases and variability, ensuring the model's reliability and generalizability in real-world applications.
- **Inherent Limitations of Single-Channel Imaging:** The single-channel approach, while beneficial for reducing complexity and computational load, inherently limits the detail and type of cellular information that can be captured. This thesis explores the trade-offs involved, discussing how certain cellular processes and components may be obscured or less distinct without the color-coding typically provided by multi-channel fluorescence tags. Strategies to overcome these limitations include enhancing image processing algorithms and integrating contextual data analysis to infer missing details effectively.

These compounded challenges highlight the intricacy of developing an effective single-channel analysis system for breast cancer cells. However, the methods developed here push the boundaries of what is achievable with minimal data inputs, setting a foundation for future advancements in automated cellular analysis.

1.3 Methodological Innovations

The methodological innovations of this thesis center around the use of a **Unet model** with a pretrained **backbone encoder** for **semantic segmentation**, a novel approach in the context of cell cycle analysis. The Unet model, a type of convolutional neural network developed primarily for biomedical image segmentation [RFB15], features a symmetric encoding and decoding path. This architecture enhances localization precision in the image segmentation process. Utilizing a pretrained backbone encoder within the Unet model improves its capability for feature extraction, making it highly effective for the detailed tasks required in this research. For a detailed view of the Unet architecture, refer to Figure 2.

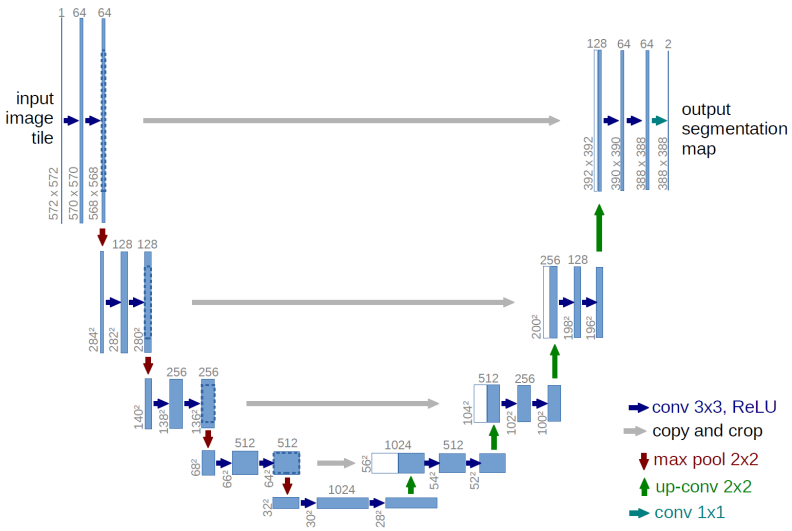

Figure 2: The Unet model architecture. Unet is a convolutional neural network designed for biomedical image segmentation. It features a symmetric encoder-decoder structure, which enhances localization precision and utilizes a pretrained backbone encoder to improve feature extraction capabilities.

At the core of these innovations is **deep learning**, a subset of machine learning that involves artificial neural networks learning from large amounts of data [LBH15]. In this thesis, deep learning is applied to process and analyze medical images, enabling the automatic learning of optimal features for tasks such as image classification and segmentation without the need for manual feature extraction. By employing deep learning, we streamline the analytical process, reducing the necessity for extensive manual annotation and multiple imaging channels, which are typically required in traditional methods.

Semantic segmentation is a crucial technique used in this context. It classifies each pixel in an image into a category or class label, segmenting the image into regions of interest [CPK+14]. This pixel-level classification provides a detailed understanding of the image content, which is essential for identifying cell cycle phases from microscopy images. For an example of semantic segmentation applied to cell microscopy, see Figure 3.

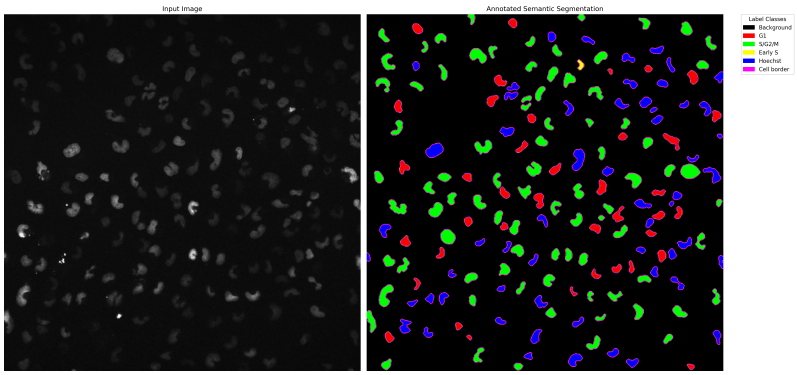

Figure 3: Semantic Segmentation of Microscopy Images: The figure illustrates the application of semantic segmentation to microscopy images of cells, with each color representing a different phase or structure. Red represents the G1 phase, green represents the S/G2/M phases, yellow indicates the Early S phase, and blue is used for Hoechst staining, which highlights the DNA content of the cells and is crucial for identifying all cell cycle phases, including the G0 resting phase. Magenta outlines the cell borders. This detailed pixel-level classification aids in precise analysis of cell cycle phases and morphological features.

Integral to the Unet model is the use of a **backbone encoder**, a pre-trained network that acts as the feature extraction portion of the model [Dom12]. In the Unet architecture, the backbone encoder processes input images into a rich set of features before these features are used by the remainder of the network to perform tasks such as segmentation. This is particularly beneficial in medical image analysis, where high levels of detail and accuracy are required. The pretrained backbone encoder significantly enhances feature extraction capabilities, thereby improving the accuracy and efficiency of the segmentation and classification processes.
Another critical innovation is the approach to **dataset annotation**. Annotating a dataset involves labeling images of cells with their corresponding cell cycle phases, a crucial step for training accurate machine learning models. The absence of ground truth data necessitated the development of a unique workflow to automate this process. This workflow combines predictions from preliminary models with expert validation, ensuring high-quality training data. Automating the dataset annotation process overcomes the challenge of lacking ground truth and ensures the production of robust, reliable machine learning models.
These methodological innovations not only address the challenges of single-channel imaging but also lay the foundation for future advancements in the field. By improving the efficiency and accuracy of cell cycle phase identification, these methods can accelerate research in cellular biology, cancer research, and drug development. The integration of deep learning techniques, semantic segmentation, and advanced feature extraction capabilities represents a significant departure from traditional techniques, offering a more efficient yet equally effective alternative to conventional multi-channel methods.

1.4 Research Objectives

This thesis aims to significantly enhance the accuracy and efficiency of cell cycle phase identification through a novel application of deep learning technology. The specific objectives are:

- To develop a Unet-based deep learning model with a pretrained backbone encoder capable of performing semantic segmentation to accurately identify cells and classify them into the correct cell cycle phases.
- To focus on cell cycle analysis using single-channel DNA staining, thereby streamlining the process and reducing the complexity and data requirements traditionally associated with multi-channel fluorescence microscopy.
- To demonstrate the practical utility of the model in real-world biological research settings, highlighting its potential for broader applications in cellular biology and medical diagnostics.

1.5 Thesis overview

This master thesis, developed under the guidance of Lu Cao from the Leiden Institute of Advanced Computer Science (LIACS) and Sylvia Le Dévédec from the Leiden Academic Center for Drug Research (LACDR), delineates a novel pathway toward streamlining cell cycle analysis. The manuscript is structured as follows: Section 1 provides an introduction to the challenges and solutions presented in this thesis. Section **??** delves into the definitions crucial for understanding this research project; Section 2 reviews the work related to this study, highlighting the gaps this thesis aims to fill. In Section 3, we explain how the data has been obtained and prepared in order to conduct the experiments in Section 4, which describes the experimental setup, and results, underscoring the efficacy of the proposed approach; finally, Section 5 concludes the thesis with a summary of findings and potential avenues for future research. This structure is designed to lead the reader through a comprehensive examination of the innovative strategies for identifying cell cycle phases using advanced deep learning techniques, highlighting a substantial contribution to the field of time-lapse cell cycle analysis in computer vision and biology.

