# Microaneurysm Detection on Retinal Fundus Images with Multi-resolution GlobalNet to Support the Diagnosis of Diabetic Retinopathy

Fadhila Ramadiastri
*Biomedical Engineering Department*
*Bandung Institute of Technology*
Bandung, Indonesia
18318006@std.stei.itb.ac.id

Astri Handayani
*Biomedical Engineering Department*
*Bandung Institute of Technology*
Bandung, Indonesia
a.handayani@itb.ac.id

*Abstract*—**Diabetic retinopathy, one of the main causes of blindness in the world, can be diagnosed early by utilizing microaneurysms (MA). The existence of microaneurysms can be detected by performing semantic segmentation on the retinal fundus image. In this paper, deep learning method testing was carried out using the GlobalNet model at various input image resolution values to see the effect of resolution alteration on MA detection in retinal fundus images. The experiment provided AUPR values of 0.391 ± 0.026, 0.387 ± 0.035, and 0.394 ± 0.050, along with F1 scores of 0.361 ± 0.034, 0.361 ± 0.022, and 0.360 ± 0.022. Evaluation done at pixel and lesion levels shows that the difference in resolution of the input images does not cause a significant change in the AUPR and F1 scores. However, the resolution alteration does affect the total number of false positive lesions in outputs. In addition, a combination method is developed, providing a better trade-off in sensitivity and precision compared to the GlobalNet segmentation model.**

*Keywords — microaneurysm, GlobalNet, multi-resolution, diabetic retinopathy*

## I. INTRODUCTION

Diabetic retinopathy (DR), a complication of diabetes mellitus, is largely to blame for the rise in the number of visually impaired people worldwide. According to the World Report on Vision in 2019, diabetic retinopathy is one of the leading causes of blindness in the world [1]. Blindness in DR patients is often attributed to a delayed diagnosis that occurs when the disease is detected after the patient has experienced damage from visual loss, which will be difficult to repair. Hence, to reduce the occurrence of diabetic retinopathy-related blindness, early detection of the disease is essential. Retinal features found on each DR stage, which consists of the Proliferative Diabetic Retinopathy (PDR) stage and the Non-proliferative Diabetic Retinopathy (NPDR) stage, can be used to aid the early diagnosis of DR [2].

A sac-like formation which bulges from the retina's blood vessels called a microaneurysm (MA) develops in the mild NPDR stage and is often used to support the initial detection of DR. A Microaneurysm appears as the result of blood vessel obstruction caused by a high glucose level in DR patients and can be found as a red spot or a small circular structure with a diameter of 10-100 μm on a fundus image [3], and sometimes also as an irregular shape. The presence of one or more microaneurysm lesions enables the ophthalmologist to make the diagnosis without considering the size of the lesion, since the lesion's size does not accurately reflect the severity level

of DR. The diagnosis can also help the doctor as the first step in treatment to avoid the disease getting worse.

The presence of microaneurysms in the retina can be determined using semantic segmentation on retinal fundus images, which will act as the framework for lesion-level MA detection [4]. By using semantic segmentation [5], each pixel of the fundus image will be classified as a microaneurysm or non-microaneurysm. The modest size of the lesions makes them particularly vulnerable to a change in resolution values in fundus images, which is one of the challenges in segmenting microaneurysms. The lesion may be partially or completely obscured when the input's resolution is decreased. As a result, the purpose of this study is to determine how the resolution value impacts the output of the segmentation model for microaneurysm detection.

In recent days, a lot of microaneurysm segmentation methods have been developed to support the diagnosis of diabetic retinopathy. However, some of the existing methods have certain limitations, such as the model's inability to identify tiny lesions near blood vessels [4] and its incapacity to identify lesions with irregular forms [6].

