M-GAN: Retinal Blood Vessel Segmentation by Balancing Losses Through Stacked Deep Fully Convolutional Networks

KYEONG-BEOM PARK, SUNG HO CHOI, AND JAE YEOL LEE, (Member, IEEE)
Department of Industrial Engineering, Chonnam National University, Gwangju 61186, South Korea
Corresponding author: Jae Yeol Lee (jaeyeol@jnu.ac.kr)

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education under Grant 2019R1I1A3A01059082, and in part by the Korea Health Technology Research and Development Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare under Grant HI19C0642.

ABSTRACT Until now, the human expert segments retinal blood vessels manually in fundus images to inspect human retinal-related diseases, such as diabetic retinopathy and vascular occlusion. Recently, many studies were conducted for automatic retinal vessel segmentation from fundus images through supervised and unsupervised methods to minimize user intervention. However, most of them lack in segmentation robustness and cannot optimize loss functions so that results of the segmentation have made lots of fake or thin branches. This article proposes a new conditional generative adversarial network called M-GAN to conduct accurate and precise retinal vessel segmentation by balancing losses through stacked deep fully convolutional networks. It consists of a newly designed M-generator with deep residual blocks for more robust segmentation and an M-discriminator with a deeper network for more efficient training of the adversarial model. In particular, a multi-kernel pooling block is added between the stacked layers to support the scale-invariance of vessel segmentations of different sizes. The M-generator has down-sampling layers to extract features and up-sampling layers to make segmented retinal blood vessel images from the extracted features. The M-discriminator also has a deeper network similar to the down-sampling of the M-generator, but the final layer is constructed as a fully connected layer for decision making. We conduct pre-processing of the input image using automatic color equalization (ACE) to make the retinal vessels of the input fundus image more clear and perform post-processing that makes the vessel branches smooth and reduces false-negatives using a Lanczos resampling method. To verify the proposed method, we used DRIVE, STARE, HRF, and CHASE-DB1 datasets and compared the proposed M-GAN with other studies. We measured accuracy, the intersection of union (IoU), F1 score, and Matthews correlation coefficient (MCC) for comparative analysis. Results of comparison proved that the proposed M-GAN derived superior performance than other studies.

INDEX TERMS Retinal blood vessel segmentation, medical imaging, generative adversarial networks (GAN), conditional GAN, stacked deep fully convolutional networks.

I. INTRODUCTION
The development of medical imaging technologies has made it possible to find diseases fast. Besides, many studies have been conducted to automatically process medical images without the intervention of human experts using computer vision technologies [1]. In particular, the segmentation of retinal blood vessels is very important to retinal diseases, such as retinal vascular occlusion and diabetic retinopathy, which are the most common reasons for blindness in Europe [2]. Currently, an expert manually separates blood vessels from the patient’s retinal image to inspect retinal diseases. However, this process takes much time and causes human mistakes and slips. In addition, analysis results may lack in terms of objectivity as different experts may generate different results. Therefore, it is essential to perform retinal blood vessel segmentation automatically by minimizing the direct intervention of experts [3].
The fundus image is the projection of the inner surface of the human eye, including the retina, vascular tree, fovea, and optic disc [4], as shown in Figure 1(a). As the color of the retinal surface is similar to that of the retinal blood vessel, it is challenging to segment retinal blood vessel images robustly, as shown in Figure 1(b). Many studies were conducted to achieve the segmentation automatically and accurately using computer vision, supervised or unsupervised machine learning algorithms [5]–[15]. Notably, studies using deep learning architectures derived better performance than existing methods [16]–[28].

Deep learning, a neural network architecture, has deeper and complicated perceptron layers [29]. Representative types of deep learning approaches include convolutional neural networks (CNN) for image analyses, such as object recognition, instance segmentation, and human pose estimation, and recurrent neural network (RNN) for sequential data analysis, such as machine translation, time-series data analysis, and speech recognition. In particular, the deep learning model with deeper CNN layers has shown superior performance in object detection than other methods [30]. Meanwhile, one of the most innovative neural networks is generative adversarial networks (GAN) that estimate new models through adversarial learning [31]. GAN simultaneously trains two models: a generator for creating new synthetic images and a discriminator for distinguishing human-annotated vessel maps (real label) or machine-generated vessel maps (fake label).

Medical image segmentation was conducted for detaching core parts of CT, MRI, and microscopic images [3], [32]–[34]. A fully convolutional network (FCN) has shown excellent performance in semantic segmentation [35]. FCN performs feature extraction through convolutional layers for down-sampling and then generates segmented images through transposed convolutional layers for up-sampling. U-Net based on FCN showed good performance in biomedical image segmentation [32]. U-Net is composed of a path for capturing context and a symmetric path for enabling localization. A deeper network was proposed, called FusionNet [16], which includes summation-based skip connections to allow for the segmentation. However, the previous methods using the deeper FCN have an imbalanced learning problem such as a higher false-positive rate causing fake branches or higher false-negative rate, causing thinner branches than the ground truth [22], [36]. In other words, there are still inherent problems in retinal blood vessel segmentation, such as wrong segmentation of pathological information and low microvascular segmentation [27].

This article proposes a new approach called M-GAN to conduct robust retinal vessel segmentation by balancing losses through stacked deep fully convolutional networks, which can not only support precise segmentation but also reduce low microvascular segmentation. The proposed approach takes three steps for the accurate retinal blood vessel segmentation. Firstly, a pre-processing using automatic color equalization (ACE) [37] is conducted to the input image to make the retinal vessels of input fundus image more clear. Then, M-GAN based on a conditional GAN [38] with an ‘M’ network structure is applied to segment retinal blood vessels from the pre-processed images. Finally, post-processing is applied to make the segmented vessel branches be smooth and reduce false-negatives further using a Lanczos resampling method [39].

M-GAN can support a deeper network and more precise segmentation with the help of short-term skip connections and the long-term residual connections between residual blocks [40]. M-GAN consists of a newly designed M-generator with deep residual blocks and an M-discriminator with a deeper network. The M-generator is designed as two-stacked deep FCNs by copying and pasting the same network for better segmentation. In particular, a multi-kernel pooling (MKP) block is added between the stacked layers to support the scale-invariance of vessel segmentations of different sizes. The M-generator has down-sampling layers to extract features and up-sampling layers to make segmented retinal blood vessel images from the extracted features. Each residual block has another three convolutional layers with a short-term skip connection. The M-discriminator also has a deeper network similar to the down-sampling of the M-generator, but the final layer is designed as a fully connected layer for decision making. The M-generator is trained based on the binary cross-entropy (BCE) loss function and the false-negative (FN) loss function not only to increase the accuracy but also to decrease the false-negative decision in segmentation. The discriminator is trained to minimize the least square (LS) loss [41]. A ground truth image is trained to be classified as a real label, and a generated-segmented image is trained to be classified as a fake label. Meanwhile, when training the M-generator, the generated-segmented image is trained to be classified as a real label.

To verify the proposed approach, we used publically available DRIVE [33], STARE [42], CHASE-DB1 [43], and HRF [44] datasets and compared the proposed M-GAN with previous studies. We measured recall (sensitivity), precision, specificity, accuracy, the intersection of union (IoU), the area
under curve (AUC), F1 score, and Matthews correlation coefficient (MCC) for comparative analyses. Comparative analyses proved that the proposed M-GAN derived superior performance than other studies.

