Universiteit Leiden
The Netherlands

Bachelor Computer Science & Bioinformatics

Improving Early S Phase Detection in Hoechst-Based Cell-Cycle Segmentation Using Temporal Context

Laurent Elfrink

Supervisors:
Lu Cao & Maarten Paul

BACHELOR THESIS

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

01/07/2026

Abstract

Early S phase detection is a challenging part of Hoechst-based cell-cycle segmentation because early S is rare, visually subtle, and represents a transitional state between G1 and later S/G2/M. This thesis investigates whether temporal context from time-lapse microscopy can improve early S detection compared with single-frame prediction.

Starting from an existing U-Net baseline with an EfficientNet-B5 encoder, several temporal approaches were tested. First, predicted cell instances were extracted, tracked across consecutive frames, and used to construct temporal label windows. A rule-based method and a logistic regression classifier were then evaluated using predicted class labels from $t-1$, t, and $t+1$. Second, image-based temporal U-Net models were trained using three consecutive Hoechst frames as input to predict the segmentation mask of the center frame.

The temporal label-based classifier performed almost identically to the center-frame baseline on the matched cell-instance subset. A temporal U-Net trained from scratch performed poorly, especially for early S. Initializing the temporal U-Net from the trained single-frame baseline improved performance compared with training from scratch, but did not outperform the inherited single-frame baseline on the same temporal test center frames. These results suggest that temporal context is not automatically sufficient to improve early S detection. More careful temporal model design, training strategy, and evaluation using consistent data splits are needed to determine whether temporal information can improve Hoechst-based cell-cycle segmentation.

Contents

1 Background and theory **3**
- 1.1 Cell Cycle . . . 3
- 1.2 Fluorescence-based cell-cycle imaging . . . 3
- 1.3 Hoechst-based prediction without FUCCI input . . . 4
- 1.4 Deep learning methods for biomedical segmentation . . . 5
- 1.5 Related work in cell and nuclei segmentation . . . 5
- 1.6 Challenges in early S phase detection . . . 6
- 1.7 Ordered nature of the cell cycle . . . 6
- 1.8 Thesis motivation . . . 7

2 Methods **7**
- 2.1 Dataset . . . 7
- 2.2 Overview of experimental pipeline . . . 8
- 2.3 Baseline model . . . 9
- 2.4 Instance extraction . . . 10
- 2.5 Cell tracking . . . 10
- 2.6 Temporal window construction . . . 11
- 2.7 Temporal label-based approach . . . 11
  - 2.7.1 Rule-based relabeling . . . 11
  - 2.7.2 Logistic regression . . . 12
- 2.8 Temporal image-based approach . . . 13
  - 2.8.1 Temporal U-Net trained from scratch . . . 13
  - 2.8.2 Baseline-initialized temporal U-Net . . . 13
- 2.9 Evaluation . . . 14
- 2.10 Implementation and hardware . . . 14

3 Results **15**
- 3.1 Temporal dataset construction . . . 15
- 3.2 Temporal label-based approaches . . . 15
- 3.3 Temporal U-Net trained from scratch . . . 16
- 3.4 Baseline-initialized temporal U-Net . . . 16
- 3.5 Comparison with the single-frame baseline . . . 16
- 3.6 Example segmentation results . . . 17
- 3.7 Summary of findings . . . 18

4 Discussion **19**
- 4.1 Temporal label-based approaches . . . 19
- 4.2 Temporal U-Net models . . . 19
- 4.3 Comparison with the single-frame baseline . . . 20
- 4.4 Limitations . . . 20
- 4.5 Future work . . . 20

5 Conclusion **21**

1 Background and theory

1.1 Cell Cycle

The cell cycle is an ordered sequence of events during which a cell grows, duplicates its genetic material, and divides into two daughter cells. The cell cycle generally consists of four main phases: G1, S, G2 and M. During G1 phase, the cell grows and prepares for DNA replication. In S phase, the cell synthesizes a complete copy of its DNA. This is followed by G2 phase, in which the cell continues to grow and prepares for mitosis. Finally, during M phase, the duplicated chromosomes are separated and the cell divides [Sch98].

This process is regulated by cyclins, cyclin-dependent kinases (CDKs), and checkpoint mechanisms [Bir03]. Cyclins are proteins whose concentrations rise and fall during specific stages of the cell cycle, while CDKs are enzymes that become activated when bound to cyclins and drive progression into the next phase. Specific cyclin-CDK complexes are needed for the cell to transition from one phase to another. For example, when cyclin D expression increases, CDK4/6 form a complex with cyclin D, which stimulates the cell to progress through G1 phase and it supports re-entry of the cell cycle. Cyclin E forms a complex with CDK2 to promote entry into S phase. Cyclin B forms a complex with CDK1 which is needed for entry into M phase [VVBB03].

There are three key checkpoints to ensure important events are completed correctly. The G1/S checkpoint monitors whether the cell is sufficiently prepared for DNA replication and whether any DNA damage is present. The G2/M checkpoint ensures that DNA replication has been completed correctly before the cell enters mitosis. During mitosis, the spindle checkpoint verifies that chromosomes are properly attached to the mitotic spindle before they are separated. Together, these regulatory mechanisms help preserve genomic stability and prevent damaged or incompletely replicated DNA from being passed on to daughter cells [Nig95].

Although the cell cycle is commonly described as a series of discrete phases [HCG+22, EKB+17, NSC+20, DC24, Cui25], progression through these phases is in reality gradual and continuous. Transitions between phases may therefore be difficult to define sharply, particularly around the entry into S phase [DC24, Cui25]. This biological continuity is important to consider when interpreting microscopy-based cell-cycle labels.

1.2 Fluorescence-based cell-cycle imaging

The identification of cell-cycle phases is crucial for studying cell proliferation, cancer progression, and drug response [Sch98, ZE15]. A widely used approach for identifying cell-cycle phases in living cells is fluorescence-based cell-cycle imaging. These methods monitor cell-cycle progression by linking molecular events to fluorescent signals.

One of the most established methods for fluorescence-based cell-cycle imaging is the Fluorescent Ubiquitination-based Cell Cycle Indicator (FUCCI) system, introduced by Sakaue-Sawano et al. [SSKM+08]. FUCCI uses two fluorescent markers linked to two specific proteins, Cdt1 and Geminin, present during different phases of the cell cycle. Cdt1 is mainly present during G1 phase, while Geminin is mainly present during S and G2 phases. As a result, cells show different fluorescence signals depending on their current cell-cycle state. This allows researchers to distinguish different phases based on distinct fluorescent patterns and their transitions over time [SSKM+08, ZE15].

An important strength of FUCCI is that researchers can directly observe how individual cells

progress through the cell cycle in space and time. This has made FUCCI a versatile and widely adopted tool across different model systems, including mammalian cells, mice, fish, flies, and plants [ZE15]. The method has contributed to studies of developmental patterning, stem cell behavior, quiescence, and the relationship between cell-cycle progression and cell fate decisions [ZE15].
During the transition from G1 to S phase, FUCCI signals are not always sharply separated. In the original FUCCI system, cells in G1 phase are labeled red, whereas cells in S, G2, and M phases are labeled green. However, during a short period around the G1/S transition, both markers can be present at the same time, causing cells to appear yellow [ZE15, SSKM+08]. This is particularly relevant for early S phase, which represents the start of DNA replication and therefore forms a transitional state between G1 and later S phase. As a result, early S phase does not always appear as a clearly distinct category, but rather as an intermediate state that shares characteristics with neighboring phases. This biological overlap is important for this study, because it suggests that the boundary between G1 and early S phase may be inherently gradual rather than sharply defined.

