2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

# Multi-task Learning Based Ocular Disease Discrimination and FAZ Segmentation Utilizing OCTA Images

Zhonghua Wang[1,†], Li Lin[1,†], Jiewei Wu[1] and Xiaoying Tang[1,*]

*Abstract*— **In this paper, we proposed and validated a multi-task based deep learning method for simultaneously segmenting the foveal avascular zone (FAZ) and classifying three ocular disease related states (normal, diabetic, and myopia) utilizing optical coherence tomography angiography (OCTA) images. The essential motivation of this work is that reliable predictions on disease states may be made based on features extracted from a segmentation network, by sharing a same encoder between the classification network and the segmentation network. In this study, a cotraining network structure was designed for simultaneous ocular disease discrimination and FAZ segmentation. Specifically, we made use of a classification head following a segmentation network's encoder, so that the classification branch used the feature information extracted in the segmentation branch to improve the classification results. The performance of our proposed network structure has been tested and validated on the FAZID dataset, with the best Dice and Jaccard being 0.9031±0.0772 and 0.8302±0.0990 for FAZ segmentation, and the best Accuracy and Kappa being 0.7533 and 0.6282 for classifying three ocular disease related states.**

*Clinical Relevance*— **This work provides a useful tool for segmenting FAZ and discriminating three ocular disease related states utilizing OCTA images, which has a great clinical potential in ocular disease screening and biomarker delivering.**

## I. INTRODUCTION

Optical coherence tomography angiography (OCTA) is a diagnostic imaging technique based on optical principles. It makes use of the blood flow information of the retinal vasculature, displaying blood vessels from the inner limiting membrane layer to the choroid layer at the capillary level [1]. The foveal avascular zone (FAZ) is a capillary-free area in the center of the macula, which has received significant research interest in fluorescein angiographic analysis [2]. FAZ's functionality in vision has been investigated, and its size has been identified to be related to visual ability.

The size and shape of the FAZ in three ocular diseases related states, namely normal, diabetic, and myopia, have been intensively investigated. The FAZ area has shown statistically significant enlargement in diabetic eyes compared to healthy ones [3]. Similarly, in myopic eyes, especially those with a high degree, enlarged FAZ areas have been identified, with relatively reduced blood vessel diameters [4]. In such context, the FAZ morphology may provide essential biomarkers for diabetic and myopia, which has great potential for clinical utility, especially when using OCTA images (non-invasiveness, fast scanning speed, and high imaging resolution). A necessary prerequisite for FAZ morphology quantification is to have the FAZ region segmented out from OCTA images.

Manually delineating the FAZ region is subjective and tedious. Furthermore, due to the fuzzy edge information in OCTA, it may lead to intra- and inter-variability between experts. As such, automated FAZ segmentation methods are needed. A representative automatic FAZ segmentation algorithm employed vascular edge identification and morphological closure, and the high correlation between manual extracted and automated extracted area sizes identifies its effectiveness [5]. However, its performance may be severely affected by the presence of vessels. Deep learning-based methods may possess better robustness and reliability, especially in biomedical image segmentation tasks [6].

In clinical practice, the OCTA images can be used for not only FAZ segmentation but also ocular disease discrimination [4] [7]. However, relevant work has been relatively rare in utilizing OCTA images for automatically discriminating three ocular disease related states, namely normal, diabetic, and myopia.

Recently, multi-task learning has been found to achieve superior performance in joint segmentation and classification tasks, including such tasks on histopathology images [8] and fundus images [9]. The reported superior performance on data of different types has also revealed its data generalization ability. Multi-task learning can effectively explore the commonness and difference between different tasks.

In this paper, we propose a novel multi-task cotraining network for joint FAZ segmentation and ocular disease classification. We adopt an encoder-decoder structure for our segmentation architecture, and a classification head is added right after the encoder to predict disease states. We validated the effectiveness of our proposed framework by comparing its performance with that of each single task, namely conducting segmentation and classification separately.

