U-Net : Convolutional Networks for Biomedical Image Segmentation

CABRERA DECHIA Angel Joel ANGEL.JOEL_JOEL@ETU.SORBONNE-UNIVERSITE.FR
Master Ingénierie des Systèmes Intelligents
Sorbonne Université
Paris, France

CHELFAT Aya AYA.CHELFAT@ETU.SORBONNE-UNIVERSITE.FR
Master Ingénierie des Systèmes Intelligents
Sorbonne Université
Paris, France

DE BRION MATTHIEU MATTHIEU.DE-BRION@ETU.SORBONNE-UNIVERSITE.FR
Master Ingénierie des Systèmes Intelligents
Sorbonne Université
Paris, France

PAPAKOSTAS Konstantinos KONSTANTINOS.PAPAKOSTAS@ETU.SORBONNE-UNIVERSITE.FR
Master Ingénierie des Systèmes Intelligents
Sorbonne Université
Paris, France

Editor: Machine Learning Avancé (2025-2026)

Abstract

Semantic segmentation is a fundamental task in computer vision, with critical applications in biomedical imaging such as cell and tissue analysis. Recent advances in deep learning, particularly fully convolutional networks, have significantly improved segmentation accuracy by enabling dense, pixel-wise predictions. In this work, we revisit the U-Net architecture, a reference model for biomedical image segmentation, and reproduce its original experiments using a modern PyTorch implementation. We evaluate the model on three widely used microscopy datasets—EM-SEG, DIC-HeLa, and PHC-U373—and analyze its performance using standard metrics including Intersection over Union (IoU) and Dice coefficient. Beyond reproducing the original results, we conduct additional experiments to assess the impact of architectural and training choices, including multi-dataset training and the removal of skip connections. Our findings show that U-Net continues to be highly effective for dataset-specific semantic segmentation, while also revealing that architectural elements, such as skip connections, can have effects that vary across datasets. Furthermore, we compare U-Net with the general-purpose Cellpose segmentation model, showing that task-specific training provides superior performance for cell segmentation tasks. Overall, this study confirms the robustness and flexibility of U-Net for practical biomedical semantic segmentation and emphasizes the importance of adapting network architectures to the characteristics of the target data rather than relying on a single universal design. **Code:** `https://github.com/MattDBrR/unet-mla`

Keywords: semantic segmentation, biomedical imaging, U-Net, fully convolutional networks, convolutional neural networks,

1. Introduction

Deep learning has made significant breakthroughs in recent years, especially in the field of computer vision, where convolutional neural networks (CNNs) have become the most widely used approach

for solving complex visual tasks [1, 2]. Among these tasks, image segmentation plays a crucial role, particularly in biomedical imaging [3, 4].

Biomedical image segmentation is a challenging problem due to several inherent difficulties. First, obtaining large amounts of annotated data is often expensive and time-consuming, as it requires expert knowledge. Second, biomedical images frequently exhibit high variability in shape, scale, and appearance, as well as low contrast and noise. These challenges make classical image processing techniques insufficient and motivate the use of deep learning–based methods capable of learning rich hierarchical representations directly from data.

U-Net, introduced by Ronneberger et al., is a fully convolutional network and a standard model for biomedical image segmentation. Its symmetric encoder–decoder design captures both global context and fine details, with skip connections linking corresponding layers to improve segmentation accuracy even with limited data.

The objective of this project is to closely reproduce the experiments presented in the original U-Net paper. Our goal is to understand the main design choices of the U-Net architecture and to verify its performance on biomedical image segmentation tasks. In addition, we compare our results not only with those reported in the paper, but also with results obtained using publicly available pre-trained U-Net models, in order to provide a clearer and more meaningful comparison.

2. The U-Net model

U-Net is an extended and modified version of the fully convolutional network (FCN) introduced in [5] for semantic segmentation. The architecture is named U-Net due to its characteristic U-shaped structure. It follows an encoder–decoder paradigm: the encoder (contracting path) progressively downsamples the input image through successive convolutional operations, enabling the network to capture high-level contextual information. The decoder (expanding path) then upsamples the feature maps to recover spatial resolution, while concatenating them with corresponding feature maps from the encoder via skip connections. This symmetric structure allows the model to combine contextual information with precise localization. By processing the entire image end-to-end, U-Net outperforms traditional sliding-window CNN approaches in semantic segmentation tasks.