For the present study, FUCCI is particularly important because it provides the reference labels used for cell-cycle phase annotation. However, the aim of this work is to predict cell-cycle phase from the Hoechst channel alone, rather than from dedicated cell-cycle reporter channels. This is useful because FUCCI requires genetically encoded fluorescent reporters and additional fluorescence channels, which may not always be available in other experiments. In contrast, Hoechst is a commonly used nuclear DNA stain and provides a simpler single-channel input. Reducing the number of required imaging channels may also reduce experimental complexity, image acquisition time, data storage requirements, and the amount of fluorescence exposure during live-cell imaging. This is relevant because phototoxicity is an important consideration in live fluorescence microscopy, where illumination can affect sample physiology [IWWN17]. If cell-cycle phase can be predicted from Hoechst alone, FUCCI can be used for annotation during model development while not being required as input during prediction. FUCCI therefore serves as ground truth during model development, while the prediction task itself focuses on Hoechst-based cell-cycle detection.

1.3 Hoechst-based prediction without FUCCI input

An alternative to dedicated cell-cycle reporters is to detect cell-cycle phases from simpler imaging signals, such as DNA staining with Hoechst. Hoechst is a cell-permeable fluorescent dye that binds to DNA at A/T-rich regions in the minor groove, and is widely used as a nuclear stain in fluorescence microscopy [BLG18]. Because Hoechst labels nuclear DNA, its fluorescence intensity is related to DNA content, which changes across the cell cycle [BLG18]. This makes Hoechst relevant for cell-cycle analysis, even though it is not a phase-specific reporter such as FUCCI. For the present study, this is valuable because it enables the prediction of cell-cycle phase from a single Hoechst channel, reducing reliance on dedicated cell-cycle labeling while still retaining useful biological information.
Unlike FUCCI, Hoechst does not directly label specific cell-cycle phases. Instead, phase-related information must be detected indirectly from differences in DNA content and nuclear signal. This makes the task more challenging, especially for phases that are close together in the cell cycle. At the same time, it also makes Hoechst-based prediction attractive, because it allows cell-cycle analysis to be performed from a single staining channel. Deep learning methods are therefore a suitable approach for detecting subtle differences in Hoechst images, because they can learn complex

visual patterns directly from image data.

1.4 Deep learning methods for biomedical segmentation

Deep learning has become widely used in biomedical image analysis because it can learn relevant image patterns directly from data. This is especially useful in microscopy, where differences between structures can be subtle and difficult to capture with hand-crafted features or simple threshold-based methods. In segmentation tasks, deep learning models can learn both fine local details, such as cell boundaries, and broader contextual information from the surrounding image. This makes them well suited for pixel-level classification in biomedical images.
One of the most commonly used architectures for biomedical segmentation is U-Net [RFB15]. U-Net was created for biomedical image segmentation and has become a standard model in the field. It consists of an encoder path, which extracts increasingly abstract features from the input image, and a decoder path, which reconstructs a segmentation map at the original resolution. An important feature of U-Net is that it uses skip connections between the encoder and decoder to help preserve spatial detail, which is important when segmenting small biological objects such as nuclei. By combining detailed low-level image information with high-level contextual features, skip connections help the model localize structures more precisely.

Many later studies have extended the U-Net architecture by using stronger encoder backbones, including EfficientNet [SPRA+22]. These pretrained encoders can improve feature extraction and lead to better segmentation performance. U-Net-based models have therefore been used successfully in many biomedical tasks, including nucleus segmentation, tissue segmentation, and cell classification. In the context of cell-cycle analysis, deep learning is particularly relevant because the visual differences between phases may be subtle, especially when only a single imaging channel is available. Rather than relying on manually defined thresholds or intensity rules, a deep learning model can learn phase-related patterns directly from annotated training data. Previous work has shown that U-Net-based models with EfficientNet encoders can perform cell-cycle phase segmentation from single-channel Hoechst images, demonstrating that meaningful phase information can be extracted even without dedicated cell-cycle reporters [DC24, Cui25]. For the present study, this makes deep learning-based segmentation a suitable foundation for investigating whether difficult transitional states, especially early S phase, can be detected more accurately.

1.5 Related work in cell and nuclei segmentation

Besides U-Net-based models, several deep learning methods have been developed specifically for cell and nuclei segmentation. One important example is Cellpose, which was designed as a generalist cell segmentation algorithm that can work across different microscopy image types and cell morphologies [SWMP21]. Instead of relying only on a model trained for one specific dataset, Cellpose aims to generalize to many types of cellular images. This makes it useful in biological imaging settings where cell shape, size, and imaging conditions can vary strongly between experiments. In the context of this thesis, Cellpose is relevant because it shows how deep learning can be used to segment cells robustly across diverse microscopy datasets.
Another widely used method is StarDist, which is designed for instance segmentation of cells and nuclei [SWBM18]. StarDist represents objects using star-convex polygons, which allows the method

to separate individual nuclei even when they are close together or touching. This is particularly relevant for microscopy images, where neighboring nuclei can overlap or be difficult to distinguish using simple thresholding or semantic segmentation alone. Compared with standard semantic segmentation, instance segmentation methods such as StarDist are especially useful when the goal is to identify separate cell objects rather than only classify each pixel.
Although Cellpose and StarDist are important methods for cell and nuclei segmentation, they were not used in this thesis. The present study builds on an existing U-Net-based semantic segmentation baseline from previous work.

1.6 Challenges in early S phase detection

Although deep learning has shown strong potential for Hoechst-based cell-cycle segmentation, the detection of early S phase remains particularly challenging. Previous work has reported that early S is one of the most difficult classes to identify accurately in single-channel Hoechst images [DC24, Cui25]. This is partly because early S is a short transitional phase between G1 and later S/G2/M, rather than a clearly separated and stable state. As a result, cells in early S may share characteristics with both neighboring phases, which makes the class boundary less distinct.
Furthermore, early S is typically underrepresented in the data. Because the phase is relatively short compared with other parts of the cell cycle, fewer cells are observed in early S at any given time. This leads to class imbalance, meaning that the model is trained on substantially fewer early S examples than on classes such as G1 or S/G2/M. In supervised learning, such imbalance can make the model more biased toward the majority classes and reduce its ability to learn the more subtle characteristics of rare classes [Cui25]. In practice, this can result in early S cells being misclassified as one of the neighboring phases.
Another important issue is that evaluation results may depend strongly on how the data are split. Cui [Cui25] showed that random splits can lead to overly optimistic results because images from the same experimental conditions or time-lapse sequences may appear in both training and test sets. In contrast, replicate-aware or sequence-aware splits provide a more realistic assessment of model generalization. Under these more stringent conditions, the detection of early S remains difficult, which further suggests that the problem is not only one of model capacity, but also of class rarity, subtle visual differences, and biological continuity between phases.

Taken together, these factors make early S phase detection a particularly challenging problem in Hoechst-based cell-cycle segmentation. The class is rare, visually subtle, and biologically intermediate, which increases the likelihood of confusion with neighboring phases. These challenges motivate the search for modeling strategies that may better capture the transitional nature of the cell cycle and improve the detection of early S phase.

1.7 Ordered nature of the cell cycle

The cell cycle is not simply a set of unrelated categories, but an ordered biological process. Cells progress through the phases in a fixed sequence, moving from G1 to S phase, then to G2 and M phase, before returning to G1 after division [Sch98]. This means that neighboring phases are biologically closer to each other than phases that are further apart. Early S phase is a clear example

of this, because it lies between G1 and later S/G2/M and therefore represents an intermediate state.
This ordered structure is relevant for this study because it shows that cell-cycle phase should not only be viewed as a static label, but also as part of a progression over time. Around transition points, the appearance of a cell may change gradually rather than abruptly. As a result, a single image may not always contain enough information to distinguish a transitional phase clearly from its neighboring phases. This is especially true for early S, which may share characteristics with both G1 and later S/G2/M.
For this reason, the ordered nature of the cell cycle provides an important motivation for considering temporal context in cell-cycle prediction. If the same cell can be followed over multiple frames, its current state may be interpreted in relation to the states before and after it. In this way, the sequential nature of the cell cycle suggests that temporal information may be particularly useful for identifying transitional phases such as early S.