2 Related Work

2.1 Deep Learning models in biomedical image analysis

Deep learning has revolutionized the field of biomedical image analysis, providing tools for more accurate and efficient processing of complex datasets. Various deep learning-based algorithms have been developed for 2D and 3D cell segmentation, offering significant improvements over traditional methods. For instance, Al-Kofahi et al. [AKZG+18] presented a single-channel whole-cell segmentation algorithm that combines deep learning with watershed-based segmentation, achieving high accuracy in diverse microscopy conditions. Buggenthin et al. [BBHea17] utilized deep learning for the prospective identification of hematopoietic stem cells in large-scale single-cell imaging data, demonstrating the potential of these techniques in high-throughput biological applications.

2.2 Understanding the U-Net Architecture

The U-Net architecture, originally developed for biomedical image segmentation, represents a milestone in deep learning applications in the medical field. Its design is fundamentally a convolutional network that excels at tasks requiring precise localization, a common requirement in medical image analysis, especially in microscopic images [RFB15]. The architecture is characterized by its symmetric shape, which includes a contracting path to capture context and a symmetric expanding path that enables precise localization. This structure is particularly well-suited for medical image processing, where the detailed segmentation of intricate structures is crucial.

2.3 Encoder-Decoder Configurations

Moving beyond the basic U-Net architecture, the choice of the encoder, or the backbone, significantly impacts the network's performance, computational efficiency, and parameter efficiency. The encoder's primary role is to downsample the input image and create a set of feature maps that encapsulate the input's essential information, which the decoder then uses to construct the output segmentation map.

- **ResNet** represents a profound shift in the architecture of deep neural networks. Introduced to tackle the vanishing gradient problem, ResNet incorporates "skip connections" or "shortcut connections" that allow gradients to flow through the network without diminishing as they pass through numerous layers. This is accomplished by connecting the input of a set of layers directly to their output, which facilitates the training of much deeper networks by essentially allowing layers to learn residual functions with reference to the layer inputs instead of learning unreferenced functions [HZRS16]. The ResNet architecture is depicted in Figure 4.

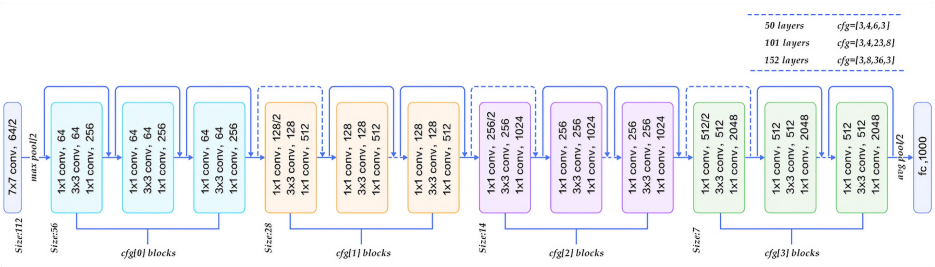

Figure 4: Schematic diagram of the ResNet architecture. Image from [Dev21].

The figure illustrates the sequential arrangement of convolutional blocks at varying depths. The architecture starts with an initial convolutional layer followed by a maximum pooling layer. This is followed by four main stages (cfg[0] to cfg[3] blocks), each containing a variable number of blocks with 3x3 and 1x1 convolutions. The stages are configured to progressively increase the number of filters while reducing the spatial dimensions of feature maps. Residual connections (skip connections) are depicted as dashed lines, enabling direct data flow from one layer to another across the blocks. The diagram shows configurations for different ResNet variants, detailing the number of layers (50, 101, 152) with corresponding block configurations shown at the top right.

- **EfficientNet** stands out due to its systematic scaling of networks, which uses a compound coefficient to uniformly scale network width, depth, and resolution in a principled way. The architecture is based on a baseline model that is scaled up to obtain a family of models that achieve a better trade-off between latency and accuracy [TL19]. The EfficientNet scaling methods are illustrated in Figure 5.

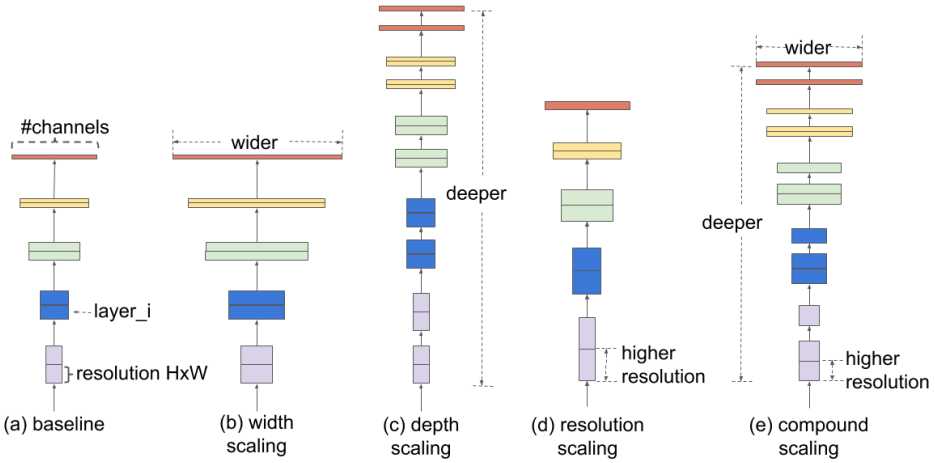

Figure 5: Illustration of scaling methods in EfficientNet. Image from original paper [TL19].

Panel (a) shows the baseline configuration with standard layer width, depth, and resolution. Panel (b) depicts width scaling, where the network's width (number of channels) is increased, enhancing the model's capacity without altering layer depths or input resolution. Panel (c) illustrates depth scaling by increasing the number of layers, potentially improving feature

extraction capabilities. Panel (d) shows resolution scaling where the input resolution is increased, enhancing the model's ability to recognize finer details. Panel (e) represents the compound scaling method, a combination of increasing depth, width, and resolution proportionally, which balances the network's computational efficiency with performance across different dimensions.

- **VGG** (Visual Geometry Group) architecture is noteworthy for its simplicity, primarily utilizing small 3x3 convolutional filters throughout the entire network, which are stacked in increasing depth. Despite its architectural simplicity, VGG has shown a remarkable capacity for feature extraction. Its design involves successive convolutional layers that increase in depth, followed by max-pooling layers, which help in reducing spatial dimensions and focusing on the most relevant features [SZ14]. The VGG network architecture is detailed in Figure 6.

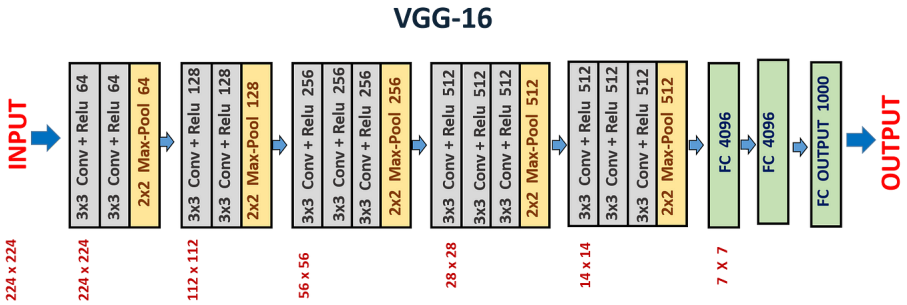

Figure 6: Diagram of the VGG network architecture. Image from [AI22].

The architecture consists of a sequence of convolutional layers, primarily using 3x3 convolutional filters, highlighted by the different colors representing each stage. The network begins with two convolutional layers with 64 filters each, followed by a max pooling layer, continuing this pattern with increasing filter counts. Each block is color-coded: orange for 64 filters, red for 128, purple for 256, green for 512, and blue for fully connected layers. The progression through deeper layers with more filters increases the network's capacity to capture more complex features, culminating in three fully connected layers that lead to the final classification output. The architecture is known for its simplicity and depth, which allows it to capture detailed texture information critical to many visual recognition tasks.

