Improving Model Performance in Biomedical Image Segmentation: A Comprehensive Analysis of the U-Net Approach

Zuhair Ahmed | Pritom Debnath | Sadia Sultana Ali | Md. Ali Ahnaf

Department of Computer Science | Aiub

*Abstract*— UNet's main purpose is to carry out precise semantic segmentation of images. Semantic segmentation is achieved within the UNet architecture by predicating a segmentation diagram that labels each pixel in the input image. When fed an image, the UNet model outputs a segmentation map with labeled pixel positions of the same dimensions as the original picture. This paper discusses potential issues and solutions that may arise if the kernel size is increased in an attempt to fix the overfitting problem, therefore enhancing the model's performance.

Index Terms— Unet, Overfitting Issue, Image Segmentation

# **Introduction**

## **Background Information**

**U**Net architecture is a prominent deep-learning model for semantic image segmentation. Since its introduction by Ola Ronneberger, Philipp Fischer, and Thomas Brox in 2015, it has garnered popularity as a model for various image segmentation issues.

The name of the structure is derived from its U-shaped structure. It comprises of a path that expands and a path that contracts. The contracting path, which consists of a series of convolutional layers each followed by a pooling operation, captures more context information and down-samples the feature maps. The expansive path consists of multiple convolutional layers, each of which is followed by an up-sampling procedure to help recover the spatial resolution lost during the contracting path and to up-sample the feature maps.

Since then, the UNet architecture has been implemented in numerous other disciplines, such as satellite photography, natural photography, and more. It was initially designed for segmenting biological images, such as dividing cells and nuclei. Numerous enhancements and modifications have been suggested for the model, which has performed admirably on a variety of benchmarks and remains a subject of continuing research.

## **Overview of this report**

There can be noticeable differences in performance between UNet architectures with different kernel sizes utilized in the convolutional layers. This paper proposes a different approach to reducing overfitting by adjusting the kernel size. However, this would entail fewer filters being utilized in the convolutional layers. However, cutting down on the size of the kernel will assist simplify the model by reducing the number of parameters. Reduced feature extraction caused by a smaller kernel size can prevent the model from overfitting by preventing it from learning too much from the training data.

## **Literature review**

In 2015, Olaf Ronerberger et al. [1] achieved improved accuracy in image segmentation using Fully Convolutional Networks (FCN). UNet builds on FCN by excavating multi-scale features of an image through its backbone. Seo et al. added layers to a contracting network, replacing pooling with up-sampling to increase output resolution. The high-resolution features from the contracting path are combined with the up-sampled ones for localization. A convolution layer then assembles a more precise output using this information.

In 2020, Kandel et al. [2] describes that a higher batch size does not usually achieve high accuracy, and the learning rate and the optimizer used will have a significant impact as well. A smaller filter size will be catching local patterns and more local patterns will be accumulated as the filter size goes down. On the other hand, a larger filter size is computationally costly and also, looks at the larger picture. So, in our implementation we used the Unet with smaller filters and kernel size.

In 2016, Çiçek et al. [3] introduces an extension which was made to the U-Net design that allows it to process 3D volumetric data. The scientists presented a network known as the 3D U-Net, which could process volumetric medical pictures by utilizing 3D convolutions and max-pooling layers. The 3D U-Net was able to produce accurate segmentation results despite only having little annotation, making it useful for applications such as the segmentation of organs in medical imaging.

In 2018, Zhou et al. [4] implemented U-Net++, an improved version of U-Net that makes use of a tiered design to capture more fine-grained data. The authors suggested the use of dense skip connections between related encoder and decoder blocks in order to enhance the flow of information. When compared to the original U-Net as well as other approaches that are considered to be state-of-the-art, the performance of U-Net++ was shown to be better in numerous medical picture segmentation tasks.

# **Methods**

## **Description of the solution**

For segmenting biomedical images, the proposed solution employs a U-Net architecture. For processing, the input images have been resized to 128 by 128 by 3 pixels. The U-Net architecture is comprised of an encoder and a decoder interconnected by skip connections.

The encoder processes the input images through a series of convolutional layers. Conv1 and Conv2 each have 16 filters with a 3x3 kernel dimension, forming the initial set of convolutional layers. To regularize the model, a dropout layer with a rate of 0.1 is employed between these convolutional layers. The spatial dimensions are then reduced to 64x64x16 via Max pooling. Conv3 and Conv4 layers, each with 32 filters and a 3x3 kernel size, are then utilized. These layers encapsulate the encoded representation's more complex characteristics. The objective of the decoder section is to reconstruct the segmentation map using the transmitted features. The encoder's up-sampled feature maps are concatenated with the skip connections' corresponding feature maps. This information fusion enables the decoder to utilize spatial details from earlier encoding stages. The concatenated characteristics are then processed by Conv5 and Conv6 layers with 64 filters and a 3x3 kernel size. These layers aid in refining the map's segmentation.

As the output layer, a 1x1 convolutional layer with a sigmoid activation function is utilized. It produces the final segmentation map with a shape of 128x128x1, where each pixel represents the predicted class probability.

|  |
| --- |
| Figure-1: Architecture of the model |

The proposed U-Net architecture with skip connections and transpose layers enables accurate biomedical image segmentation by facilitating efficient feature extraction, information propagation, and spatial preservation.

