Semantic Segmentation of Histopathological Images with Fully and Dilated Convolutional Networks

Huseyin Turhan, Gokhan Bilgin
Department of Computer Engineering
Yildiz Technical University
Davutpasa Campus 34220 Istanbul, Turkey
hseyintrhan@gmail.com, gbilgin@yildiz.edu.tr

***Abstract*—Nowadays, the segmentation of different components in medical images is a major subject of study, and parallel to this, numerous image segmentation methods are still being developed. This study aimed to assess image segmentation methodologies utilizing deep learning models, due to the success of deep learning models in image processing applications. Firstly, starting from the introduction, a literature review on semantic segmentation and medical image segmentation is introduced in this study. In addition, pre-processing steps and techniques, models used, evaluation criteria, and the reasons for their preference are also explained. In the methods section, SegNet, U-Net, and DeepLabV3+ model architectures are introduced, and the architectures of these models are visualized at a basic level. The application results section includes all evaluation results with the metrics used in measuring accuracy. The comparison of the evaluation results and the evaluations on these results are included in the results and discussion section. In addition to these, visualized prediction results are also presented under the application results section.**

***Index Terms*—Semantic segmentation, medical image segmentation, deep learning, fully and dilated convolutional networks.**

I. Introduction

Computer vision is a branch of science that studies how computers can extract high levels of meaning from digital images and videos. When evaluated in point of engineering, it is concluded that the human visual system tries to automate the tasks that it can do. Thanks to its comprehensive and wide usage area, it has led many academicians to integrate and collaborate with various scientific disciplines [1], [2]. There are several levels at which computers can understand images. Almost every one of these levels has a defined problem in computer vision, such as image classification, object detection, semantic segmentation, and instance segmentation. In computer vision and image processing, image segmentation is crucial. In this study, semantic segmentation techniques are explained using various deep learning models [3], [4].

Advances in health technologies have also benefited from computer vision substantially. Computer vision algorithms can aid in the automation of tasks like detecting malignant areas in images and identifying symptoms on x-ray and MRI scans. Radiologists' analysis can be enhanced by computers, in other words, machines and the time it takes to run diagnostic tests can be cut down.

978-1-6654-3663-2/21/ $31.00 © 2021 IEEE

While using Deep Convolutional Neural Networks for the semantic segmentation task, one of the two problems is the decreased feature resolution produced by successive pooling operations or convolution striding. However, dense prediction tasks requiring specific spatial information may be hindered as a result of this. The use of dilated (atrous) convolution, which has been demonstrated to be successful for semantic image segmentation, is suggested as a solution to this problem. Dilated convolution presents a new dilation rate parameter to convolutional layers. The dilated convolution of a signal $x(i)$ is specified as $y_i = \sum_{k=1}^{K} x[i+rk]\omega[k]$ (r is the dilation rate demonstrating a spacing between the weights of the kernel w). In the field of real-time segmentation, dilated convolutions have become prominent, and many papers have reported on their application.

Hesamian et al. [5] conducted a literature study on network architecture and model training methods and presented an overview of the most recent methods using deep learning in medical image segmentation; more focused on applied machine learning techniques, in-depth looked at their structures and methods, and their strengths and weaknesses were analyzed. Karimi et al. [6] reviewed the literature of the latest technologies for label noise handling techniques in deep learning and evaluated available approaches for segmentation and classification tasks in medical image processing on three medical imaging datasets. Zhou et al. [7] introduced the general principles of deep learning and multi-mode medical image segmentation taking a look at deep learning-based techniques, analyzed the proposed fusion strategy techniques, and evaluated the compared results. Göceri [8] discussed the difficulties affecting the performance of these methods associated with training and the architecture of deep neural networks for medical image segmentation and explained the current solution methods for data deficiency and class imbalance problems. Tajbakhsh et al. [9] represented a literature review that addresses the difficulties of annotations in the segmentation task where obtaining both data and annotations is costly, particularly in the medical imaging field, and includes detailed solutions.

