Intra-retinal cyst segmentation in optical coherence tomography images via transfer learning

Lavsen Dahal (ld258), Fong Chi Ho (fh82), Zion Sheng (zs144)
Dept. of Electrical and Computer Engineering
Duke University

Abstract

Diabetic macular edema (DME) is a complication of diabetic retinopathy, characterized by the retina cysts imaged with Optical Coherence Tomography (OCT). The development of machine learning improved the automatic retinal cyst segmentation accuracy, but a further improvement is hampered by the limited medical data size. In this study, three transfer learning models pre-trained in a different domain were compared and evaluated across various architectures and configurations for retinal cyst segmentation. The models were trained and tested on Duke OCT and UMN OCT dataset. Swin model pre-trained on scene-centric data with unfrozen backbone showed the best performance with data augmentation of random crop and rotation, achieving a mean Dice score of 0.41 on Duke OCT dataset and 0.81 on UMN OCT dataset. This illustrated the potential of transfer learning in domain-shifted segmentation tasks, with limited OCT data for retinal cyst segmentation training. Code available at `https://github.com/zs144/23fa-ece661-final-proj`.

1 Introduction

Fluid-filled retinal cysts are recognized as hallmarks of several ocular diseases, such as age-related macular degeneration (AMD), ocular inflammation, and diabetes mellitus (DM), which is a widespread sight-threatening retinopathy [1]. DM-induced diabetic retinopathy, especially Diabetic Macular Edema (DME), significantly risks vision loss and requires early detection to prevent blindness [2]. Optical Coherence Tomography (OCT) is pivotal for identifying DME by revealing the microstructure of retina and the potential cysts in high resolution. [3].

However, the manual retinal cyst segmentation has high variability due to doctor subjectivity, time constraints, and low repeatability. To address these issues, multiple automated cyst segmentation algorithms has been proposed using Deep Learning (DL). To address the limited performance due to limited OCT dataset for training, transfer learning was proposed in this study to refine the pre-existing DL models trained in another domain, to improve the efficiency and accuracy of OCT cyst segmentation, thereby contributing to more effective DME treatment planning and prognosis assessment.

2 Related works

Multiple studies have focused on segmenting retinal fluid from OCT images, but it remains a challenging task due to image noise and historically low segmentation accuracy reported in the literature. Stephanie J. Chiu identified the retinal layers and fluid in OCT images by a kernel-based regression classifier followed by a graph theory dynamic programming framework, achieving a dice coefficient of 0.53 [4]. Cecilia S. Lee segmented the fluid-filled spaces in OCT images using a convolutional neural network (CNN) with 18 convolutional layers modified from the U-net autoencoder architecture, resulting in a mean dice score of 0.728 [5]. Thomas Schlegl presented a semantic segmentation-based method using autoencoder to analyze intraretinal and subretinal fluid

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

segments of AMD, DME, and retinal vein occlusion (RVO), respectively and presented an average area under the curve (AUC) of 0.94 [6]. Abhijit G. Roy segmented the retinal cyst from the macular OCT images with a fully convolutional network (ReLayNet) that used autoencoder structure and a jointed logistic regression and Dice overlap loss functions, achieving a Dice coefficient of 0.77 [7].

Taimur Hassan developed a deep learning approach that utilized CNN and structure tensor graph search-based segmentation framework (CNN-STGS) to segment and identify the location of retinal layers and cysts from the scan, showing a Dice score of 0.906 [8]. Zhenhua Wang proposed a segmentation of ME in OCT images based on the transfer learning of DeepLab framework (OCT-DeepLab). ResNet-101 and Atrous Spatial Pyramid Pooling (ASPP) were implemented to extract major and scale features respectively. The features were used for segmentation using the DeepLab model, achieving an AUC of 0.963 [9]. Xiaoming Liu segmented the multi-class retinal fluid regions using a semi-supervised segmentation with GAN Network (SGNet). The model involved a U-Net alike segmentaation model comprising multi-class cross entropy, dice loss, adversarial loss and the semi-supervised loss in training, and another U-Net alike discriminator network with entropy loss function. This generative model achieved a Dice score of 0.803 [10]. Da Ma segmented the DME fluid regions using a dual-branch deep neural network (LF-UNet), which replaced the encoder of the U-Net with a dilated network, and combined it with a cascaded network framework for anatomical awareness. This model achieved a Dice score of 0.568 [11].