The proposed approach has the following contributions.
1) A novel deep learning architecture, M-GAN, is proposed to generate the segmentation of retinal blood vessels more accurately and precisely by combining the conditional GAN and deep residual blocks. The M-generator consists of two-stacked deep FCNs with short-term skip connections and long-term residual connection and the multi-kernel pooling block for supporting the scale-invariance of vessel features between the two-stacked FCNs. The M-discriminator has also a deeper neural network.
2) We have re-designed loss functions that include BCE, LS, and FN losses to derive better performance. In particular, the FN loss function can reduce the false-negative rate of previous studies so that it can balance precision and recall.
3) We have also improved the performance of the proposed approach by taking pre- and post-processing using computer vision algorithms before and after conducting M-GAN.
4) The proposed M-GAN is compared with other studies concerning many factors such as accuracy, AUC, IoU, F1 score, and MCC. The comparative analysis proved that M-GAN outperformed other studies.

The rest of the paper is organized as follows. Section II discusses related work on retinal blood vessel segmentation. Section III presents the proposed M-GAN. Section IV compares the M-GAN with previous studies based on public datasets. Finally, we will conclude the paper with future work.

II. RELATED WORK
We analyze related studies on retinal blood vessel that used computer vision, neural network, wavelet transform, etc. We will also analyze related works using different deep learning architectures.

A. COMPUTER VISION-BASED AND TRADITIONAL MACHINING LEARNING-BASED SEGMENTATION
Akram et al. and Soares et al. performed retinal blood vessel segmentation through the wavelet transform [5], [6]. They used the 2D Gabor wavelet and sharpening filter such as Gaussian mixture to enhance and sharpen the vascular pattern. Marín et al. performed vessel segmentation using a neural network for pixel classification and computed a 7D vector of gray-level and moment invariants-based features [7]. Ocbagabir et al. proposed a rule-based approach called star networked pixel tracking to determine whether a processed pixel was a part of a vessel or not [8]. Noise artifacts like small vessels were filtered by the eight-direction network pixel tracking algorithm. However, these studies could not show high-performance measures because they used simple architectural algorithms.

Some studies conducted retinal vessel segmentation using more advanced algorithms. Becker et al. proposed a fully-discriminative algorithm for curvilinear structure segmentation that learned a classifier and the features it relied on simultaneously [9]. Their method relied on the Gradient Boosting framework to learn discriminative convolutional filters at each stage and operated on raw image pixels in addition to additional data sources. Ganin and Lempitsky proposed a combination of convolutional neural networks with the nearest neighbor search for image processing, such as thin object segmentation or natural edge detection [10].

Unsupervised or filter-based approaches were also introduced. Roychowdhury et al. presented an unsupervised iterative segmentation method performing the iterative enhancement of the segmented vessels [11]. Annunziata et al. proposed an unsupervised approach to vessel detection in retinal images using Hessian-based enhancement with an inpainting technique [12]. Zhang et al. proposed an automatic filter-based approach to retinal blood vessel segmentation via locally adaptive derivative frames in orientation scores [13]. Neto et al. presented an unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images [14]. Their methodology combined Gaussian smoothing, vessel contrast enhancement, and a morphological top-hat operator. Soomro et al. proposed a sequential process for retinal blood vessel segmentation using independent component analysis (ICA) that could be used to determine three color channels as three enhanced independent components channels with potentially noise-free contents [15].

Despite the unsupervised or filter-based approaches, it derived high accuracy measurement at that time, but their performance is much lower than the performances of recent studies using deep learning-based retinal vessel segmentation methods.

B. DEEP LEARNING-BASED SEGMENTATION METHODS
Several studies using deep neural networks were conducted for retinal blood vessel segmentation [17]–[25]. Soomro et al. presented a comprehensive review of the principle and application of deep learning in retinal image analysis [17]. They focused on recent advances in deep learning methods for retinal image analysis. Liskowski and Krawiec presented a supervised segmentation method with deep neural networks after performing data augmentation with global contrast normalization and zero-phase whitening using geometric transformations and gamma corrections [18]. Zhang et al. proposed a CNN-based architecture for retinal blood vessel segmentation. Their method used low-level features with high-level features, which involved atrous convolution to get multi-scale features [19]. Fu et al. and Orlando et al. performed vessel segmentation using conditional random fields (CRF) to model the long-range interactions between pixels [20], [21]. Maninis et al. presented a framework of retinal data analysis called Deep Retinal Image Understanding (DRIU) by providing both retinal vessel and optic disc segmentation [22]. DRIU showed more consistent
results with a gold standard than a second human annotator used as a control. However, it derived high sensitivity, but the similarity was significantly low. Zhang et al. proposed a generic medical segmentation method, called Edge-aTention guidance Network (ET-Net), which embeds edge-attention representations to guide the process of segmentation. An edge guidance module is utilized to learn the edge-attention representations and preserve the local edge characteristics in the early encoding layers [24]. Soomro et al. proposed a fully convolutional model consisting of an encoder part and a decoder part where the pooling layers are replaced with strided convolutional layers. The morphological mappings along with the principal component analysis (PCA)-based pre-processing steps are used to generate contrast images for the training dataset [25].

Besides, there were some studies that performed image processing in different areas using deep CNN. Xie and Tu performed image-to-image prediction using a deep learning method that leveraged fully convolutional neural networks and deeply supervised networks [45]. Although it automatically learned rich hierarchical representations for resolution ambiguity in edge and object boundary detection, it could not detect thin vessel branches. Ronneberger et al. proposed a network architecture called U-Net that consisted of a down-sampling path for capturing context and a symmetric up-sampling path for enabling localization in segmenting neuronal structures in electron microscopic stacks [32]. U-Net is widely used in various medical applications. Soomro et al. analyzed the impact of an image enhancement technique on the CNN model. Pre-processing is applied to the input image based on fuzzy logic and image processing tactics. Then, a stride encode-decoder CNN model was used to generate the segmented images. The model was based on a deeper U-net [26]. Pan et al. also proposed an improved deep learning U-net model for retinal vessel segmentation by enhancing input images [27]. Gu et al. proposed a dense atrous convolution (DAC) block that could capture wider and deeper semantic features with multi-scale atrous convolutions and a residual multi-kernel pooling (RMP) motivated from spatial pyramid pooling. It relies on the DAC block and the RMP block to get more abstract features and preserve more spatial information to boost the performance of medical image segmentation [28].

Recently, some studies dealt with fundus images with high quality. The whole image cannot be handled effectively in deep learning due to computational complexity and cost. For this reason, the whole image is divided into small patches, those patches are trained, and finally outputs of the patches are combined together for retinal blood vessel segmentation. Gao et al. proposed a method for retinal blood vessel segmentation that combined U-net with Gaussian matched filter [46]. 48 × 48 patches were used for training in U-net. The Gaussian filter is used to detect retinal thin vessels. Yan et al. proposed a method to balance thick vessels and thin vessels in the training phase [47]. For training, they cropped each fundus image into 128 × 128 patches. Alom et al. proposed a recurrent residual U-net for retinal blood vessel segmentation [48]. 48 × 48 patches were used for training. Jin et al. presented Deformable U-net (DUNet) that utilizes the retinal vessels’ local features with a U-shape network. 48 × 48 patches were also used for training [49]. Patch-based segmentation methods are widely used. However, the patch-based segmentation takes much time in training and computation.

These previous works showed good performance in retinal blood vessel segmentation, but it is still challenging to conduct more optimized segmentation concerning various factors such as accuracy, IoU, and F1 score overall.

