FINE VESSEL SEGMENTATION WITH REFINEMENT GATE IN DEEP LEARNING ARCHITECTURES

Ali Q Saeed[1,2*], Siti Norul Huda Sheikh Abdullah[1*], Jemaima Che-Hamzah[3], and Ahmad Tarmizi Abdul Ghani[1]

[1]Center for Cyber Security, Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia Selangor, Malaysia.

[2]Computer Center, Northern Technical University
Ninevah, Iraq

[3]Department of Ophthalmology, Faculty of Medicine, Universiti Kebangsaan Malaysia, Cheras Kuala Lumpur, Malaysia

Emails: ali.qasim@ntu.edu.iq[1]*, snhsabdullah@ukm.edu.my[2]*, jemaima@ppukm.ukm.edu.my[3], atag@ukm.edu.my[4]

ABSTRACT

Automated vessel segmentation is essential in diagnosing eye-related disorders and monitoring progressive retinal diseases. State-of-the-art methods have achieved excellent results in this field, but very few have considered the post-processing of feature maps. As a result, there is often a lack of small and fine vessels or discontinuities in segmented vessels. To address this issue, this study introduces a novel post-processing technique called the refinement gate, which works with a deep learning model during training. The refinement gate enhances contextual information to extract important features from feature maps better. The proposed technique is applied with U-net architecture and placed after every convolution block in the encoder path. Visual and statistical comparisons demonstrate the robustness of the proposed method using three publicly available datasets, namely: the DRIVE DB, the STARE DB, and CHASE_DB1 datasets, showing significant improvements to segment weak and tiny vessels. The reported results confirm the potential of the model to be used as a segmentation tool in the medical field. This study is the first to propose such a gating mechanism without additional trainable parameters or standalone networks as in other literature.

Keywords: Image segmentation; Deep learning; Fundus image; Blood vessels.

1.0 INTRODUCTION

In the medical field, precise vessel segmentation is critical for ophthalmologists to diagnose and estimate pathological eye disorders. Accurate vessel segmentation can enhance the diagnostic capabilities of ophthalmologists [1]. However, segmenting fine and tiny vessels poses a significant challenge due to various factors, including low image contrast, complex branching patterns, pathological issues, and overlapping pixels, making them difficult to distinguish [2], [3].

Numerous studies have investigated segmented vessel trees using two-dimensional retinal images [4], [5], aiming to aid physicians in diagnosing and preventing vision loss. Previous studies in retinal segmentation are categorized as feature extraction-based [6], [7], pattern recognition-based [8], [9], matched filtering-based [10], and classical machine learning-based [11]. However, these methods mainly rely on feature engineering, where handcrafted features are extracted, and a classifier is trained to segment the image. Feature engineering is challenging, expensive, and time-consuming, limiting the results and not always matching the practitioner's expertise.

Recently, deep learning (DL) has shown excellent performance in medical imaging tasks like classification [12], [13], [14], vessel segmentation [15], [16], and language processing [17], using various deep learning architectures, such as U-net [18] and Convolutional Neural Networks (CNN). These architectures can automatically learn features and surpass an expert's ability to segment retinal images [19], [20], [21].

In standard CNN, gradual downsampling of feature maps assists in capturing semantic contextual information. Although the maps generated from this process contain high-level information with complex features, some features may disappear due to low resolution or low shape variability, especially for tiny and fine vessels [22]. As a result, this process indirectly deteriorates the primary diagnostic features of retinal images and weakens the

ability to regain contextual information effectively. See Figure 1. Hence, many state-of-the-art studies suffer from low segmentation accuracy of fine vessels [23], [24], [25], in addition to the difficulties of segmenting vessels that are close to each other [26]. Many attempts tackled these drawbacks, such as [27], [28], [29], whose method consists of two phases: the localization phase and the vessels segmentation phase using multi-stage CNNs. Also, Tian et al. [15] designed a multi-path CNN for segmenting retinal blood vessels.

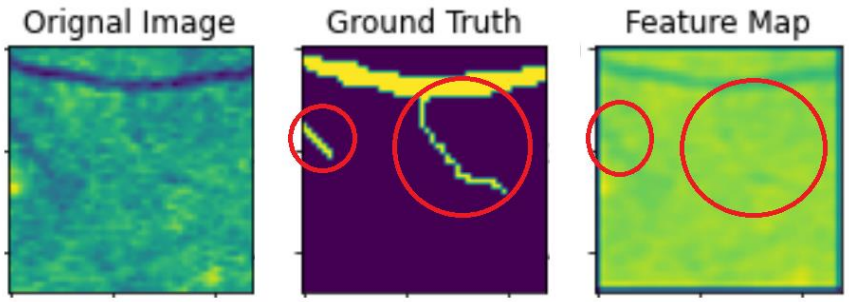

Fig. 1: Vanishing of fine vessels in a feature map due to a downsampling process.

Another study by [30] uses a standalone refinement network as a post-processing network to enhance the segmented masks. However, the study did not achieve the expected level of improvement compared to the significant architecture they proposed. According to the literature, the main difficulty of segmenting blood vessels is depicted in Figure 2.

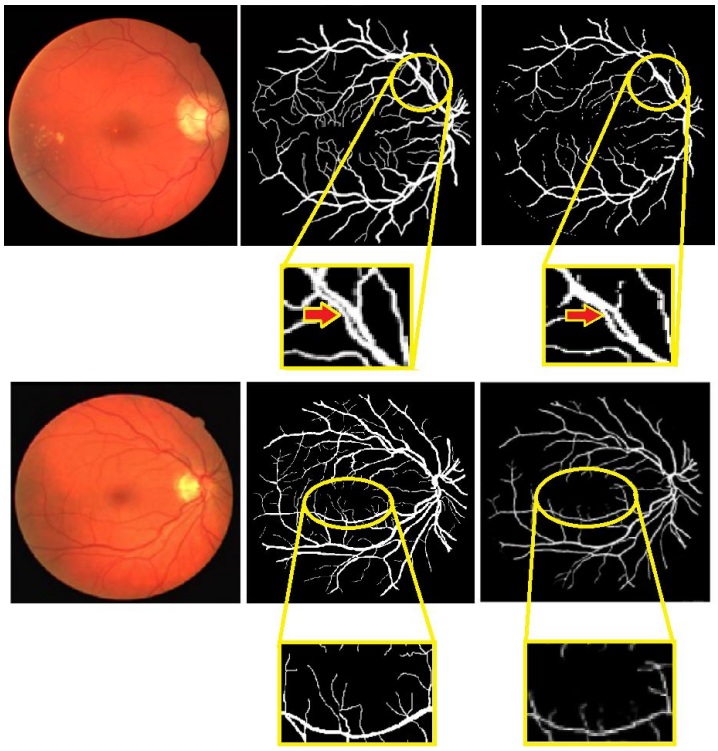

Fig. 2: Two images from the DRIVE dataset alongside their corresponding ground truth (in the middle column) and the segmented mask (in the right column). The top row represents the segmentation output produced by [26], while the bottom row illustrates the segmentation output generated by [15].

Our objective is to overcome the limitations mentioned earlier. To do that, this paper introduces a new refinement gate mechanism to emphasize pixels' values in the feature maps' low contrast region. The main concept of the proposed approach involves identifying small vessels in low-contrast regions deteriorated by the consecutive downsampling process. The significant contributions presented in this paper are:

1. A new refinement gate is proposed on top of a modified U-net to produce plausible vessel segmentation results. Even though the network architecture is more shallow (consisting of just two layers) than what is presented in other similar studies, the refinement gate serves as a guide for the network to accurately identify and segment tiny vessels.
2. This study first incorporates the refinement gate as an attached layer to a deep learning model rather than using a standalone network to refine the segmented vessels, which requires a complete training stage.

3. Several experiments were conducted using multiple public datasets to demonstrate the refinement gate's efficacy. The statistical and visual comparisons demonstrated remarkable improvement in the segmented vessels compared to literature works. Furthermore, we attained the best AUC score compared to other literature studies.
4. The proposed framework can inspire other domains that utilize semantic segmentation tasks, particularly medicine.

Utilizing the proposed gate, the model can capture weak vessels that are eligible to vanish during the consecutive downsampling process. The proposed gate accelerates vessel segmentation accuracy by restoring feature information presented in the initial layers. Unlike previous approaches that rely on multi-stage CNN models or separated networks with many trainable parameters and lengthy training time, our refinement gate can be used in any deep learning block and enhances the visibility of vessel trees within the feature maps.
The paper is organized as follows: Section 2.0 reviews DL-based methods that utilize post-processing methods to enhance their performance. Section 3.0 details the implementation steps of the proposed method. Section 4.0 explains the environment setup, the incorporated datasets, the framework architecture, the training strategy, and the metrics used to assess the method's performance. Section 5 discusses the statistical and visual comparisons with literature studies. Lastly, study limitations are explained in Section 6. Lastly, Section 7 contains the conclusion.

2.0 RELATED STUDIES

The success of DL techniques attracted researchers' attention to employ various DL architectures to segment retinal images [31]. Among these architectures, Convolutional Neural Networks (CNNs) is particularly promising. Authors in [15] proposed a multi-path CNN vessels segmentation method. The first path is used to characterize the overall features of the vascular tree, while the second path utilizes low and high Gaussian pass filters to characterize the detailed local features. However, this method did not significantly improve accuracy and had low specificity and weak generalization capability. To enhance segmentation performance, fully connected conditional random fields (CRFs) are occasionally employed as post-processing techniques in medical imaging. [32] proposed a CNN-CRF segmentation method to integrate CNN's prediction ability with the CRF's segmentation ability, aiming to improve small vessel segmentation performance. However, the model could not handle imbalanced classes of vessels and backgrounds because CRF is typically employed in pixel-wise label prediction. As a result, there was no significant increase in the accuracy rate. Additionally, if the training and testing datasets differ, CRF can harm the model's performance, hindering training speed and inference rather than enhancing it [33].