1.8 Thesis motivation

Previous work has shown that Hoechst-based cell-cycle segmentation is possible using deep learning, but early S phase remains a difficult class to detect reliably. In single-frame prediction, this transition may be difficult to recognize because the visual differences in the Hoechst channel can be subtle and gradual.
Most existing approaches treat each image frame independently. However, the cell cycle is an ordered process that unfolds over time. A cell that is in early S phase at one time point is expected to be preceded by a G1-like state and followed by a later S/G2/M-like state. This suggests that temporal information from neighboring frames may provide useful context for identifying transitional phases. For example, a cell that appears ambiguous in a single frame may become easier to interpret when its previous and following states are also considered.
Therefore, this study investigates whether temporal context can improve early S phase detection in Hoechst-based cell-cycle prediction. Starting from an existing single-frame U-Net baseline, individual cells are extracted, tracked across consecutive frames, and organized into short temporal windows. This makes it possible to compare the original single-frame prediction with temporal approaches that use information from the same cell over time.

2 Methods

2.1 Dataset

The dataset consists of time-lapse fluorescence microscopy images of cells progressing through the cell cycle. For each timepoint, multiple fluorescence channels were available, including a Hoechst channel and FUCCI reporter channels. Only the Hoechst channel was used as model input, while the FUCCI-derived annotations were used as ground-truth labels for the cell-cycle phases.
The data were organized into sequences corresponding to different locations and imaging sites. A sequence refers to one time-lapse image series from a fixed imaging location and site. Each sequence contains 70 consecutive timepoints of the same field of view. A sequence therefore does not refer to the trajectory of a single cell, but to the full image series in which multiple cells are visible over

time. This temporal information is important for the present study because cells can be followed across consecutive frames and their phase can be interpreted using surrounding timepoints.
The ground-truth masks contained labels for the main cell-cycle-related classes, including G1, S/G2/M, early S, and an additional Hoechst-related annotation class, as well as background and border labels. Early S was strongly underrepresented compared with the other phase classes, reflecting its role as a short transitional state and making it a particularly challenging class for prediction.

2.2 Overview of experimental pipeline

Figure 1 gives an overview of the experimental pipeline used in this thesis. The single-frame U-Net baseline was inherited from previous work and used as the starting point. Two temporal extensions were then explored. First, a label-based temporal approach was tested as an exploratory experiment by extracting predicted cell instances, tracking them across frames, and applying temporal rule-based relabeling and logistic regression. Second, image-based temporal U-Net models were trained using three consecutive Hoechst frames as input. This included a temporal U-Net trained from scratch and a temporal U-Net initialized from the trained single-frame baseline. The final comparison focused on pixel-level performance of the single-frame baseline, the temporal U-Net from scratch, and the baseline-initialized temporal U-Net.

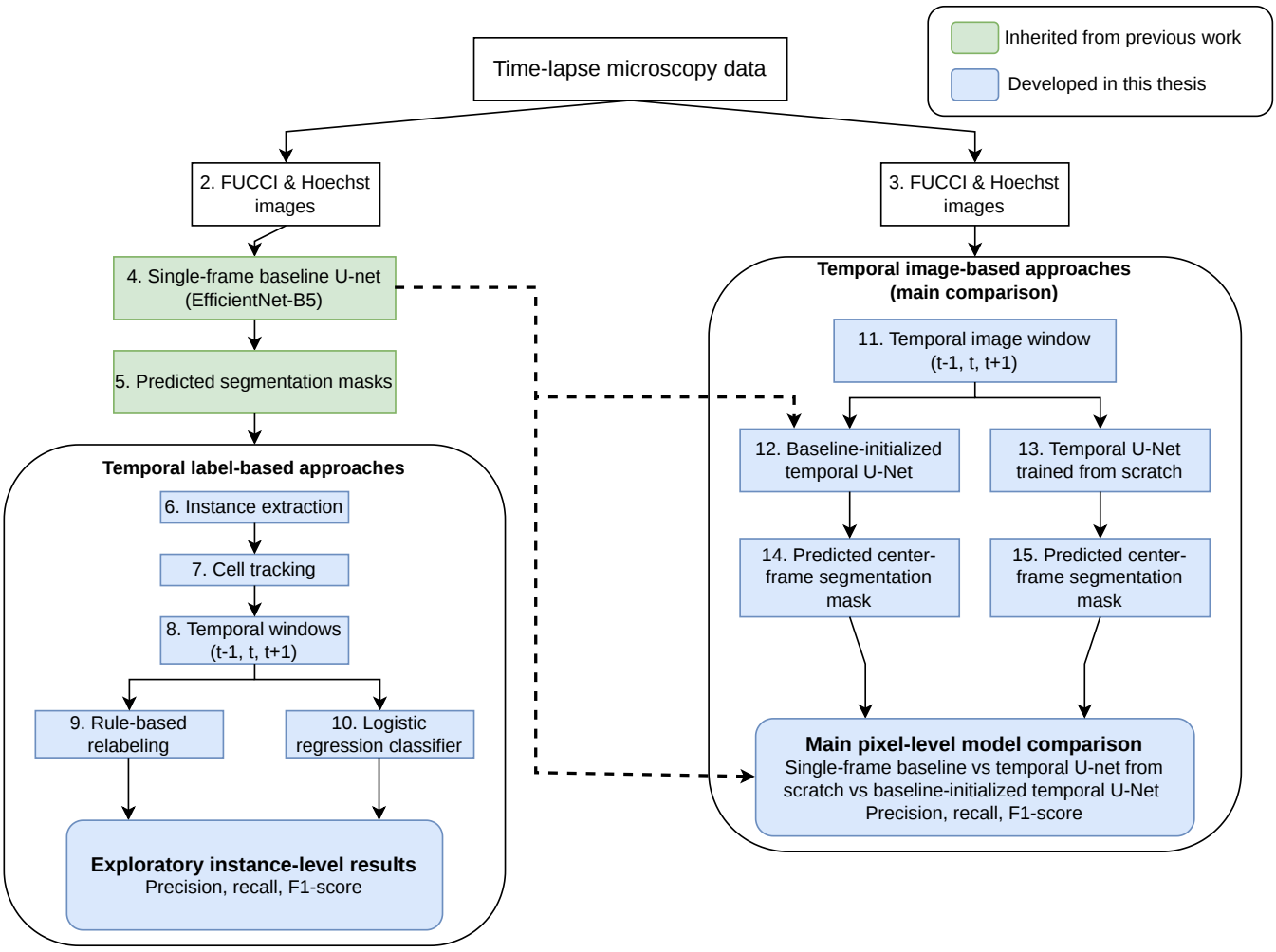

Figure 1: Overview of the experimental pipeline. Green components indicate parts inherited from previous work, while blue components indicate methods and analyses developed in this thesis. The diagram was recreated in diagrams.net based on an earlier diagram draft generated with the assistance of a generative AI tool.

2.3 Baseline model

The baseline model used in this study is based on the Hoechst-based cell-cycle segmentation pipeline developed in previous work [DC24]. The model performs single-frame semantic segmentation, meaning that each input image is processed independently and each pixel is assigned to one of the predefined classes. The input to the model is a single-channel Hoechst image, while the output is a multiclass segmentation mask containing the predicted cell-cycle-related labels.