Semantic segmentation may be thought of as a classification of the pixel problem using semantic tags. Semantic segmentation labels all pixels in an image at the pixel level. Semantic segmentation can also be thought of as pixel-level image classification. Therefore, it's a trickier assumption to make than image classification, which mostly forecasts a single label for the whole image. For instance, segmentation will classify all items in an image containing multiple automobiles as "car" objects. Instance segmentation, on the other hand, broadens the semantic segmentation scope by labeling specific instances of objects in the image. However, instance segmentation is not included in this study. According to the nature of their contributions, Taghanaki et al. [10] divided into six different groups of semantic image segmentation literature. In addition, they also examined the behavior of several prominent loss functions for training segmentation models in dealing with scenarios of different levels of false-negative and false-positive predictions, and following this comprehensive review they proposed important research instructions for each category. Guo et al. [11] presented a review of deep learning-based semantic segmentation methods. These methods are split into three groups according to their main component and the main ideas of these categories are mentioned in their work. These categories are region-based, FCN-based, and poorly controlled segmentation methods. Hu et al. [12] reviewed the development of semantic segmentation of the RGB-D dataset most commonly used for semantic segmentation. The current methods for RGB-D segmentation are classified as traditional machine learning-based techniques (such as conditional random fields and decision trees) and deep learning-based network architecture which recently become a notable successful model. Deep learning architectures for semantic segmentation of natural images using deep neural networks were summarized by Lateef and Ruichek [13], according to common concepts that form the basis of their architecture.. More than a hundred different models, the corpus of each model, its original architecture, test benchmarking, 33 public data sets, and the attributes of each data set were discussed. By analyzing their structural designs and their performances on the evaluated data sets, the success of these methods was emphasized.

S.Minaee et al. [14] examined in detail over a hundred deep learning-based segmentation algorithms that have been published up to 2019 and categorized these methods by model architectures. Several of these models are used for semantic segmentation in this study. The models used are U-Net, DeepLabV3 + and SegNet. The main reason for preferring these models is that these models outperformed many high-end segmentation networks in existing image segmentation studies. The Accuracy, F1 / Dice-score, and Cohen kappa score, which are some of the popular metrics used to evaluate deep learning segmentation models [15], were used in the evaluation and testing stages in this study. The results obtained with these models and metrics and their comparisons are shown in the Application Results section.

II. Methods

A. U-Net Model Architecture

U-Net was developed by Olaf Ronneberger et al. [16] for medical picture segmentation. U-Net architecture mainly includes two ways. The primary way is to use the encoder path (contraction) to get the image's context. A conventional stack of convolutional and max-pooling layers makes up the encoder path and the path has FCN-like architecture. The decoder path (symmetric expanding) is utilized to get accurate localization utilizing transposed convolutions in the second step. As a result, U-Net is an end-to-end FCN that can handle any size image.

The U-Net model architecture defined in [16] is shown in Figure 1.

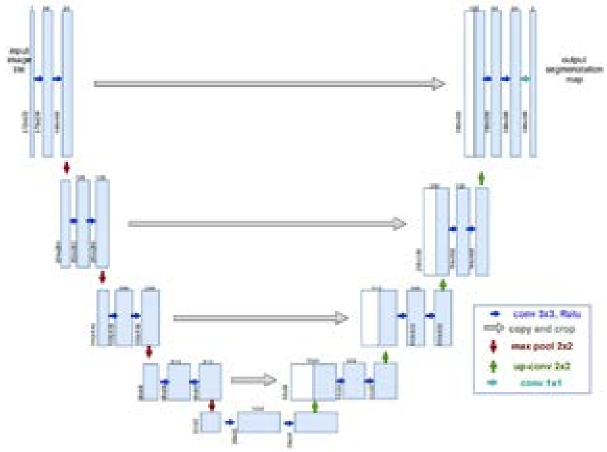

Fig. 1: U-Net Model Architecture.

B. DeepLabV3+ Model Architecture

Chen et al. [17] proposed DeepLabV3+ in 2018, which extends DeepLabv3 to [18] using an encoder-decoder structure. DeepLabV3+ encodes multiscale contextual information while optimizing segmentation results across object boundaries with the decoder module. On the PASCAL VOC 2012 [19] and Cityscapes [20] datasets, DeepLabV3+ outperformed many high-end segmentation networks.