3. Datasets

For this project, we used several datasets to train and test our U-Net segmentation model. The datasets used are as follows:

- **DIC-HeLa [6]:** This dataset contains images of HeLa cells captured using Differential Interference Contrast (DIC) microscopy. The images are accompanied by manually annotated segmentation maps, which allow us to evaluate the segmentation performance of the model. This dataset is commonly used in cell segmentation challenges, making it highly relevant to our study.
- **EM-Seg:** This dataset consists of electron microscopy images with annotated cell membrane segmentation. The EM-Seg dataset is typically used for high-resolution segmentation tasks, which aligns well with the goals of our approach.
- **PHC-U373 [7]:** This dataset contains images of glioblastoma-astrocytoma (U373) cells recorded using phase contrast microscopy. These data were used in the ISBI 2015 cell tracking challenge and are useful for evaluating the model's performance on transmitted light images in cell segmentation tasks.

The data is first preprocessed to adjust resolution and apply data augmentations (e.g., elastic deformations), which improves the model's robustness to variations in input images. Segmentation on these datasets is evaluated in terms of Intersection over Union (IoU) and Dice scores, which measure the accuracy of the segmentation for each cell.

3.1 Justification of Differences from the Reference Paper

The original U-Net paper by Ronneberger et al. (2015) reports results on the same datasets, but one key difference lies in the use of data augmentation techniques, such as elastic deformations (similar to those applied in the ISBI 2015 challenge). Even if we apply the same transformations, their random nature ensures that our model is not trained on exactly the same augmented samples as in the original work. This, however, does not pose any issue. The original U-Net uses an overlap-tile strategy where large images are partitioned into overlapping tiles. Each tile is mirror-padded and processed so that the valid central predictions can be stitched together seamlessly. Our current dataset loader applies mirror padding but does not perform explicit patch-overlap tiling and merge logic as described in the original paper. Therefore, the full overlap-tile strategy is not implemented, which could affect border consistency in large images.

Moreover, examining the original implementation, we note that in 2015 U-Net was implemented using Caffe [8], a deep learning framework developed by Berkeley AI Research, primarily in C++. Since then, more flexible and widely adopted frameworks such as PyTorch [9] and Keras [10] have become standard.

In this work, we chose to implement U-Net in PyTorch due to its ease of use, dynamic computation graph, and strong support for custom modifications and GPU acceleration.

4. Experiment evaluation

In addition to qualitative visual evaluation, we rely on quantitative metrics to allow a direct comparison with the results reported in the original paper. In particular, we use the Intersection over Union (IoU), which for segmentation masks is defined as:

$$\text{IoU} = \frac{|M_{\text{pred}} \cap M_{\text{gt}}|}{|M_{\text{pred}} \cup M_{\text{gt}}|} \tag{1}$$

where M_{pred} denotes the predicted segmentation mask and M_{gt} the ground truth mask.

In addition, we report the Dice coefficient, which is widely used in medical image segmentation and is defined as:

$$\text{Dice} = \frac{2|M_{\text{pred}} \cap M_{\text{gt}}|}{|M_{\text{pred}}| + |M_{\text{gt}}|} \tag{2}$$

The Dice coefficient is closely related to IoU but is more sensitive to small structures, making it a complementary metric for evaluating segmentation performance.

Experimental setup: All networks were trained with a batch size of 1, a learning rate of 10^{-4}, and a weight decay of 10^{-4}. A learning rate scheduler was applied during training to improve convergence.

4.1 Results

4.1.1 Experiment 1

The first experiment consisted in training three separate neural networks, each one using a distinct dataset. All networks shared exactly the same architecture and layers as the original U-Net model.

The best results were obtained at epochs 18, 37, and 89 for EM-SEG, DIC-HeLa, and PHC-U373, respectively.

The predictions of each neural network are presented in Figure 1. Visually, all networks capture the overall structure of the cells and are able to separate them correctly. The network trained on the DIC-HeLa dataset shows some holes in the masks and a tendency to over-segment the cells.

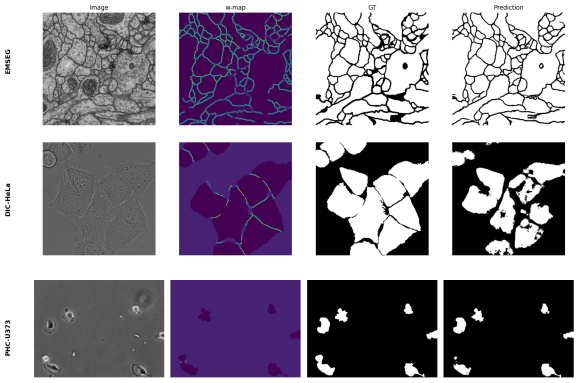

Figure 1: Results of Experiment 1 for the three neural networks.