**C. GENERATIVE ADVERSARIAL NETWORK-BASED METHODS**

Recently, one of the most interesting methods in machine learning is a generative adversarial network (GAN). GAN estimates generative models through an adversarial process, and it trains two models simultaneously: a generative model G for estimating the data distribution and a discriminative model D for capturing the probability that data came from the training data rather than G. The G and D play the following minimax two-player game with value function $V(D, G)$:

$$
\min_G \max_D V(D, G) = \mathbb{E}_{x \sim P_{data}(x)} \left[ \log D(x) \right] + \mathbb{E}_{z \sim P(z)} \left[ \log (1 - D(G(z))) \right] \quad (1)
$$

There were several studies using GAN for segmentation [50], [51], synthetic image generation [52], and image translation [53]. Concerning segmentation, Xue et al. proposed brain tumor segmentation using GAN-based architecture, called SegAN that used a fully CNN as a generator to create segmentation label maps [54]. They proposed an adversarial critic network with a multi-scale L1 loss function to make the critic and generator to learn both global and local features for capturing spatial relationships between pixels. Sato et al. proposed a method for semantic segmentation of cell membrane and nucleus by improving the Pix2Pix [53] algorithm [50]. Son et al. presented a method that generated the map of retinal vessels through generative adversarial training. They improved segmentation performance using binary cross-entropy loss when training the generator [51].

**III. PROPOSED APPROACH**

This article proposes a new conditional GAN called M-GAN, which can conduct retinal blood vessel segmentation more accurately and precisely by balancing losses and adopting a stacked deep fully convolutional network. The architecture of the proposed M-GAN is shown in Figure 2. Firstly, pre-processing using automatic color equalization (ACE) [37] is applied to the input fundus image to enhance the retinal blood vessels more clearly. Then, M-GAN is applied to the enhanced image to segment retinal blood vessels more robustly than previous studies. Lastly,
post-processing is applied to the segmented images to smooth the vessel image and to reduce false-negatives.

As shown in Eq. (2), the first phase calculates $R_c (p)$ from the input pixel $I_c (·)$ while the second phase calculates the enhanced output image $O_c (p)$ from $R_c (p)$. In the first phase, the chromic/spatial adjustment creates an output image $R$ where every pixel $p$ is recomputed concerning the image content. Each pixel in the output image $R$ is computed separately for each channel $c$ [37].

$$
R_c (p) = \frac{\sum_{j \in \text{Subset}, j \neq p} r(I_c(p) - I_c(j))}{\sum_{j \in \text{Subset}, j \neq p} \max d(p, j) - d(p, j)}
$$

(2)

where $I_c (p) - I_c (j)$ implies the lateral inhabitation mechanism, $d(·)$ is a distance function that weights the value of local or global contribution, $r(·)$ is the relative lightness appearance of the pixel, and $r_{\max}$ is the maximum value of $r(·)$.

$$
O_c (p) = \text{round}[s_c R_c(p) - m_c]
$$

(3)

For each pixel where $s_c$ is the slope of the segment $(m_c, 0), (M_c, 255)$, with $m_c = \min_p R_c(p)$ and $M_c = \max_p R_c(p)$. The enhancement result using ACE is shown in Figure 2 (left).

Details are described in reference 37.

B. ARCHITECTURE OF M-GAN

The proposed M-GAN for retinal blood vessel segmentation is designed based on a conditional GAN, such as Pix2Pix. Unlike a vanilla GAN that has no input data except random variables, the generator of Pix2Pix improves the learning performance by adding the $L_1$ loss function that performs the pixel-wise comparison between the ground truth image and the generated image. However, because Pix2Pix is designed to solve a general-purpose problem for image-to-image translation such as colorizing images from gray images or reconstructing objects from edge maps, it cannot be directly applied to retinal blood vessel segmentation that requires higher IoU and $F_1$ values.

A novel architecture of M-GAN consists of a newly designed M-generator and an M-discriminator (Figure 3). The M-generator has a deep network with residual blocks, short-term skip connections, and long-term residual connections. Each residual block (yellow block in Figure 3) consists of three convolutional layers with a short-term skip connection. The short-term skip connection conducts identity mapping, and its output is added to the output of stacked layers [16], [40]. Besides, the M-generator has long-term residual connections between layers of the down-sampling layer and the up-sampling layer. Those connections support a longer deep network and reduce the gradient vanishing problem.

The proposed M-generator consists of two stacked deep FCNs, each of which is composed of down-sampling and up-sampling of features in the image. Each FCN has the same network structure, and weights are shared between the two stacked networks so that no additional computation memory is needed. In particular, a multi-kernel pooling (MKP) block is added between them to support the scale-invariant features.
The network architecture of the proposed M-GAN.

The MKP is adopted from a residual multi-kernel pooling (RMP) [28] proposed by Gu et al. RMP was originally used between down-sampling and up-sampling processes in a deep FCN. However, MKP was used between two deep FCNs to improve segmentation accuracy and precision quantitatively and qualitatively. The process for multi-kernel pooling is shown in Figure 4. MKP pooling is applied to the output of the up-sampling of the first FCN using three different sizes of kernels of $2 \times 2$, $4 \times 4$, and $5 \times 5$ to learn the scale-invariance of the features. At the same time, three channels were reduced to one channel through $1 \times 1$ Conv to extract features of size $320 \times 320$, $160 \times 160$, and $128 \times 128$. Finally, they are converted to the original size through a bilinear up-sampling and concatenated to existing $640 \times 640 \times 3$ features, which are input to the second FCN.

In particular, during training the M-generator, a binary cross-entropy (BCE) loss instead of the $L_1$ loss is utilized to increase the quality of the generated image that is divided into mask pixel values and background pixel values. Note that we have found a problem that the false-negative rate is higher than the false-positive rate. For this reason, we added a false-negative loss function to decrease the false-negatives by imposing a penalty for the false-negative error. Through the FN loss function, it can improve segmentation performance by keeping lower both the false-negative and false-positive rates together. The M-discriminator includes a deeper convolutional network consisting of residual blocks, and it can reduce the gradient vanishing problem by replacing the vanilla GAN loss function with the Least Square GAN (LSGAN) [41] loss function.

Both the image segmented by the M-generator and the ground truth fundus image are concatenated and fed into the M-discriminator for training. The learning process continues until the machine-generated vessel map is enough to be judged as a real label. The discriminator is similar in architecture for the down-sampling of the M-generator. However, the...
The BCE loss function makes the segmented image fit to the image with ground truth by pixel-wise comparison. The loss function that compares the generated segmentation ground truth.

study, it was set between 0 and 1.

of the generated segmented image, and \( N \) value of the ground truth image, \( r \) number of positive pixels in the ground truth image, and \( p \) of positive pixels in the ground truth image.

For the detailed description of loss functions, we divide them into two parts: 1) GAN loss function for adversarial learning of the M-generator and the M-discriminator and 2) new loss functions of the M-generator to improve segmentation performance. The loss functions used in the proposed M-generator and M-discriminator are described in Eq. (4) and Eq. (5), respectively. We also adopt the LS loss function for the GAN loss [41] instead of the \( L_1 \) loss in Eq. (6) and Eq. (7).

\[
M\text{-generator Loss} = L_GAN (G) + L_{BCE} + L_{FN} \tag{4}
\]

\[
M\text{-discriminator Loss} = L_GAN (D) \tag{5}
\]

\[
\min_G L_GAN (G) = \frac{1}{2} E_i \left[(1 - D(x, G(x)))^2 \right] \tag{6}
\]

\[
\min_D L_GAN (D) = \frac{1}{2} E_{x,y} \left[(1 - D(x, y))^2 \right] + \frac{1}{2} E_x \left[ (D(x, G(x)))^2 \right] \tag{7}
\]

where \( x \) is a fundus image, and \( y \) is the mask image of the ground truth.

