Comparison of Fully Convolutional Networks and U-Net for Optic Disc and Optic Cup Segmentation

Zixiao Jin

College of Engineering, University of California, Santa Barbara, 93107, United States

Abstract. Glaucoma, the leading cause of irreversible blindness, must be diagnosed early and thus treated in time. However, it has no noticeable symptoms in its early stages and may not be detected easily. This paper aims to compare two well-known convolutional neural network (CNN) structures, namely Fully Convolutional Networks (FCNs) and U-Net for the segmentation of the optic disc (OD) and optic cup (OC) from retinal fundus images which play an important role in glaucoma diagnosis. The performance of both models is assessed using qualitative parameters such as the Dice coefficient, Jaccard index, and cup-to-disc ratio (CDR) error. In our experiment, the U-Net model yields more accurate segmentation results with 0.9601 average pixel accuracy and 0.9255 dice score for OD segmentation, outperforming the FCNs model with 0.9560 average pixel accuracy and 0.9132 dice score for OD segmentation. However, FCNs have a shorter inference time of 0. 0043 seconds against U-net's 0. 0062 seconds making FCNs more suitable for real-time applications. The restrictions related to this study include biases from using only one dataset acquired from particular imaging devices, dependency on mask-based cropping techniques, and comparison being restricted to two fundamental architectures. This work presents the contribution of the deep learning models in improving glaucoma screening and therefore helping in avoiding blindness.

1 Introduction

Glaucoma, a progressive optic neuropathy, is the leading cause of irreversible blindness worldwide, which severely lowers patients' living standards [1]. While available treatment can be conducted in the early stages of the disease, its asymptomatic nature makes it hard to detect on time [1]. In 2020, it is estimated that around 52 million people worldwide are affected by primary open-angle glaucoma (POAG), the most common form of glaucoma, among which 43 million cases are undetected [1]. Moreover, this number is projected to increase by 53% per cent by 2040, reaching 67 million undetected cases if no significant improvements in early detection are made [1]. Such an urgent situation underscores the critical need for accurate early detection of glaucoma and timely treatment to reduce the detriment caused by the disease.

Corresponding author: zixiao_jin@ucsb.edu

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

An enlarged optic cup (OC) compared to the optic disc (OD) due to the loss of optic fibers caused by glaucoma is one of the critical indicators used in diagnosing the disease [2]. This relationship is measured by cup-to-disc ratio (CDR), the ratio between the area of the OD and the OC, which tends to increase as glaucoma develops. However, traditional methods of extracting the OD and OC from retinal fundus images require experts to manually draw the regions, which is extremely laborious and subjective. Because of these limitations, a quicker automatic OD and OC segmentation method is needed to address these challenges. Advanced deep learning and computer vision algorithms, especially those utilizing convolutional neural networks (CNNs), have become widely used in various kinds of image classifications and segmentations. These technologies are then introduced to accurately segment the OD and OC from retinal fundus images, enabling the analysis of OD and OC regions, and the diagnosis of glaucoma with minimal human intervention. State-of-the-art models like the M-Net with polar transformation and multi-task deep learning models have shown reliable results in OD, OC segmentation, and glaucoma prediction [3,4]. These models are capable of learning complex features and patterns from retinal fundus images and thus improve the efficiency and accuracy of glaucoma detection remarkably.

This paper focuses on two fundamental CNN architectures in computer vision areas: FCNs and U-Net architecture and employs them to perform OD and OC segmentation tasks for glaucoma detection. The study evaluates the performance of FCNs and U-Net on the segmentation task using retinal fundus images. By analyzing and comparing the accuracies of the predictions and computational efficiency of the two models, this paper tries to investigate the advantages and disadvantages of each model in segmentation tasks, especially in the specific context of glaucoma screening. Additionally, this paper examines the structural differences and explores the reasons behind their performance variations, providing insights that could guide the development of more effective segmentation models for segmentation tasks and glaucoma diagnosis.