## II. METHODOLOGY

### A. Dataset

The experiment was conducted using the Indian Diabetic Retinopathy Image Dataset (IDRID) [7]. The fundus images have a resolution of 4288 x 2848 pixels and were captured using a Kowa VX-10α fundus camera with a field of view of 50°. Along with 81 RGB fundus images in JPG format, the dataset also contains pixel-level annotations for each image in TIF format. Observers trained by ophthalmologists annotated each lesion on the dataset with the help of software developed by ADCIS.

### B. Architecture

In this paper, microaneurysm segmentation was done by modifying the architecture of U-Net [8] to a model called GlobalNet [9]. This model was adopted by Yan et al. by changing the number of the pooling layer in U-Net [8], which was used for the backbone architecture, and applying weighted cross entropy as the loss function instead of cross entropy. The basic U-Net architecture consists of a 3x3 convolutional layer with the activation function ReLu, followed by a 2x2 pooling layer at the end of each

convolutional layer. Instead of having four max pooling layers like basic U-Net, GlobalNet consists of three max pooling layers. In this paper, the model was also complemented by a dropout layer at the end of each convolutional layer. While U-Net used a resolution of 572 x 572 pixels for its input, GlobalNet by Yan et al. utilized a resolution of 640 x 640 pixels for its input. To determine the best resolution value to do a microaneurysm segmentation, we utilized three different resolution values for the model input.

## C. Training

Dataset preparation was performed before the training to optimize the segmentation results. The background of the fundus image was removed using the crop center method, and the dataset images were downsized to a value of 640 x 640 pixels, 960 x 960 pixels, and 1280 x 1280 pixels. The downsized images will serve as the input for each segmentation model.

During training, bicubic interpolation was used to do downsampling and upweighting of the model input image. The bicubic interpolation method will predict the pixel value in the enlarged or cropped image by looking at the nearest 16 pixels or 16 neighboring pixels from the pixel of the predicted image [10]. This method performs downsampling and upweighting with the least amount of interpolation distortion.

During preparation, data on MA pixel distribution, total, and the size distribution of MA per image were also collected to divide the dataset into folds. Before downsampling, the dataset was divided into folds, where each fold contained a similar percentage of the total pixels of microaneurysms per image. This was done to reduce bias in the segmentation model. The dataset was divided into five different folds, with 16 or 17 images on each fold.

TABLE I. HYPERPARAMETER TRAINING OF GLOBALNET MULTIRESOLUTION

| Hyperparameter | Value | Note |
|---|---|---|
| Dropout | 0.1 | 0.1 for fold 1-4 |
| | 0.2 | 0.2 for fold 5 |
| Optimizer | Adam | - |
| Batch size | 1 | |
| Initial learning rate | 0.0002 | - |
| Maximum epochs | 450 | When the validation loss remains unchanged after 20 epochs, the training comes to an end |

Segmentation model GlobalNet was trained using three different resolution values. Hyperparameter optimization was performed for selected hyperparameters such as dropout, batch size, and epochs as shown in Table I, and their outputs were evaluated at the lesion and pixel levels. Weighted cross entropy was also used as the loss function for the training in this paper. The utilization of dropouts in the segmentation model serves to prevent its overfitting.

## D. Multiresolution Combination Method

Due to their small size, the pixels of microaneurysms are very susceptible to changes in image resolution. Based on our findings on the IDRID dataset, 5.157% of microaneurysm pixels were missing on images with a resolution of 640 x 640 pixels, 7.027% of microaneurysm pixels were missing on images with a resolution of 960 x 960 pixels, and 10.421% of pixels were missing on images with a resolution of 1280 x 1280 pixels. In this paper, we propose a combination method to optimize the output result from the GlobalNet segmentation model where Figure 1 describes the flow of the combination methods. The approach was developed by taking into consideration the impacts of the input resolution on the presence of microaneurysms in images.
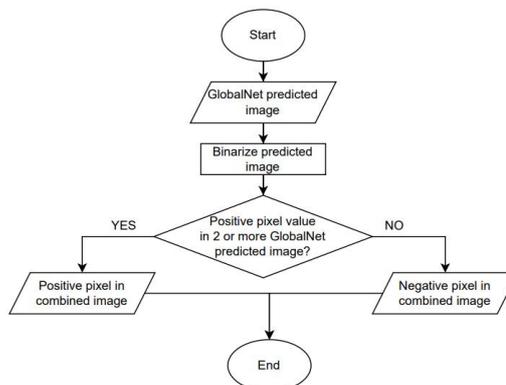


