# Detecting Mitosis Cells using U-Nets

1st Sadiq Qara *Faculty of Natural Sciences and Engineering*
*Sabanci University*
Istanbul, Türkiye
sadig.gara@sabanciuniv.edu

***Abstract*—Mitosis detection in histopathological images is a pivotal task in cancer diagnosis and prognosis, as the mitotic count directly correlates with tumor aggressiveness and potential for metastasis. Traditional manual methods for mitosis detection are labor-intensive and subject to inter-observer variability, necessitating the development of automated and reliable techniques. This paper presents a novel approach utilizing a UNet-based model for the detection of mitotic cells in histopathological images. The UNet architecture, known for its efficacy in biomedical image segmentation, is leveraged to accurately localize and segment mitotic cells. Our model is trained and evaluated on a comprehensive dataset of annotated histopathological images, demonstrating superior performance in terms of precision and recall compared to existing methods. The proposed approach not only improves the accuracy of mitosis detection but also significantly reduces the time and effort required for analysis. This advancement holds promise for enhancing the diagnostic workflow in pathology, leading to more accurate and timely treatment decisions for cancer patients.**

***Index Terms*—Mitosis detection, UNet, Biomedical image segmentation, Deep learning, Digital pathology**

## I. Introduction

Mitosis detection is a critical task in histopathology, significantly influencing the prognosis and treatment decisions in cancer patients. The accurate identification and counting of mitotic cells in histological slides provide essential insights into the aggressiveness of tumors and their potential for metastasis. Traditional methods for mitosis detection rely heavily on the expertise and meticulousness of pathologists, making it a labor-intensive and time-consuming process prone to variability and human error.

The advent of digital pathology and the increasing availability of large-scale annotated datasets have paved the way for the application of machine learning and deep learning techniques to automate mitosis detection. Among these techniques, convolutional neural networks (CNNs) have shown remarkable performance in image analysis tasks due to their ability to learn hierarchical features from raw pixel data. However, the complexity and variability of histopathological images pose significant challenges for traditional CNN architectures.

UNet, a type of convolutional neural network designed for biomedical image segmentation, has emerged as a powerful tool for this task. Originally proposed by Ronneberger et al. in 2015 for segmenting neuronal structures in electron microscopic stacks, the UNet architecture has proven to be highly effective in various medical imaging applications. Its distinctive architecture, characterized by a symmetric encoder-decoder structure with skip connections, enables precise localization and robust segmentation even with limited training data.

This paper presents a novel approach for mitosis detection utilizing a UNet-based model. By leveraging

the strengths of the UNet architecture, our model aims to achieve high accuracy and reliability in identifying mitotic cells in histopathological images. We explore the historical context of mitosis detection, the evolution of digital pathology, and the advancements in deep learning that have culminated in the development of our proposed solution. Our work contributes to the ongoing efforts to enhance diagnostic accuracy and efficiency in pathology through innovative computational methods.

## II. Background

Mitosis detection in histopathological images has long been a critical task in pathology, providing key insights into the aggressiveness and progression of various cancers. Traditionally, pathologists manually examine histological slides under a microscope to identify and count mitotic figures, a process that is not only time-consuming but also prone to inter-observer variability and subjectivity. The advent of digital pathology has introduced the potential for automation in this domain, leveraging advancements in image processing and machine learning to enhance accuracy and efficiency.

Early automated methods for mitosis detection relied on handcrafted features and classical machine learning techniques, which often struggled with the complex and heterogeneous nature of histopathological images. The emergence of deep learning, particularly convolutional neural networks (CNNs), has revolutionized image analysis by enabling models to learn and extract hierarchical features directly from raw pixel data. Despite their success, traditional CNN architectures faced challenges in achieving precise localization of mitotic cells due to the intricate patterns and small size of these cells in high-resolution images.

The introduction of the UNet architecture by Ronneberger et al. in 2015 marked a significant breakthrough in biomedical image segmentation. UNet's distinctive encoder-decoder structure with skip connections allows for detailed spatial information to be preserved and effectively utilized in segmentation tasks. This architecture has demonstrated exceptional performance across various medical imaging challenges, including organ segmentation and lesion detection. Recognizing its potential, researchers have adapted UNet for mitosis detection, aiming to harness its ability to deliver precise and reliable segmentation in the context of histopathological analysis.