The baseline architecture consists of a U-Net model with an EfficientNet-B5 encoder. The model was trained to predict six classes: background, G1, S/G2/M, early S, Hoechst, and border. The border class is included to help separate neighboring or touching nuclei.

The baseline model serves as the single-frame reference method. Its predictions are used to evaluate

performance without temporal context and to provide baseline class labels for the temporal label-based approaches. In addition, the trained baseline model is used to initialize the temporal U-Net model. This allows the temporal model to start from a model that has already learned Hoechst-based segmentation features, while extending the input from a single frame to multiple neighboring frames.

2.4 Instance extraction

After baseline inference, the predicted segmentation masks were converted into individual cell instances. This step was necessary because the baseline model produces semantic segmentation masks, while the temporal part of this study requires individual cells to be followed across frames. Therefore, each predicted mask was processed to extract separate connected components.
For each frame, the background and border classes were excluded from instance extraction. The remaining phase-related classes were processed separately, so that connected regions belonging to G1, S/G2/M, early S, and Hoechst were identified as candidate cell instances. Very small connected components were removed using a minimum-area threshold of 20 pixels.
The same instance extraction idea was applied to both predicted and ground-truth masks, but the masks were stored in different formats. The predicted masks were saved as color-coded images, so class regions were identified by their RGB color. The ground-truth masks were stored as class-index images, where each pixel value corresponded to a class label. Therefore, ground-truth instances were extracted directly from the numeric pixel values. Both procedures produced an instance table containing the frame name, class label, centroid, area, and bounding box for each detected object.
To allow evaluation at the cell-instance level, predicted instances were matched to ground-truth instances within the same frame. Matching was based on centroid distance, with each ground-truth instance assigned to at most one predicted instance. Matches were only accepted when the distance between the predicted and ground-truth centroids was at most 20 pixels. The resulting matched instances were used to compare predicted class labels with the corresponding ground-truth labels and to attach ground-truth labels to temporal samples in later steps.

2.5 Cell tracking

After extracting individual cell instances from the predicted masks, cells were linked across consecutive frames within each time-lapse sequence. Here, a sequence means the 70-frame image series belonging to one fixed imaging location and site. Tracking was performed separately for each sequence so that cells were only matched between frames from the same field of view. This prevents cells from being incorrectly linked across different imaging locations. The resulting tracks represent approximate single-cell trajectories within a sequence.
For each pair of consecutive frames, candidate matches were computed based on the Euclidean distance between cell centroids. A cost matrix was constructed in which each row represented a cell instance in the previous frame and each column represented a cell instance in the current frame. The value in the matrix corresponded to the centroid distance between the two instances. The Hungarian algorithm was then used to find an assignment that minimized the total matching cost between the two frames.
A match was accepted only if the centroid distance was at most 30 pixels. If a cell in the current frame could not be matched to a cell in the previous frame within this threshold, it was assigned a

new track identity. In this way, each tracked cell was given a track ID that could be followed across frames for as long as valid matches were found.
The resulting tracks were used to construct temporal samples in later steps. Because this tracking procedure is based only on centroid distance, it should be interpreted as a simple tracking approach rather than a full cell-lineage reconstruction method. It is sufficient for creating short temporal windows, but it may lose tracks when cells move too far, overlap, disappear, or are segmented differently between frames.

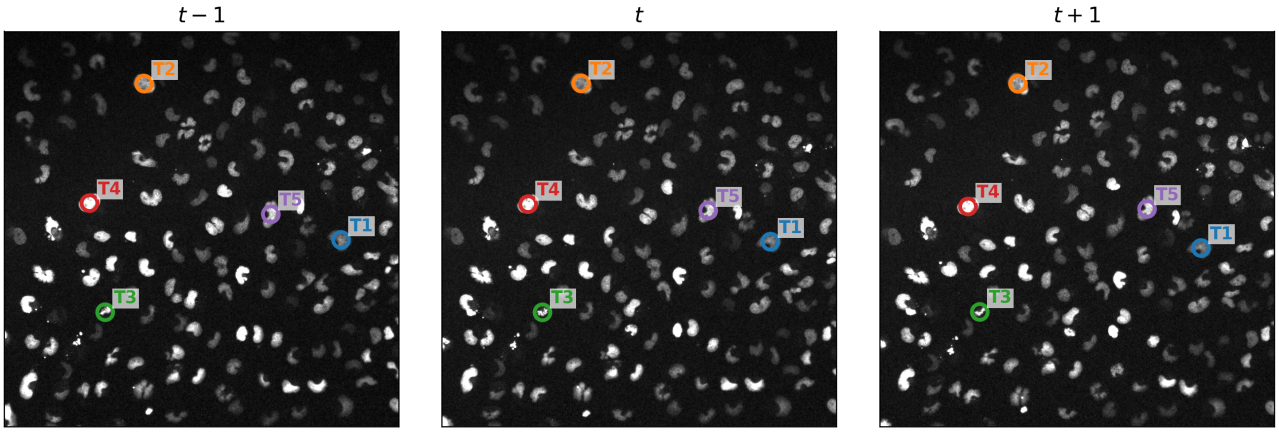

Figure 2: Example of centroid-based cell tracking across three consecutive frames. The same tracked cells are marked with consistent track IDs (T1-T5) in frames $t-1$, t, and $t+1$.

2.6 Temporal window construction

After cell tracking, the resulting tracks were used to construct short temporal samples. Each temporal sample was centered around one cell instance at a specific timepoint. For this study, a three-frame window was used, consisting of the previous frame, the current frame, and the next frame of the same tracked cell. This can be written as $t-1$, t, and $t+1$, where t is the center frame. Only strictly consecutive windows were included. This means that a sample was only created when the same track contained detected instances at $t-1$, t, and $t+1$ without missing timepoints. The class label predicted by the baseline model was stored for each of the three frames.
The temporal windows were constructed so that the phase prediction of the center frame could use information from the predicted phases of the same tracked cell in the neighboring frames. For example, a cell predicted as G1 before the center frame and as S/G2/M after the center frame may indicate that the center frame lies near the G1/S transition. This setup was used to test whether neighboring timepoints provide useful context for identifying early S phase.

2.7 Temporal label-based approach

2.7.1 Rule-based relabeling

Within the label-based temporal approach, two variants were tested. The first variant was a simple rule-based relabeling method, while the second variant used a learned logistic regression classifier. Both variants used the predicted class labels of the same tracked cell at $t-1$, t, and $t+1$.

The rule-based method was designed as a simple first test of whether transition patterns could be used to identify early S phase. If the predicted labels of the neighboring frames suggested a transition from G1 toward S/G2/M, the center frame was relabeled as early S. For example, if the previous frame was predicted as G1 and the following frame was predicted as S/G2/M, the center frame was considered a possible early S candidate. This method was included as an interpretable baseline for using temporal information.

2.7.2 Logistic regression

In addition to the rule-based approach, a logistic regression classifier was trained on the temporal samples. The input features were the predicted class labels at $t-1$, t, and $t+1$. Because the input features were categorical class labels, they were converted into binary indicator variables using one-hot encoding before training the logistic regression classifier. The target label was the ground-truth class of the center frame. In this way, the classifier learned from the training data which combinations of neighboring predicted labels were informative for predicting the true center-frame class.

Figure 3 illustrates how a temporal sample was represented for the logistic regression classifier. The baseline-predicted class labels of the same tracked cell at timepoints $t-1$, t, and $t+1$ were used as input. These categorical labels were converted into a one-hot encoded feature vector, which was then used to predict the ground-truth class of the center frame t.