Figure 2 depicts the DeepLabV3+ model architecture.

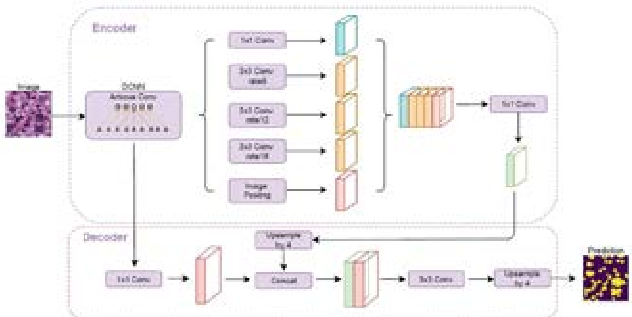

Fig. 2: DeepLabv3+ Model Architecture.

C. SegNet Model Architecture

SegNet, a convolutional encoder-decoder architecture for image segmentation, was developed by Badrinarayanan et al.

[21]. SegNet is a FCN. SegNet is made up of a pixel-based classification layer, an encoder network, and a matching decoder network. In terms of the number of trainable parameters, SegNet is also much less than rival designs.

SegNet model architecture from [21] is shown in Figure 3.

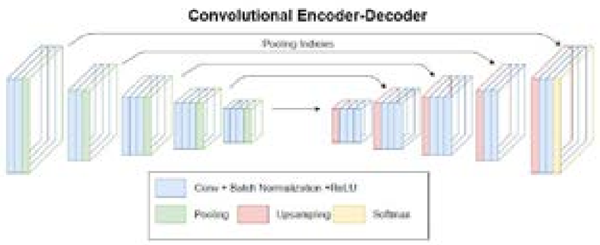

Fig. 3: SegNet Model Architecture.

III. Application Results

A. Experimental Dataset

1) Dataset Size and Nuclei Classes: PanNuke dataset has been generated semi-automatically with 6 different exhaustive nuclei labels over 19 different tissue types for nuclei instance segmentation and classification. The dataset contains 481 visual fields, most of which were selected at random from over 20K whole slide images at various magnifications. In all, there are over 205 thousand labeled nuclei in the dataset, each with its own instance segmentation mask. The labeled nuclei classes are Neoplastic, Inflammatory, Connective, Dead, Epithelial, and Non-Nuclei [22], [23].

2) Dataset Images and Masks: $256 \times 256 \times 3$ RGB images were used as the model input, while 256×256 nuclei masks were used as control labels to evaluate the model output. In fact, the original masks folder contains 6 layers of depth as labeled nuclei classes. However, because we worked on semantic segmentation here, the mask layer we used is the "non-nuclei" layer. In other words, it became the "all-nuclei" layer after pre-processing.

3) Visualization of the Dataset: Within the scope of dataset visualization, an example image is shown in Figure 4, and the mask layers of this image in Figure 5.

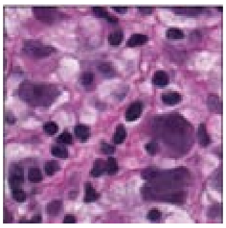

Fig. 4: Original 256×256 RGB sample image.

The selected mask layer for model training and its image after pre-processing is visualized in Figure 6.

B. Results

Using the Python Tensorflow Keras library, the above-mentioned models are used to achieve the best success. In the training phase of this work, models were run between 15 and 50 epochs if the loss was suitable for recovery. In case the loss could not be improved, model training was completed after a few additional epochs and the evaluation and test phase was initiated. After the training phase was completed, the weights of the models with the highest success were reloaded to the models and thus the test phase was completed. The results are shown in Table I.