Subsequently, a more sophisticated architecture called U-net was presented by [18]. The primary objective of U-net is to produce more precise segmentation using very few annotated images. To achieve this, U-net expands on and modifies the fully convolutional networks by incorporating two paths: a contracting path (the encoder) for feature extraction and a symmetric expanding path (the decoder) for image reconstruction. Additionally, skip connections are used between layers of the same level to combine low and high-level features, resulting in precise localization outcomes. [34] presented a modified U-net architecture to improve the segmentation accuracy of the retinal discs. Their performance was less remarkable than others due to the limited size of the training images. Authors in [35] propose a multiscale feature interaction network (MFI-Net) using a pyramid squeeze-and-excitation (PSE) module and coarse-to-fine (C2F) module to encourage the interaction among the features extracted at multiple scales and incorporate both modules, together with deep supervision and feature fusion, into a U-shaped backbone. Although their method effectively segments vessels with variable width, there is still potential to improve sensitivity by incorporating contextual information from the neighborhood of vessel pixels. A different technique for segmenting fine vessels is employed by [36], where the image is magnified to help the deep learning model capture very small and fine details. Based on an encoder-decoder design, their lightweight model demonstrated good segmentation performance. However, it is susceptible to domain shift and requires optimization in terms of architectural design, such as the number of layers, channels, and variations in skip connections.
Recently, [37] proposed a promising deep learning architecture called the Generative Adversarial Networks (GANs). GANs are highly effective in various applications, such as image generation [38], natural language processing [39], and image-to-image translation [40]. The GANs consist of two deep learning architectures, the generator (G) and the discriminator (D). Both G and D train jointly in a competitive manner against each other in a min-max game. GANs aim to learn the distribution of the training dataset to generate images similar to training distribution, making it useful for fine vessel segmentation. The remarkable achievement of GAN in computer vision has led to numerous image segmentation studies [41]. Vessels are the most commonly used feature in GAN-based segmentation methods [24], [25], [42], [43], while researchers occasionally use the optic disc/cup features, and the RNFL and RL are rarely used [44]. On the other hand, [23] proposed an MI-GAN framework for vessel segmentation and generation. The generator network utilized the cross-entropy loss function and a combination of

other loss functions namely, the style loss, the content loss, and total variation loss, to train model's weights, while the discriminator network used adversarial loss only. The authors followed a specific training regime by updating the generator to double the discriminator, saving the training time and ensuring faster convergence. The results were promising, with fewer false positive rates in the fine vessels than in other methods. However, their approach tends to provide a low probability of pixels in ambiguous regions, resulting in more false negatives near vessel edges and terminals. Authors in [45] introduced a GAN-based approach for two landmark segmentation tasks (the blood vessels and the optical disc), incorporating the U-net generator and multiple discriminatory features in the discriminator. Their method effectively addressed false positives, prevented overfitting, and sped up the training process. They also demonstrated that using GAN architecture for retinal vessel segmentation improves segmentation accuracy for thin and fine vessels compared to using a standalone U-net. However, their approach under-segmented some fine vessels that an annotator could identify with anatomical knowledge. [46] leveraged the benefits of GAN-based U-net to enhance visual feature extraction through a large receptive field. They employed residual blocks to alleviate the degradation issue of deep networks and dilated convolution to get a larger receptive field with the same computational complexity. While their method outperforms [45] in feature extraction and recognition, it falls slightly behind [23] in specificity and AUC. They could not handle unbalanced classes of vessel and background pixels and may require further network depth to capture deeper features. Another study by [47] utilized the GANs and Unet and incorporated residual attention mechanisms to segment retinal arteries from the fundus background. The residual attention mechanism helps the network focus on salient regions while maintaining contextual information, which enhances the model's overall performance.

Although several studies [15], [26], [30], [48], [49], [50], [51] worked on retinal segmentation of blood vessels and showed impressive performance, few have performed post-processing to enhance the segmentation performance of vessels. This study proposes a novel post-processing technique called the refinement gate (RG) that enhances vessels in low-quality regions during model training without requiring additional trainable parameters.

3.0 IMPLEMENTATION OF THE REFINEMENT GATE

Generally, each feature map x_L obtained from a convolutional layer L consists of a 2D array of pixels, where each pixel's value corresponds to a particular color in $x_{i,c}$, where i and c refer to the spatial and channel dimensions, respectively. According to [52], the differences in pixel characteristics between images contain valuable information that can be used to describe the image content. To this end, we examine the variations among pixel values in each $\boldsymbol{x_{i,c}^{L}}$ and observe that the primary cause of low image contrast is the slight difference in pixel values between foreground and background regions. This difference is typically due to the activation functions, such as Relu, which rescale pixel values between 0 and 1 after each convolution process. Our proposed approach aims to identify foreground pixels from the surrounding, unnecessary pixels and assign them higher weights while suppressing all other background pixels. To achieve this, we need to increase the gap between the pixels' values to extract the pixels of interest efficiently. We achieve this by performing element-wise multiplication of the same feature map using the following equation:

$$\ddot{\mathrm{X}}_{i,c}^{L} = x_{i,c}^{L} \,.\, x_{i,c}^{L} \qquad (1)$$

Next, we group the values in the resulting feature map into a set of quartiles, which we refer to as bins. Each bin encompasses a particular range of values identified by an integer index. By categorizing the pixels into a predetermined set of bins, we can efficiently detect background pixels that consistently fall within a narrow range of values. This enables us to effectively suppress the bins corresponding to the background region and shift the model's focus to pixels with higher weights (i.e., foreground pixels) during model training. The boundary of each bin is calculated using the following equation:

$$bin_{(i)} = \sum_{i=1}^{n}\left[\, bin_{[i-1]} \,,\; bin_{[i-1]} + \; \delta\right] \qquad (2)$$

The starting boundary of $bin_{(0)}$ is set to zero, while (i) represents an integer index that starts at 1 and corresponds to the current bin(i). The interval value that differentiates the current bin from the previous one is denoted by (δ). For this experiment, we selected a value of 0.01 for (δ), which yielded the best results compared to other tested values. Figure 3 illustrates a comparison among different values of (δ).

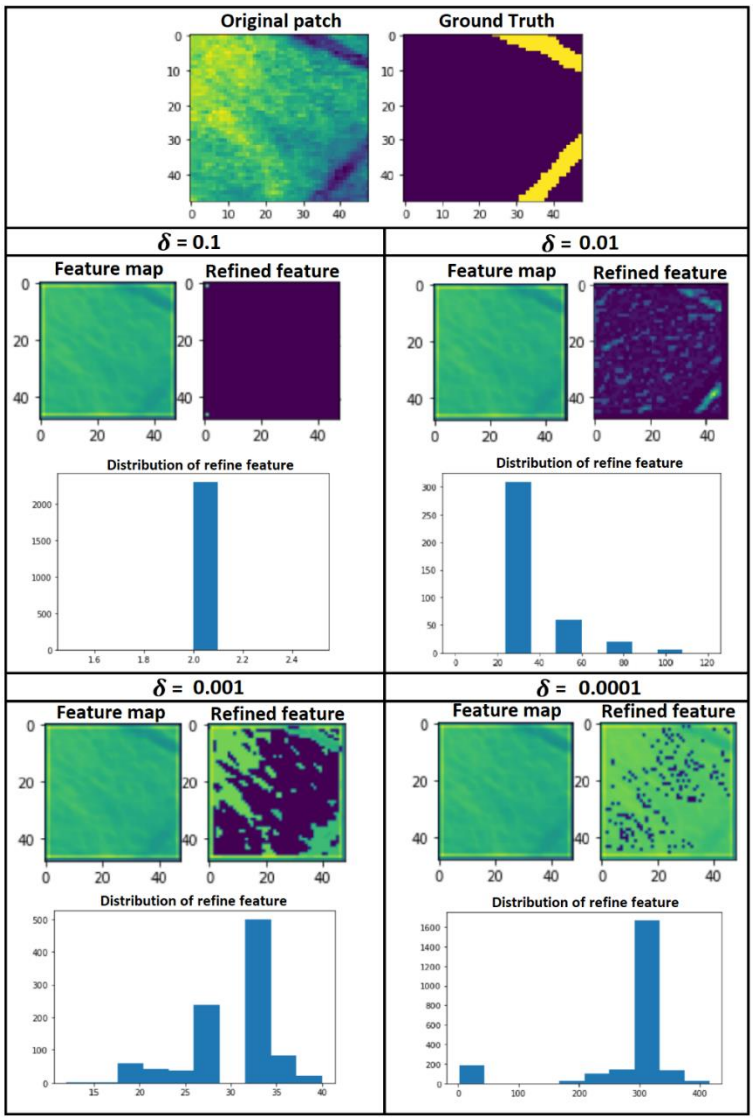

Fig. 3: experimental results of using different values of (δ).

If the value is 0.1, the feature map will be grouped into a single bin due to the large variation between bins compared to pixel values, resulting in a blackened feature map. Conversely, if the value is 0.0001, the feature map will have too many bins, making it difficult to identify which ones belong to the background region to suppress them. When the value is 0.001, the foreground pixels are retained well, but noise remains in the refined map, affecting the segmentation process. However, with a value of 0.01, most backgrounds are suppressed, and only significant pixels are preserved, achieving good performance.

Once the boundaries for each bin are defined, the bins will replace every pixel in the feature map with an integer index corresponding to the bin that contains the pixel's value. If x is a pixel value in the feature map $\ddot{X}_{i,c}^{L}$, then $\{\forall x : x \in \ddot{X}_{i,c}^{L}, x = \ddot{x}\}$, where c represents the channel dimensions and $\ddot{x}$ is:

$$\ddot{x} = \begin{cases} i\ , if\ x\ within\ the\ range\ of\ bin_{(i)} \\ x\ , \qquad\qquad otherwise \end{cases} \tag{3}$$

The effect of converting pixel values of a feature map into discretized integers is illustrated in Figure 4. The resulting feature map will contain integer values rather than float values, making the vessel boundaries sharper and easier to identify by the model, see Figure 4.

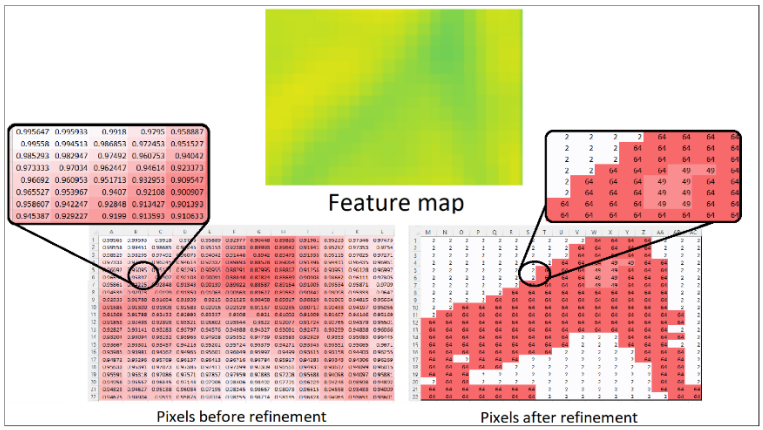

Fig. 4: A feature map with its corresponding pixel values before and after discretizing pixel values into bins.

The primary objective of converting pixels' values into integer bins is to obtain the histogram distribution of the given feature map and locate the bin with the highest peak. This helps identify the background pixels that occupy most of an image's pixels and assigns these pixels a low value. Similarly, the other pixels that belong to the foreground region are reassigned with higher values. Figure 5 displays samples of easy, moderate, and challenging cases refined by the suggested approach and their histogram distribution before and after implementing the proposed technique.