## II. METHOD

Our overall framework builds upon a segmentation architecture together with a classification head. The baseline segmentation network can be any one with an encoder-decoder structure, such as U-Net [10], U-Net++ [11] and Deeplabv3+ [12]. Each baseline segmentation network is also

This study was supported by the National Natural Science Foundation of China (62071210), the Shenzhen Basic Research Program (JCYJ20190809120205578), the National Key R&D Program of China (2017YFC0112404), and the High-level University Fund (G02236002). The authors declare that they have no competing financial interests.

[†] Both authors contribute equally to this work.

[*] Correspondence to Dr. Xiaoying Tang: tangxy@sustech.edu.cn

[1] Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

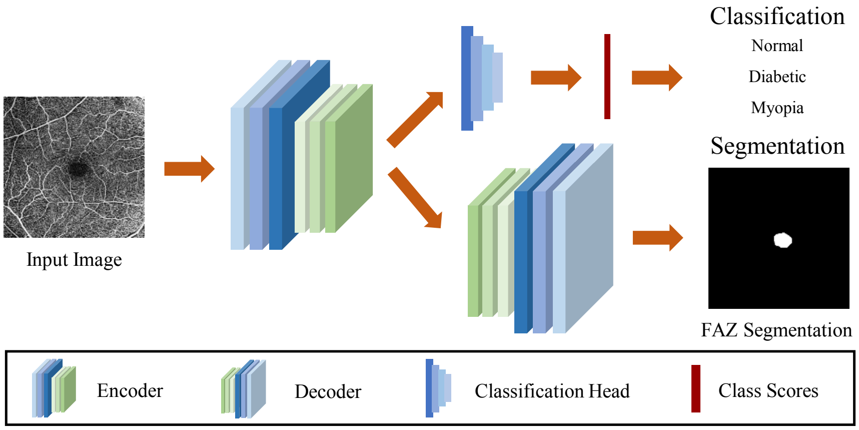

Fig. 1. The overall framework of our proposed cotraining approach.

used for the single-task segmentation. The classification head is connected after the encoder of the segmentation network. For single-task classification, we choose major image recognition networks including ResNet50 [13], ResNext50 [14] and ResNest50 [15] for verfication. The overall framework is shown in Figure 1.

### A. *Classification Head*

The classification head is a module following the encoder of the segmentation network, designed specifically for classification purpose. The last several layers that output the predicted categories in a recognition network, such as ResNet, consist of Global Average Pooling, Flatten, Linear, and Activation layers. Our original classification head is also designed based on this structure. However, the classification head for the recognition network should use features extracted from the encoder. For the cotraining network, the features extracted from the encoder need to accommodate the segmentation task, which might not be suitable for the original classification head. As such, we propose a new classification head for the cotraining network, which consists of one Global Average Pooling, one Flatten, two Max Pooling, three Linear, and one Activation layers. Compared to the original design, the proposed classification head has more robust generalization and recognition abilities.

### B. *Multi-task Learning Architecture*

For the segmentation task, the encoder focuses on features of the FAZ region, including size, shape, and position. These features are also useful for classification of our three ocular disease related states. We add the classification head to the end of the encoder to make categorical predictions. To exploit the feature extraction ability, we replace the original encoder of the segmentation network with the feature extraction part of the recognition network. This creates the new encoder of the cotraining network. The combination of the encoder and the classification head builds up the main structure for classification. The decoder of the network remains the same as that used in the baseline segmentation network. In this way, the classification process is able to use the feature information for segmentation. Meanwhile, the segmentation process also utilizes the between-category differences used for optimizing the classification process. Compared to each single task, multi-task learning makes better use of complementary information by accommodating two tasks simultaneously.

During the multi-task learning process, we initiate the classification process after the segmentation process converges to avoid over-weighting of the encoder. The objective function for cotraining consists of two losses: Dice loss for segmentation, represented as $L_{seg}$, and Cross-entropy loss for classification, represented as $L_{cls}$. The total loss is a weighted sum of the aforementioned two losses:

$$L_{total} = \alpha L_{seg} + \beta L_{cls}, \tag{1}$$

where $\alpha$ and $\beta$ are two parameters that adjust the relative weights. The initial values are set to be $\alpha = 1$ and $\beta = 0$ to make the segmentation process converge. Then the weights are changed to be $\alpha = 0.5$ and $\beta = 0.5$ to update both processes.