Despite the prevalent use of pre-trained networks to enhance training efficiency and segmentation accuracy in various models, the impact of transfer learning on OCT retinal fluid segmentation has not been extensively explored. Given the limited availability of OCT image datasets, this paper aims to investigate and compare the efficacy of transfer learning using multiple pre-trained models for the precise segmentation of retinal fluid.

3 Methodology

3.1 Data

In this work, we used two datasets, Duke OCT and UMN OCT to evaluate the transfer learning on medical imaging datasets. Duke OCT includes ten DME subjects (patients) with lateral and azimuthal resolutions ranging from 10.94 to 11.98m/pixel and 118 to 128m/pixel, respectively. This dataset is available online and includes automated and manual segmentation results[1]. UMN OCT [16] dataset contains OCT images from 29 DME subjects with 25 Bscans per subject. The datasets are split into training, validation, and testing sets following [17]. A 5-fold cross-validation was employed for the Duke dataset, alongside standard training, validation, and testing splits, to reduce outlier effects due to its limited sample size. Upon discovering the most effective model with Duke OCT dataset, it used subject #1 - #6 for training, #7 and #8 for validation, and #9 and #10 for test to cohere with the work of [18], [19], and [20] for a consistent comparison. Table 1 presents the train, validation and testing split of two datasets.

Table 1: Data Split for Training, Validation and Testing

Dataset	# Training	# Validation	# Testing
Duke OCT	50	12	16
UMN OCT	475	125	125

3.2 Model Architecture

In our research, we aimed to evaluate the effectiveness of different deep learning architectures for the task of segmenting fluid-filled areas from Optical Coherence Tomography (OCT) images. The study utilized the MM-Segmentation [12] framework, and incorporated three distinct architectures: two based on Convolutional Neural Networks (CNNs) and one utilizing a Transformer-based approach. Specifically, we employed the Swin[13] Transformer as the Transformer architecture, and Deeplabv3Plus[14] and MobileNetv3[15] as the CNN architectures. The selection of these architectures was intentional; the Swin Transformer was chosen to represent a Transformer-based model,

[1] http://duke.edu/ sf59/Chiu˙BOE˙2014˙dataset.htm

while Deeplabv3 Plus served as a representative of a widely-used CNN architecture. Furthermore, MobileNetv3 was selected for its compact nature and reduced parameter count.

Our methodology included a transfer learning approach, implemented in two distinct ways. In the first method, all layers of the models were fine-tuned without freezing. In the second method, the backbone of each model was frozen during training. The dataset used in this study, the Duke OCT Dataset, presented a challenge in terms of available ground truth data; only 78 images from 10 patients were labeled with `manual_fluid_1` annotations. To address this limitation and ensure robust evaluation, a 5-fold cross-validation approach was employed.

For each of the three architectures, five models were trained per fold, with both backbone freezing and without freezing approaches. Consequently, for the transfer learning segment of the study, a total of 30 models (3 architectures × 2 methods × 5 folds) were trained and evaluated. Additionally, in the training-from-scratch scenario, we trained five models per architecture using 5-fold cross-validation, amounting to 15 models. Therefore, the study comprehensively trained and evaluated a total of 45 models (30 transfer learning models + 15 trained-from-scratch models) for the Duke OCT Dataset. Upon discovering of the most effective model, it was applied with three different data augmentation techniques (simple - random crop & rotation, gaussian blur, gaussain blur with noise) and a loss function combining cross entropy and Dice, accordingly, to evaluate the best fine-tuning techniques among them. To further investigate the generalizability of the particular model, it was decided to train and test on the UMN OCT dataset [16].