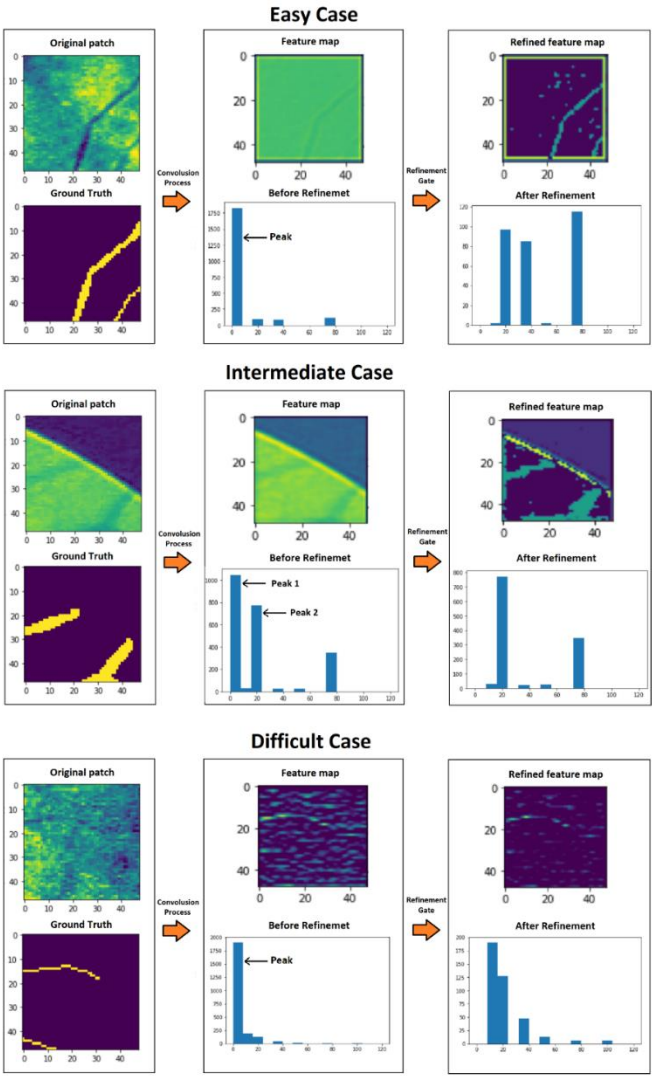

Fig. 5: Three distinct samples of feature maps with different severity, displayed before and after the execution of the proposed method to surpass the majority of background pixels.

The top row (easy case) demonstrates that vessels can be easily distinguished from the background, even in conditions of low image contrast. Conversely, the bottom row (difficult case) shows that the vessels are skinny, in similar contrast to some background pixels. In such cases, the proposed gate can suppress 75% of the background pixels and preserve only the vessels. In the middle row (moderate case), the challenge lies in differentiating the area inside and outside the FOV. When faced with this situation, the distribution graph will exhibit two peaks since the image is divided into two regions, making it difficult to determine which part to suppress. Consequently, it is possible to retain pixels outside the FOV rather than actual fundus pixels. After binarizing pixel values into bins, we exclude the distribution peak corresponding to the background pixels, increasing the values of other bins belonging to the foreground pixels. By doing this, we can identify which bins contain the pixels of interest and assign them higher weights, facilitating the model's ability to capture them easily. Prior to determining the corresponding bin to the peak of the distribution, the following equation is utilized:

$$px = argmax(bincount(\ddot{X}_{i,c}^{L})) \qquad (4)$$

The variables i and c denote the spatial and channel dimensions, respectively. Any pixel value in the feature map that matches the output of equation (4) will be replaced with a low value to suppress them, while other pixel values will be duplicated. See image x presented in Table 1, where Table 1 (A) and (B) represent the image before and after initializing background pixels with low values while assigning the pixels of interest with high values.

Table 1: The impact of implementing the refinement gate on a feature map. Part (A) of the table displays the original feature map values, while part (B) exhibits the feature map with refined pixel values.

0.904694	0.911244	0.91935	0.92555	0.931775	0.940256	0.951419	0.9605	0.966306
0.855275	0.864431	0.875556	0.884944	0.894325	0.904944	0.917975	0.9308	0.942688
0.827469	0.831044	0.837294	0.842338	0.847775	0.855331	0.868931	0.88535	0.902519
0.836425	0.828206	0.823756	0.819381	0.8179	0.820556	0.830138	0.841963	0.854738
0.880313	0.8623	0.847044	0.831838	0.821338	0.815469	0.81585	0.817794	0.820438
0.927813	0.909344	0.890188	0.868956	0.852356	0.839744	0.831294	0.823363	0.8164
0.957613	0.944156	0.927481	0.906506	0.888269	0.871156	0.856144	0.842881	0.832975
0.953594	0.942575	0.929956	0.914694	0.901438	0.886869	0.873506	0.863206	0.85555
0.921344	0.910119	0.901269	0.893238	0.887644	0.878656	0.86905	0.860956	0.855094
0.880506	0.872656	0.870281	0.871344	0.873788	0.871181	0.86675	0.861781	0.858038
0.865725	0.865388	0.869419	0.875819	0.881919	0.882969	0.881725	0.877756	0.874319
0.883975	0.888538	0.896113	0.905475	0.913869	0.9174	0.91745	0.91375	0.909994
0.922344	0.924769	0.929038	0.935088	0.941644	0.945438	0.945575	0.942169	0.938631

(A)

9	9	9	9	9	9	9	2	2	2	2	2	2	2
9	9	9	9	9	9	9	9	9	9	9	2	2	2
9	9	9	9	9	9	9	9	9	9	9	9	9	9
2	9	9	9	9	9	9	9	9	9	9	9	9	9
2	2	2	2	9	9	9	9	9	9	9	9	9	9
2	2	2	2	2	2	9	9	9	9	9	9	9	9
2	2	2	2	2	2	2	2	9	9	9	9	9	9
2	2	2	2	2	2	2	2	2	9	9	9	9	9
2	2	2	2	2	2	9	9	9	9	9	9	9	9
2	2	9	9	9	9	9	9	9	9	9	9	9	9
2	9	9	9	9	9	9	9	9	9	9	9	9	9
9	9	9	9	9	9	9	9	9	2	2	2	2	9
9	9	9	2	2	2	2	2	2	2	2	2	2	2

(B)

Assume that $bin_{(i)}$ cover the foreground pixels. In this case, the index (i) of the relevant bin replaces the pixel values and duplicates to preserve these features during subsequent downsampling processes. On the other hand, background pixels will be assigned a low value and will not be duplicated (refer to Table 1 (B)). Consequently, the resulting feature map will display clear and distinct differences between foreground and background pixels. Additionally, the refinement gate reallocates pixel weights within the feature map as appropriate, assigning greater importance to certain pixels. Figure 6 presents a flowchart to summarize the implementation steps of the proposed refinement gate. Figure 7 compares the output of certain intermediate layers before and after the refinement gate is applied. It is evident that, after the downsampling process, small vessels disappeared. However, the vessels are retained when the refinement gate is employed, while the background pixels are suppressed.

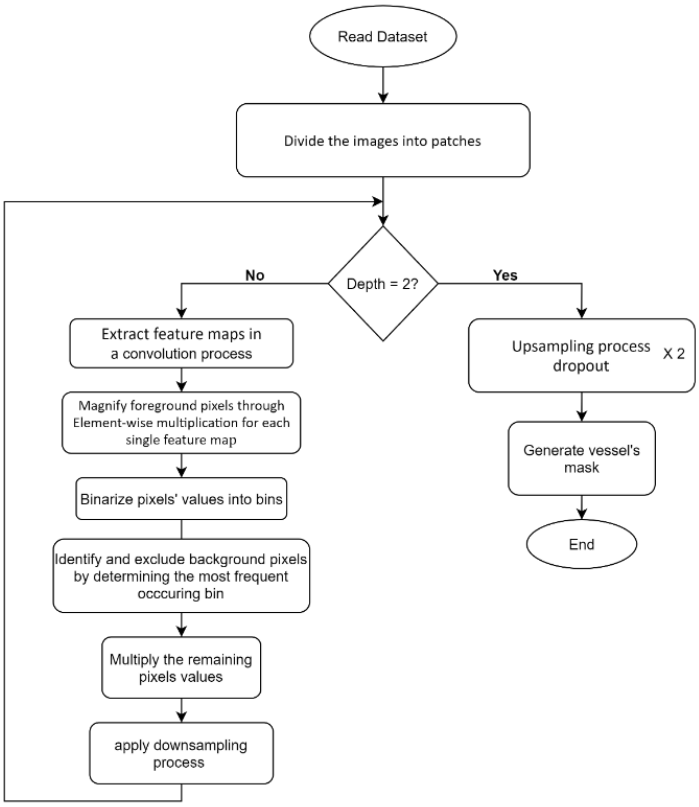

Fig. 6: Three distinct samples of feature maps with different severity, displayed before and after the execution of the proposed method to surpass the majority of background pixels.

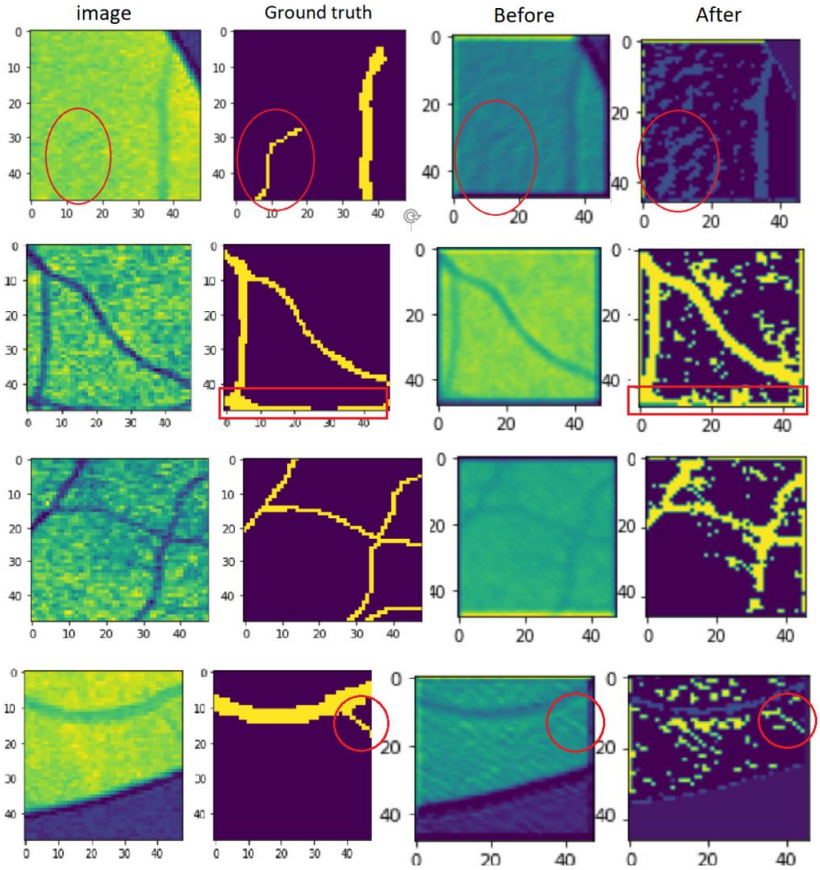

Fig. 7: The implementation of the refinement gate on some intermediate layers. The red circle indicates the vessels that disappeared and retained using the refinement gate.