Fig. 1. Flow of our combination method

To begin with, the outputs of the segmentation models from the previous step were enlarged to a size of 3328 x 2816 pixels, this image will be used as the input for the combination methods. Following this, the pixel values of the input were transformed from 0-255 to 0-1. Then, the input images were binarized using a threshold value of 0.5. The process of combining the segmentation images with different resolutions begins with creating an empty array with the same size as the input (3328 x 2816). In the next stages, each pixel value of the input will be checked, and each pixel will be grouped based on the rules as shown in Figure 1. Once this has been completed, the pixel value will be assigned to the empty array to create a new image based on the combination of the segmentation results from GlobalNet.

The combination method utilized a threshold value of 0.5 for binarization of all models output to make the result easier to interpret. In addition, the algorithm of the combination method does not require additional training, which contributes to the light computational load of the model.

## E. Evaluation

The evaluation of the result was measured both at the pixel and lesion levels. Pixel-level evaluation was done by calculating the area under precision recall (AUPR) and F1 Score, while lesion-level evaluation was done by calculating the sensitivity. The evaluation for lesion level was executed on the resolution of 3328 x 2816 pixels, while the pixel level evaluation was performed on the model size (640 x 640 pixels, 960 x 960 pixels, and 1280 x 1280 pixels).

## III. RESULTS

### A. Deep learning multi-resolution segmentation model GlobalNet

Table II displays the results of the performance of the GlobalNet segmentation model with various resolution inputs. In this experiment, an ANOVA statistical analysis was carried out for the AUPR value and F1 score to determine the impact of the resolution value on the AUPR and F1 score. The p value was checked for a significance threshold of 0.05. From the statistical analysis, the p value for the AUPR of the segmentation model was found to be 0.955, while the p value for the F1 score was 0.999. The value denotes that the varying input resolution settings have no impact on the AUPR value or F1 score and suggests that the AUPR value or F1 score of the three models' outputs is not statistically significant.

TABLE II. AUPR AND F1 SCORE OF GLOBALNET MULTI-RESOLUTION MODEL (PIXEL LEVEL)

| Input resolution (pixel) | AUPR | F1 Score |
|---|---|---|
| 640 x 640 | 0.391 ± 0.026 | 0.361 ± 0.034 |
| 960 x 960 | 0.387 ± 0.035 | 0.361 ± 0.022 |
| 1280 x 1280 | 0.394 ± 0.050 | 0.360 ± 0.022 |

Because of their size, the pixels of microaneurysms are prone to a change in image resolution, resulting in a low score of evaluation at the pixel level for AUPR and F1 score. For a more accurate assessment, an evaluation at the lesion level is needed, which will be discussed in the next section. Meanwhile, Figure 2 shows the output images of GlobalNet segmentation models. True positive lesions are shown as yellow spots; false positive lesions are shown in red; and false negative lesions are shown in green.
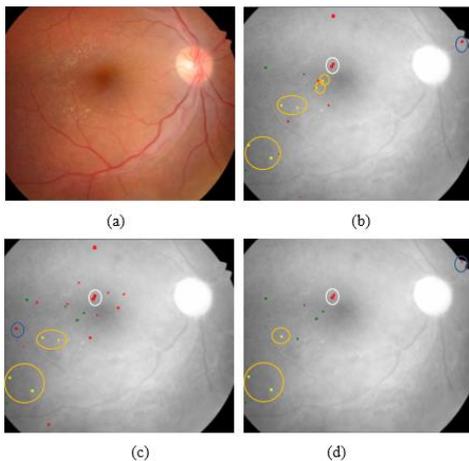