The table 2 offers a summary of the number of layers present in each architecture and the status of layer freezing:

Table 2: Architecture, Layer Counts, and Backbone Freezing Status

Architecture	**# Layers**	**# Backbone Layers**	**Backbone Freeze**	**# Parameters**
Swin Transformer	322	257	No	59,865,425
Swin Transformer	322	257	Yes	32,344,727
Deep Labv3 Plus	222	147	No	43,580,644
Deep Labv3 Plus	222	147	Yes	20,053,380
MobileNetv3	273	249	No	3,282,225
MobileNetv3	273	249	Yes	310,273

3.3 Evaluation Metrics

Intersection over Union (IoU), Dice Similarity Coefficient (DSC), Sensitivity and Specificity are used as evaluation metrics. IoU measures the overlap between the predicted segmentation X and the ground truth Y, defined as:

$$IoU = \frac{|X \cap Y|}{|X \cup Y|}, \tag{1}$$

where X is the set of pixels in the predicted segmentation and Y is the set of pixels in the ground truth segmentation. DSC evaluates the similarity between two sets of data, X and Y.

$$DSC = \frac{2 \times |X \cap Y|}{|X| + |Y|}, \tag{2}$$

where X and Y typically denote the pixels in the ground truth and the predicted segmentation, respectively. Sensitivity further characterize the accuracy of binary segmentation:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3}$$

indicating the model's ability to identify positive (foreground) and negative (background) classes.

4 Results and Discussions

4.1 Model Evaluation

Figure 1, 8, and 2 respectively showed the DSC score, IoU (Appendix A), and sensitivity of 45 models (3 architectures × 3 methods × 5 folds). Across each model architecture, the highest DSC,

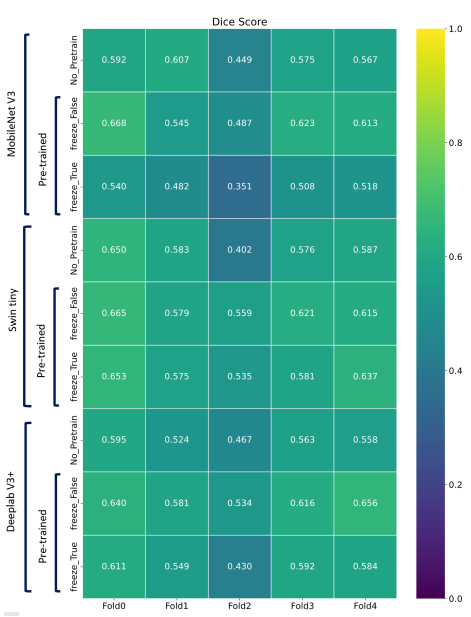

Figure 1: Dice Similarity Coefficient (DSC) of models on each cross-validation fold

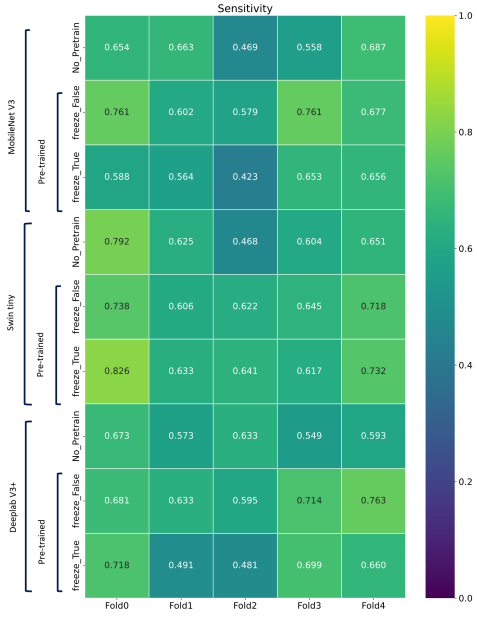

Figure 2: Sensitivity of models on each cross-validation fold

IoU and sensitivity were achieved in Fold 0 with configuration of pre-trained parameters and unfrozen backbone, except the highest sensitivity in Swin and Deep Labv3 Plus is attained by pre-trained parameters and frozen backbone. The highest and the lowest performance of the three metrics among the model architectures saturated at Fold 0 and fold 2, respectively. It demonstrated a visible discrepancy in performance across folds due to limited data, and high sensitivity in the split of training, validation and test sets.