The loss optimization of the M-generator includes two functions for the pixel-wise calculation. We adopt the BCE loss function that compares the generated segmentation image with ground truth image by pixel-wise comparison. The BCE loss function makes the segmented image fit to the ground truth image more accurately than the \( L_1 \) loss function. \( \lambda_{BCE} \) is a weight parameter of the BCE loss function in Eq. (8). In this study, we set its value to 10. \( r_i \) is the pixel value of the ground truth image, \( f_i \) represents the pixel value of the generated segmented image, and \( N \) implies the total number of pixels in the image.

\[
L_{BCE} = -\frac{1}{N} \sum_{i}^{N} \lambda_{BCE} (r_i \log (f_i) + (1 - r_i) \log (1 - f_i)) \tag{8}
\]

Meanwhile, we have found that there is a problem during the evaluation of the retinal blood vessel segmentation with the loss functions mentioned before. As shown in the left side of Figure 5, there are many false-negatives, which recognizes the true mask pixel (white) as the background pixel (black). To reduce the false-negative error, the FN loss function is added to the M-generator loss function in Eq. (9).

\[
L_{FN} = \frac{1}{N_p} \sum_{i}^{N_p} \lambda_{FN} (1 - p_i)^2, \quad p_i = \begin{cases} 1, & \text{if } p_i \geq 0.5 \\ p_i, & \text{if } p_i < 0.5 \end{cases} \tag{9}
\]

\( \lambda_{FN} \) is a weight parameter of the FN loss function. In this study, it was set between 0 and 1. \( N_p \) represents the number of positive pixels in the ground truth image, and \( p_i \) is the \( i \)th pixel (0~1) in the generated image. If \( p_i \) is greater than or equal to the threshold, 1 ~ \( p_i \) becomes zero, which implies that \( p_i \) is equal to the \( i \)th pixel value in the ground truth. Otherwise, \((1-p_i)^2 \) will be loss value, which means that the predicted value is different from the ground truth. Therefore, the generated-segmented image must have a lower FN rate to minimize \( L_{FN} \). The result of adding the proposed FN loss function to the segmentation is shown at the right of Figure 5. It confirms that the false-negative rate is reduced, and a more accurate retinal vessel segmentation image is generated.

D. POST-PROCESSING USING INTERPOLATION FILTER

Post-processing based on a Lanczos resampling method is applied to the generated image of M-GAN. The image resampling filter makes the extracted vessel image smooth that fits the ground truth and thus reduces false-negatives.

It was confirmed that the segmentation performance was improved by applying the post-processing using interpolation filter or resampling methods. For image resampling or interpolation methods, bicubic, Lanczos, and bilinear methods are widely used [55]. To select the method that shows the best performance as a result of post-processing, comparative evaluation between different methods was conducted as shown in Table 1. The evaluation shows that the post-processing using Lanczos resampling increases 0.042% with respect to IoU. Therefore, the Lanczos resampling method was selected as the algorithm for post-processing in retinal vessel segmentation. Details for the Lanczos resampling method are described in reference 39.

E. ABLATION STUDY ON THE PROPOSED ARCHITECTURE WITH DIFFERENT LOSS FUNCTIONS

To train the proposed M-GAN, we used an Adam optimizer for the M-generator and discriminator with the learning rate \( = 0.0002 \), beta1 = 0.5, beta2 = 0.999, the batch size = 2, and the number of the epoch is 60. Also, we initialized weights with the Gaussian normalization (mean: 0, STD: 0.02) in convolutional layers.

To confirm the effectiveness and advantage of the proposed M-GAN architecture and its loss functions, we first compared...
TABLE 1. Comparison of different image resampling methods for post-processing.

| Method      | IoU       | Difference |
|-------------|-----------|------------|
| NON         | 0.71152 ± 0.00028 | -          |
| LANCZOS     | 0.71194 ± 0.00030 | +0.042%   |
| BILINEAR    | 0.70636 ± 0.00031 | -0.516%   |
| BICUBIC     | 0.71162 ± 0.00029 | +0.010%   |
| HAMMING     | 0.70908 ± 0.00031 | -0.244%   |

The results with/without the LS loss and FN loss functions in addition to the BCE function. Then, we compared the proposed deeper network with shallow architecture without residual blocks. Finally, we compared the two-stacked FCN architecture with the one-stacked FCN architecture. We evaluated performance measurements such as IoU, $F_1$ score, and accuracy for comparative evaluation (also in Section IV.B). Due to the characteristics of deep learning, different results can be derived for each training with the random initialization of weights. Thus, we trained M-GAN five times and measured the performance on average. To conduct the comparative evaluation, we used the well-known DRIVE dataset [33].

We assume that the proposed M-GAN is based on the two-stacked deep FCNs, and only loss functions are different. We compared three different types of loss functions: 1) BCE loss function with the basic GAN loss, 2) BCE + LS loss functions, and 3) BCE + LS + FN loss functions, as shown in Table 2. The BCE loss function calculates the binary cross-entropy between the ground truth image and the generated segmented image by pixel-wise comparison, and the LS loss function calculates the least square error between them instead of the basic GAN loss for both the discriminator and M-generator. Finally, the FN loss function is used to decrease the false-negative error by imposing a penalty to the error.

TABLE 2. Comparison of the results depending on the additional loss functions proposed in this article.

| Loss Function | IoU       | $F_1$ score       |
|---------------|-----------|-------------------|
| BCE           | 0.6854 ± 0.00030 | 0.8133 ± 0.0021 |
| BCE + LS      | 0.7061 ± 0.0005  | 0.8277 ± 0.0003  |
| BCE + LS + FN | 0.7119 ± 0.0003  | 0.8317 ± 0.0002  |

We measured IoU and $F_1$ score for comparative analysis and measured precision (Pr) and recall (Re) to verify the robustness of the proposed loss functions.

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (10)$$

$$\text{Pr} = \frac{TP}{TP + FP} \quad (12)$$

$$\text{Re} = \frac{TP}{TP + FN} \quad (14)$$

where $TP$ is the number of pixels corresponding to true positive, $FP$ the number to false positive, $TN$ the number to true negative, and $FN$ is the number to false negative.

The result using all three functions (BCE + LS + FN) showed higher IoU and $F_1$ scores than the others. In the case of using only the BCE loss function, the learning did not proceed uniformly as the standard deviation fluctuates too much. We have found that the LS loss function is used to lower the standard deviation error and to make the training more consistent and efficient while maintaining high accuracy. However, there is a difference in the ratio of precision and recall, as shown on the left of Figure 6. Precision is higher than recall so that the false-negative rate that recognizes the mask pixel (true) as the background pixel (false) is measured higher than the false-positive rate that recognizes the background pixel (true) to the mask pixel (false). To reduce the false-negative error and to improve recall, we added the FN loss function. As a result, precision and recall having a trade-off relationship become similar, and we can confirm that accuracy is also improved as shown on the right side of Figure 6.