- **Inception** architecture introduces an innovative approach by incorporating heterogeneous convolutional kernels within the same network layer. This design allows the model to capture information at various scales simultaneously, which optimizes computational resource usage and increases the network's ability to handle diverse image content [SLJ+15]. The Inception V4 architecture module is shown in Figure 7.

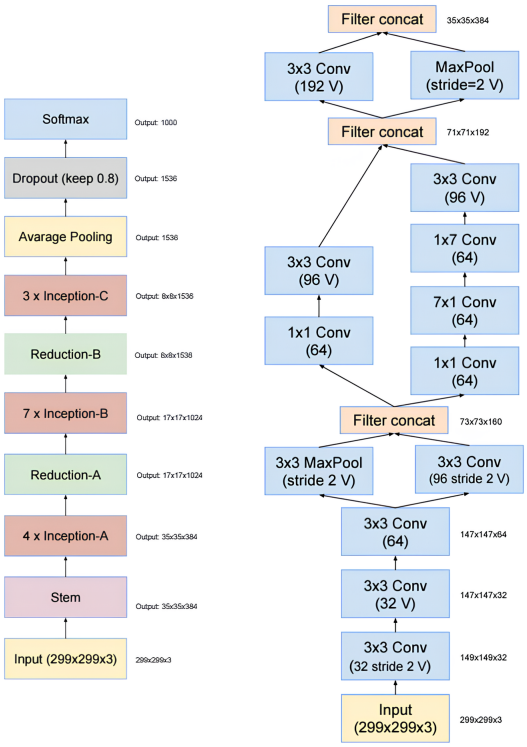

Figure 7: Schematic of an Inception V4 module. The image on the left represents the overall network architecture, while the image on the right details the stem module. Image from original paper [SLJ+15].

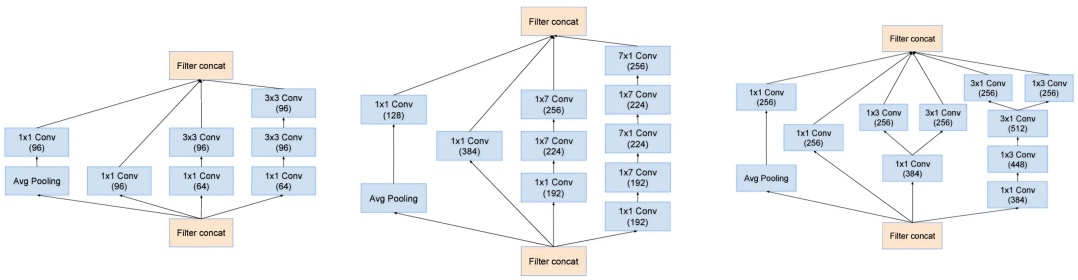

Figure 8: Detailed view of the Inception modules A, B and C in order from left to right. Image from original paper [SLJ+15].

This diagram highlights the innovative use of multiple convolutional filter sizes and concatenation points to capture information at varying scales within a single network layer. The module starts with a typical 299x299x3 input, processed through successive convolutional layers of varying filter sizes (1x1, 3x3) and strides to reduce dimensionality and increase depth,

depicted as different stages in blue. Each path uses convolutional layers to extract features at different granularities, which are then concatenated into a single output feature map, enhancing the network's ability to capture complex patterns without significantly increasing computational cost. Notable components include max pooling and strides for dimensionality reduction and multiple "filter concat" steps where the outputs of various convolutions are combined, showcasing the modular and scalable nature of the Inception architecture. The notation "V" (e.g., 96 V, 32 stride 2 V) denotes the use of valid padding in convolutional layers, meaning no padding is added to the input and the output size is reduced, which helps in maintaining the spatial resolution while increasing the depth of the network.

2.4 Balancing Computational Requirements and Performance in Encoder Selection

The selection of an encoder within the U-Net architecture is not merely a technical choice but a strategic decision that balances computational requirements with performance. This balance is especially critical in environments where resources are constrained. Efficiency is paramount in such settings, as the computational cost of processing and segmenting medical images directly impacts the practicality and scalability of medical imaging solutions. Each encoder brings a unique set of capabilities and trade-offs that need to be carefully considered.

ResNet is known for its deep network capabilities facilitated by residual learning, allowing it to handle more complex models and larger datasets effectively. It is suitable for situations where the depth of feature extraction is crucial and there is a modest allowance for computational resource usage. Its robustness makes it ideal for high-stakes medical imaging tasks where error margins are minimal. **EfficientNet**, on the other hand, is particularly designed for efficiency. It uses a compound scaling method that systematically adjusts network depth, width, and resolution, which can be finely tuned to match hardware constraints without significantly sacrificing performance. Its ability to deliver high accuracy with a lower computational and memory footprint makes it a prime candidate for applications where efficiency is critical.

The **VGG** model, despite having a higher parameter count that implies greater computational demand, benefits from its architecture that captures fine-grained details in images. In settings where image detail is more critical than computational efficiency, such as detailed diagnostic imaging in well-resourced labs, the depth and simplicity of VGG offer significant advantages. Lastly, the **Inception** model, with its mixed convolutional kernel sizes, is adept at handling images of varying scales and complexities. It provides a good balance between detail recognition and computational efficiency, making it well-suited for tasks where images contain diverse features that vary greatly in size, such as pathological analysis or analyzing cellular structures under different magnifications.

In this subsection, the focus shifts from a broad comparison of technical specifications to an emphasis on how these specifications translate into real-world application efficiencies. By prioritizing efficiency alongside performance, this analysis aids in selecting the most appropriate encoder that aligns with the specific constraints and requirements of biological image processing tasks. This strategic approach ensures that the deployment of U-Net-based models is not only theoretically effective but also practically feasible and resource-conscious.

3 Methodology

This section details the acquisition and processing of the dataset used in this study, including the development of ground truth annotations essential for training our model. We will also outline the preparatory steps taken before conducting the experiments discussed in Section 4.

3.1 Dataset Acquisition

The dataset was acquired from an experiment conducted to facilitate imaging for ground truth, alongside the use of compounds to induce cell cycle arrest. The experiment was performed by Xuesong Wang on January 31, 2024.

3.1.1 Reagents and Materials

The materials used for the experiment included:

- Phosphate-buffered saline (PBS).
- Trypsine for detaching cells.
- RPMI 1640 culture medium enhanced with L-glutamine, 25mM Hepes, and supplemented with 10% fetal calf serum (FCS) and penicillin-streptomycin (pen/strep).
- Dimethyl sulfoxide (DMSO).
- Culture-treated, flat-bottom 96-well plates.
- Counting slides for cell quantification.

3.1.2 Experimental Details

The cell lines utilized were Hs578t WT and Hs578t FUCCI, with passage numbers of 5 and 13, respectively, both exhibiting 90% confluence in T25 flasks. The seeding density was set to 6,000 cells per well.

3.1.3 Protocol

Day 1 - Preparation and Cell Seeding
The experiment commenced with the preparation of collagen at a concentration of 20 µg/ml. This solution was added to each well of the 96-well plate and incubated for one hour. Following the incubation, the wells were washed three times with PBS, with the last wash left in the wells until further use.
Cells were then washed with 5 mL of PBS, detached using 0.5 mL of trypsine, and incubated at 37°C under a 5% CO2 atmosphere. Post detachment, cells were suspended in 2 mL of culture medium, centrifuged at 1000 rpm for 5 minutes, and the supernatant was discarded. The pellet was resuspended in 0.5 mL of medium, and cells were counted using a 10 µL sample of the cell suspension.

The cell suspensions were then diluted to achieve a density of 6,000 cells per well. A total of 100 µL of this cell mix was plated in each well and incubated for 24 hours at 37°C in a 5% CO2 environment. The outer wells of the plate were filled with 100 µL of PBS to minimize evaporation.

