U-net for Cell Segmentation in Brightfield Microscopy Imaging

Alexander Gannon  
Department of Integrative Neuroscience  
Fordham UniversityNew York, USA  
[agannon5@fordham.edu](mailto:agannon5@fordham.edu)

<https://github.com/alexei-gannon/CellSegmentation_MLfinal_2025>

*Abstract*— —Cell segmentation is a fundamental step in biomedical imaging pipelines, particularly in applications involving high-throughput microscopy of bacteria. Manual segmentation is infeasible at scale, and classical algorithmic approaches are often imprecise, especially when dealing with overlapping or irregularly shaped cells. In this study, we develop and tailor a U-Net-based fully convolutional neural network for the segmentation of *Escherichia coli* and *Bacillus subtilis* cells in grayscale brightfield microscopy images. The model is trained on fewer than 1,000 annotated images and incorporates domain-specific post-processing criteria to automatically exclude filamentous or truncated cells. To improve generalization and boundary precision, we implement data transformations, boundary-weighted loss functions, and architectural modifications such as residual layers. We compare the effects of kernel size, network depth, and image transforms on model performance. Our final model achieves a low test loss and accurately recapitulates exclusion criteria, demonstrating that a lightweight, domain-adapted U-Net can be effectively trained on small bacterial microscopy datasets to assist in automated microscopy pipelines.

Keywords—cell segmentation, U-net, biomedical image analysis, deep learning, microscopy

# Introduction

Segmentation, partitioning an image into meaningful regions, is a foundational task in computer vision. In recent years, advances in deep learning have significantly improved segmentation performance, particularly for semantic and instance segmentation tasks in natural images [1]. These improvements have extended into biomedical imaging, where segmentation enables the extraction of critical features from complex visual data. Biomedical segmentation tasks often involve challenging scenarios such as low-contrast imaging, overlapping structures, and sparse annotations, making them ideal candidates for convolutional neural networks (CNNs), which can learn hierarchical representations from limited supervision [2], [3].

Within the domain of biomedical segmentation, cell segmentation represents a particularly important subproblem, especially in the context of microscopy-based studies. Accurate cell segmentation facilitates quantitative analysis of cellular morphology, tracking of cell dynamics, and spatial localization of molecular markers. This is especially crucial in high-throughput imaging pipelines, where manual segmentation is infeasible due to scale and variability [4]. In bacterial microscopy, challenges such as variable cell shapes, occlusions, low signal-to-noise ratio, and subtle boundaries further complicate segmentation, making automated methods a necessity for efficient downstream analysis [5].

The Thrall Lab at Fordham College Rose Hill studies the molecular mechanisms of DNA replication in gram positive and gram negative model organisms, Bacillus Subtilis and Escherichia Coli bacteria. By researching DNA replication, they seek to develop antibiotics that also prevent the development of antibiotic resistance, a growing threat to modern medicine. Attaching fluorescent tags to molecular machines of interest, they track the molecular dynamics of DNA replication with high-throughput microscopy. To characterize molecular dynamics within the cell independent of the movement of the entire cell, tags must be localized to the cells in which they are located.

In this way, cell segmentation is a key step in their microscopy pipeline. Given each image contains up to one-hundred bacteria, manual cell segmentation is infeasible. The lab currently utilizes an algorithmic approach, but this approach is imprecise and therefore requires time-consuming manual corrections. In particular, given the scope of the lab’s research, cells that are filamentous (overly long) or not completely in frame must be excluded. Therefore, this semi-automated cell segmentation process remains a rate-limiting step in this lab’s time-sensitive research.

This study seeks not only to investigate the use of deep learning for object segmentation but also to tailor a model to the practical needs of the Thrall Lab. We set out to train a model on a dataset of <1000 samples that performs accurate cell segmentation, automatically applies the exclusion criteria of filamentous and out of frame cells, and is not overly compute-intensive.