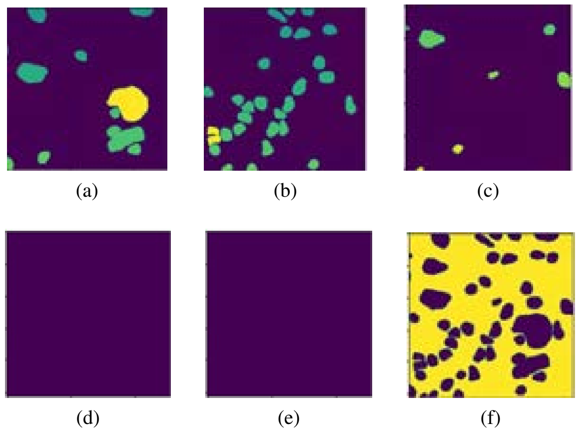

Fig. 5: PanNuke dataset, (a-f) image masks of 256×256 for all 6 layers. For this example please note that there are no nuclei pixels belonging to these classes in layers 4 and 5.

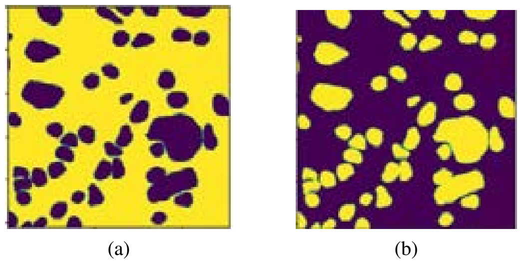

Fig. 6: Belonging to the 6th "non-nuclei" layer selected to be used; (a) raw image, (b) image transformed as "all-nuclei" after pre-processing.

TABLE I: Model based numerical results.

Model	Accuracy	F1/Dice-score	Kappa
DeepLabV3+	0,922	0,802	0,753
U-Net	0,916	0,786	0,667
SegNet	0,812	0,554	0,436

Basically, accuracy measures how many observations are correctly classified, both positive and negative. F1/Dice-score calculates the harmonic mean between the two by combining precision and recall values into a single metric. Cohen Kappa tells how much the model is better than a random classifier that predicts based on class frequencies. As the data set was imbalanced, F1 / Dice-score is preferred as the main evaluation metric here. Besides, the DeepLabV3+ with an F1 / Dice-score of 80.2% was measured as the model with the highest success in this work.

C. Visualization of the Results

A few examples of masks predicted with trained models from the test images of the PanNuke dataset are visualized in Figure 7, Figure 8 and Figure 9. Please remember the first row of all figures is shown as ground truth masks, while the second row is shown as a prediction based on each model.

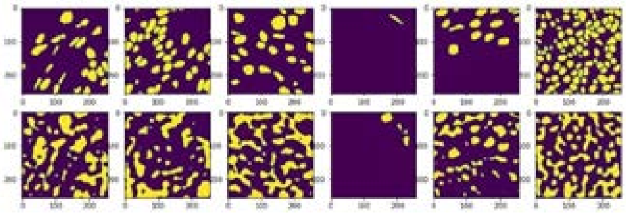

Fig. 7: SegNet Model Predictions.

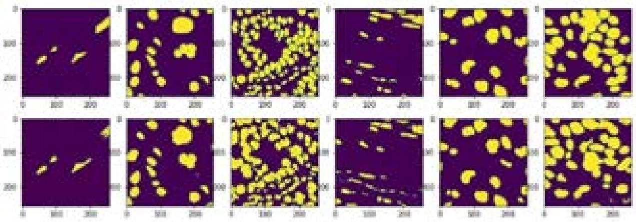

Fig. 8: U-Net Model Predictions.

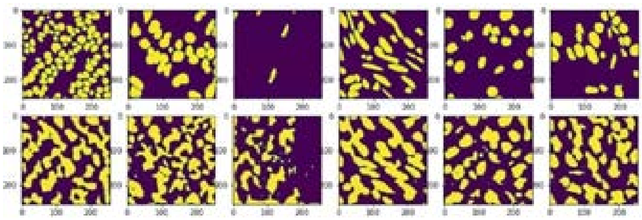

Fig. 9: DeepLabv3+ Model Predictions.