4.0 METHOD

This section discusses the environmental setup used in the proposed method. Additionally, this section provides information on the datasets used, the preprocessing techniques, the evaluation metrics, and the structure of the proposed model.

4.1 Experimental Setup

The Keras library that operates on top of TensorFlow version 2.6.0 was used and performed on Python version 3.8.10 as an implementation platform to execute our approach. All stages of training were performed on a Linux server equipped with a Xeon® CPU (2.20 GHz) and 26 GB of RAM, along with a NVIDIA Tesla P100-PCIE-16GB GPU (Graphical Processing Unit), and CUDA Toolkit 11.2.

4.2 Datasets

In this work, extensive experimentation was conducted on three public datasets, namely: the DRIVE DB [53], the STARE DB [10] and the CHASE DB1 [54]. All these datasets offer a diverse range of retinal images with vascular structures but differ in image size, volume, and split strategy. For the DRIVE dataset, there are a total of 40 images available, each with the corresponding masks and ground truth annotations of the vessels. The 40 images are split into equal parts: 50% for training and 50% for testing. Only one manual annotation is given for each image as a gold standard in the training set. However, independent human graders gave each image two annotated masks in the test set.

The STARE DB comprises twenty images and two manual annotation sets provided by two independent experts, Valentina Kouznetsova (VK) and Adam Hoover (AH). We have considered the annotation by VK to be the reference standard. On the other hand, the CHASE_DB1 dataset contains images acquired from 14 children of different ages and ethnic backgrounds, and the images were taken from both eyes. The 28 fundus images are paired with corresponding standard annotations of vessels, which two annotators provided. Unlike the DRIVE DB and STARE DB, the images in the CHASE_DB1 dataset exhibit low contrast vessels and uneven background illumination.

As the STARE DB and CHASE_DB1 DB do not have any standard split for the train and test sets, we split the data in a 70:30 ratio, following the methodology of [55]. In this study, we considered the manual annotations of the first annotator as the ground truth for all the datasets, as performed in other studies by [15], [48]. Additionally, since the FOV masks were not provided in both the STARE and CHASE_DB1 datasets, we created them. Figure 8 showcases images from the DRIVE DB, STARE DB, and CHASE_DB1 DB, along with their respective ground truths.

Fig. 8: samples of retinal images with their ground-truth masks from the DRIVE dataset, the STARE dataset, and the CHASE_DB1 dataset from left to right, respectively.

The datasets used in this study had varying resolutions. For instance, the DRIVE dataset has images with 565×584 resolution and 45° FOV, whereas the STARE and CHASE-DB1 datasets have images with resolutions of 700×605 pixels and 999×960 pixels, with 35° and 30° FOVs, respectively. The method proposed by [56] is followed to prepare the images for the training stage. All images are transformed from three channels (RGB) to one channel (grayscale) to reduce the complexity of the model and improve the generalization capability. Then, the contrast-limited adaptive histogram equalization (CLAHE) was applied to adjust vessels and the contrast in the background. The CLAHE usually balances inhomogeneous illuminations and depends on the histogram equalization process. Then, after removing the artifacts of the tile borders, bilinear interpolation is applied, and gamma adjustment is performed to improve the segmentation performance. Finally, a set of 48×48 patches was extracted from the image. Figure 9 shows the preprocessing stages applied to images from the DRIVE dataset.

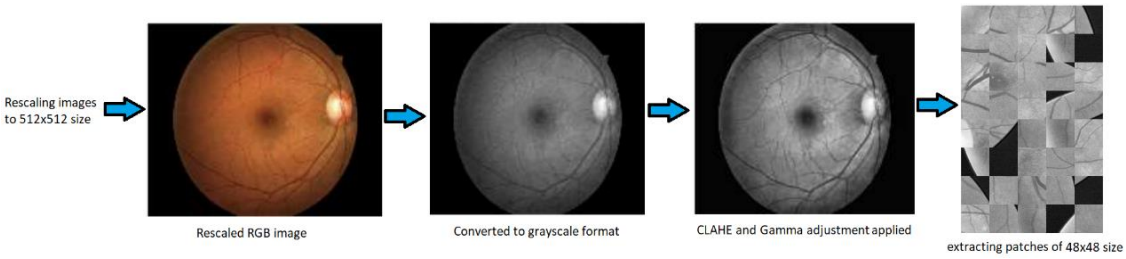

Fig. 9: preprocessing stages performed on image sample from the DRIVE dataset.

4.3 Architecture Design

The proposed method used a modified version of the U-net architecture as a default architecture. The modified U-net had two down-sampling stages and two up-sampling stages, unlike the original U-net. Each stage includes two convolutional layers with a dropout in between them to improve training performance. The proposed refinement gate is placed after the second convolutional layer in the encoder path, and its output is passed through the skip connection to the corresponding layer in the decoder path. The ReLu activation function was employed in both the down-sampling and up-sampling stages. The SoftMax activation function was used in the output layer of the model. The modified U-net has a smaller network size as two levels were removed compared to the original design, resulting in 32 feature maps at the first layer, increasing gradually to 128 at the bottleneck stage. The architecture of the proposed model is illustrated in Figure 10.

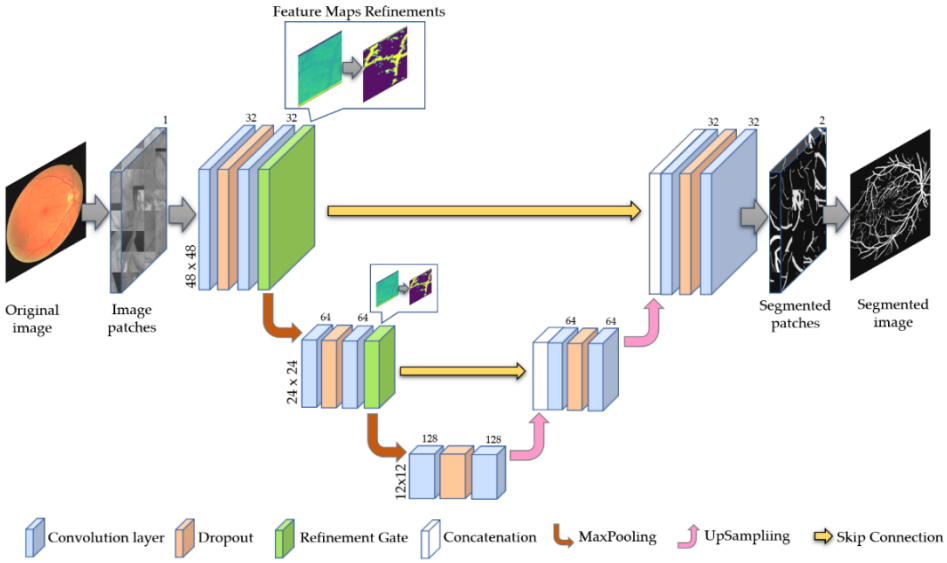

Fig. 10: Architecture of the proposed method.

During the training stage, the proposed method randomly extracted 9500 patches from each image, resulting in 190,000 patches from the 20 images in the DRIVE dataset. Of these, 171,000 patches were used for training and 19,000 for validation. These patches were of size 48×48 and were randomly selected from inside the entire image, even if they were partially or completely outside the FOV, to help the network distinguish blood vessels from the FOV border. For the STARE dataset, 14 images were randomly selected for training, and the remaining six images were used for testing, following a similar approach as in [55]. Approximately 133,500 patches were extracted from the 14 images for the training phase, and 57,000 patches were extracted from the six images in the testing set. Similar to the DRIVE dataset, 90% of the training patches (120.150) were used for training, and 10% (13.350) were used for validation. The same splitting strategy was applied to the CHASE_DB1 dataset, and the detailed summary of the training and testing sets with the number of extracted patches for all three datasets is provided in Table 2.

The hyperparameters used in this work are the cross-entropy loss function and the stochastic gradient descent as an optimization with a learning rate of 0.01, momentum of 0.3, and decay of 1e-6. The Rectifier Linear Unit (ReLU) activation function was used after every convolutional layer, and the dropout value was 0.2. The model was trained for 150 epochs with a batch size of 32 on a server specified in the experimental setup. Figures 11, 12, and 13 depict the training and validation loss and accuracy of the proposed model on the DRIVE DB, STARE DB, and CHASE_DB1 datasets, respectively.

Table 2: Configurations of the datasets used in this study.

	DRIVE DB (40 images)	STARE DB (20 images)	CHASE_DB1 (28 images
Splitting strategy	50% for training 50% for testing	70% for training 30% for testing	70% for training 30% for testing
Number of training and testing images	20 images for training 20 images for testing	14 images for training 6 images for testing	20 images for training 8 images for testing
Number of training patches	190,000 patches Training 90% Validation 10% 171,000 19,000	133,500 patches Training 90% Validation 10% 120.150 13.350	190,000 patches Training 90% Validation 10% 171,000 19,000
Number of testing patches	190,000 patches	57,000 patches	76,000 patches
Training time	150 sec per epoch	89 sec per epoch	143 sec per epoch

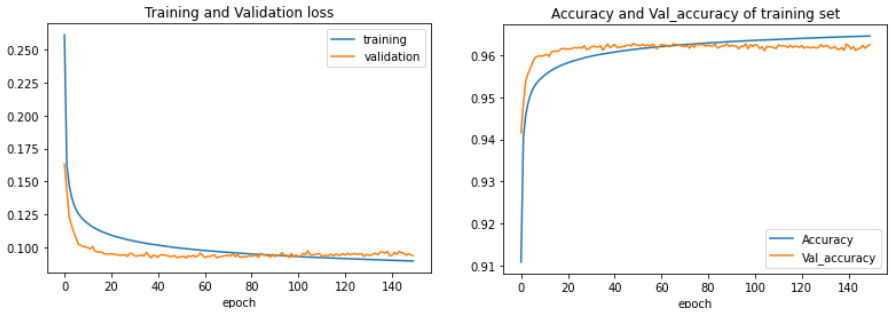

Fig. 11: The loss/validation loss and accuracy/validation accuracy while model trained on DRIVE dataset.

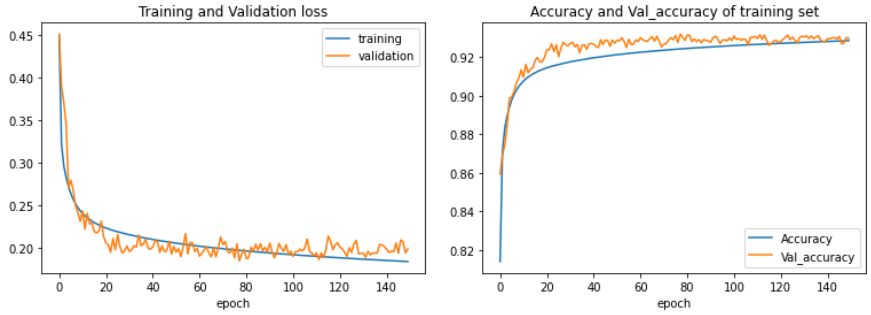

Fig. 12: The loss/validation loss and accuracy/validation accuracy while model trained on STARE dataset.