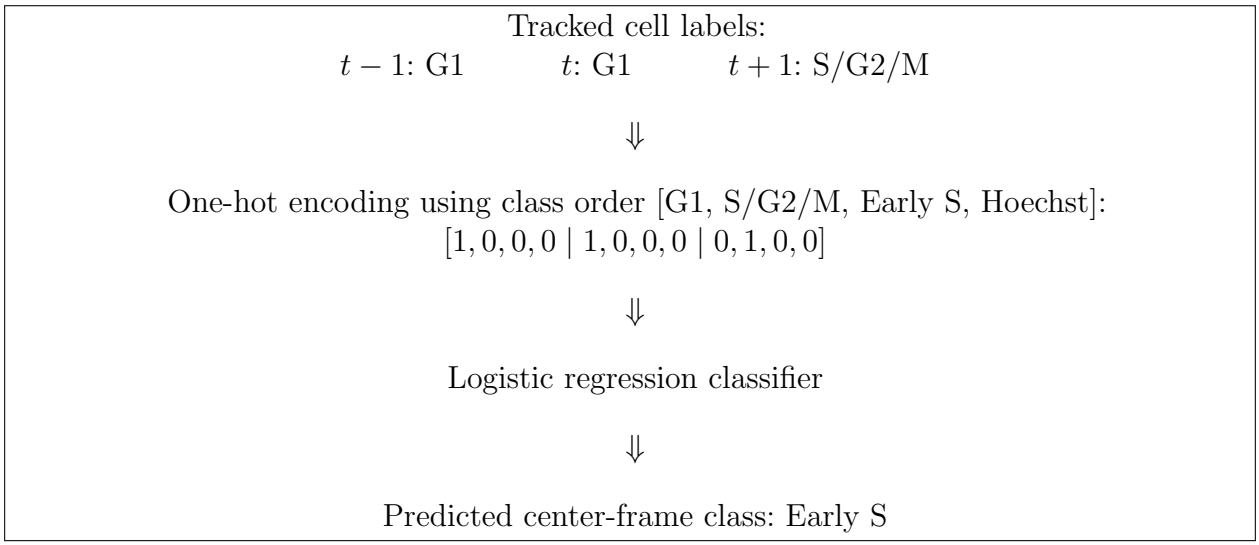

Figure 3: Example of the feature representation used for the temporal logistic regression classifier.

The temporal classifier was trained on the training split and evaluated on the validation and test splits. Its performance was compared with the original center-frame baseline prediction on the same test samples. This allowed the temporal approach to be evaluated under the same conditions as the baseline prediction, while using additional information from neighboring timepoints.

2.8 Temporal image-based approach

The label-based temporal methods described in the previous section incorporated temporal information after the baseline segmentation by using the predicted class labels from neighboring frames. In contrast, the following approach integrates temporal information directly into the segmentation model itself. Instead of using predicted labels as input, consecutive Hoechst images $(t-1, t, t+1)$ are provided to a temporal U-Net, allowing the network to learn temporal image features during segmentation.

Two variants of this image-based temporal approach were tested: one temporal U-Net trained from scratch and one temporal U-Net initialized from the inherited single-frame baseline. Both variants used the same three-frame input and center-frame target; only the training initialization differed.

2.8.1 Temporal U-Net trained from scratch

A temporal U-Net was trained from scratch using the same three-frame input setup. This model used the Hoechst frames at $t-1$, t, and $t+1$ as input and predicted the ground-truth mask at t, but did not use the trained single-frame baseline weights. This model was first tested with standard cross-entropy loss and was then also tested with weighted cross-entropy loss. This experiment was used as a comparison to determine whether baseline initialization was necessary for the temporal image-based model.

The model was trained using the Adam optimizer with a learning rate of 1×10^{-4} and a batch size of 1. The model predicted six segmentation classes. The weighted cross-entropy loss used class weights $[0.05, 1.0, 1.0, 10.0, 1.0, 0.5]$ for background, G1, S/G2/M, early S, Hoechst, and border, respectively. Training was performed for 10 epochs, and the checkpoint with the lowest validation loss was selected for evaluation.

2.8.2 Baseline-initialized temporal U-Net

The baseline-initialized temporal U-Net used the same three-frame temporal input windows as the temporal U-Net trained from scratch. Thus, the input consisted of the Hoechst frames at $t-1$, t, and $t+1$, stacked as three input channels, and the target output remained the ground-truth segmentation mask of the center frame t.

The temporal U-Net architecture was based on the same U-Net with EfficientNet-B5 encoder as the baseline model. To make the comparison with the baseline more direct, the temporal model was initialized from the trained single-frame baseline checkpoint. Since the baseline model used one input channel and the temporal model used three input channels, the first convolutional layer had to be adapted. The original one-channel weights were copied across the three temporal input channels and divided by three, so that the initial activation scale remained similar. All other compatible weights were loaded directly from the baseline checkpoint.

The model was fine-tuned using the Adam optimizer with a learning rate of 1×10^{-5} and a batch size of 1. A weighted cross-entropy loss was used with class weights $[0.05, 1.0, 1.0, 5.0, 1.0, 0.5]$ for background, G1, S/G2/M, early S, Hoechst, and border, respectively. As with the temporal U-Net trained from scratch, training was performed for 10 epochs, and the checkpoint with the lowest validation loss was selected for evaluation.

2.9 Evaluation

The temporal label-based approaches and the temporal U-Net models were evaluated using different procedures, because they produced different types of output. The label-based approaches produced class predictions for tracked cell instances, while the temporal U-Net models produced full segmentation masks.
For both evaluation settings, precision, recall, and F1-score were computed for each class, with particular attention to the early S class. For a given class, these metrics were defined as:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Here, TP denotes true positives, FP denotes false positives, and FN denotes false negatives for the class being evaluated. Since early S is underrepresented, the F1-score was used as the main summary metric for early S performance, while precision and recall were also reported to show whether the model mainly missed early S examples or overpredicted them.

For the temporal label-based approaches, evaluation was performed at the cell-instance level using the matched predicted and ground-truth instances described in Section 2.4. For each temporal sample, the predicted class of the center-frame instance was compared with the ground-truth class of the matched center-frame object. A cell instance was considered correctly classified when the predicted class label matched the ground-truth class label.
For the temporal U-Net experiments, evaluation was performed at the pixel level on the test split. The predicted segmentation mask for the center frame was compared with the ground-truth mask of the same frame. The comparison with the single-frame baseline was performed on the same test center frames used for the temporal U-Net evaluation. The baseline received only the Hoechst image at timepoint t as input, while the temporal U-Net received the three-frame window $(t-1, t, t+1)$. Both models were evaluated against the same ground-truth masks at t.
However, the baseline checkpoint was inherited from previous work and was trained using a different data split than the temporal models in this study. Therefore, some images or sequences in the current temporal test set may have been included during the original training of the baseline model. For this reason, the baseline comparison should not be interpreted as a fully independent held-out test result.

2.10 Implementation and hardware

The experiments were implemented in Python using PyTorch. The original baseline model was implemented with the `segmentation_models_pytorch` library, using a U-Net architecture with an EfficientNet-B5 encoder. Temporal image models were trained and evaluated on the ALICE high-performance computing cluster at Leiden University. GPU jobs were submitted using Slurm. The temporal U-Net experiments were run using NVIDIA GPUs available on ALICE, including NVIDIA L4, RTX 2080 Ti, and A100 GPUs depending on scheduling availability. Training was performed with a batch size of 1 because of the large (1024×1024) input images. Model checkpoints were selected based on validation loss and evaluated on the held-out test split.