While transformers have recently been shown to perform well on biomedical segmentation tasks, they are compute-intensive and can require large datasets to perform well. The classical approach would instead point to the utility of fully convolutional neural networks (FCN) [6]. A FCN employs a series of convolutional encoder and decoder blocks to perform segmentation and classification tasks. A particular FCN architecture, the U-net, has been the gold standard in biomedical segmentation [7]. This architecture is based on a “U-shaped” symmetric set of encoders and decoders with skip connections between layers of equivalent dimension, combining low-level and high-level features for pixelwise classification.

Given the ability of this model to perform biomedical segmentation tasks and incorporate high-level features with low computational cost at test time, this study set out to tailor a U-net model for the Thrall Lab’s use case. For this purpose, the Thrall Lab provided a dataset of just under 1000 labeled microscopy images. Seeking to optimize our performance on a small dataset, this study assesses the value of various treatments on the dataset, model architecture, and training protocol in developing a precise and generalizable model.

# Methods

## Dataset

The dataset consisted of [how many?] grayscale brightfield microscopy images of B. Subtilis or E. Coli cells, each paired with a manually segmented binary mask identifying cell regions. All samples were imaged at a fixed resolution of 512×512 pixels. Images that had high levels of background fluorescence (e.g. noise) were removed from the dataset. Cells that were filamentous or truncated at the image boundary were excluded during annotation to reflect the lab’s post-processing criteria.

## Model Architecture

All models were FCNs with a U-net architecture. Each stage consisted of two convolutional layers with batch normalization and ReLU activations. Skip connections between encoder and decoder blocks allowed the network to combine high-level semantic features with fine-grained spatial detail.

Given this segmentation task requires the incorporation of high-level features to meet the exclusion criteria, various architectural changes were tested to determine how to best capture relevant information. Increasing the kernel size increases the receptive field of the model, trading fine-detail for high-level information. Therefore, both 3x3 and 5x5 convolutional kernels were tested. Furthermore, it was hypothesized that a deeper network could identify high-level patterns more accurately without losing low-level information, so both 4-layer and 5-layer networks were tested.

## Training

All models were trained using the Adam optimizer with learning rates of 1e-4 or 1e-3. A batch size of 8 was chosen due to compute constraints. The loss function was binary cross-entropy (BCE) with logits, which trains the model to output prediction values that, when input into a sigmoid function, give the estimated probability that each pixel is a cell. Each epoch, the model would test on the validation set. If validation loss failed to decrease from the previous epoch, training was concluded.

The dataset was split 80%/20% into train and test sets, with 20% of the train set used for validation. Samples were normalized from -1 to 1 based on their deviation from the mean. To increase generalization, each sample was randomly rotated in a multiple of 90 degrees when loaded from the training set.

As some images contain dense colonies of bacteria close together, it was assumed that predicting clear boundaries between cells could become difficult. Therefore, the loss function during training was augmented to weigh cell boundaries more heavily. In particular, a binary boundary mask was computed, and the BCE loss was multiplied elementwise by a weight map of the form

(1 + 0.5 \* boundary)

The performance of this approach was evaluated against standard BCE loss.

## Testing

During evaluation, model predictions were passed through a sigmoid activation with a prediction threshold of 0.5 to obtain binary masks. Segmentation quality was further evaluated using the Dice coefficient, which measures overlap between the model’s binary mask and the ground truth masks.

# Results

## Image Transforms

At a learning rate of 1e-3 with 3x3 kernel and four layers, models with and without image transforms were evaluated on the validation set. While both models converged quickly, image transforms were shown to have a robust positive effect on the generalization of the model. Therefore, image transforms were adopted for the remainder of the study.

1. Image Transform Effect

| Minimum Validation Set Loss | | |
| --- | --- | --- |
| Architecture | No Transform | With Transform |
| Kernel: 3, lr: 1e-3, Layers: 4 | 0.1536 | 0.0749 |

## Learning Rate

Given a learning rate of 1e-3 resulted in convergence in only two epochs, it was suspected that lower learning rates would improve performance. Therefore, a lower learning rates of 1e-4 was tested on the standard 4 layer U-net architecture with 3x3 kernels and image transforms. This resulted in smoother convergence and better performance, so a learning rate of 1e-4 was adopted for the remainder of the study.