To evaluate the overall performances of 3 model architectures in 3 methods, the mean DSC, IoU and sensitivity were calculated across 5 folds and displayed in Figure 3, 9 and 4, respectively. In Figure 3, all three model architectures demonstrated the highest intra-model mean DSC using pre-trained parameters and unfrozen backbone, which resembled to the results found in Figure 1. Swin showed the highest inter-model and intra-model mean DSC of 0.608, compared to the second highest of 0.605 in Deep LabV3 Plus model. Similar results were demonstrated in Figure 9 in Appendix A, where the highest and the second highest intra- and inter-model mean IoU of 0.463 and 0.460 were performed by Swin model and Deep LabV3 Plus using pre-trained weights and unfrozen backbone, respectively. Regarding the sensitivity shown in Figure 4, Swin using pre-trained parameters with frozen backbone showed the highest value of 0.690 among all methods and architectures.

A inter-model performance in DSC, IoU and sensitivity improved from MobileNet V3 to Deep LavV3 Plus and to the Swin was revealed in Figure 3, 9 and 4. A intra-model performance across metrics for methods of pre-trained with frozen backbone and no pre-train were unclear and showed comparable results, while method of pre-trained parameters with unfrozen backbone revealed the best performance among them.

4.2 Auxiliary training techniques on Swin

Being the best performing model among the 45 models, Swin with pre-trained parameters and unfrozen backbone were further investigated by applying with different auxiliary training techniques. The performance of mean DSC and mean IoU were illustrated in Figure 10 (in Appendix). The use of different auxiliary training techniques merely affects the performance as they all derive similar DICE and IoU scores. Among the techniques, data augmentation using random crop and rotation showed the highest mean DSC of 0.47, and highest mean IoU of 0.33, equivalent to that of data augmentation using gaussian blur.