3 Results

3.1 Temporal dataset construction

Temporal samples were constructed in two different ways, depending on the type of temporal model. For the cell-tracking-based approaches, individual predicted cell instances were linked across frames and used to create short temporal windows centered around one cell instance. This resulted in 550,860 temporal cell-level samples, of which 475,227 had a valid ground-truth match for the center frame. These matched samples were used for the temporal label-based experiments.
For the image-based temporal U-Net experiments, temporal windows were constructed at the full-image level. Each sample consisted of the Hoechst images at $t-1$, t, and $t+1$, with the ground-truth segmentation mask at t as the target. This resulted in 2,176 training samples, 272 validation samples, and 816 test samples. The temporal train, validation, and test splits were constructed at the sequence level, meaning that all temporal windows derived from the same 70-frame time-lapse sequence were assigned to the same split. Unlike the cell-level temporal samples, these full-image samples did not require explicit cell tracking.

Table 1: Number of temporal image samples used for the temporal U-Net experiments.

Split	**Number of samples**
Training	2,176
Validation	272
Test	816

3.2 Temporal label-based approaches

The learned temporal classifier performed very similarly to the inherited single-frame baseline on the same temporal subset. On the test set, the original center-frame baseline reached an early S precision of 0.9383, recall of 0.9060, and F1-score of 0.9219. The temporal classifier reached an early S precision of 0.9382, recall of 0.9045, and F1-score of 0.9210. Therefore, using only the hard predicted labels from neighboring frames did not improve early S detection in this subset.
This result suggests that the temporal label-based representation may have been too limited. Since the classifier only received the final predicted class labels, it could not use more detailed visual information from the Hoechst images or uncertainty in the baseline model predictions. In addition, the tracked temporal subset appeared to be relatively clean, which may have made it difficult to improve over the center-frame prediction.

Table 2: Cell-instance-level Early S performance of the temporal label-based classifier compared with the center-frame baseline on the same matched temporal subset.

Method	**Precision**	**Recall**	**F1-score**
Center-frame baseline	0.9383	0.9060	0.9219
Temporal logistic regression classifier	0.9382	0.9045	0.9210

3.3 Temporal U-Net trained from scratch

The initial temporal U-Net trained from scratch performed poorly, especially for the early S class. With standard cross-entropy loss, the model did not detect early S pixels. Therefore, the final from-scratch experiment used the weighted cross-entropy setting described in Section 2.8.1.
Although the weighted loss caused the model to predict some early S pixels, performance remained very low. On the test set, the weighted temporal U-Net trained from scratch reached an early S precision of 0.0191, recall of 0.0085, and F1-score of 0.0118. The other foreground classes also remained relatively weak, indicating that training the temporal model from scratch was not sufficient for this dataset.

3.4 Baseline-initialized temporal U-Net

The baseline-initialized temporal U-Net performed substantially better than the temporal U-Net trained from scratch. On the test set, the model reached an early S precision of 0.2379, recall of 0.6908, and F1-score of 0.3539. This shows that the model was able to detect a large proportion of early S pixels, although precision remained relatively low. In other words, the model became more sensitive to early S, but also produced many false positive early S predictions.

Table 3: Pixel-level test performance of the baseline-initialized temporal U-Net.

Class	**Precision**	**Recall**	**F1-score**	**Support pixels**
Background	0.9966	0.9411	0.9680	707,848,130
G1	0.6264	0.7656	0.6890	20,450,265
S/G2/M	0.7134	0.8880	0.7912	60,609,447
Early S	0.2379	0.6908	0.3539	1,383,757
Hoechst	0.7352	0.8307	0.7801	29,234,753
Border	0.4875	0.6711	0.5647	36,111,664

These results suggest that initialization from the trained single-frame baseline was important for the temporal U-Net. The model trained from scratch almost completely failed to detect early S, while the baseline-initialized temporal model recovered meaningful early S predictions. However, this does not yet show that temporal input improves over the original single-frame baseline. For that reason, the baseline model was also evaluated on the same test frames.

3.5 Comparison with the single-frame baseline

To determine whether temporal image input improved over the original single-frame model, the trained baseline model was evaluated on the same 816 test center frames as the temporal U-Net. Both models were compared against the same ground-truth masks at timepoint t. The difference between the models was the input: the single-frame baseline used only the center Hoechst frame t, while the temporal U-Net used the three-frame input window $(t-1, t, t+1)$.
On this same-test-frame pixel-level comparison, the single-frame baseline performed better than both temporal U-Net models. For early S, the single-frame baseline reached a precision of 0.8354, recall of 0.8067, and F1-score of 0.8208. In comparison, the baseline-initialized temporal U-Net reached a precision of 0.2379, recall of 0.6908, and F1-score of 0.3539. Although the temporal

model detected a substantial proportion of early S pixels, its much lower precision shows that it overpredicted the early S class. Therefore, in this experiment, adding neighboring frames as input did not improve early S detection over the original single-frame baseline.

Table 4: Pixel-level Early S performance on the same temporal U-Net test center frames.

Method	**Precision**	**Recall**	**F1-score**
Single-frame baseline	0.8354	0.8067	0.8208
Temporal U-Net trained from scratch	0.0191	0.0085	0.0118
Baseline-initialized temporal U-Net	0.2379	0.6908	0.3539

Because the baseline checkpoint was trained using a different data split, this comparison should be interpreted with caution and mainly serves to compare the inherited single-frame model and the temporal U-Net under the same pixel-level evaluation procedure.

3.6 Example segmentation results

In addition to the quantitative evaluation, qualitative examples were inspected visually. Figure 4 shows representative segmentation results for the single-frame baseline and the baseline-initialized temporal U-Net. The first row shows an example where the temporal U-Net gives a relatively good prediction. The second row illustrates a case where the temporal U-Net overpredicts early S, visible as additional yellow regions compared with the ground truth. The third row shows an example where the single-frame baseline is closer to the ground truth than the temporal U-Net.
These examples are consistent with the quantitative results. The temporal U-Net is able to predict early S regions, but it also tends to assign early S to too many pixels. This explains why the temporal U-Net achieved relatively high early S recall but low precision compared with the single-frame baseline.