2 Data and Method

2.1 Datasets

The study utilized the Retinal Fundus Glaucoma Challenge (EE) dataset, an often-used benchmark for the area, to compare and evaluate FCNs and U-Net models for OD and OC segmentation tasks [5]. This dataset includes 400 2124 × 2056 training images, 400 1634 × 1634 validation images, and 400 1634 × 1634 testing images with OD and OC masks accurately annotated by proficient expert ophthalmologists. Moreover, both glaucoma-affected and healthy retinal images are included in the three sets captured with two different devices to cover various retinal appearances and lighting conditions for comprehensive model evaluation. Figure 1 presents some of the example images and masks from the dataset.

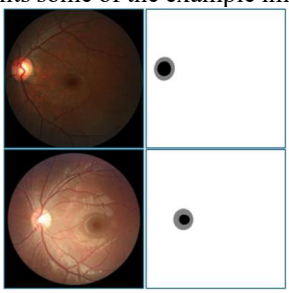

Fig. 1. Retinal fundus images and their corresponding masks from the REFUGE dataset with different lighting conditions (Photo/Picture credit : Original).

2.2 Method

This paper utilizes the FCNs and the U-Net architectures described below. The FCN architectures, introduced by Evan Shelhamer et al., transformed from the traditional CNNs for image classification tasks by replacing the top fully connected layers with convolutional layers as shown below in Figure 2 [6].

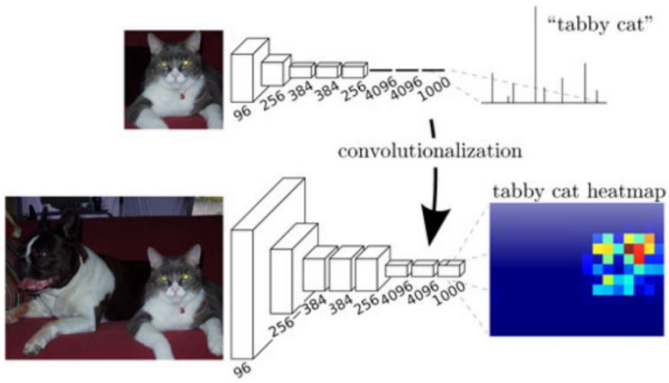

Fig. 2. Transformation of traditional CNNs to FCNs [6].

This strategy enables FCNs to produce outputs with the same spatial dimension as the input, and thus perform pixel-to-pixel dense predictions such as semantic segmentation [6]. Moreover, pre-and post-processing methods like superpixels and patchwise training used in traditional CNNs can be avoided, which greatly improves the computational efficiency of FCNs [6].

The U-Net model proposed by Olaf Ronneberger et al. is a more complex model based on the FCNs [7]. This network has a symmetric architecture with a contracting downsampling path and an expanding upsampling path as shown in Figure 3 [7].

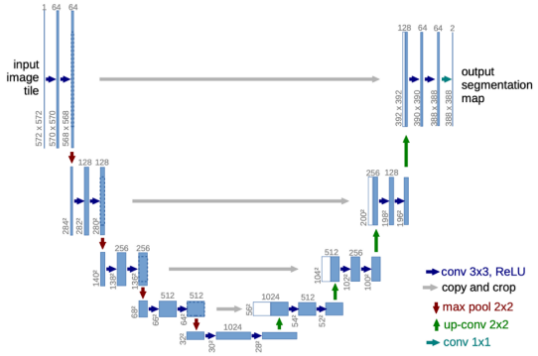

Fig. 3. U-Net Architecture [7].

It can also be observed from the figure that a technique, later called skip connection shown as grey arrows, is also used to concatenate the cropped feature maps from the downsampling path, represented by white blocks, directly with the upsampled blue blocks of the corresponding layers in the upsampling path [7]. This innovation enables U-Net to recapture spatial details from the earlier layers, allowing for more accurate segmentations.