We also performed a comparative evaluation using three types of architectures of the proposed M-GAN. Three architectures are a basic shallow GAN architecture without residual blocks, one-stacked FCN architecture using residual blocks, and the proposed two-stacked deep FCN architecture using residual blocks. All the architectures used BCE + LS + FN loss functions. The results of the comparative evaluation are shown in Table 3. The two-stacked deep FCN architecture derived the best results than the others.
We performed a comparative evaluation of the M-GAN with/without pre- and post-processing as shown in Table 4. The method including both processing methods derived the higher performance. Furthermore, segmentation accuracy and precision according to the presence or absence of the discriminator, which is the core part of the GAN structure, was evaluated. As shown in Table 5, the proposed method with the adversarial discriminator showed better segmentation performance than the method without the discriminator. In particular, as shown in Figure 7, when learning without the discriminator, although the loss decreases, the $F_1$ score decreases from the middle of the learning process, which implies overfitting. On the other hand, when learning with the discriminator, the $F_1$ score gradually increases as the learning progresses due to adversarial learning, which overcomes overfitting.

### TABLE 3. Comparison of different architectures of M-GAN.

| Architecture  | IoU          | $F_1$ score |
|---------------|--------------|-------------|
| No residual   | 0.6923 ± 0.0006 | 0.8182 ± 0.0004 |
| One-stacked   | 0.7035 ± 0.0005 | 0.8259 ± 0.0003 |
| Two-stacked   | 0.7119 ± 0.0003 | 0.8317 ± 0.0002 |

### TABLE 4. Comparison of pre- and post-processing in M-GAN.

| Processing      | IoU          | $F_1$ score |
|-----------------|--------------|-------------|
| Without pre- &  post- | 0.7067 ± 0.0003 | 0.8282 ± 0.0002 |
| Pre-            | 0.7115 ± 0.0003 | 0.8314 ± 0.0002 |
| Pre- & post-    | 0.7119 ± 0.0003 | 0.8317 ± 0.0002 |

### TABLE 5. Comparison of methods without and with discriminator in M-GAN.

| Discriminator | IoU          | $F_1$ score |
|---------------|--------------|-------------|
| without       | 0.7040 ± 0.0005 | 0.8263 ± 0.0004 |
| with          | 0.7119 ± 0.0003 | 0.8317 ± 0.0002 |

### IV. EVALUATION

We compared M-GAN with previous studies to confirm the advantages of the proposed approach to the retinal blood vessel segmentation. We used publically available databases of fundus images such as DRIVE [33], STARE [42], CHASE-DB1 [43], and HRF [44].

#### A. DATASET

The DRIVE dataset has 40 images taken from a diabetic retinopathy screening program [33]. The image has 565 × 584 pixels with 8 bits per color channel. It also has the segmented blood vessel images corresponding to the ground truth produced by human experts. In this study, we use 20 images for training and the remaining for testing. Since the number of training images was too small, we augmented the training data through transforming images using flipping and rotating with intervals at 10°, and thus, we created 1,440 training images.

The STARE database is composed of 20 fundus images and segmented blood vessel images corresponding to the ground truth [42]. Each image has a 605 × 700 pixels with 24 bits per pixel (standard RGB). In this study, we used 10 images for training and the remaining 10 images for testing. Similar to the DRIVE dataset, we augmented the training data through transforming images through flipping and rotation with intervals at 10°, and thus, we could create 720 training images.

The CHASE-DB1 database consists of 28 retinal fundus images captured from multiethnic school children [43]. The images have 960 × 999 pixels. In this study, we used the first 8 images for training datasets and the remaining 20 images for test datasets. Also, we augmented the training data through transforming images using flipping and rotation with intervals at 10°, and thus, we could create 576 training images.

The HRF dataset consists of healthy retinas, glaucomatous retinas, and DR retinas of fundus images [44]. Each contains colored retinal images (3 sets × 15 images = total 45 images) with the corresponding manually segmented images and mask images. The images have a size of 3,504 × 2,336 pixels. In this study, we used the first 30 images for training and the remaining 15 images for testing. Also, we augmented the training data through transforming images using flipping and rotation with intervals at 10°, and thus, we could create 2,160 training images.

The fundus images of the four databases have different sizes. For this reason, we resized the image into the size of 640 × 640, which is easy to down-sample fundus images and to up-sample segmented blood vessel images. In the case of the DRIVE dataset, we increased the size through padding.
TABLE 6. Comparison of segmentation results using the DRIVE dataset.

| Methods               | Re(Se) | Pr  | Sp   | Acc  | AUC  | IoU  | F₁   | MCC  |
|-----------------------|--------|-----|------|------|------|------|------|------|
| Odstrcilik et al. [44]| 0.7060 | -   | 0.9693 | 0.9340 | 0.9519 | -    | -    | -    |
| Becker et al. [9]     | 0.8505 | 0.6541 | 0.9568 | 0.9475 | 0.9466 | 0.5867 | 0.7395 | 0.7184 |
| Ganin & Lempitsky [10]| 0.7370 | 0.8554 | 0.9880 | 0.9661 | 0.9774 | 0.6554 | 0.7918 | 0.7760 |
| Xie & Tu [45]         | 0.9419 | 0.5344 | 0.9212 | 0.9230 | 0.9795 | 0.5174 | 0.6819 | 0.6753 |
| Ronneberger et al. [32]| 0.8300 | 0.7856 | 0.9783 | 0.9653 | 0.9794 | 0.6767 | 0.8072 | 0.7885 |
| Soares et al. [6]     | 0.5258 | 0.8829 | 0.9933 | 0.9524 | 0.9539 | 0.4915 | 0.6591 | 0.6600 |
| Maninis et al. [22]   | 0.9477 | 0.6046 | 0.9405 | 0.9411 | 0.9861 | 0.5850 | 0.7382 | 0.7296 |
| Isola et al. [53]     | 0.7339 | 0.8041 | 0.9828 | 0.9610 | 0.9424 | 0.6226 | 0.7674 | 0.7471 |
| Gao et al. [46]       | 0.7802 | **0.8948** | 0.9876 | 0.9636 | 0.9772 | -    | -    | -    |
| Son et al. [51]       | 0.8014 | 0.8509 | 0.9865 | 0.9703 | 0.9859 | 0.7027 | 0.8254 | 0.8096 |
| Thangaraj et al. [23] | 0.8014 | -   | 0.9753 | 0.9606 | 0.8884 | -    | -    | -    |
| Yan et al. [47]       | 0.7653 | -   | 0.9818 | 0.9542 | 0.9752 | -    | -    | -    |
| Alom et al. [48]      | 0.7661 | -   | 0.9807 | 0.9613 | 0.9793 | -    | -    | -    |
| Soomro et al. [26]    | 0.802  | -   | 0.974  | 0.959  | 0.948  | -    | -    | -    |
| Jin et al. [49]       | 0.7963 | -   | 0.9800 | 0.9566 | 0.9802 | -    | 0.8237 | -    |
| Pan et al. [27]       | 0.8310 | -   | 0.9863 | 0.9650 | 0.9811 | -    | -    | -    |
| Gu et al. [28]        | 0.8309 | -   | -    | 0.9545 | 0.9779 | -    | -    | -    |
| M-GAN                 | 0.8346 | 0.8302 | 0.9836 | **0.9706** | **0.9868** | **0.7129** | **0.8324** | **0.8163** |

around the background with the black color. However, in the case of the other database, we downsized the fundus images to 640 × 640.