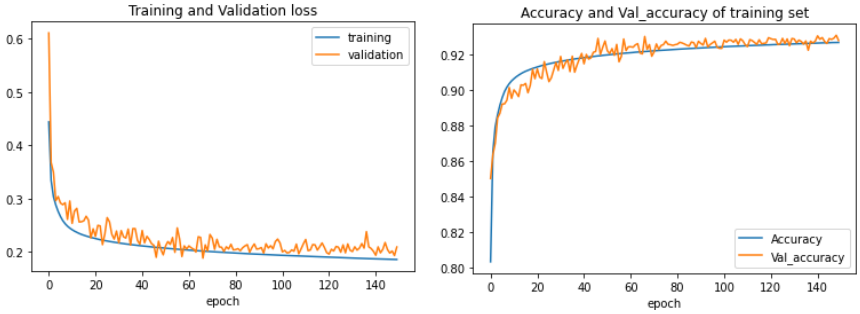

Fig. 13: The loss/validation loss and accuracy/validation accuracy while model trained on CHASE_DB1 dataset.

4.4 Evaluation Metrics

Evaluating a model's performance on imbalanced datasets with no standard splitting of training and testing sets, relying solely on a single evaluation metric like accuracy (Acc) may not accurately assess the classifier's performance. If the dataset has a small number of infected images and the model fails to detect them, the accuracy rate will be high, leading to an over-optimistic evaluation of the model's performance. Therefore, additional evaluation metrics are necessary to validate the results [57]. Specificity (Se), sensitivity (Sp), and area under the

receiver operating characteristic curve (AUC) are recommended as a minimum in artificial intelligence studies [58]. This study evaluates the proposed model using four metrics: Sp, Se, AUC, and Acc. Specifically (Sp) measures the ratio of pixels correctly classified as non-vessels (True Negative (TN)) to the total number of non-vessel pixels regardless of whether they were correctly classified as TN or incorrectly classified as False Positive (FP). The notation and equation for Sp are defined as follows:

$$Specificity = \frac{TN}{TN+FP} \quad (5)$$

Sensitivity (Se) is a measure that calculates the proportion of pixels correctly identified as vessels (True Positives (TP)) in relation to the total number of vessels pixels and non-classified vessels pixels (False Negatives (FN)). This concept is represented mathematically in the following equation:

$$Sensitvity = \frac{TP}{TP+FN} \quad (6)$$

Accuracy (Acc) is a measure that determines the proportion of all classified pixels, whether vessels or non-vessels, which were correctly identified TP, and TN compared to the overall classified pixels. This concept can be expressed mathematically in the following equation:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (7)$$

Additionally, the area under the curve (AUC) is an appropriate metric to accurately assess the performance of a model when dealing with imbalanced datasets in a segmentation task. AUC is computed by comparing the true positive rate for vessel classification (Se) and the true negative rate for non-vessel classification (Sp) through the following equation:

$$AUC = \frac{Se + Sp}{2} \quad (8)$$

5.0 RESULTS AND DISCUSSION

This study compared the results to the state-of-the-art studies that used DRIVE DB, STARE DB, and CHASE_DB1 databases. Furthermore, the annotations of the second expert were also included in the comparison. Table 3 presents the comparisons on three datasets, along with the results of the state-of-the-art methods. The optimal score of each evaluation metric is highlighted in bold. Results on the DRIVE dataset showed that the proposed model outperforms other methods presented in the literature and even surpasses human capability in all metrics except for Se, where the sensitivity is slightly lower than expert by 0.03 margin. However, for Sp and Acc, the proposed method outperforms experts by 0.012 and 0.007, respectively. The proposed method reports the highest AUC score, equivalent to the score of [59]. However, the proposed method is superior by 0.005 and 0.001 margins in terms of Sp and Acc, respectively, but inferior by 0.029 in Se.

The work proposed by [30] reported the highest Se score compared to other studies, and their score is 0.17 higher than ours. However, their Sp and Acc metrics performance is worse than ours by 0.054 and 0.011, respectively. The reason for the high Se score reported by (J. Yang et al., 2020) and the second-highest Se score achieved by [60] can be attributed to a large architecture with four networks and five layers in depth. Additionally, they utilized a generative adversarial network to enlarge the size of training images, a pre-trained VGG19 network, and a standalone refinement network to enhance the segmentation output. This huge architecture improved one metric (Se), but it did not lead to an improvement in Sp and Acc metrics.

In other words, the expected improvement did not meet the desired level, given the extensive computations, memory consumption, trainable parameters, and time invested in their architecture. In comparison, the proposed architecture has only two layers in-depth, 471k trainable parameters, and no additional standalone networks. Nevertheless, our method surpasses literature studies in terms of (Sp) and (Acc) metrics.

Additionally, the work proposed by [30] did not include the AUC, a preferred metric in the segmentation task of unbalanced classes. Lastly, although the degraded performance of the proposed method in terms of the (Se) metric, it still produced results comparable to other cutting-edge techniques and established new state-of-the-art scores in terms of Sp, Acc, and AUC.

Regarding the STARE dataset, the proposed model achieved the highest performance in Sp and AUC, close to the best score reported by [48] in Acc, with only a 0.004 difference. However, [49] achieved the closest Se score to ours, outperforming the proposed method by around 0.07.

Since the second annotator paid more attention to tiny vessels than the first observer, our Se metric results were lower than the second annotator's, with a 0.195 margin. Nevertheless, our results are more competitive in terms of Sp and Acc, indicating a lower false-negative rate.

The best score for Se was achieved by [26], who proposed a Dense U-net with a patch-based learning strategy. Unlike the proposed method, they used random transformations as a data augmentation strategy and surpassed human accuracy by 0.079. However, the proposed model performed better in terms of Sp and Acc. The low Se and Acc scores of our proposed model reported in this dataset do not necessarily indicate reduced capability in detecting small vessels. However, the model performance is influenced by the number of pathological images randomly chosen for the training set. The network performance will decrease if images with pathological issues are included. Study [20] demonstrated that the significant decrease in sensitivity metric was mainly due to the

inclusion of challenging images in the training setup. Nevertheless, the results obtained from this dataset indicate that our model can detect blood vessels accurately, and the results on the DRIVE and CHASE_DB1 datasets proved that.

It is worth noting that the AUC metric, unlike Se, Sp, and Acc metrics, does not depend on a threshold to produce the final segmentation. Most researchers, including [26], did not provide the AUC metric. Considering this, we found that our method achieved the best AUC score compared to other studies, aligning with the score of [61]. However, our method performed better in terms of Sp and Acc by margins of 0.009 and 0.004, respectively, and scored lower Se with 0.072 margin.

Table 3: Comparison of the proposed method with literature studies on the DRIVE, STARE, and CHASE_DB1 datasets. The symbol "-" means no contribution in the relevant metric in the referred dataset.

Dataset	DRIVE				STARE				CHASE_DB1			
	SE	SP	SE	SE	SE	SE	ACC	AUC	SE	SE	ACC	AUC
[62]	-	-	-	0.945	-	-	-	-	-	-	-	-
[63]	0.753	0.969	0.926	-	-	-	-	-	-	-	-	-
[64]	0.746	0.917	0.948	0.831	0.748	0.922	0.947	0.835	-	-	-	-
[61]	0.725	0.983	0.952	0.962	0.772	0.973	0.952	0.969	0.720	0.982	0.953	0.953
[65]	0.739	0.978	0.949	0.967	0.732	0.980	0.956	0.967	0.762	0.958	0.947	0.962
[66]	-	-	0.923	-	-	-	-	-	-	-	-	-
[67]	0.716	0.978	0.944	0.848	0.776	0.954	0.943	0.865	-	-	-	-
[60]	0.808	0.967	0.947	-	0.807	0.976	0.959	-	0.755	0.975	0.952	-
[68]	0.766	0.970	0.944	0.961	0.772	0.970	0.956	0.950	0.759	0.959	0.939	0.949
[19]	0.760	-	0.952	-	0.741	-	0.959	-	0.713	-	0.949	-
[69]	0.776	0.977	0.950	0.972	-	-	-	-	-	-	-	-
[70] **(LID-OS)**	0.747	0.976	0.947	0.952	0.768	0.976	0.955	0.961	0.756	0.968	0.946	0.957
[71]	0.757	0.982	0.953	0.974	-	-	-	-	0.751	0.979	0.958	0.972
[72]	-	-	-	0.973	-	-	-	-	-	-	-	-
[59]	0.777	0.979	0.953	0.975	-	-	-	-	-	-	-	-
[73]	-	-	-	0.965	-	-	-	-	-	-	-	-
[74]	0.739	0.956	0.948	0.844	0.748	0.962	0.947	0.855	-	-	-	-
[26]	0.798	0.973	0.951	0.974	**0.974**	0.972	0.954	-	-	-	-	-
[48]	0.763	-	0.954	-	0.742	-	**0.960**	-	0.782	-	0.959	-
[49]	0.756	0.970	0.942	-	0.720	0.973	0.948	-	-	-	-	-
[15]	-	-	-	-	-	-	-	-	**0.878**	0.968	0.960	0.958
[30]	**0.918**	0.930	0.943	-	-	-	-	-	-	-	-	-
2nd observer	0.776	0.972	0.947	-	0.895	0.938	0.935	-	0.743	0.979	0.954	-
Ours	0.748	0.984	**0.954**	**0.975**	0.700	**0.982**	0.956	**0.969**	0.704	**0.982**	**0.964**	**0.975**

Furthermore, despite the challenging images of the CHASE_DB1 datasets, the proposed method performed very well compared to other studies. Images in this dataset contain vessels with bright strips running down the middle of thick vessels and along their length, forming a central vessel reflex. Furthermore, the dataset includes low contrast and heterogeneous background illumination. The proposed method reported the best scores in terms of SP (0.982), Acc (0.964), and AUC (0.975). In terms of Se, the reported results were comparable to those of human observers, with a margin lower by 0.039. The highest Se value was reported by [15], who proposed a multi-path CNN that uses Gaussian low-pass and high-pass filters to capture local and global details of the image. They outperformed human graders by a margin of 0.135. However, their Sp, Acc, and AUC scores were lower than ours by margins of 0.074, 0.004, and 0.017, respectively.

In summary, the proposed method is capable of accurately segmenting blood vessels in three different datasets with sensitivity values that are very close to those of human experts. Additionally, our straightforward model architecture achieved the highest accuracy in the DRIVE DB and CHASE_DB1 and approached the best performance in the STARE dataset. These results confirm that the proposed method is robust and has the potential to be used in various clinical applications after testing on different data distributions.