2.3 Implementation details

To train the models, various data augmentation techniques, including flipping, random rotation, and color modification, were applied to the training set and corresponding masks. Flipping and rotation help the models to become effective in various object orientations, while color modification allows the models to generalize better to different lighting conditions [8]. This approach enhances the diversity of the dataset and thus reduces overfitting during the training process and improves model robustness. After the augmentation for the training set, all the images and masks were then padded as needed and cropped to 512 × 512 pixels with OD centered based on the center of mass calculated from the retinal fundus mask. This approach ensures the region of interest is well-positioned for the segmentation task, most of the irrelevant areas are removed and therefore reduces the computational cost.

Both the FCNs model and the U-Net model used in the study were built on pre-trained ResNet-18 as feature extractors for consistency, with the final average pooling layer and the fully connected layer for classification purposes removed [9]. A 1 × 1 convolutional layer was added for the FCNs model, to convert the number of output channels to the number of classes, for classification purposes. A transposed convolutional layer was appended then to enlarge the image back to its original size. For the U-Net model, on the other hand, symmetric upsampling layers with skip connections were constructed. Both networks were implemented using Python with the PyTorch package. All experiments are carried out in the Google Colab environment using an L4 GPU with a memory of 22.5 GB. The Adam optimizer with an initial learning rate of 1 × 10-4 and a batch size of 16 for the training set with 120 epochs was used for both training processes.

The loss function used was calculated from the Dice score for each class:

$$\mathrm{Dice_i} = \frac{2\times\sum(\mathrm{pred_i}\cdot\mathrm{true_i})+\varepsilon}{\sum\mathrm{pred_i}+\sum\mathrm{true_i}+\varepsilon} \tag{1}$$

Where Dicei, predi, and truei each stand for the Dice score, predicted probability, ground truth probability of class i, and ε is a smoothing constant, set as 1 × 10-6, to avoid division by zero. A macro-average was then taken to calculate the final loss:

$$\mathrm{L} = 1 - \frac{\sum_{i=1}^{n}\mathrm{Dice_i}}{n} \tag{2}$$

where n is the number of classes. In the study, the background class is ignored, and OC, despite having a different class label from OD, was considered a part of OD when calculating the Dice score for the OD class, since the OD region overlaps the OC region.

During the training processes, the models with the best performance based on the loss observed from the validation set were saved to prevent overfitting, and global pixel accuracy for both the training and the validation set was also recorded. After the training processes, the OD and OC segmentation performances were measured on the testing set separately based on popular evaluation metrics, including Dice coefficient (F-Measurement), Jaccard (overlapping), accuracy, sensitivity, and specificity defined below:

$$\mathrm{Dice} = \frac{2\times\mathrm{tp}}{2\times\mathrm{tp}+\mathrm{fp}+\mathrm{fn}} \tag{3}$$

$$\mathrm{Jaccord} = \frac{\mathrm{tp}}{\mathrm{tp}+\mathrm{fp}+\mathrm{fn}} \tag{4}$$

$$\mathrm{Accuracy} = \frac{\mathrm{tp}+\mathrm{tn}}{\mathrm{tp}+\mathrm{tn}+\mathrm{fp}+\mathrm{fn}} \tag{5}$$

$$\mathrm{Sensitivity} = \frac{\mathrm{tp}}{\mathrm{tp}+\mathrm{fn}} \tag{6}$$

$$\text{Specificity} = \frac{\text{tn}}{\text{tn+fp}} \tag{7}$$

where tp, tn, fp, and fn stand for true positive, true negative, false positive, and false negative [10]. Moreover, the error in CDR is also calculated as:

$$E_{CDR} = |\text{pred}_{CDR} - \text{true}_{CDR}| \tag{8}$$

where predCDR and TrueCDR stand for the CDR of the prediction and the CDR of the ground truth. Since CDR is a critical factor in glaucoma diagnosis.