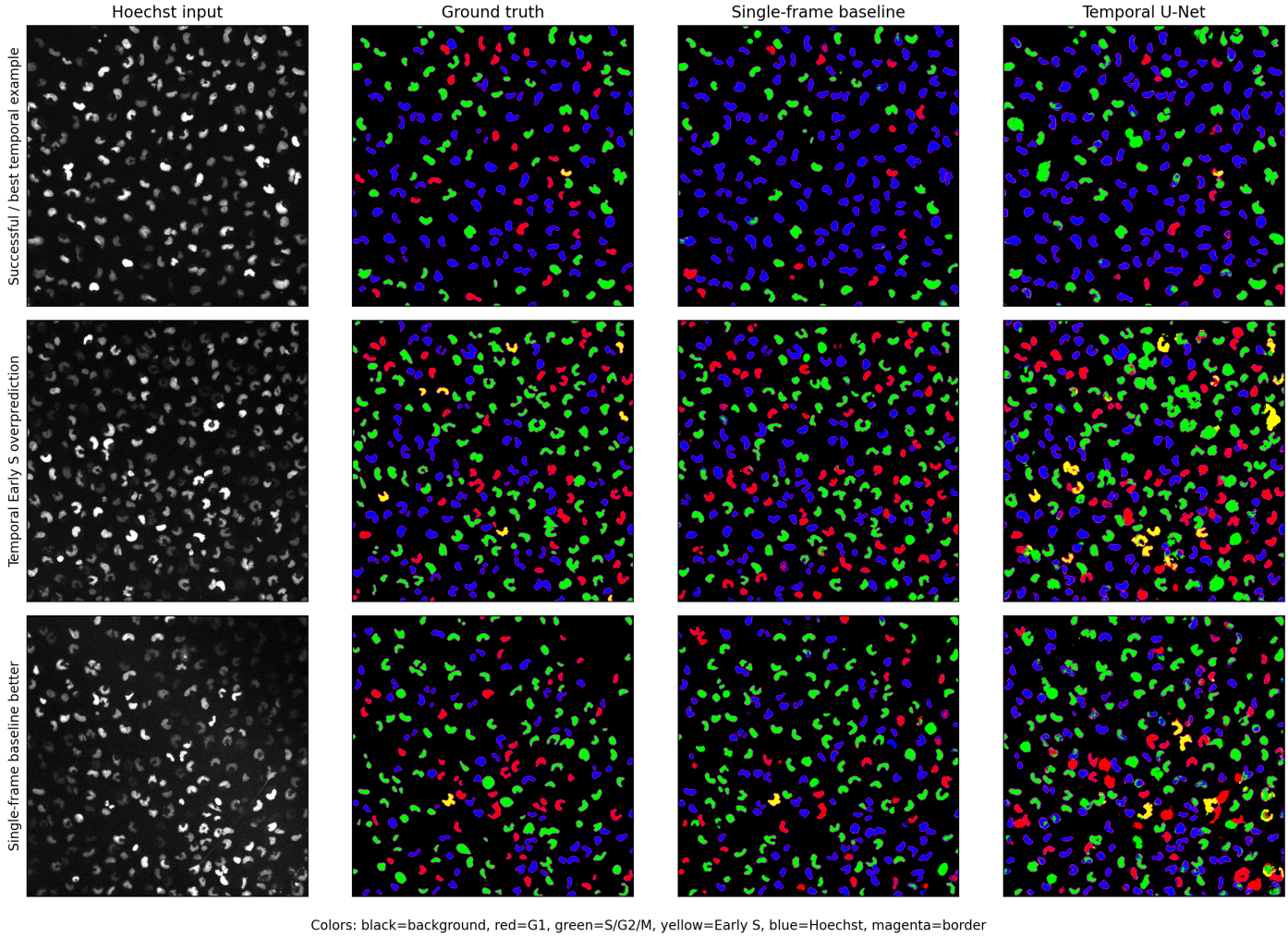

Figure 4: Example segmentation results for the single-frame baseline and the baseline-initialized temporal U-Net. Each row shows one test frame. The columns show the Hoechst input image, the ground-truth segmentation mask, the single-frame baseline prediction, and the temporal U-Net prediction.

3.7 Summary of findings

Overall, the experiments show that the tested temporal approaches did not improve early S detection over the inherited single-frame baseline. The temporal label-based classifier performed almost identically to the center-frame prediction on the tracked cell-instance subset, suggesting that hard predicted labels from neighboring frames did not add enough useful information. The temporal U-Net trained from scratch performed poorly and almost completely failed to detect early S. Initializing the temporal U-Net from the trained baseline checkpoint improved performance compared with training from scratch, but the model still performed worse than the single-frame baseline on the same test frames.

These results suggest that temporal context may still be useful, but the temporal approaches tested in this study did not successfully exploit it. In particular, the baseline-initialized temporal U-Net

became more sensitive to early S but also produced many false positives, leading to a lower F1-score than the original baseline model.

4 Discussion

This study investigated whether temporal context could improve early S phase detection in Hoechst-based cell-cycle segmentation. Since early S is a short and visually subtle transition between G1 and later S/G2/M, the main expectation was that neighboring timepoints could provide useful information that is not available in a single frame. Several temporal approaches were tested, including temporal label-based classification and temporal U-Net models using image windows. Overall, the results show that the tested temporal methods did not improve early S detection over the inherited single-frame baseline, although some temporal approaches were more successful than others.

4.1 Temporal label-based approaches

The first temporal approaches used predicted class labels from tracked cells at $t-1$, t, and $t+1$. These methods were based on the idea that a cell transitioning from G1 toward S/G2/M might be easier to recognize when its neighboring predicted labels are also considered. However, the logistic regression classifier performed almost identically to the original center-frame prediction on the same temporal subset. This suggests that the hard predicted labels from neighboring frames did not contain enough additional information to improve early S detection.

One likely explanation is that the temporal classifier only used the final predicted class labels of the baseline model. This means that uncertainty in the baseline prediction was lost. For example, a cell that is visually ambiguous between G1 and early S is still represented only as one discrete predicted label. The classifier therefore could not use softer information, such as class probabilities or image features, which may have been more informative. In addition, the temporal subset created by tracking may have been relatively clean, since only cells with valid consecutive detections and ground-truth matches were included. This may have made the center-frame baseline already difficult to improve upon.

4.2 Temporal U-Net models

The temporal U-Net models used the Hoechst images directly instead of only using predicted labels. This allowed the model to receive visual information from neighboring timepoints. The temporal U-Net trained from scratch performed poorly, especially for early S. With standard cross-entropy loss, the model did not detect early S pixels, and even after applying class weighting the early S F1-score remained very low. This indicates that training the temporal model from scratch was not sufficient for this dataset.

The poor performance of the model trained from scratch is likely related to the difficulty of the segmentation task and the strong class imbalance. Background pixels dominate the images, while early S pixels are rare. As a result, a model trained from scratch may learn to focus on the larger and easier classes while failing to learn the subtle visual features of early S.

The baseline-initialized temporal U-Net performed substantially better than the temporal U-Net trained from scratch. This suggests that initialization from the trained single-frame baseline was useful. The baseline model had already learned relevant Hoechst-based segmentation features, and the temporal model could build on these features instead of learning the task completely from the beginning. However, although the baseline-initialized temporal U-Net detected a larger proportion of early S pixels than the from-scratch model, its precision remained low. This means that the model became more sensitive to early S, but also predicted early S too often.

4.3 Comparison with the single-frame baseline

When the original single-frame baseline was evaluated on the same temporal U-Net test center frames, it performed better than the temporal U-Net models. For early S, the baseline reached a much higher F1-score than the baseline-initialized temporal U-Net. This shows that, in the current experimental setup, adding neighboring image frames as input did not improve early S detection. This result does not necessarily mean that temporal information is not useful for this problem. Instead, it suggests that the temporal information was not successfully used by the tested models. The temporal U-Net may have been affected by the training setup, the class weighting, or the limited three-frame input window. The low precision of the baseline-initialized temporal U-Net suggests that the weighted loss may have encouraged the model to overpredict early S. Therefore, the model detected more early S pixels, but at the cost of many false positives.

4.4 Limitations

There are several limitations to this study. First, the temporal label-based approaches depended on the quality of the baseline predictions and the tracking procedure. If the baseline prediction was already wrong, the temporal classifier only received this incorrect label as input. In addition, the tracking method was based only on centroid distance and was designed for short temporal windows, not for full cell-lineage reconstruction. This may have caused errors when cells moved, overlapped, disappeared, or were segmented inconsistently across frames.
Second, the temporal U-Net experiments used only a short three-frame window. This may not have been enough to capture the gradual progression into early S phase. A longer temporal window could provide more context, but would also increase memory usage and make training more difficult.
Third, the temporal U-Net was evaluated at the pixel level, while the temporal label-based methods were evaluated at the cell-instance level. These evaluation settings are useful for their respective methods, but they make direct comparison between all experiments more difficult.
Finally, the comparison with the inherited baseline model is limited by the use of different data splits. Because the baseline checkpoint was not retrained using the same split as the temporal models, the baseline performance on the current temporal test frames may be overly optimistic. This limits the strength of the conclusion that can be drawn from the direct baseline comparison.