1. Learning Rate Effect

| Minimum Validation Set Loss | | |
| --- | --- | --- |
| Architecture | Lr: 1e-3 | Lr: 1e-4 |
| Kernel: 3, Layers: 4, Image Transforms | 0.0749 | 0.0505 |

## Kernel size and U-net Depth

To improve the ability of the model to capture high-level features, both kernel size and network depth were increased and compared to the optimal 3x3 kernel, 4 layer model above. Surprisingly, both a 5 layer model and a 5x5 kernel worsened model performance.

Figure I. Kernel Size & U-net Depth Effect

A graph of loss of a graph

Description automatically generated with medium confidence

The poor performance of the 5x5 kernel is likely attributable to a loss in detail in trade-off with the larger receptive field. The worse performance in the deeper model was more surprising. To gain further intuition as to why a deeper model would result in worse performance, outputs model were visualized on the validation set and compared to ground truth masks. This suggested that the deeper model struggled to assess clusters of cells, with predictions bleeding in between nearby bacteria.

Figure II. 5 Layer U-net performance

A collage of images of different types of bacteria

Description automatically generated

Therefore, it appeared that the limits of our model may lay in its inability to predict clear boundaries between bacteria. We therefore set out to alter our model to better predict cell boundaries.

## Residual Layers & Weighted Edges

On the one hand, it was hypothesized that the ability to incorporate more fine-grained information would assist in the ability to define cell boundaries. Therefore, an additional skip layer called a residual was added to each convolution block. Here, a single convolution would be added to the output of each block before the activation function to provide further source information and assist in learning identity relationships.

On the other hand, it seemed likely that the loss function was not significantly encouraging the learning of clean cell boundaries. Therefore, boundary maps around each cell were calculated from the ground truth masks. These were used to multiply the loss of pixels around the boundaries of cells, penalizing the model extra for poor boundary prediction. However, the loss on the validation set was measured without weighted edges.

Residual layers were shown to result in much faster convergence with slightly less loss on the validation set. However, adding weighted edges onto the model actually increased minimum loss on the validation set.

Figure III. Residual and Weighted Edges Effects

A graph with blue line and green line

Description automatically generated

## Test Set Results

Given the superiority of the 4 layer U-net with residuals, this was chosen as the best model and applied to the test set. Validation loss matched the test loss closely, indicating that they held similar distributions, as desired.

1. Final Model Test Loss

| Architecture | Validation | Test |
| --- | --- | --- |
| 4 Layers, 3x3 Kernel, Image Transforms, Residual Layers | 0.0417 | 0.0427 |

Model performance was then visualized to determine how useful this would be in applications. While overall a good fit, model performance still appeared to be lacking in some instances, as below.

Figure IV. Final Model Performance (Representative)

A collage of microscopic images

Description automatically generated

On the one hand, extremely large filamentous bacteria are clearly excluded from the dataset, as are partially occluded cells. This suggests that this model has learned the exclusion criteria broadly. On the other hand, the model still struggles to differentiate clusters of cells from filamentous cells, which is visible by inconsistent predictions around clusters.

# Discussion

We have demonstrated that a U-net architecture can produce generally accurate cell segmentations from a dataset of less than 1000 labeled images. However, further work must be done create a completely automated segmentation pipeline. These results are consistent with prior findings that U-Net architectures perform well on biomedical segmentation tasks, even with limited training data [7], [8].

## Annotation Iconsistencies

It is possible the dataset itself has inconsistently labeled segmentations. From observation, it would appear that the decision as to whether two sufficiently proximal cells will be excluded as filamentous included as two cells can be inconsistent within the same image. A poorly defined segmentation rule in the training set would naturally result in poorer predictions in the model. I plan to reach out to the Thrall Lab to see if they have a subset of segmentations that may be more consistently labeled. That said, even incorporating this model into the current research pipeline could accelerate the rate of research, producing more samples to train future models.