Fig. 2 (a) Retina fundus image IDRiD_57 (b) Predicted image of segmentation model GlobalNet 640 x 640 pixels (c) Predicted image of segmentation model GlobalNet 960 x 960 pixels (d) Predicted image of segmentation model GlobalNet 1280 x 1280 pixels

Based on Figure 2, the output of segmentation models with a resolution input of 1280 x 1280 pixels produced a lower number of true positive lesions than the other two models, which were shown on the yellow circle. From Figure 2, the image output from the segmentation model with a resolution of 640 x 640 pixels produced a segmentation output with the highest number of true positive lesions compared to other output, which is indicated by the number of yellow spots surrounded by a yellow circle in Figure 2b. While the yellow circle highlights the true positive lesions, the dark blue circle denotes the false positive lesions that appeared on one segmentation output. On Figure 2b, 2c, and 2d, the location of the blue circle differs for each image because of the various locations of the false positive lesion from each output. Meanwhile, the white circle that can be seen at the same location for Figure 2b, 2c, and 2d indicates the false positive lesions that appeared across all outputs of the segmentation models.

Based on the output results, each segmentation model produces unique segmentation results for microaneurysms, with certain lesions being recognized in all three segmentation models and others being detected in just one or two models. Using this circumstance, it was possible to merge the data from all three outputs of the segmentation model and produce a composite image.

### B. Multi-resolution combination method

In our combination method, a pixel with a positive value (a pixel with a value of 1) only in one image output will be discarded, whereas a pixel with a value of 1 on two or more image outputs will be considered a positive pixel in the combined image of the three segmentation models. Figure 3 shows the comparison between the output images of both the deep learning GlobalNet segmentation method and the combination method. Like Figure 2, true positive lesions are shown in yellow, false positive lesions in red, and true negative lesions are shown as green lesions.

Figure 3 displays that the yellow circle region in Figure 3a and 3c denotes the existence of two true positive lesions. On the other hand, Figure 3b shows the presence of one true positive lesion and one false negative lesion within the yellow circle area. Since two true positive lesions were detected in the two outputs of the segmentation model, two true positive lesions appeared in the same exact area in the combined image, as shown in Figure 3d.
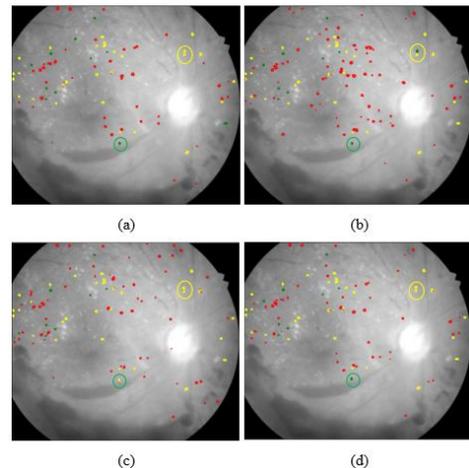


Fig. 3 (a) Lesion map of predicted image GlobalNet 640 x 640 pixels (b) Lesion map of predicted image GlobalNet 960 x 960 pixels (c) Lesion map of predicted image GlobalNet 1280 x 1280 pixels (d) Lesion map of combination method

The green circle in Figure 3a and 3b indicates the presence of one false negative lesion. However, the same green circle in Figure 3c shows the presence of one true positive lesion. As the true positive lesion was detected in only one of the outputs, the combination method identified the related lesion as a false negative lesion in the concatenated image, which can be seen in Figure 3d. A comparison of the predicted image from each method shows the difference in the total amount of true

positive, false negative, and false positive lesions between the segmentation output of the deep learning segmentation model with GlobalNet and the segmentation result of the combination method. Therefore, we also compare the assessment outcomes for each approach at the lesion level. Table III displays evaluation outcomes at the lesion level for the combination method and the deep learning model, which were done at a binary threshold level of 0.5.