4.1.2 Experiment 2

In the second experiment, a single neural network was trained using all three datasets. The network was trained for 100 epochs.

Figure 2 shows the segmentation results. The performance is lower compared to Experiment 1, and some structural differences between datasets are apparent in the predicted masks.

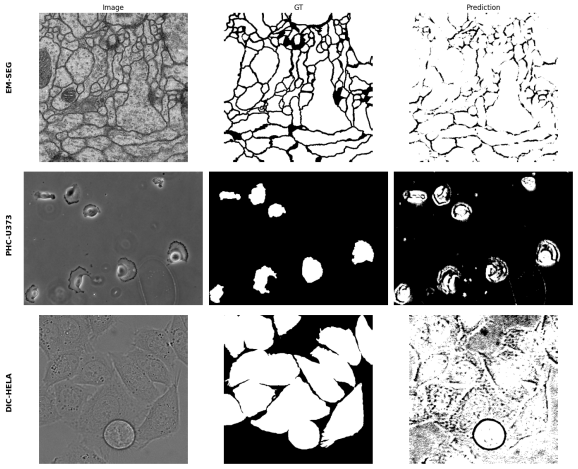

Figure 2: Results of Experiment 2 using a single neural network trained on all three datasets.

4.1.3 Experiment 3

In the final experiment, the network architecture was modified by removing the skip connections.

Results are visible on Figure 3. For the DIC-HeLa dataset, the predicted masks show fewer holes and smoother contours. For the PHC-U373 dataset, some cells are not segmented correctly.

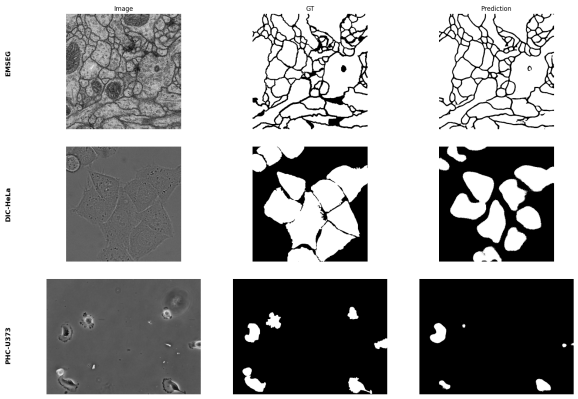

Figure 3: Results of Experiment 3 without skip connections.

Table 1 reports IoU and Dice scores for Experiments 1 and 3, along with IoU results from the original U-Net reported in the ISBI Cell Tracking Challenge 2015.All metrics are computed over the entire test set by averaging the per-image scores.

Dataset	Experiment 1		Experiment 3		U-Net (2015)
	IoU	Dice	IoU	Dice	IoU
EMSEG	0.8952	0.9447	0.8734	0.9324	–
DIC-HeLa	0.3176	0.4821	0.3731	0.5434	0.7756
PHC-U373	0.6934	0.8189	0.2703	0.4255	0.9203

Table 1: Segmentation performance comparison between Experiment 1 (standard U-Net), Experiment 3 (U-Net without skip connections), and the original U-Net architecture reported in the ISBI Cell Tracking Challenge 2015.

4.1.4 Comparison with Cellpose

In addition to the U-Net experiments, we evaluated *Cellpose* [11], a publicly available model for cell segmentation. Cellpose produces instance segmentation masks, which were converted to binary masks for fair comparison.

The predicted masks were evaluated using IoU and Dice:

- Average Dice coefficient: 0.8432
- Average IoU score: 0.7308

Since Cellpose outputs instance segmentation masks, all instances were merged into a single binary foreground mask before computing IoU and Dice scores.

4.2 Discussion

The experimental results highlight several important aspects regarding the generalization capacity of U-Net-based architectures and the role of skip connections in cell segmentation tasks.

First, the comparison between Experiments 1 and 2 shows that training a single neural network on heterogeneous datasets leads to a degradation in performance. While the unified model is exposed to a broader variety of data, the structural differences between datasets such as EMSEG, DIC-HeLa and PHC-U373 appear to limit its ability to learn discriminative features that generalize well across all scenarios. This suggests that, for cell segmentation tasks, dataset-specific characteristics may be better captured by specialized models rather than a single shared architecture, unless additional strategies such as domain adaptation or longer training are employed.