## III. Architecture

The UNet architecture, specifically designed for biomedical image segmentation, consists of a symmetric encoder-decoder structure that effectively captures and processes spatial information. The encoder path, or contraction path, progressively reduces the spatial dimensions of the input image while increasing the number of feature channels. This is achieved through a series of double convolutional layers followed by max-pooling layers. In this implementation, the encoder has three stages, starting with an initial input channel size of 3 (for RGB images) and progressing through feature sizes of 32, 64, and 128 channels.

At the bottleneck of the network, the feature representation is further refined with additional convolutional layers before entering the expansive path, or decoder. The decoder path upsamples the feature maps using transposed convolution layers, restoring the spatial dimensions while reducing the number of feature channels. Each upsampled feature map is concatenated with the corresponding feature map from the encoder path via skip connections, allowing the network to recover fine-grained details that might have been lost during the downsampling process.

The UNet's decoder also consists of three stages, mirroring the encoder, with feature channels decreasing

from 256 to 128, 128 to 64, and finally 64 to 32. Each stage employs a double convolutional block to refine the concatenated features. The final output layer uses a 1x1 convolution to map the 32-channel feature map to the desired number of output classes, facilitating precise segmentation. This design, combined with techniques like batch normalization and dropout to prevent overfitting, enables the UNet to achieve high accuracy and robustness in segmenting complex biomedical images.

### *A. Encoder Path*

The encoder path of the UNet architecture is responsible for extracting and compressing the feature representation of the input image. It consists of a series of double convolutional blocks followed by max-pooling layers to progressively reduce the spatial dimensions while increasing the depth of the feature maps.

- **Double Convolutional Block:** Each double convolutional block in the encoder path comprises two consecutive convolutional layers, each followed by a batch normalization layer and a ReLU activation function. This block is designed to capture complex features and patterns in the input image. Additionally, a dropout layer is included to prevent overfitting by randomly dropping out a fraction of the neurons during training.
- **Max-Pooling Layer:** After each double convolutional block, a max-pooling layer with a kernel size of 2 and a stride of 2 is used to downsample the feature maps. This layer reduces the spatial dimensions by half, allowing the network to learn hierarchical features while maintaining computational efficiency.

In this specific implementation, the encoder path consists of three stages, with the following configuration:

- **Stage 1:** The input image with 3 channels (RGB) is passed through a double convolutional block to produce feature maps with 32 channels, followed by a max-pooling layer.
- **Stage 2:** The 32-channel feature maps are further processed by another double convolutional block to produce 64-channel feature maps, followed by a max-pooling layer.
- **Stage 3:** The 64-channel feature maps undergo another double convolutional block to produce 128-channel feature maps, followed by a max-pooling layer.

The output of the encoder path is a set of 128-channel feature maps with reduced spatial dimensions, which are then passed to the bottleneck of the network for further processing.

### *B. Decoder Path*

The decoder path of the UNet architecture is responsible for reconstructing the spatial dimensions of the feature maps while reducing the depth, ultimately producing the segmented output. It mirrors the encoder path and utilizes transposed convolution layers for upsampling and double convolutional blocks to refine the feature maps.

- **Transposed Convolution Layer:** Each stage in the decoder path begins with a transposed convolution layer (also known as a deconvolution layer) that upsamples the feature maps by a factor of 2. This layer increases the spatial dimensions while reducing the depth of the feature maps.
- **Skip Connections:** The upsampled feature maps from the transposed convolution layer are concatenated with the corresponding feature maps from the encoder path. These skip connections allow the network to combine high-level semantic information from the decoder with low-level spatial information from the encoder, improving the accuracy of the segmentation.

- **Double Convolutional Block:** After concatenation, the feature maps are passed through a double convolutional block, similar to those in the encoder path, consisting of two convolutional layers, batch normalization, ReLU activation, and dropout. This block refines the concatenated feature maps and prepares them for further upsampling.