4.5 Future work

Future work should first retrain the single-frame baseline using the same train, validation, and test split as the temporal models. This would make it possible to compare the baseline and temporal models under fully controlled conditions. In addition, temporal methods could be improved by

using class probabilities instead of hard predicted labels, since probabilities preserve information about uncertainty. This may be especially useful for early S, where cells may be ambiguous between neighboring phases. Further experiments could also compare pixel-level and instance-level evaluation more directly, since these metrics emphasize different aspects of performance.
Another possible direction for future work is to investigate the use of brightfield images. While this thesis used only the Hoechst channel as input, brightfield microscopy can provide morphological information that is useful for automated cell image analysis, such as cell outlines, shape, and boundary structure [SRP+09]. This information may be complementary to the DNA-related signal captured by Hoechst staining. Combining Hoechst and brightfield images as multimodal input could therefore be a useful direction for distinguishing subtle transitional states such as early S phase, especially when the Hoechst signal alone is ambiguous. This could be tested for both single-frame models and temporal models.

Temporal architectures. For the image-based models, future work could explore architectures that are designed more explicitly for sequential or spatiotemporal data. In this thesis, the temporal U-Net used neighboring frames by stacking $t-1$, t, and $t+1$ as input channels. This is a simple way to provide temporal context, but it does not explicitly model how cell appearance changes over time. A ConvLSTM-based model could be useful because ConvLSTM extends recurrent neural networks with convolutional operations, allowing the model to preserve spatial information while updating its representation over time [SCW+15]. With the current data, this could be implemented by passing a sequence of Hoechst frames from the same imaging location through a convolutional encoder and using the temporal representation to predict the segmentation mask of the center frame.
Another option would be to use 3D convolutional models. 3D convolutional networks can learn features across both spatial and temporal dimensions, which may help capture gradual changes in nuclear appearance around the G1/S transition [TBF+15]. A segmentation-specific version of this idea is 3D U-Net, which extends U-Net-style dense prediction to 3D inputs [ÇAL+16]. With the current dataset, short clips such as five or seven consecutive Hoechst frames could be treated as a small spatiotemporal volume, while the target mask remains the center timepoint.
Transformer-based temporal models could also be explored. Video transformer models such as TimeSformer use self-attention over space and time, allowing the model to learn which spatial regions and timepoints are most informative for the prediction [BWT21]. For early S detection, this could help the model use frames before and after the transition more flexibly than a fixed three-frame channel stack. However, transformer-based models usually require more data and careful regularization, so they may need to be combined with pretrained image encoders or patch-based training. Overall, these architectures could provide a more explicit way to model temporal progression while still using the same Hoechst image sequences and FUCCI-derived ground-truth masks available in the current dataset.

5 Conclusion

This thesis investigated whether temporal context could improve early S phase detection in Hoechst-based cell-cycle segmentation. Early S phase is difficult to detect because it is rare, visually subtle, and represents a transitional state between G1 and later S/G2/M. The main idea of this study

was therefore that information from neighboring timepoints could provide additional context for identifying early S more accurately than a single-frame prediction.

Several temporal approaches were tested. The temporal label-based methods used predicted class labels from tracked cells at $t-1$, t, and $t+1$, but performed almost identically to the center-frame prediction. This suggests that hard predicted labels from neighboring frames did not contain enough additional information to improve early S detection. A temporal U-Net trained from scratch also performed poorly, especially for early S. Initializing the temporal U-Net from the trained single-frame baseline improved performance compared with training from scratch, showing that the baseline model provided useful Hoechst-based segmentation features.

However, the baseline-initialized temporal U-Net did not outperform the inherited single-frame baseline on the same temporal test center frames. The temporal model became more sensitive to early S, but also produced many false positive early S predictions, resulting in a lower F1-score than the single-frame baseline. Therefore, the temporal approaches tested in this study did not successfully improve early S detection over the existing baseline model.

Overall, these results show that temporal context is not automatically sufficient to improve Hoechst-based cell-cycle segmentation. Temporal information may still be useful, but more careful model design and evaluation are needed. Future work should retrain the single-frame baseline and temporal models using the same data split, evaluate them on a fully independent test set, and explore temporal methods that can better model changes over time.

References

[Bir03] R Curtis Bird. Role of cyclins and cyclin-dependent kinases in g1 phase progression. *G1 Phase Progression. Kluwer Academic/Plenum, New York*, pages 40–57, 2003.

[BLG18] Jonas Bucevičius, Gražvydas Lukinavičius, and Rūta Gerasimaitė. The use of hoechst dyes for dna staining and beyond. *Chemosensors*, 6(2):18, 2018.

[BWT21] Gedas Bertasius, Heng Wang, and Lorenzo Torresani. Is space-time attention all you need for video understanding? In *Proceedings of the 38th International Conference on Machine Learning*, pages 813–824, 2021.

[ÇAL+16] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 424–432. Springer, 2016.

[Cui25] Di Cui. Generalisation and imbalance in deep learning for cell-cycle imaging. Master's thesis, Leiden University, 2025.

[DC24] Vicent Descals Carbonell. Identifying cell cycle phase using dna staining in time-lapse microscopy. Master's thesis, Leiden University, 2024.

[EKB+17] Philipp Eulenberg, Niklas Köhler, Thomas Blasi, Andrew Filby, Anne E Carpenter, Paul Rees, Fabian J Theis, and F Alexander Wolf. Reconstructing cell cycle and disease progression using deep learning. *Nature communications*, 8(1):463, 2017.

[HCG+22] FeiMing Huang, Lei Chen, Wei Guo, Tao Huang, and Yu-dong Cai. Identification of human cell cycle phase markers based on single-cell rna-seq data by using machine learning methods. *BioMed Research International*, 2022(1):2516653, 2022.

[IWWN17] Jaroslav Icha, Michael Weber, Jennifer C. Waters, and Caren Norden. Phototoxicity in live fluorescence microscopy, and how to avoid it. *BioEssays*, 39(8):1700003, 2017.

[Nig95] Erich A. Nigg. Cyclin-dependent protein kinases: Key regulators of the eukaryotic cell cycle. *BioEssays*, 17(6):471–480, 1995.

[NSC+20] Yukiko Nagao, Mika Sakamoto, Takumi Chinen, Yasushi Okada, and Daisuke Takao. Robust classification of cell cycle phase and biological feature extraction by image-based deep learning. *Molecular biology of the cell*, 31(13):1346–1354, 2020.

[RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[Sch98] K. A. Schafer. The cell cycle: A review. *Veterinary Pathology*, 35(6):461–478, 1998. PMID: 9823588.

[SCW+15] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

[SPRA+22] Nahian Siddique, Sidike Paheding, Abel A Reyes Angulo, Md Zahangir Alom, and Vijay K Devabhaktuni. Fractal, recurrent, and dense u-net architectures with efficientnet encoder for medical image segmentation. *Journal of Medical Imaging*, 9(6):064004–064004, 2022.

[SRP+09] Jyrki Selinummi, Pekka Ruusuvuori, Irina Podolsky, Adrian Ozinsky, Elizabeth Gold, Olli Yli-Harja, Alan Aderem, and Ilya Shmulevich. Bright field microscopy as an alternative to whole cell fluorescence in automated analysis of macrophage images. *PloS one*, 4(10):e7497, 2009.

[SSKM+08] Asako Sakaue-Sawano, Hiroshi Kurokawa, Toshifumi Morimura, Aki Hanyu, Hiroshi Hama, Hatsuki Osawa, Saori Kashiwagi, Kiyoko Fukami, Takaki Miyata, Hiroyuki Miyoshi, et al. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell*, 132(3):487–498, 2008.

[SWBM18] Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell detection with star-convex polygons. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 265–273. Springer, 2018.

[SWMP21] Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, 18(1):100–106, 2021.

[TBF+15] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4489–4497, 2015.

[VVBB03] Katrien Vermeulen, Dirk R Van Bockstaele, and Zwi N Berneman. The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell proliferation*, 36(3):131–149, 2003.

[ZE15] N. Zielke and B. A. Edgar. Fucci sensors: powerful new tools for analysis of cell proliferation. *WIREs Developmental Biology*, 4(5):469–487, 2015.