The results of Experiment 3 further emphasize the dataset-dependent nature of architectural choices. Removing skip connections leads to contrasting outcomes across datasets. For DIC-HeLa, both IoU and Dice scores improve, and the segmentation masks exhibit smoother and more coherent regions, with the disappearance of holes inside the cells. This behavior suggests that skip connections may propagate high-frequency noise or fine-grained artifacts that are detrimental for this dataset. In contrast, for PHC-U373, the absence of skip connections causes a significant drop in performance, indicating that preserving high-resolution spatial information is crucial for accurately segmenting cells with more complex or elongated structures.

Interestingly, the predicted contours in the absence of skip connections tend to be more rounded, which points to a loss of detailed edge information. While this smoothing effect can be beneficial for certain datasets, it becomes harmful when precise boundary delineation is required. These observations are consistent with recent findings [12], which demonstrate that skip connections are not universally beneficial and that their effectiveness strongly depends on the properties of the data.

The comparison with Cellpose confirms that task-specific training with U-Net generally outperforms a generalist instance segmentation model, particularly for datasets with small or tightly packed cells. The differences in performance can be attributed to the model objectives, dataset-specific training, and architectural optimizations.

Finally, when comparing our results with those reported for the original U-Net on the ISBI Cell Tracking Challenge 2015, we observe that the standard U-Net architecture remains highly competitive, particularly for the PHC-U373 dataset. However, the improved performance on DIC-HeLa without skip connections indicates that alternative architectural configurations may be more suitable for specific imaging modalities. Overall, these results underline the importance of adapting network architectures to the characteristics of the target dataset rather than relying on a one-size-fits-all design.

5. Conclusion

In this project, we successfully reimplemented the original U-Net architecture and reproduced its main experiments for biomedical image segmentation. Using a PyTorch-based implementation, we carefully followed the architectural design, training strategy, and evaluation metrics described in the original paper. Our experiments on three different datasets (EM-SEG, DIC-HeLa, and PHC-U373) confirm that U-Net remains a strong and reliable model for cell segmentation tasks, even years after its introduction.

The results show that training separate models for each dataset leads to better performance than using a single model trained on all datasets. This highlights the importance of dataset-specific characteristics and suggests that U-Net benefits from being adapted to the imaging modality and data distribution. Overall, our results are consistent with those reported in the original U-Net paper, while also revealing variation related to data augmentation, framework differences, and implementation details.

As future work, several improvements could be explored. These include experimenting with more advanced data augmentation techniques, performing more extensive hyperparameter tuning, and evaluating modern variants of U-Net. In conclusion, this work confirms the robustness and flexibility of the U-Net architecture.

References

[1] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *Nature* 521.7553 (2015), pp. 436–444. DOI: 10.1038/nature14539. URL: https://doi.org/10.1038/nature14539.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira et al. Vol. 25. Curran Associates, Inc., 2012. URL: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV]. URL: https://arxiv.org/abs/1505.04597.

[4] Geert Litjens et al. "A survey on deep learning in medical image analysis". In: *Medical Image Analysis* 42 (2017), pp. 60–88. DOI: 10.1016/j.media.2017.07.005. URL: https://www.sciencedirect.com/science/article/pii/S1361841517301135.

[5] Jonathan Long, Evan Shelhamer, and Trevor Darrell. "Fully Convolutional Networks for Semantic Segmentation". In: *arXiv* (2015). eprint: 1411.4038 (cs.CV). URL: https://arxiv.org/abs/1411.4038.

[6] G. van Cappellen. *HeLa cells on a flat glass*. https://data.celltrackingchallenge.net/training-datasets/DIC-C2DH-HeLa.zip. Accessed: 2026-01-09.

[7] Sanjay Kumar. *Glioblastoma-astrocytoma U373 cells on a polyacrylamide substrate*. https://data.celltrackingchallenge.net/training-datasets/PhC-C2DH-U373.zip. ISBI Cell Tracking Challenge Dataset. 2015.

[8] Yangqing Jia et al. "Caffe: Convolutional Architecture for Fast Feature Embedding". In: *arXiv preprint arXiv:1408.5093* (2014).

[9] Adam Paszke et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. 2019. arXiv: 1912.01703 [cs.LG]. URL: https://arxiv.org/abs/1912.01703.

[10] François Chollet et al. *Keras*. https://github.com/fchollet/keras. 2015.

[11] Marius Pachitariu, Michael Fariden, and Carsen Stringer. "Cellpose-SAM: superhuman generalization for cellular segmentation". In: *bioRxiv* (2025). preprint. DOI: https://doi.org/10.1101/2025.04.28.651001.

[12] Amith Kamath et al. "The impact of U-Net architecture choices and skip connections on the robustness of segmentation across texture variations". In: *Computers in Biology and Medicine* 184 (2025), p. 109364.