3 Results

The loss and accuracy updates of FCNs and U-Net are shown in Figure 4 and Figure 5 respectively. For both the FCNs model and the U-Net model, the training loss decreases rapidly during the first 20 epochs and convergences during the 40 to 50 epochs, and the training global pixel accuracies increase steadily during the first 10 and 15 epochs and stabilizes at around 20 epochs. The validation loss and accuracy for both models show the same overall trend, however, more oscillation and instability can be seen during the validation process for the FCNs model.

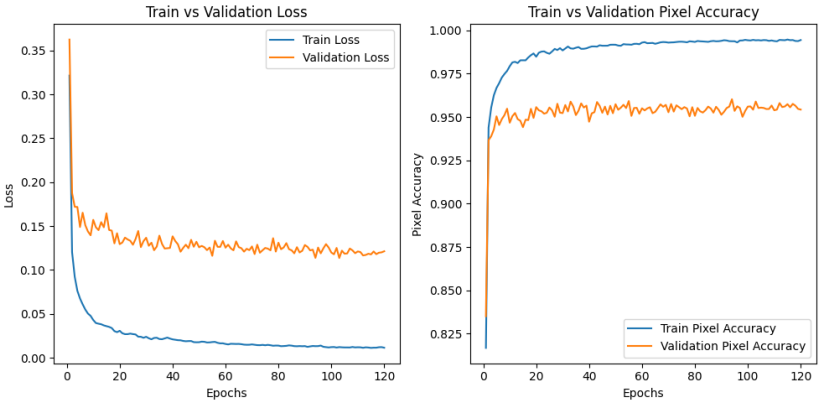

Fig. 4. Loss and accuracy assessment results for the FCN model(Photo/Picture credit : Original).

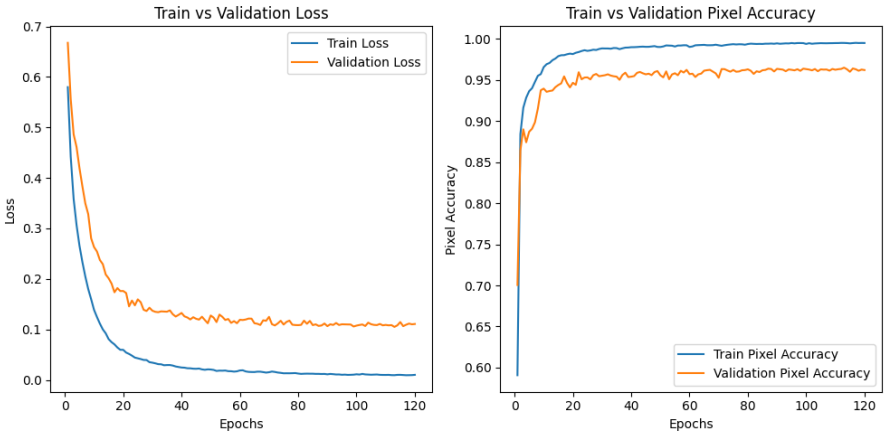

Fig. 5. Loss and accuracy assessment results for the U-Net model(Photo/Picture credit : Original).

As for model complexity shown below in Table 1, the FCNs model has fewer parameters (11.21 million) and floating point operations (FLOPs) (38.4 billion) compared to the U-Net model, which has 14.33 million parameters and 43.6 billion FLOPs. This results in a faster training time, inference time (the time cost to make a single prediction), and significantly less memory usage for the FCNs model.

Table 1. Model Complexity for the FCNs model and the U-Net model.

Model	Parameters	FLOPs	Training time	Memory usage	Convergence epochs	Inference time
FCNs	11.21M	38.4B	37.67 min	2.44 GB	42	0.0043 sec
U-Net	14.33M	43.6B	42.81 min	6.34 GB	48	0.0062 sec

Figure 6 below shows the segmentation results of a sample image from the testing set using the two models, where the black region represents OC and the grey region together with the black region represents OD. By comparing Figure 6 (a) and Figure 6 (b), it can be observed that both models can locate the OC and OD regions. However, both models failed to segment the OC boundary accurately. For the OD segmentation, on the other hand, it seems that U-Net performed a bit better by producing output with shape and area closer to the true mask.