Day 2 - Staining and Compound Treatment

On the following day, cells were stained using Hoechst and Annexin V. Hoechst was diluted 1:5000, and Annexin V was diluted 1:2000 in the medium. After removing the previous day's medium, 50 µL of Hoechst was added to each well and incubated for one hour at 37°C.

Post-Hoechst staining, various compound dilutions were prepared in a deep-well plate either with medium or Annexin V solution. The Hoechst stain was then removed from each well, and 100 µL of these compounds were added per well. The treatments included DMSO control, Nocodazole (100 ng/ml), Palbociclib (10 µM), Flavopiridol (1 µM), Bosutinib (1 µM), and Cisplatin (1 µg/ml), each designed to induce specific cell cycle arrests or effects such as apoptosis.

3.1.4 Imaging Setup

The imaging was performed using an ImageXpress system with a 20x objective lens. Four different channels were used:

- **Channel w1:** DAPI for Hoechst stain detection (G0 phase).
- **Channel w2:** FITC for visualization of Geminin-GFP (S/G2/M phase).
- **Channel w3:** Texas Red for Cdt1-RFP imaging (G1 phase).
- **Channel w4:** Cy5 for Annexin V (apoptosis).

Images were captured every 20 minutes over a cycle of 70 intervals, providing a comprehensive dataset for subsequent deep learning model training. This detailed temporal-spatial imaging allowed for the precise tracking of cell cycle progression and effects of the treatment compounds, which was crucial for developing robust ground-truth data for the experiment.

3.2 Preparation of Ground Truth Data

The preparation of ground truth data begins with the pre-processing of fluorescent microscopy images, captured using three distinct channels: **blue (DAPI), green (Geminin-GFP), and red (Cdt1-RFP)**. Each image set undergoes a normalization process to enhance contrast and ensure consistent intensity scales, crucial for addressing variability in lighting conditions and staining intensities.

Following normalization, the **Cellpose model**, a machine learning tool specifically designed for **cell segmentation** in microscopy images, is employed to identify individual cells within the images. This automated segmentation process is key to distinguishing individual cellular boundaries and structures. To enhance the accuracy of this segmentation, manual corrections are integrated when needed, allowing for refined segmentation that accurately captures cellular morphologies (see Figure 9).

Post-segmentation, the images often contain residual noise and artifacts, which could potentially interfere with model training. To address this, a series of morphological operations, including erosion, dilation, and the removal of small objects, are applied. These operations are essential for

refining the segmentation masks, smoothing the edges of larger cells, and eliminating irrelevant small artifacts, ensuring that the masks accurately represent the true cell structures.
With clean segmentation masks in hand, the procedure moves to **feature extraction**. This step involves adjusting the segmentation masks to accurately delineate essential cellular features, a critical process for downstream tasks such as **phenotype classification** and quantification of expression levels.
Once individual features have been extracted, the next step involves integrating the segmentation masks from the various channels. This integration is vital for experiments requiring the analysis of **cell cycle stages** or specific cellular conditions indicated by the presence of multiple fluorescent markers.
Finally, the processed images and their corresponding masks are meticulously organized and saved. This structured data organization ensures that each image and mask is correctly tagged and indexed according to the experimental conditions and time points, facilitating easy access and management during the training phase of the machine learning model.
These comprehensive pre-processing steps are designed to prepare the image data meticulously for robust analysis and model training, laying a solid foundation for accurate and reliable experimental results.

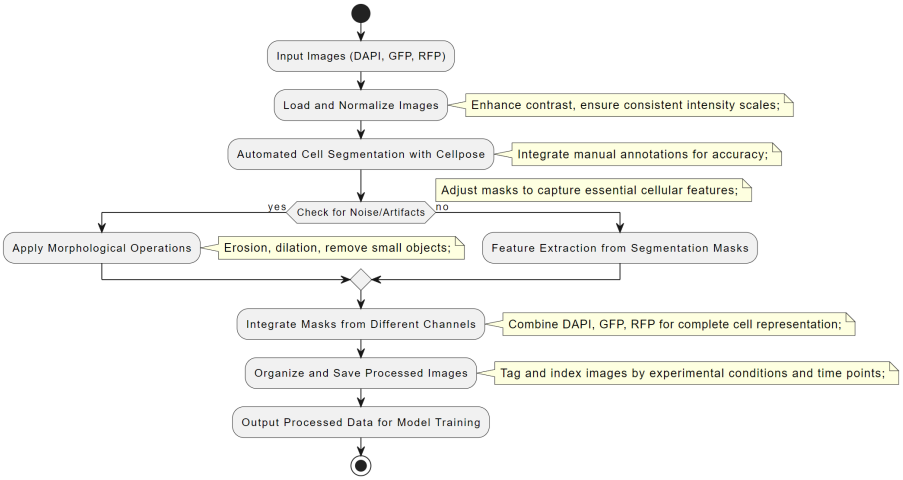

Figure 9: Flow diagram illustrating the pre-processing steps for image data used in the study. The diagram shows the sequential processes from image loading and normalization, through cell segmentation, noise reduction, and feature extraction, to the final organization and storage of processed images.

3.3 Image Augmentation

In the development of robust machine learning models for medical imaging, particularly those involved in image segmentation tasks like cell cycle phase detection, data augmentation plays a crucial role in enhancing model performance by artificially increasing the diversity of training data. This section outlines the specific augmentation techniques employed in our study to preprocess the dataset and prepare it for effective training of our U-Net models with various EfficientNet encoders.

Augmentation Techniques and Implementation The data augmentation pipeline was implemented using the `albumentations` library, a powerful tool for performing transformations in a manner that is both efficient and tailored for computer vision tasks. The choice of augmentations was strategically selected to simulate realistic variations that could occur in medical images, thereby improving the model's ability to generalize across unseen data during the testing phase.
The following transformations were applied to both the input images and their corresponding segmentation masks:

1. **Rotation**:
    - **Limit**: 180 degrees
    - **Probability**: Applied to every image (`p=1`)
    - This transformation rotates the images and their masks up to 180 degrees, covering all possible orientations. Such rotations are crucial for medical images, where orientation can vary significantly.
2. **Elastic Transformation**:
    - **Alpha**: 1
    - **Sigma**: 50
    - **Alpha Affine**: 50
    - **Probability**: 0.5
    - Elastic transformations introduce realistic deformations that mimic the variability in shape and size of biological structures in biological imaging, such as cell deformations under different microscopic settings.
3. **Piecewise Affine**:
    - **Probability**: 0.5
    - This transformation applies local affine transformations to different parts of the image. It is particularly useful for modeling the non-rigid nature of biological tissues and their variability in different samples.

By augmenting the dataset offline and saving the transformed images and masks, the training process becomes significantly more efficient. It eliminates the need for on-the-fly augmentation during each epoch, thus reducing computational overhead and allowing the model to train faster. This tailored augmentation setup, by effectively increasing the size and diversity of our training dataset, is expected to lead to improved model robustness and accuracy, making it better suited for practical deployment in medical image analysis.
With our dataset now meticulously prepared and annotated, we have laid solid groundwork for the advanced stages of our research. The refined data is ready to be utilized for training the model, ensuring that we have a robust system for the experimental validations ahead. The following section, 4, will detail how this trained model is applied and tested, demonstrating the effectiveness of our approaches and discussing the outcomes of our experiments.

4 Experiments

The experiments were conducted using a high-performance computing setup to ensure efficient training and evaluation of the model. The computational resources utilized included an Intel Core i7-12700 processor, 48GB of RAM, and an Nvidia RTX 3080 GPU with 10GB of VRAM. This setup facilitated rapid data processing and model training, enabling the handling of large datasets and complex computations required for accurate semantic segmentation.

4.1 Introduction to Evaluation Metrics for Semantic Segmentation Model

In evaluating the performance of a UNet model for multiclass semantic segmentation, particularly for segmenting and classifying cells from microscopy images, it is essential to use a set of comprehensive and robust metrics. The following metrics have been chosen for their effectiveness in assessing segmentation quality: Intersection over Union (IoU), Dice Score, Precision, Recall, and F1 Score. Each metric provides unique insights into different aspects of the model's performance. Below are detailed explanations and formulas for each metric.