Aside from the statistical comparisons, visual comparisons are also performed with state-of-the-art studies, as presented in Figures 14, 15, and 16. The figures present image samples from the DRIVE DB, STARE DB, and CHASE_DB1 datasets. To assess the improvements properly, several analyses were performed as follows:

Two images from each dataset were selected within each figure, and two local areas within each image were compared with other studies. A yellow rectangle bounds the first local area, while the second is bounded by a blue rectangle. Rows in the middle and bottom of each figure zoomed in on the yellow and blue rectangles, respectively. We use the studies of [30] and [15] as benchmarks for the DRIVE dataset to evaluate our model performance. As illustrated in Figure 14, our proposed method produces a more natural and accurate segmentation of the vessels' tree structure compared to other literature works. This is primarily due to the refinement gate, which focuses the model's attention on vessel pixels and suppresses unwanted pixels. Our segmentation results for thick vessels are

clearer and smoother than other methods. In contrast, [30] and [15] fail to segment tiny vessels, while the proposed method successfully detects these small details with the assistance of the refinement gate, as visually demonstrated by the four magnified patches in Figure 14.

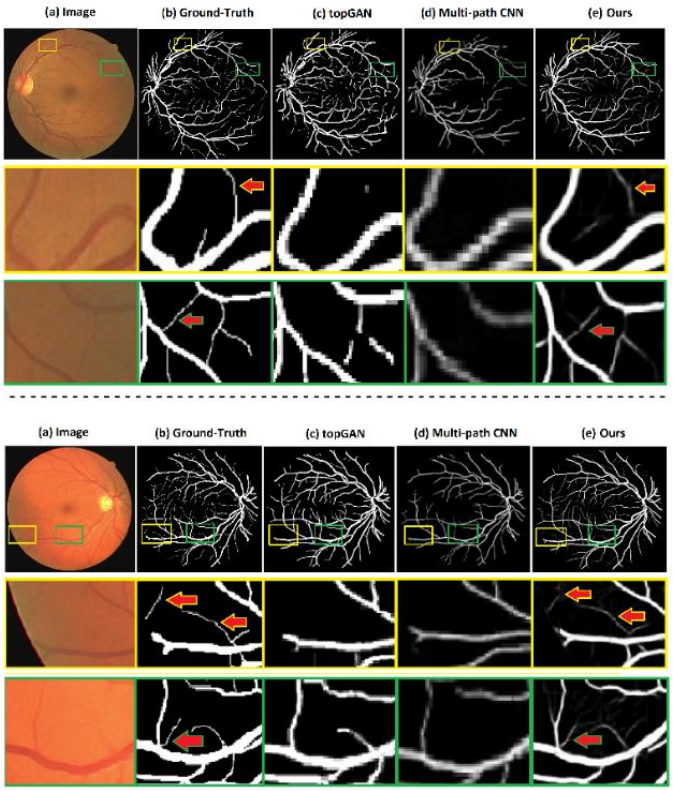

Fig. 14: A visual comparison using the DRIVE dataset. (a) original image, (b) the ground-truth annotations, (c) the result of the topGAN method by [30], (d) the result of the Multi-path CNN method by [15], and (e) the results of the proposed method. The yellow and green spots are magnified in the second and third rows.

The methods by [26] and [49] were compared in the STARE dataset. According to [75], STARE is the best dataset set used to validate the performance of a model because nearly half of the images are challenging due to the existence of pathological issues. Our proposed method identified more small vessels than those shown in Figure 15 and performed well in segmenting vessels with multiple branches and maintaining vessel connectivity, which are difficult tasks in the current literature [66]. Dense U-net [26] performed better than fusion multiple features [49], as shown in Figure 15. However, our method outperformed both studies regarding connectivity and segmentation of small vessels. The second image further validates the performance of our model against the Dense U-net architecture.

Figure 16 exhibits the efficiency of the refinement gate in segmenting vascular branches that are too close to each other, which is a difficult task, especially in low-contrast areas [59]. The proposed method successfully segmented these normal and other small vessels, as shown in yellow and green spots in the first image of Figure 16. In contrast to other studies [71] and [15], which used cross-modularity learning and multi-path CNN methods, they segmented these vessels as a single branch and failed to detect some thin and tiny vessels. In contrast, the proposed method was better at segmenting vessels close to each other. Further analysis tests the proposed method on another image used by [71] regardless [15]. The second image in Figure 16 presents the capability of our method in segmenting small vessels that could not be segmented by the cross-modularity learning method. In conclusion, these visual experiments confirm that our proposed method can handle complex images with even illumination and central light reflex.

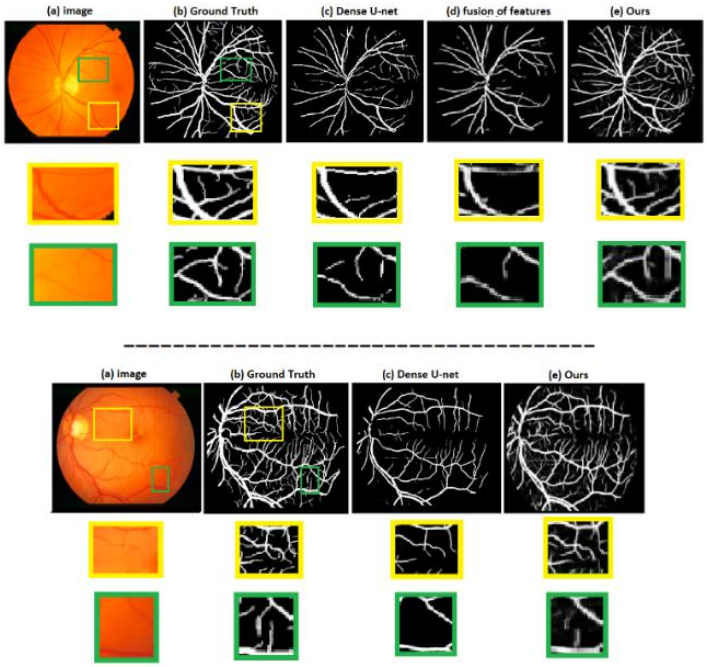

Fig. 15: A visual comparison using STARE dataset. (a) original image, (b) the ground-truth annotations, (c) the result of Dense U-net method [26], (d) the result of the feature fusion method [49], and (e) the results of our method. The yellow and green spots are magnified in the second and third rows.

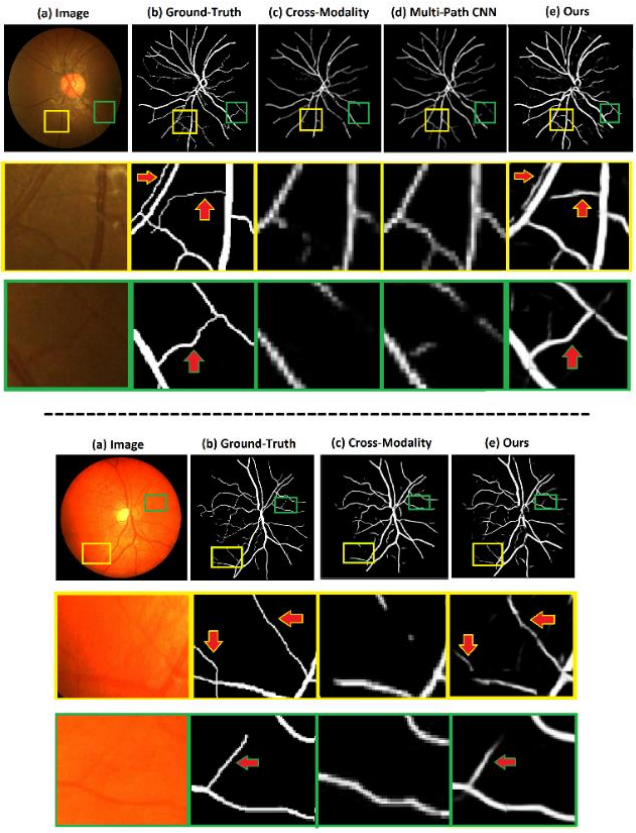

Fig. 16: A visual comparison using CHASE_DB1 dataset. (a) Original image, (b) ground-truth annotations, (c) the segmentation result of Dense U-net [71], (d) the segmentation result of the feature fusion method [15], and (e) the results of our method. The yellow and green spots are magnified in the second and third rows.

6.0 LIMITATIONS OF THE STUDY

Both qualitative and quantitative comparisons confirm the efficiency of our method, exhibiting accurate segmentation performance. Nonetheless, several limitations have been observed. Firstly, while attempting to segment small vessels in low-contrast regions, some background pixels are mistakenly segmented as vessel pixels,

affecting the sensitivity metric and degrading performance. Secondly, the method was tested on relatively small datasets without data augmentation methods. Augmentation techniques can significantly improve model performance, as demonstrated in literature studies [76], [77], [78]. Lastly, the proposed model has not experimented with different data distributions, highlighting the need for additional implementation, such as cross-dataset training, to effectively demonstrate its efficacy in real-world applications.

7.0 CONCLUSION

This research aims to present a new method for segmenting fine and tiny vessels located in the low-contrast region of fundus images. The proposed method employs a modified U-net as the backbone architecture and the refinement gate is introduced to enhance the segmentation performance by highlighting small vessels in low-contrast areas that may have vanished due to the downsampling process. This gate helps the model to differentiate between vessel pixels and background pixels by accentuating small vessels' values located in low-contrast regions or vanished due to the subsequent downsampling processes. Notably, this is the first study that emphasizes the pixels of interest while surpassing the background pixels without utilizing any additional standalone network or additional trainable parameters.

The quantitative and visual experiments on the three datasets (DRIVE DB, STARE DB, and CHASE_DB1) demonstrated that the proposed method outperformed other studies in terms of Sp, AUC, and Acc, reaching a state-of-the-art level. These promising results demonstrate the model's potential to be used as a second opinion for ophthalmologists in accurately segmenting blood vessels. Future research will investigate more complex backbone architectures to enhance model's capabilities and improve the performance. In addition to incorporating data augmentation techniques to improve generalization

ACKNOWLEDGEMENT

This work was funded by the Ministry of Higher Education (MOHE), Malaysia, under the Fundamental Research Grant Scheme (FRGS), FRGS/1/2019/ICT02/UKM/02/9. A special thanks to the Universiti Kebangsaan Malaysia for their support and resources in the Cyber Analytics Lab.

REFERENCES

[1] D. Youssef and N. H. Solouma, "Accurate detection of blood vessels improves the detection of exudates in color fundus images," *Comput. Methods Programs Biomed.*, vol. 108, no. 3, pp. 1052–1061, 2012, doi: 10.1016/j.cmpb.2012.06.006.

[2] Q. Qin and Y. Chen, "A review of retinal vessel segmentation for fundus image analysis," *Eng. Appl. Artif. Intell.*, vol. 128, Feb. 2024, doi: 10.1016/j.engappai.2023.107454.