The high sensitivity value for each method indicates that the model has succeeded in segmenting true positive lesions quite well. On the other hand, the low precision value suggested a need for improvement to reduce the false positive lesion in the segmentation output. Table III also illustrates an improvement in precision score when the input resolution is increased, which indicates that a higher resolution value will produce fewer false positives on the predicted image. It also reveals that the combination approach has the second-highest sensitivity value and the highest precision score among other methods.

TABLE III. LESION LEVEL SEGMENTATION RESULTS GLOBALNET MULTIRESOLUTION

| Methods | Sensitivity | Precision |
|---|---|---|
| GlobalNet 640 x 640 pixels | $0.689 \pm 0.063$ | $0.310 \pm 0.054$ |
| GlobalNet 960 x 960 pixels | $0.659 \pm 0.041$ | $0.327 \pm 0.030$ |
| GlobalNet 1280 x 1280 pixels | $0.647 \pm 0.057$ | $0.333 \pm 0.039$ |
| Combination method | $0.672 \pm 0.035$ | $0.394 \pm 0.014$ |

Although there is a slight difference in the sensitivity values for each method, it's necessary to see whether these differences are significant statistically. An ANOVA statistical test was carried out for a significance level of 0.05, which yielded a p-value for various parameters as shown in Table IV.

TABLE IV. ANOVA STATISTICAL TEST RESULTS

| Parameter | P value |
|---|---|
| Sensitivity | 0.672 |
| Precision | 0.027 |

Based on the statistical test, it was found that the p-value for precision is below the significance threshold. Meanwhile, the p-value of sensitivity is greater than 0.05, which implies that the differences in sensitivity for each method in the study are not statistically significant. Since precision represents the percentage of relevant samples among the recovered samples and sensitivity represents the probability of a positive result, the outcome of the ANOVA statistical test indicated a possible difference in the total number of false positive lesions from the predicted images of each method. This is because precision value is affected by the number of both true positive and false positive values, while sensitivity is affected by true positive and true negative values. Because the sensitivity values of all methods did not significantly differ from one another, it can be assumed that the combination method was more effective at removing false positives since it has the highest precision value, which is also statistically different from other methods.

The number of true positive, false positive, and true negative lesions is also affected by the binary threshold value used to assess the lesion's level. However, in this study, no variation of the binary threshold was used for binarization. Additionally, an optimal threshold might be needed for the combination method to eliminate false positive lesions while at the same time maintaining the maximum number of true positive lesions. In future studies, experiments with other threshold values can be carried out to obtain more optimal results from the segmentation.

### C. Comparison with similar studies

Evaluation comparisons at pixel and lesion levels were also conducted between the outcomes of all methods in this study and other literature, as shown in Table V and Table VI. From Table V, it was found that the multi-resolution GlobalNet model gave a higher AUPR value than the literature [11]. However, despite the model's ability to solve the drawback of Wan et al.'s [11] method in detecting small objects, the multi-resolution GlobalNet model also has a lower value of AUPR than [9].

Although both our method and Yan et al.'s method used GlobalNet for microaneurysm segmentation, a dataset augmentation to prevent overfitting of the trained model, which was carried out by [9], might be one of the potential causes of the lower AUPR value from our studies. In addition, a difference in the optimal hyperparameter value used for the loss function can also be a potential cause of the different results. Furthermore, the possibility of variability in the downsampling and upweighting methods can also be a contributing factor to the differences in segmentation results between [9] and our results.