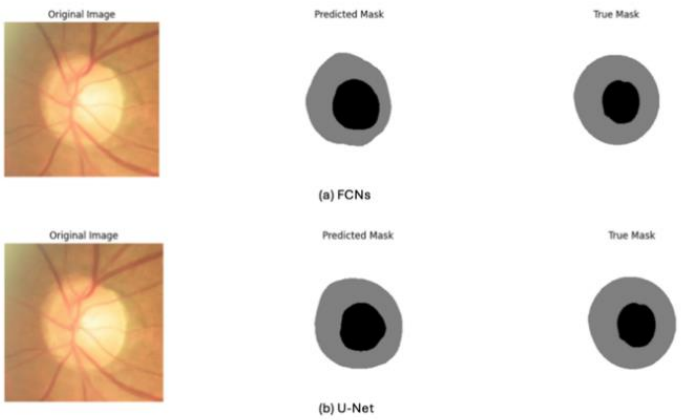

Fig. 6. Example segmentation results from the FCNs model (a) and the U-Net model (b)(Photo/Picture credit : Original).

Table 2 below presents the performance metrics for both the FCNs model and the U-Net model. The U-Net model outperforms the FCNs model in overall global pixel accuracy (0.9601 ± 0.0029 vs. 0.9560 ± 0.0045) and OD segmentation with higher Dice coefficient (0.9255 ± 0.0068 vs. 0.9132 ± 0.0111), Jaccard index (0.8613 ± 0.0118 vs. 0.8403 ± 0.0187), accuracy (0.9744 ± 0.0021 vs. 0.9697 ± 0.0038), and specificity (0.9731 ± 0.0025 vs. 0.9673 ± 0.0045). However, the FCNs model does perform slightly better than the U-Net model in terms of OC segmentation with a bit higher Dice coefficient (0.8482 ± 0.0206 vs. 0.8406 ± 0.0210), Jaccard index (0.7364 ± 0.0306 vs. 0.7250 ± 0.0312), and sensitivity scores (0.8948 ± 0.0306 vs. 0.8785 ± 0.0333). Both the FCNs model and the U-Net model perform quite well in CDR prediction, with CDR errors of 0.0156 and 0.0140 respectively, which demonstrate their strong performance on glaucoma-related predictions.

Table 2. Model Performance for the FCNs model and the U-Net model.

Metric	FCNs	U-Net
Global pixel accuracy	0.9560 ± 0.0045	0.9601 ± 0.0029
OC		
-Dice	0.8482 ± 0.0206	0.8406 ± 0.0210
-Jaccard	0.7364 ± 0.0306	0.7250 ± 0.0312
-Accuracy	0.9862 ± 0.0018	0.9856 ± 0.0019
-Sensitivity	0.8948 ± 0.0306	0.8785 ± 0.0333
-Specificity	0.9903 ± 0.0019	0.9904 ± 0.0016
OD		
-Dice	0.9132 ± 0.0111	0.9255 ± 0.0068
-Jaccard	0.8403 ± 0.0187	0.8613 ± 0.0118
-Accuracy	0.9697 ± 0.0038	0.9744 ± 0.0021
-Sensitivity	0.9822 ± 0.0062	0.9807 ± 0.0054
-Specificity	0.9673 ± 0.0045	0.9731 ± 0.0025
CDR error	0.0156 ± 0.0109	0.140 .0105

4 Discussions

As an important step in glaucoma diagnosis, the performance of FCNs and U-Net architectures for OD and OC segmentation tasks are compared. Based on the experimental outcomes, it can be seen that FCNs perform slightly better in OC segmentation, but in general, U-Net produces more accurate and consistent results in pixel accuracy globally across all classes as well as OD segmentation specifically. This overall improvement could be due to the use of innovative skip connections in the expanding path of U-Net, which helps U-Net avoid spatial feature loss in the deeper layers and therefore produces more accurate predictions.