4.1.1 Intersection over Union (IoU)

The Intersection over Union, also known as the Jaccard Index, measures the overlap between the predicted segmentation and the ground truth. It is defined as the ratio of the intersection of the predicted and ground truth masks to their union. The IoU for a single class is given by:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

where A is the set of predicted pixels, and B is the set of ground truth pixels.

4.1.2 Dice Score

The Dice Score, also known as the Sørensen-Dice coefficient, is a measure of similarity between two samples. It is particularly useful for imbalanced datasets as it gives more weight to the correctly predicted pixels. The Dice Score for a single class is calculated as:

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \tag{2}$$

where A and B are defined as above.

4.1.3 Precision

Precision measures the accuracy of the positive predictions, i.e., the proportion of true positive pixels among all pixels predicted as positive. It is calculated as:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

where TP is the number of true positive pixels and FP is the number of false positive pixels.

4.1.4 Recall

Recall, also known as Sensitivity or True Positive Rate, measures the proportion of actual positives that are correctly identified by the model. It is defined as:

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

where FN is the number of false negative pixels.

4.1.5 F1 Score

The F1 Score is the harmonic mean of Precision and Recall, providing a balance between the two. It is particularly useful when there is an uneven class distribution. The F1 Score is calculated as:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{5}$$

4.2 Encoder comparison

In the quest to optimize the performance of the U-Net architecture for segmenting cell cycle phases, we conducted experiments with several pre-trained backbone encoders, as discussed in the section 2.3. Here, we provide a detailed comparison of these models based on various performance metrics.

4.2.1 Experimental Setup:

Each encoder variant of the U-Net was trained under identical conditions to ensure that differences in performance metrics are attributable solely to the encoder's capabilities. The metrics chosen for comparison include Average Intersection over Union (Avg IOU), Dice Coefficient (Avg Dice), Precision (Avg Precision), and Recall (Avg Recall). These metrics provide a comprehensive view of model performance, particularly in the context of semantic segmentation tasks where the accuracy of boundary delineation is crucial.

4.2.2 Results

The comparative performance of the models is summarized in the table below:

Model	Avg IOU	Avg Dice	Avg Precision	Avg Recall
EfficientNet-B5	**0.656**	**0.780**	**0.772**	**0.748**
ResNet50	0.536	0.673	0.673	0.643
VGG19-BN	0.516	0.652	0.661	0.618
Xception	0.585	0.721	0.717	0.683

Table 1: Comparison of UNET Models with Different pre-trained Encoders

4.2.3 Model Performance Evaluation

The results indicate a clear performance advantage for the U-Net model utilizing the **EfficientNet-B5** encoder, which consistently outperforms the other models across all metrics. Specifically, it achieves the highest **Avg IOU** of 0.656 and an **Avg Dice** of 0.780, suggesting its superior capability in capturing the most relevant features for accurate segmentation. The high **precision** (0.772) and **recall** (0.748) further confirm its effectiveness in not only identifying the relevant areas but also in minimizing false positives and false negatives, which are critical aspects in medical image analysis. ResNet50, VGG19-BN, and Xception models also show commendable performance but are less effective compared to EfficientNet-B5. The relatively lower scores in IOU and Dice metrics for these models highlight potential areas for improvement, particularly in terms of feature extraction and boundary delineation.

The empirical evidence supports the conclusion that the choice of encoder significantly impacts the performance of semantic segmentation models. The **EfficientNet-B5** encoder, with its balanced architecture that emphasizes both depth and width, proved to be the most effective in our experiments. It not only enhances the U-Net's ability to accurately segment complex images but also ensures robustness across varied conditions. This finding underscores the importance of selecting appropriate encoder architectures for specific tasks in medical image processing and suggests potential paths for further research and development in this domain.

4.3 Finding best EfficientNet model

Given the constraints of computational resources, this subsection narrows down the scope of our investigation to three specific variants of the EfficientNet family: EfficientNet-B4, B5, and B6. These models were selected based on their computational efficiency and expected performance levels, offering a balanced view of the capabilities within the upper-middle range of the EfficientNet series. This focused comparison will help us determine the best encoder for our U-Net architecture while managing the practical limitations of our available hardware and software infrastructure.

The choice to explore EfficientNet-B4, B5, and B6 is strategic, as these models incrementally increase in complexity and capacity, potentially providing enhanced accuracy and detail in segmentation tasks without excessively straining our computational limits. By examining these specific encoders, we aim to capture a broad spectrum of performance metrics while remaining within the operable bounds of our system's capabilities.

This investigation will follow a structured approach, employing the same metrics previously used—Average Intersection over Union (Avg IOU), Dice Coefficient (Avg Dice), Precision (Avg Precision), and Recall (Avg Recall). This consistency will ensure comparability with prior experiments, allowing for an accurate assessment of each model's strengths and weaknesses in the context of segmenting cell cycle phases. The findings from this comparative analysis will not only determine the most suitable EfficientNet encoder for our current needs but also inform future resource allocation and model selection strategies in biological image analysis projects.

4.3.1 Comparison of Model Performances

This section compares the performance of two models, EfficientNet-B5 and EfficientNet-B6, across various metrics at both the pixel-level and instance-level. The pixel-level metrics evaluate the models' ability to accurately segment images at a granular level, focusing on the precision of individual pixel

classification. In contrast, the instance-level metrics assess the models' performance in accurately identifying and counting distinct cells as whole entities. These metrics provide comprehensive insight into the models' ability to segment and classify different phases of cell images.

Average Pixel-level Metrics

Metric	EfficientNet-B4	EfficientNet-B5	EfficientNet-B6
Segmentation			
IoU	96.76	**96.80**	**96.80**
Dice	98.35	**98.37**	**98.37**
Precision	**98.24**	98.23	98.23
Recall	98.46	**98.52**	**98.52**
Cell Classification			
IoU	61.87	**62.44**	62.42
Dice	76.42	**76.55**	76.42
Precision	73.11	**73.53**	72.88
Recall	72.27	73.35	**73.40**
Borders			
IoU	55.30	**55.87**	55.07
Dice	71.01	**71.48**	70.82
Precision	74.42	75.00	**75.40**
Recall	67.99	**68.36**	66.86

Table 2: Comparison of EfficientNet-B4, B5, and B6 Performance Across Different Classes

The cell classification and border metrics are relatively lower compared to the general segmentation metrics. There are two main factors contributing to these lower values.
First, in some cases, the model is able to predict in some cases the cell shape more accurately than the original ground truth (GT) annotations used to extract these metrics. Since the predicted shapes differ from the GT annotations, there is less overlap, resulting in lower values for the metrics.
Second, the inherent difficulty of the tasks impacts performance. Extracting meaningful features for detailed classification from a single grayscale image is challenging. Additionally, the dataset used is unbalanced, increasing the impact of underrepresented classes on performance.
For the cell border task, which involves identifying a 1-pixel line to separate touching cells, the performance is relatively lower. However, even though it's not perfect or close to perfection, the model's performance is sufficient to separate the cells and identify them as distinct instances. The border will be refined later with post-processing to enhance accuracy.

Average Count-based Metrics

Metric	EfficientNet-B4	EfficientNet-B5	EfficientNet-B6
G1 Phase (Red)			
Precision	**96.11**	94.96	94.42
Recall	96.69	97.48	**97.71**
F1 Score	**96.12**	95.93	95.76
S/G2/M Phases (Green)			
Precision	97.97	98.31	**98.47**
Recall	**97.70**	97.60	97.10
F1 Score	97.75	**97.88**	97.69
Early S Phase (Yellow)			
Precision	57.20	**57.67**	56.37
Recall	47.89	50.03	**50.38**
F1 Score	47.87	48.64	**48.90**
Hoechst (Blue)			
Precision	95.09	**96.17**	94.40
Recall	98.65	98.16	**98.96**
F1 Score	96.72	**97.04**	96.50

Table 3: Comparison of average count-based metrics for cell phase identification using EfficientNet-B4, B5, and B6 models. Metrics are reported as percentages to reflect accuracy in classification across different cell phases, each represented by a unique color.