TABLE V. COMPARISON AT PIXEL LEVEL

| Methods | AUPR |
|---|---|
| Yan et al. [9] | 0.484 |
| Wan et al. [11] | 0.241 |
| GlobalNet 640 x 640 pixels | $0.391 \pm 0.026$ |
| GlobalNet 960 x 960 pixels | $0.387 \pm 0.035$ |
| GlobalNet 1280 x 1280 pixels | $0.394 \pm 0.050$ |

In our study, we utilized the bicubic interpolation method for upweighting and downsampling images, which can also affect the segmentation results. The bicubic interpolation method might cause ringing artifacts due to the ringing effects, which often happen when high-frequency signals pass through a low-pass filter [12]. The ringing artifacts caused distortion of the edges of the pixels that had been expanded or reduced, which caused a change in the pixel value of the predicted image. As a result, the outcome of the microaneurysm segmentation model was affected by the appearance of ringing artifacts in the images.

Table VI displays a comparison of segmentation results between our paper and a similar study at the lesion level. According to the table, our multi-resolution GlobalNet models and combination method generated a higher value of sensitivity compared to replicated U-Net [8] by Xu et al. and their own approaches. Xu et al. [13] used U-Net as the backbone architecture in their approach and proposed a model called FFU-Net using patches as their image input. Even though their method managed to reduce the number of false

positive lesions, it still has a lower sensitivity value than our GlobalNet model. Moreover, the utilization of patches on segmentation by [13] also has disadvantages, as FFU-Net will take a lot longer and need a larger RAM to perform microaneurysm segmentation compared to GlobalNet.

TABLE VI. COMPARISON AT LESION LEVEL

| Methods | Input resolution | Sensitivity |
|---------|------------------|-------------|
| U-Net (Xu et al.) [8]ᵃ | Global image 572 x 572 pixels | 0.481 |
| Xu et al. (2021) [13] | *Patch* 256 x 256 pixels | 0.593 |
| GlobalNet 640 x 640 pixels | Global image 640 x 640 pixels | 0.689 ± 0.063 |
| GlobalNet 960 x 960 pixels | Global image 960 x 960 pixels | 0.659 ± 0.041 |
| GlobalNet 1280 x 1280 pixels | Global image 1280 x 1280 pixels | 0.647 ± 0.057 |
| Combination method | Global image 3328 x 2816 pixels | 0.672 ± 0.035 |

<sup>a.</sup> Replicated model by Xu et al.

Our approach also proved to have a higher value of sensitivity compared to a replicated U-Net, where the difference in the input resolution size may also contribute to the different performance of microaneurysm detection. A lower value of input resolution in replicated U-Net by [13] may result in a lower number of sensitivities since a change in resolution value will affect the existence of microaneurysm lesions in input images.

While most studies adopted U-Net as their backbone architecture with a few modifications in layers or pre-processing steps like [11] and [13] to perform semantic segmentation to support the diagnosis of diabetic retinopathy, different approaches, such as Fully Convolutional Neural Network (FCNN), are also used by Chudzik et al. [14]. Unlike U-Net, FCNN consists of only an encoder and a decoder without any skip connections. However, both U-Net and FCNN share similar features, allowing them to perform semantic segmentation. Both U-Net and FCNN consist of convolutional layers, which allow the models to learn the features from an image. In addition, FCNN and U-Net also generate segmentation maps by making a pixel-wise prediction on input images. Although FCNN has disadvantages in precise localization compared to U-Net, optimization done by [14] allows the model to perform microaneurysm segmentation in retinal fundus images.

## IV. CONCLUSION

In this paper, a model based on the U-Net architecture is developed and trained using various values of resolution for the input. Evaluation at the pixel level provided AUPR values of 0.391 ± 0.026, 0.387 ± 0.035, and 0.394 ± 0.050 for the GlobalNet model with input resolutions of 640 x 640 pixels, 960 x 960 pixels, and 1280 x 1280 pixels. Based on the result, it is concluded that the input size does not cause a significant change in the AUPR score, F1 score, or the total of false negative and true positive lesions on the output of the model. However, different resolution values are found to have a large

effect on the number of false positive lesions on predicted images. The total number of false positive lesions will be lower the higher the model's input resolution. On the other hand, this paper also developed a logical combination method to merge every output from the GlobalNet model. The proposed method provides a better trade-off of sensitivity at 0.672 ± 0.035 and precision at 0.394 ± 0.014 than the GlobalNet segmentation model with a resolution of 640 x 640 pixels.