B. COMPARATIVE EVALUATION

We measured sensitivity, specificity, IoU, F₁ score, and MCC for comparative analysis as defined as follows.

\[
Se = Re = \frac{TP}{TP + FN} \quad (15)
\]

\[
Sp = \frac{TN}{TN + FP} \quad (16)
\]

\[
Acc = \frac{TP + TN}{N} \quad (17)
\]

\[
MCC = \frac{TP \cdot S \cdot P}{\sqrt{P \cdot S \cdot (1 - S) \cdot (1 - P)}} \quad (18)
\]

where \( N = TP + TN + FP + FN \), \( S = \frac{TP + FN}{N} \), and \( P = \frac{TP + FP}{N} \).

In addition to IoU, F₁ score, recall, and precision used in the ablation study, sensitivity, specificity, accuracy, and MCC were added for quantitative and comparative analysis. In particular, MCC is a more meaningful measurement of pixel-wise segmentation when vessel and non-vessel pixel classes in the problem are unbalanced. For retinal vessel segmentation, only 9%-14% portion of pixels belong to the vessels, and the others belong to non-vessels [13]. Since different datasets have different sizes of images, all the images are adjusted to have 640 × 640-pixel size through padding and resizing. The output images are then converted to the original image size through padding and resizing for quantitative and comparative evaluation.

We have compared the proposed M-GAN with previous studies. The comparative evaluation using the DRIVE dataset is described in Table 6. M-GAN showed higher performance than related studies concerning IoU and F₁ score measurements. For the 2D Gabor wavelet and supervised classification [6], the precision was the highest, but IoU and F₁ scores were lower than other methods because the recall was low, in particular. For DRIU [22], the recall was the highest, but IoU and F₁ scores were derived lower because the precision was small. However, the propose M-GAN, proved not only the higher precision and recall together through the FN loss function, but also the highest IoU and F₁ measurements. This means that M-GAN outperformed the other studies so that the retinal blood vessel segmentation was achieved with the best performance. N4-Fields [10], HED [45], and Kernel Boost [9] methods couldn’t derive higher than 0.8 on the F₁ score measurements.

Next, we compared M-GAN with the other deep learning methods such as U-Net [32] and Pix2Pix [53]. U-Net performed a deep learning network by adding weights of the down-sampling network to the weights of the up-sampling network to reduce the vanishing gradient problem. However, because it only used a pixel-wise objective function, it made segmented images with blurry images and fake branches. Pix2Pix complements the disadvantage of FCNs such as U-Net by using both a generator that uses a pixel-wise objective function using the whole image. Nevertheless, the performance was poor because it used the L₁ loss.
function, and its objective functions are not suitable for image segmentation. We evaluated V-GAN [51], which uses the BCE loss function instead of the L1 loss function of the Pix2Pix. V-GAN was able to extract a segmented image with high accuracy of F1 score 0.8254, but there was a difference between precision and recall. In particular, V-GAN performed image augmentation by flipping and rotation with intervals at 3°, then it used 4,800 images for training. In contrast, the proposed M-GAN achieved better results in IoU and F1 score measurements despite using 1,440 training images (rotation with intervals at 10°).

Other studies using deep learning were also proposed by extending U-Net and showed good performance in retinal blood vessel segmentation [26]–[28], [46]–[49]. They improved recall, accuracy, and AUC. However, they have not evaluated IoU, F1 score, and MCC although the input image has unbalanced positive and negative pixel values and thus these measurements are essential. The proposed approach outperformed over previous studies with respect to the accuracy, IoU, F1 score, and MCC. In particular, previous studies made it difficult to extract high-level features as they divided the large size fundus image into smaller patches and trained them. In addition, the inference took much time [46]–[49]. However, the proposed approach does not divide the original image into smaller patches but simply resizes it for the inference, which can overcome the problems related to previous studies.

![Fundus Image Wavelets DRIU N4-Fields](image1)

![U-Net Pix2Pix V-GAN M-GAN](image2)

**FIGURE 8.** Segmentation results of an example in the DRIVE dataset.

The segmentation results of a test image in the DRIVE dataset are shown in Figure 8. The fundus input image and its ground truth image are shown on the top-left side of the figure. The others show the generated-segmented image (above) and the visual difference between the generated and the ground truth images (below). The green, white, red, and blue represents TP, TN, FP, and FN, respectively. The wavelet-based method had many blue pixels because of a high FN, and the DRIU had many red pixels because of a high FP. N4-Fields, U-Net, and Pix2Pix had also many red or blue pixels than M-GAN. In particular, it was confirmed that M-GAN could generate the branches more correctly compared with V-GAN. Furthermore, M-GAN performed training consistently with an average error of only 0.0003 in the F1 score so that stable training was possible, as shown in Table 3. Therefore, M-GAN derived a high accuracy in IoU and F1 score measurements with balanced precision and recall, as shown in Table 6.

The results of comparative analyses using the STARE dataset are shown in Table 7. Similar to the results of the DRIVE dataset, the proposed M-GAN derived the best performance in accuracy, IoU, and MCC measurements with balanced high precision and recall. The segmentation results of the STARE dataset are shown in Figure 9. When compared with V-GAN, M-GAN showed better segmentation in the dark and tiny branches of the fundus image. DRIU also found retinal vessels well in dark and blurry areas, but it showed higher FP. Although the study proposed by Alom et al. [48] showed higher AUC, their approach is based on patch-based learning, which took longer time. Although their method showed good performance in the STARE dataset, it showed worse performance in DRIVE and CHASE-DB1 datasets, which implies overfitting to the STARE dataset.

![Fundus Image DRIU U-Net](image3)

![Pix2Pix V-GAN M-GAN](image4)

**FIGURE 9.** Segmentation results of an example in the STARE dataset.

The test results of comparative evaluation using the CHASE-DB1 and HRF dataset are shown in Table 8 and
TABLE 7. Comparison of segmentation results using the STARE dataset.

| Methods             | Re(Se) | Pr  | Sp  | Acc  | AUC  | IoU  | F₁   | MCC  |
|---------------------|--------|-----|-----|------|------|------|------|------|
| Odstreilk et al. [44]| 0.7847 | -   | 0.9512 | 0.9341 | 0.9569 | -   | -   | -    |
| Xie & Tu [45]       | 0.9493 | 0.5552 | 0.9369 | 0.9378 | 0.9824 | 0.5393 | 0.7007 | 0.6986 |
| Ronneberger et al. [32] | 0.7600 | 0.8536 | 0.9892 | 0.9716 | 0.9581 | 0.6724 | 0.8041 | 0.7904 |
| Soares et al. [6]   | 0.6850 | 0.8409 | 0.9892 | 0.9659 | 0.9776 | 0.6064 | 0.7550 | 0.7413 |
| Maninis et al. [22] | 0.9403 | 0.6474 | 0.9575 | 0.9562 | 0.9815 | 0.6218 | 0.7668 | 0.7593 |
| Isola et al. [53]   | 0.7627 | 0.7995 | 0.9841 | 0.9671 | 0.9402 | 0.6403 | 0.7807 | 0.7632 |
| Son et al. [51]     | 0.7841 | -   | 0.8823 | 0.9913 | 0.9754 | 0.9882 | 0.7098 | 0.8303 | 0.8187 |
| Thangaraj et al. [23]| 0.8339 | -   | 0.9536 | 0.9435 | 0.8938 | -   | -   | -    |
| Yan et al. [47]     | 0.7581 | -   | 0.9846 | 0.9612 | 0.9801 | -   | -   | -    |
| Alom et al. [48]    | 0.8298 | -   | 0.9862 | 0.9712 | 0.9914 | -   | -   | -    |
| Soomro et al. [26]  | 0.801  | -   | 0.969  | 0.961  | 0.945  | -   | -   | -    |
| Jin et al. [49]     | 0.7595 | -   | 0.9878 | 0.9641 | 0.9832 | -   | 0.8143 | -    |
| M-GAN               | 0.8324 | 0.8417 | 0.9938 | 0.9876 | 0.9873 | 0.7198 | 0.8370 | 0.8306 |

FIGURE 10. Segmentation results of an example in the CHASE-DB1 dataset: (a) original fundus image, (b) ground truth vessel image, (c) generated vessel image, and (d) the visual difference between the ground truth and generated image.