[3] Y. Li *et al.*, "Global Transformer and Dual Local Attention Network via Deep-Shallow Hierarchical Feature Fusion for Retinal Vessel Segmentation," *IEEE Trans. Cybern.*, vol. 53, no. 9, pp. 5826–5839, Sep. 2023, doi: 10.1109/TCYB.2022.3194099.

[4] M. A. Palomera-Pérez, M. E. Martinez-Perez, H. Benítez-Pérez, and J. L. Ortega-Arjona, "Parallel multiscale feature extraction and region growing: Application in retinal blood vessel detection," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 500–506, 2010, doi: 10.1109/TITB.2009.2036604.

[5] M. Vlachos and E. Dermatas, "Multi-scale retinal vessel segmentation using line tracking," *Comput. Med. Imaging Graph.*, vol. 34, no. 3, pp. 213–227, 2010, doi: 10.1016/j.compmedimag.2009.09.006.

[6] A. Adam, A. H. A. Rahman, N. S. Sani, Z. A. A. Alyessari, N. J. Z. Mamat, and B. Hasan, "Epithelial Layer Estimation Using Curvatures and Textural Features for Dysplastic Tissue Detection," *Comput. Mater. Contin.*, vol. 67, no. 1, pp. 761–777, 2021, doi: 10.32604/cmc.2021.014599.

[7] A. S. Zaini, S. N. H. Sheikh Abdullah, A. Abdullah, N. Budin Sana, and S. Basiron, "Exploitable Edge Analysis for Free Flow Vehicle Plate Localization," *Asia-Pacific J. Inf. Technol. Multimed.*, vol. 10, no. 01, pp. 27–42, Jun. 2021, doi: 10.17576/apjitm-2021-1001-03.

[8] M. F. Nasrudin, K. Omar, C. Y. Liong, and M. S. Zakaria, "Object signature features selection for handwritten jawi recognition," *Adv. Intell. Soft Comput.*, vol. 79, pp. 689–698, 2010, doi: 10.1007/978-3-642-14883-5_88.

[9] W. Abu-Ain, S. N. H. Sheikh Abdullah, and K. Omar, "A simple iterative thinning algorithm for text and shape binary images," *J. Theor. Appl. Inf. Technol.*, vol. 63, no. 2, pp. 274–281, 2014.

[10] A. Hoover, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imaging*, vol. 19, no. 3, pp. 203–210, 2000, doi: 10.1109/42.845178.

[11] B. Bataineh, S. N. H. S. Abdullah, and K. Omar, "Adaptive binarization method for degraded document

images based on surface contrast variation," *Pattern Anal. Appl.*, vol. 20, no. 3, pp. 639–652, Aug. 2017, doi: 10.1007/s10044-015-0520-0.

[12] S. Q. H. and S. E. Raid Rafi Omar Al-Nima, "Exploiting the Deep Learning with Fingerphotos to Recognize People," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 7, pp. 13035–13046, 2020, Accessed: Feb. 06, 2022. [Online]. Available: http://sersc.org/journals/index.php/IJAST/article/view/28984

[13] N. Z. Waisi, N. Y. Abdullah, and M. T. Ghazal, "The automatic detection of underage troopers from live-videos based on deep learning," *Prz. Elektrotechniczny*, vol. 2021, no. 9, pp. 85–88, Sep. 2021, doi: 10.15199/48.2021.09.18.

[14] F. Li *et al.*, "Joint optic disk and cup segmentation for glaucoma screening using a region-based deep learning network," *Eye*, vol. 37, no. 6, pp. 1080–1087, Apr. 2023, doi: 10.1038/s41433-022-02055-w.

[15] C. Tian, T. Fang, Y. Fan, and W. Wu, "Multi-path convolutional neural network in fundus segmentation of blood vessels," *Biocybern. Biomed. Eng.*, vol. 40, no. 2, pp. 583–595, 2020, doi: 10.1016/j.bbe.2020.01.011.

[16] J. Zilly, J. M. Buhmann, and D. Mahapatra, "Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation," *Comput. Med. Imaging Graph.*, vol. 55, pp. 28–41, 2017, doi: 10.1016/j.compmedimag.2016.07.012.

[17] M. Aziz, A. Bakar, and M. Yaakub, "CoreNLP dependency parsing and pattern identification for enhanced opinion mining in aspect-based sentiment analysis," *J. ofKing Saud Univ. - Comput. Inf. Sci.*, 2024, doi: https://doi.org/10.1016/j.jksuci.2024.102035.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[19] H. Fu, Y. Xu, S. Lin, D. W. K. Wong, and J. Liu, "Deepvessel: Retinal vessel segmentation via deep learning and conditional random field," in *International conference on medical image computing and computer-assisted intervention*, Springer Verlag, 2016, pp. 132–139. doi: 10.1007/978-3-319-46723-8_16.

[20] Z. Jiang, H. Zhang, Y. Wang, and S. B. Ko, "Retinal blood vessel segmentation using fully convolutional network with transfer learning," *Comput. Med. Imaging Graph.*, vol. 68, pp. 1–15, 2018, doi: 10.1016/j.compmedimag.2018.04.005.

[21] C. Li, Z. Li, and W. Liu, "TDCAU-Net: retinal vessel segmentation using transformer dilated convolutional attention-based U-Net method," *Phys. Med. Biol.*, vol. 69, no. 1, Jan. 2024, doi: 10.1088/1361-6560/ad1273.

[22] H. Su, L. Gao, Z. Wang, Y. Yu, J. Hong, and Y. Gao, "A Hierarchical Full-Resolution Fusion Network and Topology-aware Connectivity Booster for Retinal Vessel Segmentation," *IEEE Trans. Instrum. Meas.*, 2024, doi: 10.1109/TIM.2024.3411133.

[23] T. Iqbal and H. Ali, "Generative Adversarial Network for Medical Images (MI-GAN)," *J. Med. Syst.*, vol. 42, no. 11, p. 231, Nov. 2018, doi: 10.1007/s10916-018-1072-9.

[24] J. Ma, M. Wei, Z. Ma, L. Shi, and K. Zhu, "Retinal vessel segmentation based on Generative Adversarial network and Dilated convolution," in *14th International Conference on Computer Science and Education, ICCSE 2019*, IEEE, 2019, pp. 282–287. doi: 10.1109/ICCSE.2019.8845491.

[25] W. Tu, W. Hu, X. Liu, and J. He, "DRPAN: A novel adversarial network approach for retinal vessel segmentation," in *Proceedings of the 14th IEEE Conference on Industrial Electronics and Applications, ICIEA 2019*, IEEE, 2019, pp. 228–232. doi: 10.1109/ICIEA.2019.8833908.

[26] C. Wang, Z. Zhao, Q. Ren, Y. Xu, and Y. Yu, "Dense U-net based on patch-based learning for retinal vessel segmentation," *Entropy*, vol. 21, no. 2, 2019, doi: 10.3390/e21020168.

[27] H. R. Roth *et al.*, "Hierarchical 3D fully convolutional networks for multi-organ segmentation," *arxiv.org*, 2017, Accessed: Oct. 30, 2021. [Online]. Available: http://arxiv.org/abs/1704.06382

[28] M. Khened, V. A. Kollerathu, and G. Krishnamurthi, "Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers," *Med. Image Anal.*, vol. 51, pp. 21–45, Jan. 2019, doi: 10.1016/j.media.2018.10.004.

[29] H. R. Roth *et al.*, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," *Med. Image Anal.*, vol. 45, pp. 94–107, 2018, doi: 10.1016/j.media.2018.01.006.

[30] J. Yang *et al.*, "Fully Automatic Arteriovenous Segmentation in Retinal Images via Topology-Aware

Generative Adversarial Networks," *Interdiscip. Sci. Comput. Life Sci.*, vol. 12, no. 3, pp. 323–334, 2020, doi: 10.1007/s12539-020-00385-5.

[31] Q. Li *et al.*, "A high-speed end-to-end approach for retinal arteriovenous segmentation," *Proc. - 2017 10th Int. Congr. Image Signal Process. Biomed. Eng. Informatics, CISP-BMEI 2017*, vol. 2018-Janua, pp. 1–5, 2018, doi: 10.1109/CISP-BMEI.2017.8301975.

[32] Y. Luo, L. Yang, L. Wang, and H. Cheng, "Efficient CNN-CRF network for retinal image segmentation," in *Communications in Computer and Information Science*, Springer Verlag, 2017, pp. 157–165. doi: 10.1007/978-981-10-5230-9 17.

[33] B. J. Bhatkalkar, D. R. Reddy, S. Prabhu, and S. V. Bhandary, "Improving the Performance of Convolutional Neural Network for the Segmentation of Optic Disc in Fundus Images Using Attention Gates and Conditional Random Fields," *IEEE Access*, vol. 8, pp. 29299–29310, 2020, doi: 10.1109/ACCESS.2020.2972318.

[34] M. Juneja, S. Thakur, A. Wani, A. Uniyal, N. Thakur, and P. Jindal, "DC-Gnet for detection of glaucoma in retinal fundus imaging," *Mach. Vis. Appl.*, vol. 31, no. 5, Jul. 2020, doi: 10.1007/s00138-020-01085-2.

[35] Y. Ye, C. Pan, Y. Wu, S. Wang, and Y. Xia, "MFI-Net: Multiscale Feature Interaction Network for Retinal Vessel Segmentation," *IEEE J. Biomed. Heal. Informatics*, vol. 26, no. 9, pp. 4551–4562, Sep. 2022, doi: 10.1109/JBHI.2022.3182471.

[36] F. Sultana, A. Sufian, and P. Dutta, "RIMNet: Image Magnification Network with Residual Block for Retinal Blood Vessel Segmentation," *2022 IEEE Reg. 10 Symp. TENSYMP 2022*, 2022, doi: 10.1109/TENSYMP54529.2022.9864467.

[37] I. J. Goodfellow *et al.*, "Generative Adversarial Nets," *Adv. Neural Inf. Process. Syst. 27 (NIPS 2014)*, vol. 27, pp. 2672--2680, 2014, doi: 10.1109/ICCVW.2019.00369.

[38] A. Q. Saeed, S. N. H. Sheikh Abdullah, J. Che-Hamzah, A. T. Abdul Ghani, and W. A. karim Abu-ain, "Synthesizing Retinal Images using End-To-End VAEs-GAN Pipeline-Based Sharpening and Varying Layer," *Multimed. Tools Appl.*, vol. 83, no. 1, pp. 1283–1307, Oct. 2024, doi: 10.1007/s11042-023-17058-2.