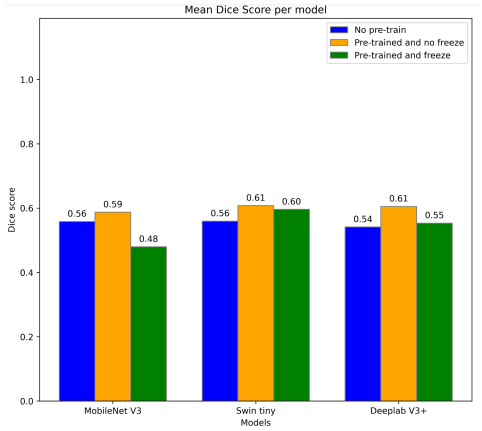

Figure 3: Mean Dice score of models averaging on all 5 cross-validation folds

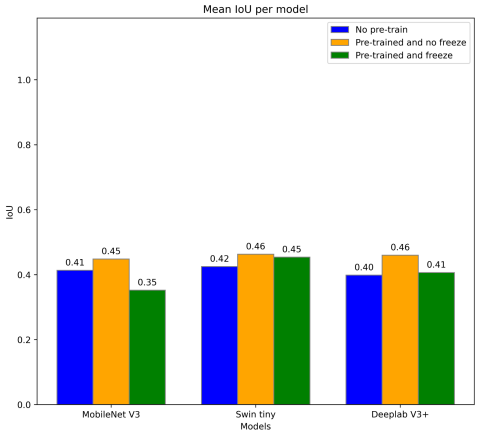

Figure 4: Mean sensitivity of models averaging on all 5 cross-validation folds

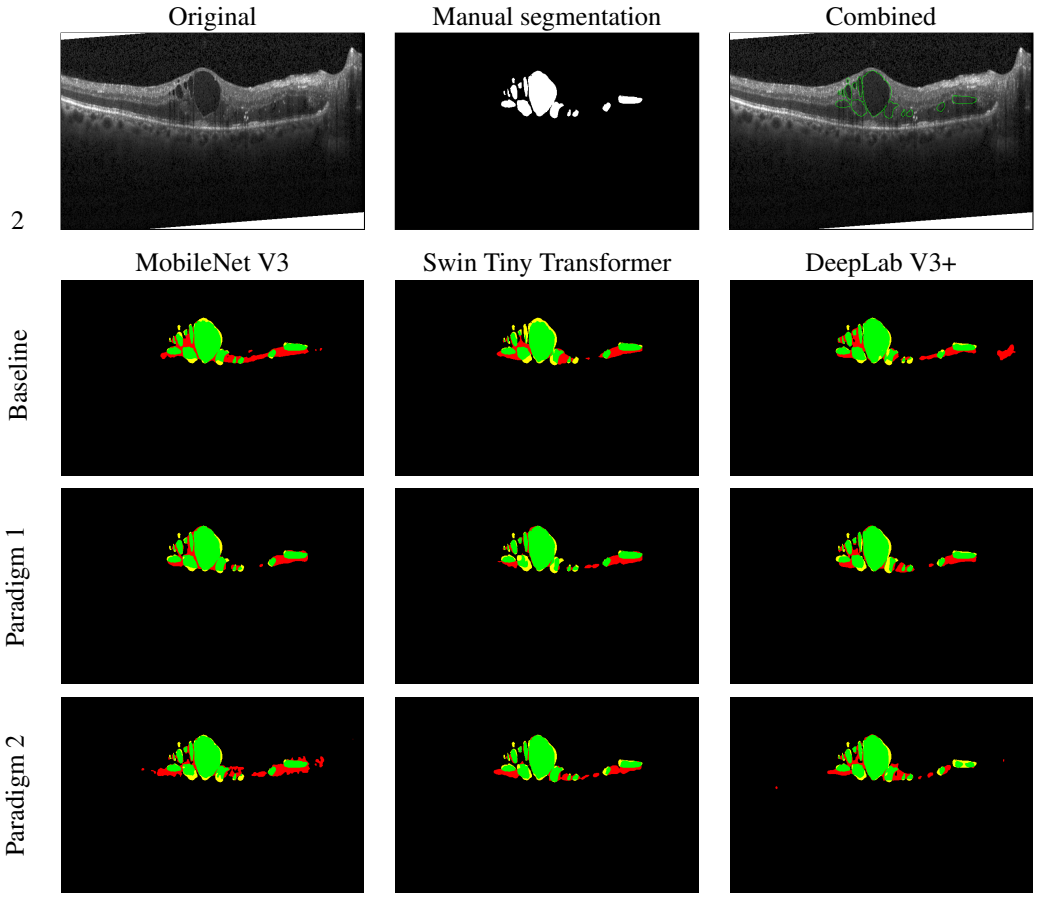

Figure 5: Manual segmentation VS Automatic segmentation using different models on an example OCT image from Duke DME dataset. Baseline is unpretrained + unfrozen. Paradigm 1 is pretrained + unfrozen. Paradigm 2 is pretrained + frozen. Green, Red, Yellow, and Black color areas are True Positive, False positive, False Negative and True Negative respectively.

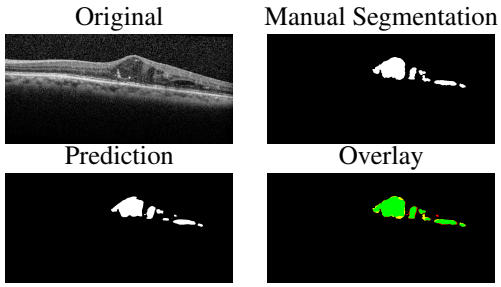

Figure 6: Example of retinal fluid segmentation in UMN Dataset

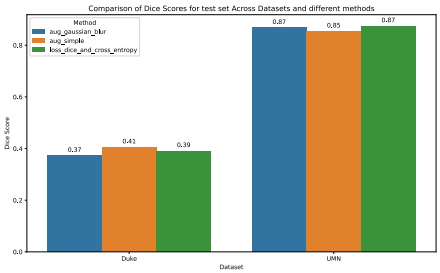

Figure 7: Comparison of Swin Model on Duke and UMN Dataset

Compared to the same model trained using 8 subjects and tested on 2 subjects, the model performance using 6 subjects with different auxiliary training techniques showed a lower mean DSC from 0.61 to 0.47, and a lower mean IoU from 0.46 to 0.33.