In this specific implementation, the decoder path consists of three stages, with the following configuration:

- **Stage 1:** The bottleneck feature maps (128 channels) are upsampled using a transposed convolution layer to produce 128-channel feature maps with doubled spatial dimensions. These are concatenated with the corresponding feature maps from the encoder path (also 128 channels) and passed through a double convolutional block to produce 128-channel feature maps.
- **Stage 2:** The upsampled feature maps from Stage 1 are further upsampled using another transposed convolution layer to produce 64-channel feature maps. These are concatenated with the corresponding 64-channel feature maps from the encoder path and passed through a double convolutional block to produce 64-channel feature maps.
- **Stage 3:** The upsampled feature maps from Stage 2 are upsampled once more using a transposed convolution layer to produce 32-channel feature maps. These are concatenated with the corresponding 32-channel feature maps from the encoder path and passed through a final double convolutional block to produce 32-channel feature maps.

The final output layer uses a 1x1 convolution to map the 32-channel feature maps to the desired number of output classes. This step ensures that each pixel in the output map is assigned a class label, resulting in precise segmentation of the input image.

The design of the decoder path, with its transposed convolutions, skip connections, and double convolutional blocks, allows the UNet to effectively reconstruct the spatial dimensions and produce accurate segmentations even in complex biomedical images.

## IV. Preprocessing

Effective preprocessing of the dataset is crucial for preparing the images and labels in a format suitable for training the UNet model. In this study, we applied several preprocessing steps to ensure that the input data was standardized and the labels accurately represented the mitotic cells.

### *A. Image Resizing*

The original histopathological images varied in size, which can pose challenges for consistent training and inference. To standardize the input dimensions, all images were resized to 512x512 pixels. This resizing ensures that the input images are uniform, which helps the model to learn effectively without being affected by varying image sizes.

### *B. Label Extraction*

To generate the training masks, we utilized a method that involves subtracting the unlabeled image from the labeled image of mitotic cells. The labeled images contain annotations of mitotic cells, whereas the unlabeled images do not. By subtracting the pixel values of the unlabeled image from the corresponding labeled image, we obtained a binary mask highlighting the regions where mitotic cells are present. This mask serves as the ground truth for training the UNet model.

The preprocessing steps can be summarized as follows:

- **Image Resizing:** All images were resized to 512x512 pixels to ensure uniformity in input dimensions.

- **Label Extraction:** The labeled image of mitotic cells was subtracted from the corresponding unlabeled image to generate a binary mask. This mask accurately represents the locations of mitotic cells and serves as the label for training the model.

## V. Dataset and Transformations

In this study, we utilized a mitosis cell dataset comprising labeled and unlabeled images of mitosis cells from five patients, resulting in a total of 35 images. The dataset was carefully curated to ensure a diverse representation of mitotic cells across different patients, providing a robust foundation for training and evaluating the proposed UNet model.

### *A. Custom Dataset Class*

To efficiently handle the dataset, we implemented a custom dataset class using PyTorch. This class, named `CustomDataset`, is designed to manage image and mask pairs, apply data augmentations, and facilitate the loading of images during training and evaluation. The key features of the `CustomDataset` class include:

- **Initialization:** The constructor accepts a list of images, a list of corresponding masks, optional transformations, and an augmentation factor. The augmentation factor allows for repeated application of transformations to increase dataset variability.
- **Length:** The `__len__` method returns the total number of samples, accounting for the augmentation factor.
- **Item Retrieval:** The `__getitem__` method retrieves an image-mask pair, applies transformations if specified, normalizes the image, and ensures the mask is in the appropriate binary format.

### *B. Data Augmentation and Transformations*

Data augmentation is a crucial technique to enhance the generalization capability of the deep learning model. We employed the `albumentations` library to define a comprehensive set of transformations, applied to both images and masks. The following transformations were applied:

- **RandomRotate90:** Randomly rotate the image and mask by 90 degrees.
- **HorizontalFlip:** Flip the image and mask horizontally.
- **VerticalFlip:** Flip the image and mask vertically.
- **Transpose:** Transpose the image and mask.
- **ShiftScaleRotate:** Apply random shifts, scaling, and rotations with specified limits.
- **GaussNoise:** Add Gaussian noise to the image.
- **RandomBrightnessContrast:** Randomly adjust the brightness and contrast of the image.
- **HueSaturationValue:** Randomly change the hue, saturation, and value of the image.