IV. Results And Discussion

In this study, different result values were obtained, as shown in Table I. Although the Overall Accuracy achieved over 90 percent, the F1 / Dice-score was the main preferred metric in the evaluation due to the imbalance of the data set. Because the amount of nuclei pixels in the total of the dataset is about $\frac{1}{5}$ relative to the non-nuclei pixels. F1 / Dice-score 80.2% was considered the best success for this work. Also due to the inconsistency, the dice coefficient metric could not be used as a loss function during model training. By solving these problems, in other words, if we train our model in which we use the F1 / Dice-score in the evaluation phase, with the Dice coefficient loss function, the results will be more reliable, and the success will be higher.

References

[1] V. Wiley and T. Lucas, "Computer vision and image processing: a paper review," *International Journal of Artificial Intelligence Research*, vol. 2, no. 1, pp. 29–36, 2018.

[2] A. Albayrak, A. Ünlü, N. Çalık, G. Bilgin, İ. Türkmen, A. Çakır, A. Çapar, B. U. Töreyin, and L. D. Ata, "Segmentation of precursor lesions in cervical cancer using convolutional neural networks," in *2017 25th Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2017, pp. 1–4.

[3] A. Albayrak, A. U. Akhan, N. Calik, A. Capar, G. Bilgin, B. U. Toreyin, B. Muezzinoglu, I. Turkmen, and L. Durak-Ata, "A whole-slide image grading benchmark and tissue classification for cervical cancer precursor lesions with inter-observer variability," *Medical & Biological Engineering & Computing*, vol. 59, pp. 1545—-561, 2021.

[4] N. Hatipoğlu and G. Bilgin, "Segmentation of histopathological images with convolutional neural networks using fourier features," in *2015 23nd Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2015, pp. 455–458.

[5] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: achievements and challenges," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 582–596, 2019.

[6] D. Karimi, H. Dou, S. K. Warfield, and A. Gholipour, "Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis," *Medical Image Analysis*, vol. 65, p. 101759, 2020.

[7] T. Zhou, S. Ruan, and S. Canu, "A review: Deep learning for medical image segmentation using multi-modality fusion," *Array*, vol. 3, p. 100004, 2019.

[8] E. Goceri, "Challenges and recent solutions for image segmentation in the era of deep learning," in *Int. Conf. on Image Processing Theory, Tools and Applications, IPTA'19*. IEEE, 2019, pp. 1–6.

[9] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," *Medical Image Analysis*, vol. 63, p. 101693, 2020.

[10] S. A. Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: A review," *Artificial Intelligence Review*, vol. 54, no. 1, pp. 137–178, 2021.

[11] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, "A review of semantic segmentation using deep neural networks," *International Journal of Multimedia Information Retrieval*, vol. 7, no. 2, pp. 87–93, 2018.

[12] Y. Hu, Z. Chen, and W. Lin, "Rgb-d semantic segmentation: a review," in *IEEE Int. Conf. on Multimedia & Expo Workshops, ICMEW'18*. IEEE, 2018, pp. 1–6.

[13] F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," *Neurocomputing*, vol. 338, pp. 321–348, 2019.

[14] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *arXiv preprint arXiv:2001.05566*, 2020.

[15] I. O. Sigirci, A. Albayrak, and G. Bilgin, "Detection of mitotic cells in breast cancer histopathological images using deep versus handcrafted features," *Multimedia Tools and Applications*, pp. 1–24, 2021.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI'15*. Springer, 2015, pp. 234–241.

[17] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *European Conference on Computer Vision, ECCV'18*, 2018.

[18] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[19] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, 2015.

[20] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR'16*, 2016, pp. 3213–3223.

[21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[22] J. Gamper, N. A. Koohbanani, K. Benet, A. Khuram, and N. Rajpoot, "Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification," in *European Congress on Digital Pathology, ECDP'19*. Springer, 2019, pp. 11–19.

[23] J. Gamper, N. A. Koohbanani, S. Graham, M. Jahanifar, S. A. Khurram, A. Azam, K. Hewitt, and N. Rajpoot, "Pannuke dataset extension, insights and baselines," *arXiv preprint arXiv:2003.10778*, 2020.