## Loss Function

We primarily used binary cross-entropy (BCE) with optional boundary weighting as the loss function. While this produced decent performance, the limitations in boundary segmentation raise the possibility that alternate loss functions may be more appropriate. The Dice coefficient loss is often preferred in biomedical segmentation because it directly optimizes for spatial overlap between predicted and ground truth masks, especially when dealing with class imbalance [9]. In fact, combined BCE + Dice loss has been shown to provide both probabilistic learning stability and segmentation precision in other segmentation tasks [10].

## Architecture

Although deeper networks or larger kernels did not improve performance in this study, future iterations could explore architectures specifically optimized for boundary prediction. For example, Attention U-Nets introduce attention gates that allow the model to focus on more relevant spatial features and have been shown to improve boundary performance in biomedical imaging [11].

Another avenue involves the use of multi-scale features. Models like DeepLabv3+ use atrous spatial pyramid pooling (ASPP) to capture information at multiple scales, which can help resolve ambiguities in tightly packed or oddly shaped objects [12]. While computationally more expensive, these architectures may be valuable in a future pipeline once a larger dataset is available or if more computational resources are secured.

## Post-Processing and Pipeline Integration

In parallel with architectural and training improvements, post-processing techniques such as watershed algorithms or connected component labeling could be employed to refine the model’s binary predictions. For example, even when clustered cells are predicted as a merged region, post-processing could separate them using intensity gradients or distance transforms [13]. This hybrid approach—model prediction followed by rule-based refinement—may prove useful to adapt to inconsistently labeled training data.

##### References

1. J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 3431-3440, doi: 10.1109/CVPR.2015.7298965. keywords: {Semantics;Training;Convolution;Image segmentation;Computer architecture;Deconvolution;Adaptation models},
2. Ciresan, D., Giusti, A., Gambardella, L. and Schmidhuber, J. (2012) Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images. *Advances in Neural Information Processing Systems*, 25, 2843-2851.
3. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, *42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005
4. Meijering, E. (2012). Cell segmentation: 50 years down the road [life sciences]. *IEEE signal processing magazine*, *29*(5), 140-145.
5. Stringer, C., Wang, T., Michaelos, M., & Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nature methods*, *18*(1), 100-106.
6. Intisar Rizwan I Haque, Jeremiah Neubert, Deep learning approaches to biomedical image segmentation, Informatics in Medicine Unlocked, Volume 18, 2020, 100297, ISSN 2352-9148, https://doi.org/10.1016/j.imu.2020.100297.
7. Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4\_28
8. Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., Dovzhenko, A., Tietz, O., Dal Bosco, C., Walsh, S., Saltukoglu, D., Tay, T. L., Prinz, M., Palme, K., Simons, M., Diester, I., … Ronneberger, O. (2019). U-Net: deep learning for cell counting, detection, and morphometry. *Nature methods*, *16*(1), 67–70. <https://doi.org/10.1038/s41592-018-0261-2>
9. Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Jorge Cardoso, M. (2017). Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. *Deep learning in medical image analysis and multimodal learning for clinical decision support : Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, held in conjunction with MICCAI 2017 Quebec City, QC,...*, *2017*, 240–248. https://doi.org/10.1007/978-3-319-67558-9\_28
10. Iglovikov, V. I., & Shvets, A. A. (2018). *TernausNet: U-Net with VGG11 encoder pre-trained on ImageNet for image segmentation.* arXiv preprint arXiv:1801.05746.
11. Oktay, O. et al. (2018). *Attention U-Net: Learning where to look for the pancreas.* arXiv preprint arXiv:1804.03999.
12. Chen, L.-C. et al. (2018). *Encoder-decoder with atrous separable convolution for semantic image segmentation.* ECCV.
13. Malpica, N., de Solórzano, C. O., Vaquero, J. J., Santos, A., Vallcorba, I., García-Sagredo, J. M., & del Pozo, F. (1997). Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry*, *28*(4), 289–297. https://doi.org/10.1002/(sici)1097-0320(19970801)28:4<289::aid-cyto3>3.0.co;2-7