Table 9. The segmentation results of the CHASE-DB1 and HRF dataset using the proposed M-GAN are also shown in Figure 10 and Figure 11. The M-GAN derived the best performance in AUC, IoU, F₁ score, and MCC measurements. In particular, the image in CHASE-DB1 has 999 × 960 pixels, and the image in HRF dataset has 3504 × 2336 pixels. The image size in both datasets is much larger than the input size of the proposed M-GAN. Instead of using patch-based training, we resized the image to have 640 × 640 pixels, and we used the resized image in M-GAN without modifying its network architecture. Finally, the output of the M-GAN is resized to the original size. As shown in Table 8 and Table 9, the proposed M-GAN outperformed previous studies.

C. DISCUSSION

Through comparative evaluation, the advantages of the proposed M-GAN are as follows. First, accurate and precise segmentation was confirmed by deriving the best performance from all four datasets. Second, M-GAN showed the state-of-the-art deep learning architecture consisting of...
TABLE 9. Comparison of segmentation results using the HRF dataset.

| Method    | Acc  | AUC   | F1    | MCC  |
|-----------|------|-------|-------|------|
| [44]      | 0.9494 | 0.9678 | -     | -   |
| [12]      | 0.9581 | -     | -     | -   |
| [13]      | 0.9556 | 0.9608 | -     | 0.7410 |
| [21]      | -     | -     | 0.7158 | 0.6897 |
| [47]      | 0.9437 | -     | -     | -   |
| [49]      | 0.9651 | 0.9831 | -     | -   |
| M-GAN     | 0.9761 | 0.9852 | 0.7972 | 0.7845 |

M-generator and M-discriminator. The proposed M-GAN with an M network structure can extract better features and perform robust retinal vessel segmentation, which uses the residual block and skip connection and repeats the same network. Third, segmentation robustness is improved greatly by using the new loss functions combining BCE, LS, and FN losses. M-GAN, is more effective in segmenting blood vessels from fundus images by using the BCE loss instead of the $L_1$ loss. Besides, it is more efficient for training by using the least square GAN loss than the basic GAN loss. Finally, the FN loss, which balances recall and precision, is able to increase $F_1$ score and reduce false-negatives.

Although the proposed method yielded better results than other studies, there were also limitations to be addressed. First, while training progresses, there is a problem that the learning capability of the M-generator and M-discriminator becomes unbalanced. This is one of the problems of the original GAN since the discriminator that performs binary decision is learned faster than the generator, resulting in an imbalance in adversarial learning between the two. Second, pre- & post-processing methods are applied before and after M-GAN. Pre- and post-processing was performed for fine-tuning. In this respect, the proposed approach was not a complete end-to-end network. Nevertheless, the end-to-end based M-GAN without pre- and post-processing showed better performance than other methods as shown in Table 4 and Table 6. In the future, it is necessary to devise an end-to-end deep learning architecture. Third, it is also necessary to reduce information loss due to resizing while segmenting high-resolution images although information loss is inevitable.

Through various ablation studies, the proposed approach outperformed previous research works. In particular, the two stacked deep FCNs in M-GAN showed better performance than one stacked deep FCN concerning quantitative and qualitative aspects. Multi-kernel pooling between deep FCNs showed the scale-invariance of vessel features of different thicknesses and sizes. Therefore, Table 3 showed better performance of M-GAN with two stacked deep FCNs quantitatively. Besides, the result from M-GAN with two stacked FCNs segmented tiny vessels more precisely and clearly and reduced false-negatives, as shown in Figure 12(a). Furthermore, two stacked FCNs could remove segmentation noises, as shown in Figure 12(b), which proves the advantage and originality of the proposed approach.

Besides, the proposed approach, M-GAN, based on the conditional GAN outperformed V-GAN. In particular, M-GAN with two stacked deep FCNs is re-designed to support residual blocks and short-term skip connection. Loss functions are also revised to reduce false-negatives and prove more segmentation accuracy and precision, as shown in Table 6 and Table 7. V-GAN was optimized based on augmentations through flipping and rotation with intervals at $3^\circ$, which requires many training data, computational memory, and cost. However, we have augmented training data based on the intervals with $10^\circ$. Regarding this difference, we have trained V-GAN with the same augmented DRIVE dataset used in this article. As shown in Figure 13, V-GAN showed worse segmentation results compared with ours and ground truth. These quantitative and qualitative analyses prove the superiority of the proposed approach to previous state-of-the-art research works including V-GAN.
V. CONCLUSION

This article proposed a new generative adversarial network called M-GAN for the precise and accurate retinal blood vessel segmentation that is essential to identify various diseases related to human eyes. The proposed M-GAN consists of a novel M-generator using deep residual blocks and an M-discriminator with a deeper network for efficiently training the adversarial model. The M-generators adds a long-term residual connection that connects each layer of the down-sampling network with that of the up-sampling network. Furthermore, we add the binary cross-entropy loss function and the false-negative loss function to improve training efficiency and to increase robustness in segmentation. In particular, the M-generator has two-stacked deep FCNs formed the ‘M’ structure by copying and pasting the same network so that it guarantees robustness in the segmentation of various datasets using the proposed approach.

To confirm the robustness of the proposed M-GAN for the retinal blood vessel segmentation, we compared M-GAN with previous studies using four public datasets. We measured recall, precision, specificity, accuracy, IoU, F1 score, and MCC for comparative analyses. According to the comparison analyses, the proposed M-GAN proved the best performance on most of the measurements than the other methods. It derived not only the balanced precision and recall together through the FN loss function but also the highest IoU and F1 score measurements. In addition, we derived better performance using a simple pre-processing through ACE algorithm and post-processing through Lanczos resampling method.

As future research works, we will improve the proposed approach by devising an end-to-end network. We will also apply the proposed M-GAN to various medical image segmentations, such as brain tumor segmentation and cell membrane segmentation. Finally, we will modify M-GAN and apply it to industrial areas such as fault detection and segmentation.

ACKNOWLEDGMENT

The authors would like to thank Dr. Minseok Kim in KISTI for advising and discussing experimental evaluation.