### C. *Evaluation Metrics*

We choose Dice and Jaccard to be our evaluation indices for segmentation. For the classification task, the most commonly used evaluation index is accuracy, representing the proportion of correct predictions. However, the overall prediction does not take into account data imbalance. To avoid this issue, we employ Kappa [16] as another evaluation index for classification.

## III. Results

### A. *Dataset*

Our proposed method was evaluated on the Foveal Avascular Zone Image Database (FAZID) [17]. This dataset contains 304 OCTA images, including 88 normal images, 107 diabetic images, and 109 myopia images, with corresponding ground truth FAZ segmentation labels. All images were divided into 60% training data, 20% validation data, and 20% testing data through five-fold cross-validation.

The original image size is 416×416, with a physical size of 6mm×6mm. Since the peripheral region might contain noise, we cropped each image to be a physical size of 3mm×3mm from the center. This 3mm×3mm size is another image size that is also commonly used in OCTA images. All images were resized to be 192×192 after center cropping. Data augmentation, including affine (with scale=1) and random flipping were applied afterward, resulting in a total of 39050 images.

### B. Implementation

The proposed pipeline was implemented by Segmentation Models PyTorch [18] on a workstation equipped with RTX 3090. The batch size during training was set to be 32. We used Adam to be our optimizer with a learning rate of 0.0001. All networks were trained for 100 epochs in total, with initial parameters pretrained on ImageNet. The classification process was set to start updating after 30 epochs to ensure the segmentation encoder had converged. The loss function chosen for the training process was Dice loss for the single-task segmentation and Cross-entropy for the single-task classification.

### C. Experimental Results

Tables 1 and 2 respectively tabulate quantitative comparisons of the proposed cotraining network with the single-task classification network and the single-task segmentation network. The "architecture" represents the segmentation network and the "encoder" indicates the recognition network that has been used to replace the original encoder of the segmentation network. We nevertheless used the original encoder for the single-task segmentation. ResNest50 was not implemented as the "encoder" when the "architecture" was Deeplabv3+ since ResNest did not support dilated convolutions.

It can be observed that the proposed method performed better than each single task approach, regardless of the architecture and the encoder chosen. Collectively considering the segmentation performance and the classification performance, the architecture being U-Net or Deeplabv3+ and the encoder being ResNet50 performed the best. However, the three-category classification task is more challenging than the FAZ segmentation task both manually and automatically, as can be also seen from the numbers listed in the two tables, we chose U-Net as the architecture and ResNet50 as the encoder to be our finally-identified cotraining network since such a combination provided the best classification performance. In other words, we consider Accuracy and Kappa to be more critical when evaluating the entire pipeline since they are more sensitive than the two segmentation metrics, as can be seen from their respective improvement degrees. This cotraining network obtained a 9.57% increase in Accuracy and a 19.14% increase in Kappa compared to the best single-task classification results, indicating the effectiveness of the proposed method for classification. Meanwhile, by adopting the proposed method, Dice also increased by 0.0115, and Jaccard increased by 0.019 over the best single-task segmentation results. These improvements suggest that the FAZ segmentation provides structure information that helps improve the classification. On the other hand, the category information helps the network have a better feature identification ability and thus enhances the FAZ segmentation performance.

TABLE I

COMPARISONS OF CLASSIFICATION RESULTS BETWEEN THE PROPOSED COTRAINING NETWORK AND THE SINGLE-TASK CLASSIFICATION NETWORK.

| | Classification Network | | Accuracy | Kappa |
|---|---|---|---|---|
| Single-Task | ResNet50 | | 0.6875 | 0.5273 |
| | ResNext50 | | 0.6711 | 0.502 |
| | ResNest50 | | 0.6382 | 0.4508 |
| | Architecture | Encoder | Accuracy | Kappa |
| Proposed | U-Net | ResNet50 | **0.7533** | **0.6282** |
| | | ResNext50 | 0.7336 | 0.596 |
| | | ResNest50 | 0.7231 | 0.5814 |
| | U-Net++ | ResNet50 | 0.7368 | 0.6019 |
| | | ResNext50 | 0.7434 | 0.6113 |
| | | ResNest50 | 0.7434 | 0.6125 |
| | Deeplabv3+ | ResNet50 | 0.7532 | 0.6265 |
| | | ResNext50 | 0.7531 | 0.6228 |
| | | ResNest50 | - | - |