These augmentations ensure that the model is exposed to a wide variety of transformations, helping it to learn robust features and improve its generalization to unseen data.

## VI. Loss Function and Training Process

### *A. Loss Function*

To effectively address the class imbalance in mitosis cell detection, we employed the Focal Loss function. This loss function is designed to focus more on hard-to-classify examples by reducing the relative loss for well-classified examples and putting more focus on hard, misclassified ones. The parameters $\alpha$ and $\gamma$ were set to 0.25 and 2, respectively, to fine-tune the balance between easy and difficult samples, thereby enhancing the model's robustness.

### *B. Training Process*

The training process utilized a 5-fold cross-validation approach to ensure the model's generalizability. The

dataset was split into five parts, with the model trained and evaluated on different subsets to provide a comprehensive performance evaluation.

For optimization, we used the Adam optimizer with an initial learning rate of 0.001. The learning rate was adjusted using a StepLR scheduler, which decreased the learning rate by a factor of 0.1 every 50 epochs to facilitate more refined learning as training progressed. Each training fold consisted of 200 epochs.

Data augmentation was applied to the training dataset to increase its size and variability artificially. Techniques such as random rotations, horizontal and vertical flips, transpositions, shifts, scales, rotations, Gaussian noise, random brightness contrast adjustments, and hue-saturation value changes were employed. These augmentations helped improve the model's robustness by exposing it to a wider range of possible variations in the input data.

During training, the model's performance was evaluated using precision, recall, and F1-score metrics. These metrics were calculated for each fold, and their mean values were reported to assess the overall effectiveness of the model. This comprehensive evaluation ensured that the model was not only accurate but also reliable across different subsets of the data.

Finally, the trained model was saved for future use, ensuring that the best-performing parameters were preserved for potential deployment and further evaluation.

## VII. Results and Discussion

The performance of the UNet model for mitosis detection was evaluated using precision, recall, and F1-score metrics. The mean values across the 5-fold cross-validation were as follows:

- **Mean Precision:** 0.5748
- **Mean Recall:** 0.3150
- **Mean F1-Score:** 0.3168

### *A. Precision*

Precision, defined as the ratio of true positive predictions to the sum of true positive and false positive predictions, reflects the model's ability to correctly identify mitotic cells without falsely labeling non-mitotic cells as mitotic. A mean precision of 0.5748 indicates that approximately 57.48

### *B. Recall*

Recall, or sensitivity, is the ratio of true positive predictions to the sum of true positive and false negative predictions. It measures the model's ability to detect all actual mitotic cells. With a mean recall of 0.3150, the model is correctly identifying about 31.50

### *C. F1-Score*

The F1-score is the harmonic mean of precision and recall, providing a single metric that balances both concerns. A mean F1-score of 0.3168 reflects the model's overall performance, balancing precision and recall. The low F1-score indicates that while the model is reasonably precise, its low recall significantly impacts its overall effectiveness in mitosis detection.

### *D. Discussion*

The results highlight several key areas for improvement in the model. The disparity between precision and recall suggests that the model is conservative in its predictions, likely leading to fewer false positives but more false negatives. This behavior might be due to the imbalanced nature of the dataset, where non-mitotic cells vastly outnumber mitotic cells.

To address this, future work could explore the following strategies:

- **Data Augmentation:** Enhance the diversity and number of mitotic cell samples through advanced

augmentation techniques to improve the model's ability to generalize.

- **Loss Function Adjustments:** Experiment with different loss functions or adjust the focal loss parameters to better handle class imbalance.
- **Model Architecture Modifications:** Investigate modifications to the UNet architecture, such as deeper layers or additional regularization techniques, to improve feature learning.
- **Ensemble Methods:** Combine predictions from multiple models to reduce variance and improve robustness.