The Adam optimizer was used to train the U-Net model. To prevent overfitting, we employed early halting based on the validation loss, monitoring it over a 4-epoch patience period. The group size was set to 16, and the model was trained for a total of 25 epochs. Utilizing tensor-board for the log provides the necessary visualization and instrumentation for machine learning experimentation.

## **Description of Dataset**

The dataset used in this study is the 2018 Data Science Bowl dataset, which was created specifically for the "Find the nuclei in divergent images to advance medical discovery" challenge. This dataset is extensively used in the field of biomedical image segmentation and is available to the public.

The dataset is comprised of an assortment of biomedical images obtained from various sources. These images depict a variety of medical conditions and were chosen to illustrate the complexity and variation of actual biomedical data. Each image within the dataset is accompanied by segmentation templates based on the ground truth. These masks include annotations at the pixel level that identify the regions of interest, namely the nuclei evident in the images. The segmentation masks serve as the benchmark by which the efficacy of segmentation algorithms is measured.

Typically, the dataset is separated into training and test sets, allowing machine learning models to be trained and evaluated. The training set consists of numerous images with their associated segmentation masks for the purpose of training the model, while the test set is used to evaluate the model's generalization performance on unseen data.

# **Results and Discussion**

## **Analytical Results**

In this study, we implemented a U-Net architecture for biomedical image segmentation, concentrating on a set of 128x128x3 biomedical images. The performance of the model was assessed by training and evaluating it with the appropriate metrics.

On the test set, the U-Net model obtained an accuracy of 0.96, as shown in Figure 2. This demonstrates that the model reliably identified and segmented the target objects in the biomedical images. The high accuracy demonstrates that the U-Net architecture is effective at capturing and exploiting the pertinent features for segmentation tasks.

Intersection over Union (IoU) was calculated to evaluate the model's efficacy further. IoU is a frequently employed metric for segmentation tasks; it measures the overlap between the predicted segmentation map and the ground truth. The calculated IoU score for the U-Net model was 0.86, indicating a high level of concordance between the predicted and ground truth segmentations.

|  |
| --- |
| Figure-2: Training and validation accuracy graph |
| Figure-3: Training and validation loss graph |

## **Discussion And Future-work**

Even with the diminished image size, the results demonstrate that the U-Net architecture can effectively learn and generalize complex patterns in biomedical images. The utilization of skip connections facilitates the transmission of information from the encoder to the decoder, allowing for the precise localization of objects in the segmentation map. The transpose layers help recover spatial details that may have been lost during encoding by up-sampling the feature maps.

Notably, the performance obtained by the U-Net model could be attributed to the correct selection of hyperparameters and the robustness of the training dataset. Additional research could examine the efficacy of the U-Net architecture on larger datasets and across a variety of biomedical imaging modalities using larger data sets. The accuracy and quality of segmentation could be enhanced by refining the model and experimenting with various architectural variations.

Further enhancements in precision and IoU may be achieved by refining the model and examining various architectural variants. The results of this study indicate that the U-Net architecture is a promising solution for biomedical image segmentation tasks. Medical diagnosis, treatment planning, and disease monitoring can greatly benefit from the precise delineation of target objects. Future research can investigate the application of this architecture to larger datasets and its performance in different biomedical imaging modalities to validate its generalizability and prospective clinical utility.

# **Conclusion**

Finally, this study applied a U-Net architecture for biomedical image segmentation to a dataset of 128x128x3 images, significantly reducing the size of each individual image. Attaining an impressive 0.96 accuracy and 0.86 Intersection over Union (IoU) score, the U-Net model performed admirably. The success of the U-Net architecture in correctly segmenting biological pictures is demonstrated by these results. Although the results only apply to low-resolution images, they still show promise for use in biomedical image segmentation using the U-Net architecture. Disease monitoring, therapy planning, and diagnosis are all greatly aided by precise segmentation.

In conclusion, the results of the study show that the U-Net architecture has great promise as a useful tool for biomedical image segmentation tasks. Delineating biomedical objects precisely aids in clinical decision-making and ultimately benefits patient care.

# **References**

[1] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.

[2] Kandel, I., & Castelli, M. (2020). The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. ICT express, 6(4), 312-315.

[3] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: learning dense volumetric segmentation from sparse annotation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19 (pp. 424-432). Springer International Publishing.

[4] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4 (pp. 3-11). Springer International Publishing.

**Contribution table:**

|  |  |  |
| --- | --- | --- |
| ID | Name | Contribution |
| 20-42414-1 | DEBNATH, PRITOM | Methodology and Result Discussion |
| 20-42378-1 | MD. ALI AHNAF | Architecture Model, Discussion, future work, Conclusion |
| 20-42386-1 | SADIA SULTANA ALI | Literature Review and Methodology |
| 20-42115-1 | ZUHAIR AHMED | Introduction and Literature Review |

Implemented code GIT repo link: <https://github.com/PritoM-Debnath/cvpr/blob/main/FINAL/Project/UNET.ipynb>