## REFERENCES

[1] World Health Organization, "World report on vision," Geneva, 2019.

[2] Ferrucci Steven and Yeh Brenda, "*Diabetic Retinopathy by the Numbers,*"*https://www.reviewofoptometry.com/article/diabetic-retinopathy-by-the-numbers#:~:text=You%20can%20categorize%20this%20version,at%20least%20one%20retinal%20quadrant.*, Aug. 09, 2022.

[3] B. K. Triwijoyo, B. S. Sabarguna, W. Budiharto, and E. Abdurachman, "2 - Deep learning approach for classification of eye diseases based on color fundus images," in *Diabetes and Fundus OCT*, A. S. El-Baz and J. S. Suri, Eds., Elsevier, 2020, pp. 25–57. doi:10.1016/B978-0-12-817440-1.00002-4.

[4] R. S. Biyani and B. M. Patre, "Algorithms for red lesion detection in Diabetic Retinopathy: A review.," *Biomed Pharmacother*, vol. 107, pp. 681–688, Nov. 2018, doi: 10.1016/j.biopha.2018.07.175.

[5] B. V, "Biomedical Image Analysis Using Semantic Segmentation," *Journal of Innovative Image Processing*, vol. 1, no. 02, pp. 91–101, Dec. 2019, doi: 10.36548/jiip.2019.2.004 .

[6] S. Long, J. Chen, A. Hu, H. Liu, Z. Chen, and D. Zheng, "Microaneurysms detection in color fundus images using machine learning based on directional local contrast," *Biomed Eng Online*, vol. 19, no. 1, p. 21, 2020, doi: 10.1186/s12938-020-00766-3.

[7] P. Porwal *et al.*, "Indian Diabetic Retinopathy Image Dataset (IDRiD): A Database for Diabetic Retinopathy Screening Research," *Data (Basel)*, vol. 3, no. 3, 2018, doi: 10.3390/data3030025.

[8] P. and B. T. Ronneberger Olaf and Fischer, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, J. and W. W. M. and F. A. F. Navab Nassir and Hornegger, Ed., Cham: Springer International Publishing, 2015, pp. 234–241.

[9] Z. Yan, X. Han, C. Wang, Y. Qiu, Z. Xiong, and S. Cui, "Learning Mutually Local-Global U-Nets For High-Resolution Retinal Lesion Segmentation In Fundus Images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 597–600. doi: 10.1109/ISBI.2019.8759579.

[10] D. Han, "Comparison of Commonly Used Image Interpolation Methods," 2013.

[11] C. Wan *et al.*, "EAD-Net: A Novel Lesion Segmentation Method in Diabetic Retinopathy Using Neural Networks.," *Dis Markers*, vol. 2021, p. 6482665, 2021, doi: 10.1155/2021/6482665.

[12] I. Aganj, B. T. T. Yeo, M. R. Sabuncu, and B. Fischl, "On Removing Interpolation and Resampling Artifacts in Rigid Image Registration," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 816–827, Feb. 2013, doi: 10.1109/TIP.2012.2224356.

[13] Y. Xu, Z. Zhou, X. Li, N. Zhang, M. Zhang, and P. Wei, "FFU-Net: Feature Fusion U-Net for Lesion Segmentation of Diabetic Retinopathy.," *Biomed Res Int*, vol. 2021, p. 6644071, 2021, doi: 10.1155/2021/6644071.

[14] P. Chudzik, S. Majumdar, F. Calivá, B. Al-Diri, and A. Hunter, "Microaneurysm detection using fully convolutional neural networks.," *Computer Methods and Programs in Biomedicine*, vol. 158, pp. 185–192, 2018, doi: 10.1016/j.cmpb.2018.02.016.