### *E. Comparison with State-of-the-Art Methods*

For a comprehensive understanding of the model's performance, we compared it with three state-of-the-art methods from the MITOS 2012 challenge:

- **ISDIA**
  - **F1-Score:** 0.7821
  - **Recall:** 0.7
  - **Precision:** 0.8861
- **IPAL**
  - **F1-Score:** 0.7184
  - **Recall:** 0.74
  - **Precision:** 0.6981
- **SUTECH**
  - **F1-Score:** 0.7094
  - **Recall:** 0.72
  - **Precision:** 0.699

These comparisons highlight the performance gap between the UNet model and other state-of-the-art methods, providing a benchmark for future improvements. The significant difference in F-measure, recall, and precision underscores the need for advanced techniques and optimization strategies to enhance the model's efficacy in mitosis detection.

## VIII. Impact of Data Augmentation

Data augmentation plays a crucial role in improving the generalization capability of deep learning models, particularly when dealing with limited datasets. In this study, we evaluated the impact of data augmentation on the performance of our UNet model for mitosis detection.

### *A. Results without Data Augmentation*

To assess the importance of data augmentation, we trained the UNet model on the same dataset but without applying any augmentation techniques. The performance metrics for the model trained on unaugmented data were as follows:

- **Mean Precision:** 0.1950
- **Mean Recall:** 0.0095
- **Mean F1-Score:** 0.0182

### *B. Comparison and Analysis*

Comparing the results with and without data augmentation highlights the significant impact of augmentation on the model's performance:

- **Precision:** The mean precision dropped from 0.5748 with augmentation to 0.1950 without augmentation. This substantial decrease indicates that the model trained without augmentation is less capable of correctly identifying mitotic cells, resulting in a higher number of false positives.
- **Recall:** The mean recall decreased dramatically from 0.3150 with augmentation to 0.0095 without augmentation. This suggests that the model trained without augmentation is unable to detect most of the actual mitotic cells, leading to a significant increase in false negatives.
- **F1-Score:** The mean F1-score, which balances precision and recall, fell from 0.3168 with augmentation to 0.0182 without augmentation. This

drastic reduction underscores the overall decline in model performance when data augmentation is not utilized.

### C. Discussion

The results clearly demonstrate that data augmentation significantly enhances the performance of the UNet model in mitosis detection. The large drop in precision, recall, and F1-score when augmentation is turned off indicates that the model struggles to generalize well to unseen data without the variability introduced by augmentation techniques.

Data augmentation helps the model to:

- **Generalize Better:** By exposing the model to a wider variety of transformations, augmentation reduces overfitting and improves the model's ability to generalize to new, unseen data.
- **Learn Robust Features:** Augmentation techniques such as rotations, flips, and brightness adjustments help the model learn more robust features, making it less sensitive to variations in the input data.
- **Handle Class Imbalance:** Augmentation can artificially increase the representation of underrepresented classes (mitotic cells), aiding the model in learning to detect these classes more accurately.

In conclusion, data augmentation is a vital component of the training process for mitosis detection using deep learning models. It substantially improves the model's performance by enhancing its ability to generalize and by mitigating issues related to limited and imbalanced datasets.

## IX. Conclusion

In this study, we presented a UNet-based approach for the detection of mitotic cells in histopathological images. The proposed model leverages the powerful UNet architecture to accurately segment mitotic cells, addressing the critical task of mitosis detection in cancer diagnosis and prognosis.

Our results demonstrated that the model achieved a mean precision of 0.5748, a mean recall of 0.3150, and a mean F1-score of 0.3168 when trained with data augmentation. These metrics indicate a reasonable performance, although there is significant room for improvement, particularly in recall.

We also explored the impact of data augmentation, comparing the performance of the model trained with and without augmentation. The stark contrast in results, with the unaugmented model showing a mean precision of 0.1950, a mean recall of 0.0095, and a mean F1-score of 0.0182, underscores the importance of data augmentation in improving model generalization and robustness.

Future work will focus on enhancing the model's recall by experimenting with advanced data augmentation techniques, refining the loss function, and exploring modifications to the UNet architecture. Additionally, incorporating ensemble methods and leveraging larger, more diverse datasets may further improve the model's performance.

In conclusion, our study highlights the potential of deep learning models, specifically UNet, in automating the detection of mitotic cells, thereby aiding pathologists in making more accurate and efficient cancer diagnoses.