In Comparison to the literature where the best performance obtained by [18], [19] and [20] with a DSC of 0.36, 0.39 and 0.56 accordingly, the mean DSC of 0.47 by the Swin model using pre-trained parameters and unfrozen backbone, comprised with random crop and rotation in data augmentation, showed comparable results.

4.3 Comparing performance of Duke & UMN OCT dataset

The Swin model with pre-trained parameters and unfrozen backbone was trained on the UMN dataset using the auxiliary training techniques. Figure 6 showed the qualitative performance of segmentation across OCT image, ground truth, model prediction and overlay image. The false positives (red region) reduced significantly compared to that in Figure 5.

Figure 7 compared the DSC of the Swin model trained on the two datasets with different auxiliary training techniques. The mean DSC improved significantly from 0.4 in Duke dataset to 0.81 in UMN dataset, achieving the DSC reported in [16]. Although a higher mean DSC of 0.91 was reported in [17], the study included the OCT images of layers without any fluid-filled areas in the performance evaluation. This study excluded these images without target from our metrics measurement because mean DSC would inflate with the number of no-target images, which easily reached a DSC of 1.

The Swin model achieved higher mean DSC on the UMN dataset because more data was available in this dataset, and the structure showed could have been easier to learn than that in the Duke dataset. The manual segmentation of two readers (manualfluid1 and manualfuild2), acted as ground truth for the training, reached a mean DSC of 0.58 between them. The disagreement in ground truth images suggested the underlying difficulty in segmenting the retinal cysts. Moreover, transfer learning is not the solution for all machine learning problems. The model performance can vary significantly across datasets and tasks. A prior knowledge of a dataset could assist in model and configuration selection for transfer learning.

5 Conclusion

This study evaluated performance of transfer learning in OCT retinal cyst segmentation task using pre-trained models in different architectures and configurations. Swin model pre-trained on scene-centric data with unfrozen backbone showed comparable performance as other works using data augmentation of random crop and rotation, achieving a mean DSC of 0.41 on Duke OCT dataset and 0.81 on UMN OCT dataset. This illustrated the potential of transfer learning in domain-shifted segmentation tasks, where the sample size in targeted domain, being OCT images in this study, was limited while data for pre-training in other domain was vastly available.

References

[1] Wei, X., & Sui, R. (2023). A Review of Machine Learning Algorithms for Retinal Cyst Segmentation on Optical Coherence Tomography. Sensors (Basel, Switzerland), 23(6), 3144. https://doi.org/10.3390/s23063144

[2] Duh, E. J., Sun, J. K., & Stitt, A. W. (2017). Diabetic retinopathy: current understanding, mechanisms, and treatment strategies. JCI insight, 2(14), e93751. https://doi.org/10.1172/jci.insight.93751

[3] Hee, M. R., Puliafito, C. A., Duker, J. S., Reichel, E., Coker, J. G., Wilkins, J. R., Schuman, J. S., Swanson, E. A., & Fujimoto, J. G. (1998). Topography of diabetic macular edema with optical coherence tomography. Ophthalmology, 105(2), 360–370. https://doi.org/10.1016/s0161-6420(98)93601-6

[4] Chiu, S. J., Allingham, M. J., Mettu, P. S., Cousins, S. W., Izatt, J. A., & Farsiu, S. (2015). Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. Biomedical optics express, 6(4), 1172–1194. https://doi.org/10.1364/BOE.6.001172