REFERENCES

[1] A. Khatami, A. Khosravi, T. Nguyen, C. P. Lim, and S. Nahavandi, “Medical image analysis using wavelet transform and deep belief networks,” Expert Syst. Appl., vol. 85, no. 15, pp. 190–198, Nov. 2017.
[2] I. Kocur and S. Resnikoff, “Visual impairment and blindness in Europe and their prevention,” Brit. J. Ophthalmo., vol. 86, no. 7, pp. 716–722, Jul. 2002.
[3] M. D. Abrahmoff, M. K. Garvin, and M. Sonka, “Retinal imaging and image analysis,” IEEE Rev. Biomed. Eng., vol. 3, pp. 169–208, Dec. 2010.
[4] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Dudinick, C. G. Owen, and S. A. Barman, “An ensemble classification-based approach applied to retinal blood vessel segmentation,” IEEE Trans. Biomed. Eng., vol. 59, no. 9, pp. 2538–2548, Sep. 2012.
[5] M. U. Akram, A. Atzaz, S. F. Aneeque, and S. A. Khan, “Blood vessel enhancement and segmentation using wavelet transform,” in Proc. Int. Conf. Digit. Image Process., Mar. 2009, pp. 34–38.
[6] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, “Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification,” IEEE Trans. Med. Imag., vol. 25, no. 9, pp. 1214–1222, Sep. 2006.
[7] D. Murin, A. Aquino, M. E. Gegúndez-Arias, and J. M. Bravo, “A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features,” IEEE Trans. Med. Imag., vol. 30, no. 1, pp. 146–158, Jan. 2011.
[8] H. Ocbagabir, I. Hameed, S. Abdulmalik, and D. B. Buket, “A novel vessel segmentation algorithm in color images of the retina,” in Proc. LISR, May 2013, pp. 1–6.
[9] C. Becker, R. Rigamonti, V. Lepetit, and P. Fua, “Supervised feature learning for curvilinear structure segmentation,” in Proc. MICCAI, 2013, pp. 526–533.
[10] Y. Ganin and V. Lempitsky, “N4-fields: Neural network nearest neighbor fields for image transformations,” in Proc. ACCV, 2014, pp. 536–551.
[11] S. Roychohdhury, D. D. Koozekanani, and K. K. Parhi, “Iterative vessel segmentation of fundus images,” IEEE Trans. Biomed. Eng., vol. 62, no. 7, pp. 1738–1749, Jul. 2015.
[12] R. Annunziata, A. Garzelli, L. Ballerini, A. Meccoci, and E. Trucco, “Leveraging multiscale hessian-based segmentation with a novel exudate imprinting technique for retinal vessel segmentation,” IEEE J. Biomed. Health Informat., vol. 20, no. 4, pp. 1129–1138, Jul. 2016.
[13] J. Zhang, B. Dashbozorg, 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. Imag., vol. 35, no. 12, pp. 2631–2644, Dec. 2016.
[14] L. C. Neto, G. L. B. Ramalho, J. F. S. R. Neto, R. M. S. Veras, and F. N. S. Medeiros, “An unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images,” Expert Syst. Appl., vol. 78, pp. 182–192, Jul. 2017.
[15] T. A. Soomro, T. M. Khan, M. A. U. Khan, J. Gao, M. Paul, and L. Zheng, “Impact of ICA-based image enhancement technique on retinal blood vessels segmentation,” IEEE Access, vol. 6, pp. 3524–3538, Jan. 2018.
[16] T. M. Quan, D. G. Hildebrand, and W. K. Jeong, “FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics,” Jul. 2016, arXiv:1607.05695. [Online]. Available: http://arxiv.org/abs/1607.05695.
[17] T. A. Soomro, A. J. Afifi, L. Zheng, S. Soomro, J. Gao, O. Hellwich, and M. Paul, “Deep learning models for retinal blood vessels segmentation: A review,” IEEE Access, vol. 7, pp. 71696–71717, Jun. 2019.
[18] P. Liskowski and K. Krawiec, “Segmenting retinal blood vessels with deep neural networks,” IEEE Trans. Med. Imag., vol. 35, no. 11, pp. 2360–2380, Nov. 2016.
[19] B. Zhang, S. Huang, and S. Hu, “Multi-scale neural networks for retinal blood vessels segmentation,” Apr. 2018, arXiv:1804.04206. [Online]. Available: http://arxiv.org/abs/1804.04206.
[20] 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 Proc. MICCAI, 2016, pp. 132–139.
[21] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko, “A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images,” IEEE Trans. Biomed. Eng., vol. 64, no. 1, pp. 16–27, Jan. 2017.
[22] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, “Deep retinal image understanding,” in Proc. MICCAI, 2016, pp. 140–148.
[23] S. Thangaraj, V. Periyasamy, and R. Balaji, “Retinal vessel segmentation using neural network,” IET Image Process., vol. 12, no. 5, pp. 669–678, May 2018.
[24] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, and L. Shao, “ET-net: A generic edge-attention guidance network for medical image segmentation,” in Proc. MICCAI, 2019, pp. 442–450.
[25] T. A. Soomro, A. J. Afifi, J. Gao, O. Hellwich, L. Zheng, and M. Paul, “Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation,” Expert Syst. Appl., vol. 134, pp. 36–52, Nov. 2019.
[26] T. A. Soomro, A. J. Afifi, Ali Shah, S. Soomro, G. A. Baloch, L. Zheng, M. Yin, and J. Gao, “Impact of image enhancement technique on CNN model for retinal blood vessels segmentation,” IEEE Access, vol. 7, pp. 158183–158197, 2019.
[27] P. Xiaosh, Q. Zhang, H. Zhang, and S. Li, “A fundus retinal vessels segmentation scheme based on the improved deep learning U-Net model,” IEEE Access, vol. 7, pp. 122634–122643, 2019.
[28] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, “CE-net: Context encoder network for 2D medical image segmentation,” IEEE Trans. Med. Imag., vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[29] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. NIPS, 2012, pp. 1097–1105.

[31] A. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” in Proc. NIPS, 2014, pp. 2672–2680.

[32] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and O. E. Viergever, “Object detection in medical images with adaptive random forest,” IEEE Trans. Med. Imag., vol. 25, no. 11, pp. 1500–1508, Nov. 2006.

[33] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Proc. MICCAI, 2015, pp. 234–241.

[34] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in Proc. CVPR, 2009, pp. 248–255.

[35] J. Son, S. J. Park, and K.-H. Jung, “Retinal vessel segmentation in fundus images with generative adversarial networks,” Jul. 2017, arXiv:1706.09318. [Online]. Available: http://arxiv.org/abs/1706.09318

[36] K.-B. Park et al.: M-GAN: Retinal Blood Vessel Segmentation by Balancing Losses

**KYEONG-BEOM PARK** received the B.S. and M.S. degrees in industrial engineering from Chonnam National University, South Korea, where he is currently pursuing the Ph.D. degree. His current research interests include deep learning-based medical application and industrial augmented reality using deep learning.

**SUNG HO CHOI** received the B.S. and M.S. degrees in industrial engineering from Chonnam National University, South Korea, where he is currently pursuing the Ph.D. degree. His current research interests include deep learning-based applications and AR-based remote collaboration.

**JAE YEOL LEE** (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in industrial engineering from the Pohang University of Science and Technology (POSTECH), South Korea, in 1992, 1994, and 1998, respectively. From 1998 to 2003, he worked as a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI), South Korea. Since 2003, he has been a Faculty Member of Chonnam National University, South Korea, where he is currently a Professor with the Department of Industrial Engineering. His current research interests include AR/VR, medical image processing, deep learning, and HCI/UX.

**Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, “SegAN: Adversarial network with multi-scale L1 loss for medical image segmentation,” Neuroinformatics, vol. 16, nos. 3–4, pp. 383–392, May 2018.

**G. Bradschi and A. Kaehler, Learning OpenCV: Computer Vision With the OpenCV Library. Newton, MA, USA: O’Reilly Media, 2008.**