However, although U-Net yielded more accurate results, FCNs performed computations faster. Because of this advantage, FCNs are more suited for real-time applications where speed is of the essence, like large-scale screenings or in situations where segmentation and instance detections must be completed quickly. The balance between speed and accuracy is important depending on the clinical settings.

Despite all the findings, there are some limitations to the current study. The REFUGE dataset is the only dataset used in the experiment where all the training images came from the Zeiss Visucam 500 device, and all the validation and testing images were taken by the

Canon CR-2 device [5]. This could impact the models' performance as images captured by the same device share similar characteristics, which could have introduced biases to the models and influenced the models' generalization abilities. Also, the cropping technique used in the data preprocessing part depends on the provided mask and thus can not be employed directly in real-world applications. Moreover, the comparison was only made between the two fundamental architectures. Future work could include images with various resolutions from different devices, automatic optic nerve region detection and cropping techniques, and hybrid multi-task learning architectures that combine the two models and more advanced preprocessing techniques in the area, such as polar transformation [3].

5 Conclusions

This paper compared FCNs and U-Net structures for the segmentation of OD and OC from retinal fundus images in glaucoma diagnosis. The analysis of the proposed models was done using the REFUGE dataset, alongside ResNet-18 as feature extractors, and Dice coefficient, Jaccard index, accuracy, sensitivity, specificity, and CDR error as evaluation metrics. The studies revealed that the results obtained by U-Net were better in terms of overall pixel accuracy and OD segmentation compared to FCNs, this could be attributed to the skip connection which helps in preserving the spatial details. However, as can be seen from the analysis of the experiment results, the proposed FCNs achieved slightly higher accuracy in OC segmentation while requiring faster computations and thus more appropriate for real-time applications. Moreover, both models are capable of producing results with satisfactory average CDR error in the prediction of glaucoma cases.

For future improvements, datasets from other imaging devices to enhance the generalizability of the models and automatic identification and cropping techniques of the OD regions could be included. More advanced hybrid multi-task learning models could also be proposed to further improve segmentation performance and glaucoma detection. These findings highlight the potential of deep learning methods in glaucoma screening and thus help eliminate blindness and its associated detriments caused by the disease.

References

1. Z. Soh, M. Yu, B. K. Betzler, S. Majithia, S. Thakur, Y. C. Tham, T. Y. Wong, T. Aung, D. S. Friedman, C.-Y. Cheng, Ophthalmology, vol. 128, no. 10, pp. 1393-1404, 2021
2. M. A. Fernandez-Granero, A. Sarmiento, D. Sanchez-Morillo, S. Jiménez, P. Alemany, I. Fondón, Journal of Healthcare Engineering, vol. 2017, no. 1, p. 5953621, 2017
3. H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, X. Cao, IEEE Transactions on Medical Imaging, vol. 37, no. 7, pp. 1597-1605, 2018
4. L. Pascal, O. J. Perdomo, X. Bost, B. Huet, S. Otálora, M. A. Zuluaga, Scientific Reports, vol. 12, no. 1, p. 12361, 2022
5. J. I. Orlando, H. Fu, J. B. Breda, K. Van Keer, D. R. Bathula, A. Diaz-Pinto, R. Fang, Medical Image Analysis, vol. 59, p. 101570, 2020
6. J. Long, E. Shelhamer, T. Darrell, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431-3440
7. O. Ronneberger, P. Fischer, T. Brox, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III, vol. 18, Springer International Publishing, 2015, pp. 234-241

8. E. Goceri, Artificial Intelligence Review, vol. 56, no. 11, pp. 12561-12605, 2023
9. K. He, X. Zhang, S. Ren, J. Sun, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778
10. B. Al-Bander, B. M. Williams, W. Al-Nuaimy, M. A. Al-Taee, H. Pratt, Y. Zheng, Symmetry, vol. 10, no. 4, p. 87, 2018