[5] Lee, C. S., Tyring, A. J., Deruyter, N. P., Wu, Y., Rokem, A., & Lee, A. Y. (2017). Deep-learning based, automated segmentation of macular edema in optical coherence tomography. Biomedical optics express, 8(7), 3440–3448. https://doi.org/10.1364/BOE.8.003440

[6] Schlegl, T., Waldstein, S. M., Bogunovic, H., Endstraßer, F., Sadeghipour, A., Philip, A. M., Podkowinski, D., Gerendas, B. S., Langs, G., & Schmidt-Erfurth, U. (2018). Fully Automated Detection and Quantification of Macular Fluid in OCT Using Deep Learning. Ophthalmology, 125(4), 549–558. https://doi.org/10.1016/j.ophtha.2017.10.031

[7] Roy, A. G., Conjeti, S., Karri, S. P. K., Sheet, D., Katouzian, A., Wachinger, C., & Navab, N. (2017). ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. Biomedical optics express, 8(8), 3627–3642. https://doi.org/10.1364/BOE.8.003627

[8] Hassan, T., Akram, M. U., Masood, M. F., & Yasin, U. (2019). Deep structure tensor graph search framework for automated extraction and characterization of retinal layers and fluid pathology in retinal SD-OCT scans. Computers in biology and medicine, 105, 112–124. https://doi.org/10.1016/j.compbiomed.2018.12.015

[9] Wang, Z., Zhong, Y., Yao, M., Ma, Y., Zhang, W., Li, C., Tao, Z., Jiang, Q., & Yan, B. (2021). Automated segmentation of macular edema for the diagnosis of ocular disease using deep learning method. Scientific reports, 11(1), 13392. https://doi.org/10.1038/s41598-021-92458-8

[10] Heisler, M., Bhalla, M., Lo, J., Mammo, Z., Lee, S., Ju, M. J., Beg, M. F., & Sarunic, M. V. (2020). Semi-supervised deep learning based 3D analysis of the peripapillary region. Biomedical optics express, 11(7), 3843–3856. https://doi.org/10.1364/BOE.392648

[11] Ma, D., Lu, D., Chen, S., Heisler, M., Dabiri, S., Lee, S., Lee, H., Ding, G. W., Sarunic, M. V., & Beg, M. F. (2021). LF-UNet - A novel anatomical-aware dual-branch cascaded deep neural network for segmentation of retinal layers and fluid from optical coherence tomography images. Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society, 94, 101988. https://doi.org/10.1016/j.compmedimag.2021.101988

[12] MMSegmentation Contributors. "MMSegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark." 2020. GitHub, https://github.com/open-mmlab/mmsegmentation.

[13] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 9992-10002.

[14] Peng, X., Yin, Z., & Yang, Z. (2020). Deeplab_v3_plus-net for Image Semantic Segmentation with Channel Compression. 2020 IEEE 20th International Conference on Communication Technology (ICCT), 1320-1324.

[15] Howard, A.G., Sandler, M., Chu, G., Chen, L., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V., & Adam, H. (2019). Searching for MobileNetV3. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 1314-1324.

[16] A. Rashno et al., "Fully Automated Segmentation of Fluid/Cyst Regions in Optical Coherence Tomography Images With Diabetic Macular Edema Using Neutrosophic Sets and Graph Algorithms," in IEEE Transactions on Biomedical Engineering, vol. 65, no. 5, pp. 989-1001, May 2018, doi: 10.1109/TBME.2017.2734058.

[17] Farshad, Azade, et al. "Y-Net: A spatiospectral dual-encoder network for medical image segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022.

[18] Maier, Heiko, Shahrooz Faghihroohi, and Nassir Navab. "A line to align: deep dynamic time warping for retinal OCT segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer International Publishing, 2021.

[19] Roy, Abhijit Guha, et al. "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks." Biomedical optics express 8.8 (2017): 3627-3642.

[20] Tran, Arianne, et al. "Retinal layer segmentation reformulated as OCT language processing." Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23. Springer International Publishing, 2020.