4.3.2 Comparison of Model Efficiency

In this section, we compare the model efficiency in terms of the number of parameters, training duration, number of epochs needed for convergence, and inference times for the EfficientNet-B4, B5, and B6 models. These models vary significantly in complexity and computational demands, as reflected in their parameter count, training times, and epochs to convergence. The inference times, crucial for real-time applications, have yet to be determined and will be updated accordingly.

Model	Parameters (M)	Training Time (h)	Epochs to Converge	Inference Time (img/s)
EfficientNet-B4	17M	8.75	35	56.89
EfficientNet-B5	28M	12.65	40	51.57
EfficientNet-B6	40M	45.6	30	42.97

Table 4: Comparison of model size, training, epochs to convergence, and inference time for EfficientNet encoders

The stopping criteria for the epochs is determined by monitoring the validation dice score, specifically

stopping the training if there has been no improvement over 5 consecutive epochs. This adaptive training approach ensures that each model is optimized without unnecessary computational expense. This table provides a snapshot of each model's computational footprint and operational efficiency, pivotal for choosing the appropriate model for specific applications in medical image analysis or other fields requiring image processing.

4.3.3 Performance evaluation

Based on the detailed comparison of the three models—EfficientNet-B4, EfficientNet-B5, and EfficientNet-B6—across pixel-level, count-based metrics, and model efficiency, the following conclusions can be drawn:

Pixel-level Metrics

- **Segmentation Performance:**
  - **IoU and Dice:** EfficientNet-B5 and B6 showed identical performance, both slightly outperforming EfficientNet-B4. EfficientNet-B5 and B6 both achieved an IoU of 96.80 and a Dice score of 98.37, indicating their superior capability in segmentation tasks.
  - **Precision and Recall:** EfficientNet-B5 and B6 excelled in Recall, achieving 98.52, marginally higher than EfficientNet-B4's 98.46. EfficientNet-B4 had a slight edge in Precision at 98.24 compared to 98.23 for both B5 and B6.
- **Cell Classification Performance:**
  - **IoU and Dice:** EfficientNet-B5 marginally outperformed the other models with an IoU of 62.44 and a Dice score of 76.55. EfficientNet-B4 lagged slightly behind.
  - **Precision and Recall:** EfficientNet-B5 had the highest Precision at 73.53, while EfficientNet-B6 had the highest Recall at 73.40.
- **Borders Performance:**
  - **IoU and Dice:** EfficientNet-B5 again showed the best performance with an IoU of 55.87 and a Dice score of 71.48. EfficientNet-B6 had the lowest scores here.
  - **Precision and Recall:** EfficientNet-B6 excelled in Precision (75.40) but had lower Recall (66.86) compared to the others. EfficientNet-B5 balanced well with the highest Recall (68.36).

Count-based Metrics

- **G1 Phase (Red):**
  - EfficientNet-B4 had the highest Precision (96.11) and F1 Score (96.12), while EfficientNet-B6 had the best Recall (97.71).
- **S/G2/M Phases (Green):**
  - EfficientNet-B6 led in Precision (98.47) but lagged slightly in Recall (97.10). EfficientNet-B5 balanced well with the highest F1 Score (97.88).

- **Early S Phase (Yellow):**
  - EfficientNet-B5 showed the best overall performance with highest Precision (57.67), Recall (50.03), and F1 Score (48.64).
- **Hoechst (Blue):**
  - EfficientNet-B5 achieved the highest F1 Score (97.04), though EfficientNet-B6 had the best Recall (98.96).

Model Efficiency

- **Parameters and Training Time:**
  - EfficientNet-B4, being the least complex with 17M parameters, trained the fastest in 8.75 hours.
  - EfficientNet-B6, the most complex model with 40M parameters, had a significantly longer training time (45.6 hours).
- **Epochs to Converge:**
  - EfficientNet-B6 converged the fastest at 30 epochs, despite its complexity. EfficientNet-B4 and B5 required 35 and 40 epochs, respectively.
- **Inference Time:**
  - EfficientNet-B4 demonstrated the highest inference speed (56.89 images/second), making it the most suitable for real-time applications.
  - EfficientNet-B5 and B6 processed fewer images per second (51.57 and 42.97 respectively), reflecting their higher computational demands.

Final Model Choice:
Based on the comparison, EfficientNet-B5 is selected as the final model choice. Although EfficientNet-B4 is more efficient in terms of training time and inference speed, EfficientNet-B5 strikes the best balance between performance and efficiency. It consistently showed the best performance across various pixel-level and count-based metrics, indicating its superior capability in segmentation and classification tasks. Moreover, EfficientNet-B5's reasonable training time and inference speed make it a practical choice for applications that require a balance of high accuracy and computational efficiency. Therefore, EfficientNet-B5 is recommended as the optimal model for high-performance tasks in medical image analysis and other image processing applications.

4.4 Evaluation of segmentation model

In this section, we delve into the detailed evaluation of our segmentation model's performance using the EfficientNet-B5 encoder, which has shown promising results in our previous experiments. We present the segmentation results for six different samples, each demonstrating the model's ability to accurately predict cell boundaries. The evaluation metrics used include precision, recall, and F1 score, providing a comprehensive assessment of the model's effectiveness in segmenting cell images.

The segmentation results are depicted through images, showing the input image, the model's prediction with borders, and the ground truth. These visualizations highlight the differences between the model's predictions and the actual cell boundaries, offering insights into the areas where the model excels and where improvements are needed. The corresponding metrics for each sample are tabulated to quantify the model's performance and facilitate a detailed comparison.

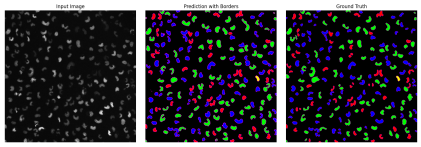

Figure 10: Sample 1

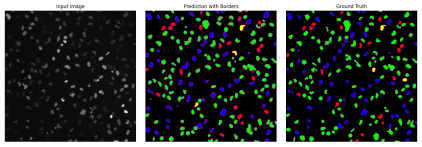

Figure 11: Sample 2

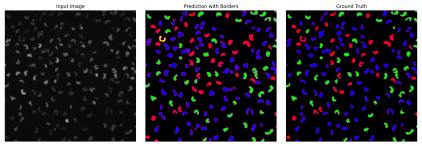

Figure 12: Sample 3

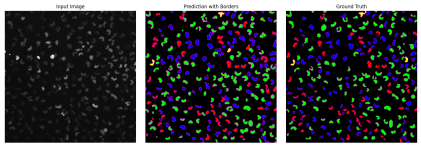

Figure 13: Sample 4

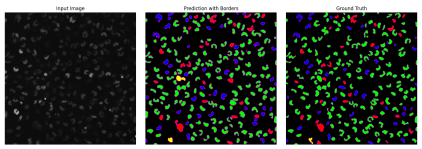

Figure 14: Sample 5

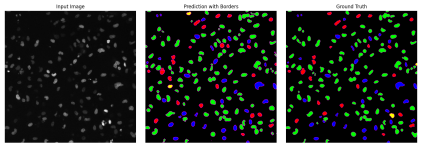

Figure 15: Sample 6

Figure 16: Segmentation results for various samples.

Sample	Precision	Recall	F1 Score
1	98.14	91.09	93.33
2	99.41	98.18	98.75
3	83.07	81.95	82.50
4	97.98	96.72	97.23
5	84.70	97.00	87.25
6	99.02	97.39	98.14

Table 5: Metrics for segmentation results of various samples.

The presented results demonstrate that the EfficientNet-B5 based segmentation model performs robustly across various samples, achieving **93.72% average precision**, **92.87% average recall**, and **93.29% average F1 score**. Notably, in samples 2 and 6, the model achieved exceptionally high precision and F1 scores, indicating its capability to accurately delineate cell boundaries with minimal false positives.