TABLE II

COMPARISONS OF SEGMENTATION RESULTS BETWEEN THE PROPOSED COTRAINING NETWORK AND THE SINGLE-TASK SEGMENTATION NETWORK.

| | Segmentation Network | | Dice | Jaccard |
|---|---|---|---|---|
| Single-Task | U-Net | | 0.8916±0.0726 | 0.8112±0.1041 |
| | U-Net++ | | 0.8898±0.0789 | 0.8093±0.1112 |
| | Deeplabv3+ | | 0.8810±0.0831 | 0.7956±0.1124 |
| | Architecture | Encoder | Dice | Jaccard |
| Proposed | U-Net | ResNet50 | 0.8983±0.0660 | 0.8211±0.0957 |
| | | ResNext50 | 0.8959±0.0798 | 0.8187±0.1029 |
| | | ResNest50 | **0.9031±0.0772** | **0.8302±0.0990** |
| | U-Net++ | ResNet50 | 0.8942±0.0825 | 0.8164±0.1064 |
| | | ResNext50 | 0.8932±0.0862 | 0.8153±0.1102 |
| | | ResNest50 | 0.8972±0.0816 | 0.8211±0.1054 |
| | Deeplabv3+ | ResNet50 | 0.9010±0.0671 | 0.8256±0.0949 |
| | | ResNext50 | 0.8978±0.0735 | 0.8213±0.1007 |
| | | ResNest50 | - | - |

During the training process, the performance of the proposed method in segmentation reached its maximum before the classification process started, and then a slight decrease appeared. This could be explained that the encoder needed to fit and balance two tasks simultaneously after initiating the classification process.

Representative examples of the FAZ segmentation results from different networks are shown in Figure 2. Clearly, the results of the proposed network contain less error information and are more correct on the boundary. We conjecture from such comparisons that with an addition of the classification branch, the encoder gained more useful boundary information which boosted the segmentation results.

## IV. CONCLUSION

In this paper, we proposed and validated a cotraining network for segmenting the FAZ region and discriminating three ocular disease related states from OCTA images. Compared to a single-task classification network or a single-task seg-

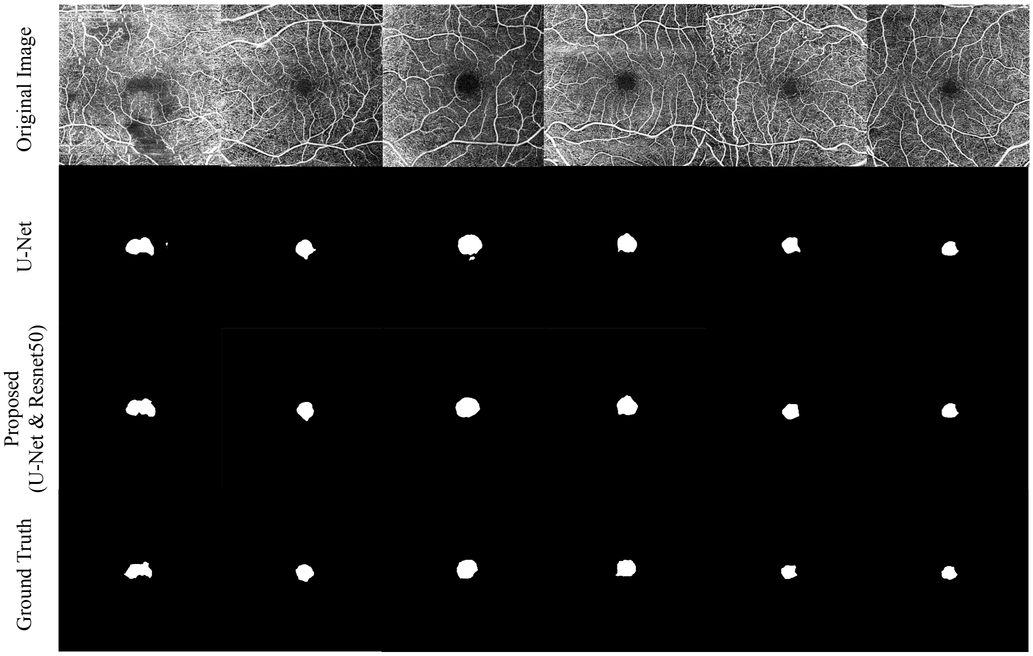

Fig. 2. Representative examples of the FAZ segmentation results, obtained from the single-task segmentation network (U-Net), the cotraining network (U-Net & Resnet50), and the manual delineation (ground truth).