[39] L. Yu, W. Zhang, J. Wang, and Y. Yu, "SeqGAN: Sequence generative adversarial nets with policy gradient," *Thirty-First AAAI Conf. Artif. Intell.*, vol. 31, no. 1, pp. 2852–2858, 2017, Accessed: Sep. 21, 2021. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/10804

[40] P. Isola, J. Zhu, A. A. Efros, B. Ai, and U. C. Berkeley, "Image-to-Image Translation with Conditional Adversarial Networks," in *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5967–5976. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017/html/Isola_Image-To-Image_Translation_With_CVPR_2017_paper.html

[41] A. Kumar, S. Juliet, and M. M. Anishin Raj, "Retinal Fundus Image Synthesis using Image Segmentation and DCGAN," *Proc. 2024 10th Int. Conf. Commun. Signal Process. ICCSP 2024*, pp. 1089–1094, 2024, doi: 10.1109/ICCSP60870.2024.10544132.

[42] S. A. Rammy, W. Abbas, N. U. Hassan, A. Raza, and W. Zhang, "CPGAN: Conditional patch-based generative adversarial network for retinal vessel segmentation," *IET Image Process.*, vol. 14, no. 6, pp. 1081–1090, 2020, doi: 10.1049/iet-ipr.2019.1007.

[43] G. Haoqi and K. Ogawara, "CGAN-based Synthetic Medical Image Augmentation between Retinal Fundus Images and Vessel Segmented Images," in *2020 5th International Conference on Control and Robotics Engineering, ICCRE 2020*, 2020, pp. 218–223. doi: 10.1109/ICCRE49379.2020.9096438.

[44] A. Q. Saeed, S. N. H. S. Abdullah, J. Che-Hamzah, and A. T. A. Ghani, "Accuracy of using generative adversarial networks for glaucoma detection: Systematic review and bibliometric analysis," *J. Med. Internet Res.*, vol. 23, no. 9, 2021, doi: 10.2196/27414.

[45] J. Son, S. J. Park, and K. H. Jung, "Towards Accurate Segmentation of Retinal Vessels and the Optic Disc in Fundoscopic Images with Generative Adversarial Networks," *J. Digit. Imaging*, vol. 32, no. 3, pp. 499–512, Jun. 2019, doi: 10.1007/s10278-018-0126-3.

[46] H. Zhao, X. Qiu, W. Lu, H. Huang, and X. Jin, "High-quality retinal vessel segmentation using generative adversarial network with a large receptive field," *Int. J. Imaging Syst. Technol.*, vol. 30, no. 3, pp. 828–842, 2020, doi: 10.1002/ima.22428.

[47] A. K. Pandey, S. P. Singh, and C. Chakraborty, "Residual attention UNet GAN Model for enhancing the intelligent agents in retinal image analysis," *Serv. Oriented Comput. Appl.*, 2024, doi: 10.1007/s11761-024-

00415-w.

[48] Y. Lin, H. Zhang, and G. Hu, “Automatic Retinal Vessel Segmentation via Deeply Supervised and Smoothly Regularized Network,” *IEEE Access*, vol. 7, no. c, pp. 57717–57724, 2019, doi: 10.1109/ACCESS.2018.2844861.

[49] Y. Yang, F. Shao, Z. Fu, and R. Fu, “Discriminative dictionary learning for retinal vessel segmentation using fusion of multiple features,” *Signal, Image Video Process.*, vol. 13, no. 8, pp. 1529–1537, Nov. 2019, doi: 10.1007/s11760-019-01501-9.

[50] Y. Prajna and M. K. Nath, “Efficient blood vessel segmentation from color fundus image using deep neural network,” *J. Intell. Fuzzy Syst.*, vol. 42, no. 4, pp. 3477–3489, Mar. 2022, doi: 10.3233/JIFS-211479.

[51] N. Yadav, “A deep data-driven approach for enhanced segmentation of blood vessel for diabetic retinopathy,” *Int. J. Imaging Syst. Technol.*, vol. 32, no. 5, pp. 1696–1708, Sep. 2022, doi: 10.1002/ima.22720.

[52] Z. Wei and G. H. Liu, “Image retrieval using the intensity variation descriptor,” *Math. Probl. Eng.*, vol. 2020, 2020, doi: 10.1155/2020/6283987.

[53] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” *IEEE Trans. Med. Imaging*, vol. 23, no. 4, pp. 501–509, Apr. 2004, doi: 10.1109/TMI.2004.825627.

[54] C. G. Owen *et al.*, “Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (caiar) program,” *Investig. Ophthalmol. Vis. Sci.*, vol. 50, no. 5, pp. 2004–2010, 2009, doi: 10.1167/iovs.08-3018.

[55] İ. Atli and O. S. Gedik, “Sine-Net: A fully convolutional deep learning architecture for retinal blood vessel segmentation,” *Eng. Sci. Technol. an Int. J.*, 2020, doi: 10.1016/j.jestch.2020.07.008.

[56] W. Xiancheng *et al.*, “Retina Blood Vessel Segmentation Using A U-Net Based Convolutional Neural Network,” in *International Conference on Data Science*, 2018, pp. 1–11. Accessed: Sep. 11, 2021. [Online]. Available: https://www.semanticscholar.org/paper/International-Conference-on-Data-Science-(-ICDS-)-A-Xiancheng-Wei/64871f99ccc255a26277f2b80b5b154a33d61627#citing-papers

[57] M. N. Bajwa *et al.*, “Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning,” *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, Jul. 2019, doi: 10.1186/s12911-019-0842-8.

[58] D. S. W. Ting *et al.*, “Deep learning in ophthalmology: The technical and clinical considerations,” *Prog. Retin. Eye Res.*, vol. 72, Sep. 2019, doi: 10.1016/j.preteyeres.2019.04.003.

[59] K. Hu *et al.*, “Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function,” *Neurocomputing*, vol. 309, pp. 179–191, 2018, doi: 10.1016/j.neucom.2018.05.011.

[60] L. Zhou, Q. Yu, X. Xu, Y. Gu, and J. Yang, “Improving dense conditional random field for retinal vessel segmentation by discriminative feature learning and thin-vessel enhancement,” *Comput. Methods Programs Biomed.*, vol. 148, pp. 13–25, 2017, doi: 10.1016/j.cmpb.2017.06.016.

[61] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, “Blood vessel segmentation of fundus images by major vessel extraction and subimage classification,” *IEEE J. Biomed. Heal. Informatics*, vol. 19, no. 3, pp. 1118–1128, 2015, doi: 10.1109/JBHI.2014.2335617.

[62] A. Lahiri, K. Ayush, P. K. Biswas, and P. Mitra, “Generative Adversarial Learning for Reducing Manual Annotation in Semantic Segmentation on Large Scale Miscroscopy Images: Automated Vessel Segmentation in Retinal Fundus Image as Test Case,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2017-July, pp. 794–800, 2017, doi: 10.1109/CVPRW.2017.110.

[63] J. H. Tan, U. R. Acharya, S. V. Bhandary, K. C. Chua, and S. Sivaprasad, “Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network,” *J. Comput. Sci.*, vol. 20, pp. 70–79, 2017, doi: 10.1016/j.jocs.2017.02.006.

[64] T. A. Soomro *et al.*, “Boosting Sensitivity of a Retinal Vessel Segmentation Algorithm with Convolutional Neural Network,” *DICTA 2017 - 2017 Int. Conf. Digit. Image Comput. Tech. Appl.*, vol. 2017-Decem, pp. 1–8, 2017, doi: 10.1109/DICTA.2017.8227413.

[65] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, “Iterative Vessel Segmentation of Fundus Images,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1738–1749, 2015, doi: 10.1109/TBME.2015.2403295.

[66] X. Xu, W. Ding, M. D. Abràmoff, and R. Cao, “An improved arteriovenous classification method for the

early diagnostics of various diseases in retinal image," *Comput. Methods Programs Biomed.*, vol. 141, pp. 3–9, 2017, doi: 10.1016/j.cmpb.2017.01.007.

[67] Y. Zhao, L. Rada, K. Chen, S. P. Harding, and Y. Zheng, "Automated Vessel Segmentation Using Infinite Perimeter Active Contour Model with Hybrid Region Information with Application to Retinal Images," *IEEE Trans. Med. Imaging*, vol. 34, no. 9, pp. 1797–1807, 2015, doi: 10.1109/TMI.2015.2409024.

[68] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable COSFIRE filters for vessel delineation with application to retinal images," *Med. Image Anal.*, vol. 19, no. 1, pp. 46–57, 2015, doi: 10.1016/j.media.2014.08.002.

[69] P. Liskowski and K. Krawiec, "Segmenting Retinal Blood Vessels with Deep Neural Networks," *IEEE Trans. Med. Imaging*, vol. 35, no. 11, pp. 2369–2380, 2016, doi: 10.1109/TMI.2016.2546227.

[70] J. Zhang, B. Dashtbozorg, E. Bekkers, J. P. W. Pluim, R. Duits, and B. M. Ter Haar Romeny, "Robust Retinal Vessel Segmentation via Locally Adaptive Derivative Frames in Orientation Scores," *IEEE Trans. Med. Imaging*, vol. 35, no. 12, pp. 2631–2644, 2016, doi: 10.1109/TMI.2016.2587062.

[71] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE Trans. Med. Imaging*, vol. 35, no. 1, pp. 109–118, 2016, doi: 10.1109/TMI.2015.2457891.

[72] Y. Guo, Ü. Budak, and A. Şengür, "A novel retinal vessel detection approach based on multiple deep convolution neural networks," *Comput. Methods Programs Biomed.*, vol. 167, pp. 43–48, 2018, doi: 10.1016/j.cmpb.2018.10.021.

[73] Y. Guo, Ü. Budak, L. J. Vespa, E. Khorasani, and A. Şengür, "A retinal vessel detection approach using convolution neural network with reinforcement sample learning strategy," *Meas. J. Int. Meas. Confed.*, vol. 125, pp. 586–591, 2018, doi: 10.1016/j.measurement.2018.05.003.

[74] T. A. Soomro, O. Hellwich, A. J. Afifi, M. Paul, J. Gao, and L. Zheng, "Strided U-Net Model: Retinal Vessels Segmentation using Dice Loss," in *2018 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2018*, IEEE, 2019. doi: 10.1109/DICTA.2018.8615770.

[75] T. A. Soomro *et al.*, "Deep Learning Models for Retinal Blood Vessels Segmentation: A Review," *IEEE Access*, vol. 7, pp. 71696–71717, 2019, doi: 10.1109/ACCESS.2019.2920616.

[76] N. Asiri, M. Hussain, F. Al Adel, and N. Alzaidi, "Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey," *Artif. Intell. Med.*, vol. 99, Aug. 2019, doi: 10.1016/j.artmed.2019.07.009.

[77] T. R. V. Bisneto, A. O. de Carvalho Filho, and D. M. V. Magalhães, "Generative adversarial network and texture features applied to automatic glaucoma detection," *Appl. Soft Comput. J.*, vol. 90, p. 106165, May 2020, doi: 10.1016/j.asoc.2020.106165.

[78] Y. Jiang, N. Tan, and T. Peng, "Optic Disc and Cup Segmentation Based on Deep Convolutional Generative Adversarial Networks," *IEEE Access*, vol. 7, pp. 64483–64493, 2019, doi: 10.1109/ACCESS.2019.2917508.