Interestingly, in some cases, such as sample 5, the model's predictions appear to be more accurate than the ground truth, capturing finer details of the cell shapes that the ground truth might have missed. This phenomenon can lead to slightly lower recall metrics, but it does not necessarily imply that the model is underperforming. Instead, it underscores the model's potential to provide more precise and reliable cell segmentation, which is crucial for downstream analysis in medical imaging.

Overall, the results confirm the efficacy of the EfficientNet-B5 encoder in enhancing the U-Net architecture for cell segmentation tasks. The high performance across different metrics and samples validates the model's robustness and suggests that it is well-suited for practical applications in medical image analysis, where accurate segmentation is paramount. Future work could focus on

further refining the model to address any remaining discrepancies and exploring its application to other segmentation tasks within the medical domain.

4.5 Case study

In this subsection, we present a detailed analysis of the experiments conducted to compare the effects of five different treatments on cell lines. The focus of this study is to observe the changes in the number of cells per cell cycle phase, allowing us to understand how different treatments influence cell behavior, specifically pushing cells into different phases of the cell cycle.

4.5.1 Experimental Setup

For this study, we used the cell line Hs578t_WT, and subjected it to five different treatments: nocodazole, cisplatin, bosutinib, flavopiridol, and palbociclib. The images were collected and processed to count the number of cells in different phases of the cell cycle, namely G1 Phase, S/G2/M Phase, Early S, and Hoechst. The results were visualized using stacked bar charts to represent the cell count over time for each treatment and cell line combination.

4.5.2 Results and Analysis

The following subsections provide an overview of the findings for each treatment.

- **Nocodazole** is known to disrupt microtubule polymerization, effectively arresting cells in the M phase. The graph shows a significant reduce in the number of cells in the Hoechst/G0 phase and arresting cells into the S/G2/M phase, indicating that nocodazole effectively halts cell division, pushing a large number of cells into mitotic arrest.

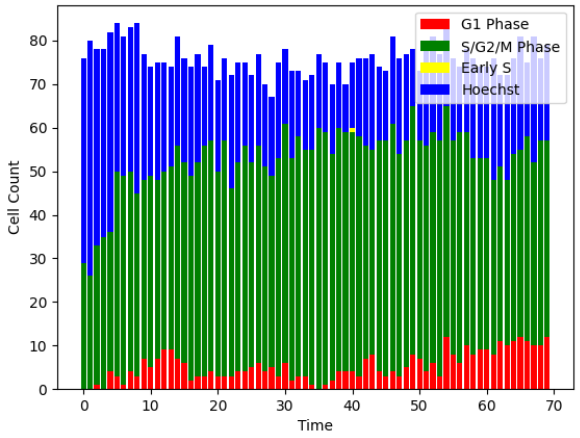

Figure 17: Nocodazole treatment on WT cells.

- **Cisplatin** is a chemotherapy drug that causes DNA crosslinking, leading to apoptosis. The graph indicates an increase in the number of cells in the G1 and S/G2/M phases, suggesting cisplatin provoques cell activation phase swtiching over time.

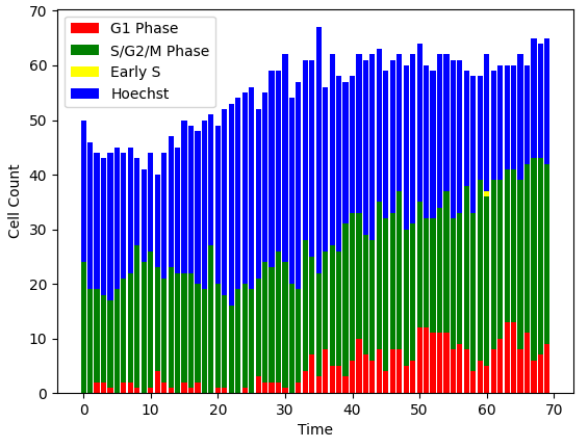

Figure 18: Cisplatin treatment on WT cells.

- **Bosutinib** is a tyrosine kinase inhibitor that targets Bcr-Abl. The graph shows a substantial number of cells accumulating in the Hoechst/G0 phase, indicating that bosutinib may cause a blockade in the cell activation preventing them to evolve properly into the cell cycle phases.

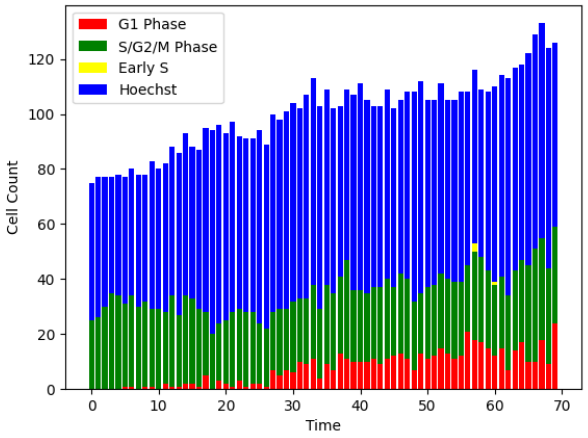

Figure 19: Bosutinib treatment on WT cells.

- **Flavopiridol** is a broad-spectrum cyclin-dependent kinase (CDK) inhibitor. The graph shows an increase in cells in the and G1 phase and reduce S/G2/M phase, suggesting that flavopiridol effectively inhibits CDK activity, preventing progression through the cell cycle and leading to cell cycle arrest at multiple points.
- **Palbociclib** specifically inhibits CDK4 and CDK6, preventing the transition from the G1 to the S phase. The graph indicates a significant accumulation of cells in the G1 phase, confirming that palbociclib effectively blocks cell cycle progression at this checkpoint.

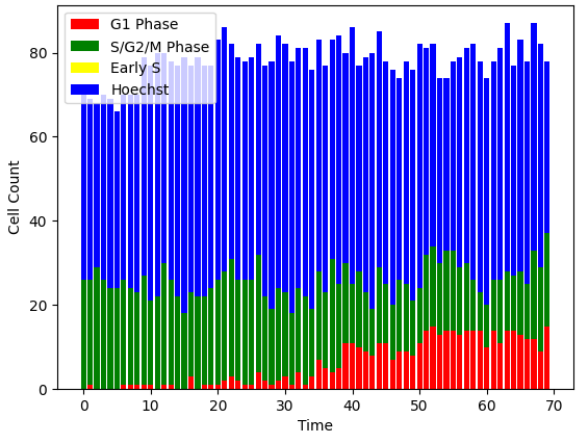

Figure 20: Flavopiridol treatment on WT cells.

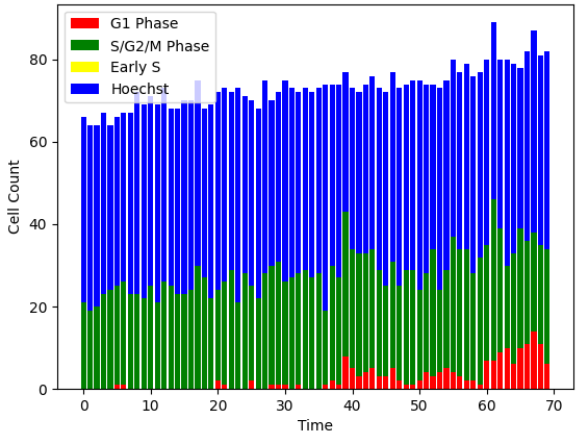

Figure 21: Palbociclib treatment on WT cells.

From these experiments, we can draw several key conclusions regarding the effects of different treatments on cell cycle progression:

- **Nocodazole** and **cisplatin** primarily induce cell cycle arrest in the M and G1 phases, respectively.
- **Bosutinib** causes a blockade in the Hoechst/G0 phase, suggesting its effect on the G1 transition.
- **Flavopiridol** and **palbociclib** both act as CDK inhibitors, with palbociclib showing broader inhibition across multiple phases, while Flavopiridol specifically targets the G1/S transition.

These findings highlight the diverse mechanisms through which different treatments can affect cell cycle progression, providing valuable insights into their potential therapeutic applications and effects on cell behavior.