mentation work, the proposed pipeline achieved significant improvements, especially for the classification task.

Our work still has limitations. With the classification branch being introduced to the cotraining network, the performance of the segmentation process decreased slightly since the encoder needed to support two tasks simultaneously. This might be improved by introducing a new encoder to the overall network architecture, employing a shared weight with the decoder for separated weight optimization. Without the interference of the classification task, the segmentation performance may not decrease. This will be our future research plan. Comparisons with other related works are not included in this paper due to dataset difference, which will nevertheless be explored in our future research work.

## References

[1] T. E. De Carlo, A. Romano, N. K. Waheed, and J. S. Duker, "A review of optical coherence tomography angiography (octa)," *International journal of retina and vitreous*, vol. 1, no. 1, p. 5, 2015.

[2] J. Conrath, R. Giorgi, D. Raccah, and B. Ridings, "Foveal avascular zone in diabetic retinopathy: quantitative vs qualitative assessment," *Eye*, vol. 19, no. 3, pp. 322–326, 2005.

[3] N. Takase, M. Nozaki, A. Kato, H. Ozeki, M. Yoshida, and Y. Ogura, "Enlargement of foveal avascular zone in diabetic eyes evaluated by en face optical coherence tomography angiography," *Retina*, vol. 35, no. 11, pp. 2377–2383, 2015.

[4] J. J. Balaji, A. Agarwal, R. Raman, and V. Lakshminarayanan, "Comparison of foveal avascular zone in diabetic retinopathy, high myopia, and normal fundus images," in *Ophthalmic Technologies XXX*, vol. 11218, p. 112181O, International Society for Optics and Photonics, 2020.

[5] M. Díaz, J. Novo, P. Cutrín, F. Gómez-Ulla, M. G. Penedo, and M. Ortega, "Automatic segmentation of the foveal avascular zone in ophthalmological oct-a images," *PLoS One*, vol. 14, no. 2, p. e0212364, 2019.

[6] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[7] D. Le, M. Alam, C. K. Yao, J. I. Lim, Y.-T. Hsieh, R. V. Chan, D. Toslak, and X. Yao, "Transfer learning for automated octa detection of diabetic retinopathy," *Translational Vision Science & Technology*, vol. 9, no. 2, pp. 35–35, 2020.

[8] H. Qu, G. Riedlinger, P. Wu, Q. Huang, J. Yi, S. De, and D. Metaxas, "Joint segmentation and fine-grained classification of nuclei in histopathology images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 900–904, IEEE, 2019.

[9] A. Chakravarty and J. Sivswamy, "A deep learning based joint segmentation and classification framework for glaucoma assesment in retinal color fundus images," *arXiv preprint arXiv:1808.01355*, 2018.

[10] 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*, pp. 234–241, Springer, 2015.

[11] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pp. 3–11, Springer, 2018.

[12] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818, 2018.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[14] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1492–1500, 2017.

[15] H. Zhang, C. Wu, Z. Zhang, Y. Zhu, Z. Zhang, H. Lin, Y. Sun, T. He, J. Mueller, R. Manmatha, *et al.*, "Resnest: Split-attention networks," *arXiv preprint arXiv:2004.08955*, 2020.

[16] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia medica: Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.

[17] A. Agarwal, R. Raman, V. Lakshminarayanan, *et al.*, "The foveal avascular zone image database (fazid)," in *Applications of Digital Image Processing XLIII*, vol. 11510, p. 1151027, International Society for Optics and Photonics, 2020.

[18] P. Yakubovskiy, "Segmentation models pytorch." https://github.com/qubvel/segmentation_models.pytorch, 2020.