5 Conclusions and Further Research

This thesis has introduced and thoroughly explored an innovative approach to cell cycle phase identification using deep learning techniques, with a focus on single-channel DNA staining in time-lapse microscopy. The research aimed to streamline the imaging process, reduce data requirements, and maintain high accuracy in segmenting and classifying cell cycle phases, particularly in breast cancer cells.

5.1 Summary of Findings

The use of **single-channel DNA staining** significantly reduces the complexity and cost associated with traditional multi-channel fluorescence imaging. This simplification has profound implications for resource allocation in research settings, making high-throughput analysis more accessible and less invasive to cells. By facilitating long-term studies and reducing the potential for **photo toxicity** and other cellular stress responses associated with multi-channel imaging, this method helps preserve cell viability and integrity.

Building on this simplified imaging approach, the application of **deep learning models**, particularly the UNet model with various pre-trained **backbone encoders** (EfficientNet-B4, B5, and B6), demonstrated that deep learning can effectively handle the reduced data complexity while maintaining high segmentation accuracy. Among these models, EfficientNet-B5 provided the best balance between performance and computational efficiency, consistently outperforming others in various metrics. This finding highlights the potential of deep learning to revolutionize cell cycle studies by automating and enhancing the accuracy of phase identification.

Experimental validation further confirmed the model's effectiveness in classifying cells into their correct cell cycle phases. The EfficientNet-B5 based UNet model demonstrated high **precision**, **recall**, and **F1 scores**, underscoring its robustness in practical applications. Detailed segmentation results and metrics validated the model's accuracy in delineating cell boundaries, which is crucial for understanding cellular behavior, disease progression, and therapeutic effects. However, some limitations were observed in cell classification and border metrics. These lower values are due to two main factors: the model sometimes predicts cell shapes more accurately than the original ground truth (GT) annotations, leading to less overlap and consequently lower metric values, and the inherent difficulty of extracting detailed features from single grayscale images, combined with an unbalanced dataset, impacts overall performance. Despite these challenges, the model effectively separates touching cells, with post-processing steps available to refine border accuracy further.

Innovative methodological contributions played a significant role in achieving these results. The integration of **automated dataset annotation**, advanced **augmentation techniques**, and strategic selection of **encoder architectures** not only addressed the challenges of single-channel imaging but also laid the groundwork for future advancements. By improving the efficiency and accuracy of cell cycle phase identification, these methods have the potential to accelerate research in cellular biology, cancer research, and drug development. This holistic approach, combining simplified imaging with powerful computational techniques, marks a significant step forward in the field.

5.2 Considerations and Limitations

While this thesis demonstrates that good results can be achieved with the proposed approach, there is room for improvement. The current methodology uses static images without accounting for motion, which limits the model's ability to capture dynamic features of cell behavior over time. The nature of cell phase classification heavily depends on tracking cell movements and understanding temporal dynamics, which are crucial for accurate classification. Biologists need to observe how cells move over time in time-lapse sequences to make precise classifications. Accurate tracking of cell movements can reveal critical insights into cell division, migration, and interactions, which are essential for understanding developmental biology, cancer metastasis, and tissue regeneration.

5.3 Future Research Directions

Future work should focus on creating a **dataset** with equal representation of all classes to enhance cell classification results. Incorporating smaller datasets where cells are exposed to different treatments will increase the model's ability to generalize to new scenarios. These improvements will likely lead to more accurate and reliable cell classification and better model performance in diverse experimental conditions.

Another important direction involves integrating **temporal features** by employing a mix of transformer-CNN architectures. Transformers can capture long-range dependencies and temporal patterns, significantly enhancing the performance and reliability of the model. This approach will allow the model to consider cell movements and changes over time, providing a more holistic view of cell cycle phases. Understanding temporal dynamics is crucial for identifying transient phases and events in the cell cycle, such as mitotic entry and exit, which are critical for accurate phase classification and understanding cellular responses to treatments.

While this study focused on **breast cancer cells**, the methodology can be extended to other cell types and conditions. Future research could explore the application of this approach to various cellular contexts, enhancing its utility in broader biological and medical research. Applying these techniques to different cell lines and tissues can help uncover cell type-specific behaviors and responses, aiding in personalized medicine and targeted therapies.

Additionally, combining the single-channel deep learning approach with other imaging modalities could provide a more comprehensive understanding of cellular processes. Future work could integrate data from different imaging techniques to enhance the model's accuracy and applicability. **Multi-modal imaging** can provide complementary information about cellular structure, function, and molecular composition, offering a more detailed and nuanced understanding of cell biology and disease mechanisms.

Given the high inference speed and accuracy of the EfficientNet-B5 based model, further research could focus on developing **real-time analysis** tools. These tools could be instrumental in dynamic monitoring and analysis of cell cycles in live-cell imaging experiments. Real-time analysis can enable immediate feedback and adjustments during experiments, improving experimental efficiency and facilitating the study of rapid cellular responses to stimuli.

Continued efforts to optimize the **computational efficiency** of deep learning models, particularly in resource-constrained environments, will be crucial. Exploring lightweight architectures and efficient training techniques can further reduce computational demands while maintaining high performance. Enhancing computational efficiency is essential for deploying these models in routine

laboratory settings and clinical diagnostics, where computational resources may be limited.
Finally, the potential **clinical applications** of this approach are significant. Future research could investigate the integration of this methodology into diagnostic workflows, potentially aiding in the early detection and monitoring of cancer and other diseases at the cellular level. Accurate and efficient cell cycle phase identification can improve the precision of diagnostic tests and enable the timely identification of pathological changes, supporting early intervention and personalized treatment strategies.

5.4 Final Remarks

In conclusion, this thesis has successfully demonstrated the feasibility and effectiveness of using deep learning for cell cycle phase identification with single-channel DNA staining in time-lapse microscopy. The approach offers a streamlined, efficient, and accurate alternative to traditional methods, with broad implications for research and clinical practice. By advancing the integration of deep learning in biological imaging, this work contributes to the ongoing efforts to harness AI for transformative insights in cellular biology and medicine. Future advancements, particularly in incorporating temporal dynamics through transformer-CNN architectures, hold the promise of further enhancing the robustness and applicability of this innovative approach. This research lays the groundwork for more comprehensive and dynamic studies of cell behavior, ultimately contributing to a deeper understanding of cellular processes and the development of more effective treatments for various diseases.

References

[AI22] Towards AI. The architecture and implementation of vgg-16, 2022. Accessed: 2024-06-27.

[AKZG+18] Yousef Al-Kofahi, Alla Zaltsman, Robert Graves, Will Marshall, and Mirabela Rusu. A deep learning-based algorithm for 2-d cell segmentation in microscopy images. *BMC Bioinformatics*, 19(365):1–11, 2018.

[AL20] J. Anderson and C. Lee. Interphase cell cycle staging using deep learning. *PubMed*, 20:5678, 2020.

[BBHea17] Florian Buggenthin, Florian Buettner, Paula S Hoppe, and et al. Prospective identification of hematopoietic lineage choice by deep learning. *Nature Methods*, 14:403–406, 2017.

[CPK+14] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.

[Dev21] DevGenius. Resnet50, 2021. Accessed: 2024-06-27.

[Dom12] Pedro Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, 2012.

[FT23] L. Fischer and I. Thievessen. Fucci reporter gene-based cell cycle analysis. *Methods in molecular biology (Clifton, N.J.)*, 2644:371–385, 2023.

[HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[JB17] K. Johnson and M. Brown. Improving cell cycle stage prediction with deep learning techniques. *PMC*, 22:112–119, 2017.

[LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[LD21] C. Lee and E. Davis. Advances in fluorescence microscopy for cell cycle analysis. *Nature*, 45:789–795, 2021.

[MG18] D. Miller and P. Green. Recent advances in fluorescence microscopy for cell cycle analysis. *Nature*, 50:200–210, 2018.

[RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[SJ21] A. Smith and B. Jones. A machine learning approach for single cell interphase cell cycle staging. *Scientific Reports*, 11:23456, 2021.

[SLJ+15] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[SZ14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[TL19] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.