



# Multi-discriminator Generative Adversarial Networks for Improved Thin Retinal Vessel Segmentation

Gabriel Tjio<sup>1</sup>(✉), Shaohua Li<sup>1</sup>, Xinxing Xu<sup>1</sup>, Daniel Shu Wei Ting<sup>2</sup>, Yong Liu<sup>1</sup>, and Rick Siow Mong Goh<sup>1</sup>

<sup>1</sup> Institute of High Performance Computing, A\*STAR, Singapore, Singapore  
gabriel-tjio@ihpc.a-star.edu.sg

<sup>2</sup> Singapore Eye Research Institute, Singapore, Singapore

**Abstract.** Retinal vessel segmentation is an important step in clinical analysis of fundus images. Low contrast and the imbalanced pixel ratios between thick and thin vessels make accurate segmentation of the thin vasculature extremely challenging. In this paper, we present a novel multiscale segmentation method named Multiple discriminator generative adversarial network (MuGAN). MuGAN contains multiple discriminators with different effective receptive fields, which are sensitive to features at different scales. These discriminators jointly teach the segmentation (generator) network to pay attention to multiscale patterns. In addition, multiple discriminators allow our model to incorporate multiple inputs, such as edge enhanced vessel images, during training. We evaluated our method on the publicly available DRIVE and STARE datasets. MuGAN achieved an overall area under the Receiver Operator Characteristic Curve (AUC) of 0.979 for DRIVE and 0.981 for the STARE dataset. On segmenting thin retinal vessels, MuGAN showed quantitative and qualitative improvements on baselines.

**Keywords:** Multiscale segmentation · Generative Adversarial Network · Fundus imaging

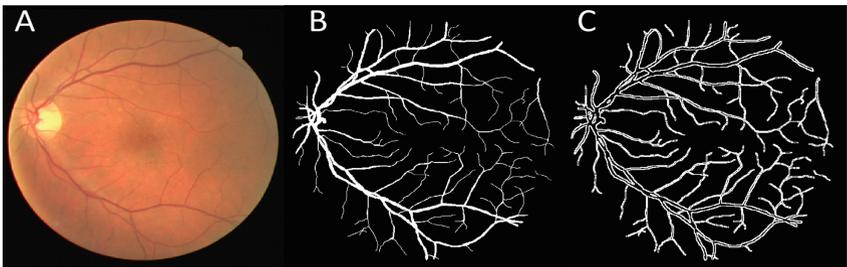
## 1 Introduction

Retinal vessels have been shown to be important features for cardiovascular disease detection [1]. Additionally, changes in retinal vessel diameters are associated with the progression of retinal diseases [2] and higher risk of cardiovascular mortality [3]. Accordingly, accurate segmentation of retinal vessels plays an important role in characterizing the patient's eye and cardiovascular health.

Recent work on retinal vessel segmentation has focused on deep learning approaches. However, accurate segmentation of thin retinal vessels remains challenging. Retinal vessel widths can range from 1 pixel to 10 pixels in diameter, making multiscale segmentation necessary for good performance. As noted in [4], large retinal vessels tend to be more accurately segmented than thin vessels because (1) the majority of vessel pixels belong to thick vessels, and (2) thick vessels typically have higher contrast than thin vessels.

This paper aims to address this problem with the following framework. We propose a novel multiscale, multi-input generative adversarial network (MuGAN) for retinal vessel segmentation. GANs [5] have been used in broad applications including image segmentation [6]. Son et al. [7] applied GANs on retinal vasculature segmentation. While [7] manifests good performance, it is not explicitly designed to detect thin vascular details. This work is inspired by the multiple discriminators approach [8], which suggests that a group of limited capacity discriminators can better capture the probability distribution of the training data compared to a single discriminator.

We build upon their work by using multiple discriminators with different architecture for improved multiscale segmentation. Firstly, we use convolutional layers of varying dilation rates for different discriminators. This preserves the resolution of the extracted features while varying the effective receptive field. Yu et al. [9] implemented dilated convolutions to combine multiscale features. We therefore suggest that introducing convolutions with different dilation rates will enable the discriminator to capture additional high level features. Additionally, multiple discriminators allow the GAN to incorporate different inputs: (1) the multiscale features extracted using dilated convolutional layers, and (2) the postprocessed outputs from the generator network and the postprocessed ground truth. We also use edge-enhanced ground truth images as additional input for training because the enhanced vessel boundaries mitigate the imbalance between thick and thin vessel pixels (Fig. 1). Finally, we use skeletal metrics based on [10] for performance evaluation because commonly used performance metrics such as AUC (Area under the receiver operator characteristic curve) may not adequately reflect the thin vessel segmentation accuracy.



**Fig. 1.** Example of (A) retinal fundus image, (B) ground truth vessel segmentation, (C) edge enhanced ground truth image. The edge enhanced ground truth increases the visibility of the thin vessel branches and mitigates the imbalance between thin and thick vessel pixels.

To assess MuGAN segmentation accuracy, we evaluated MuGAN,  $MUGAN_{noedge}$  (MuGAN without edge enhanced inputs),  $GAN_{single}$  (single generator discriminator pair) and other approaches [11, 12] on the publicly available datasets DRIVE [13] and STARE [14].

## 2 Methods

### 2.1 Generative Adversarial Networks for Retinal Vessel Image Segmentation

In Generative Adversarial Networks, the generator and discriminator are alternatively trained to minimize and maximize the objective function. The standard objective function for GAN is as follows:

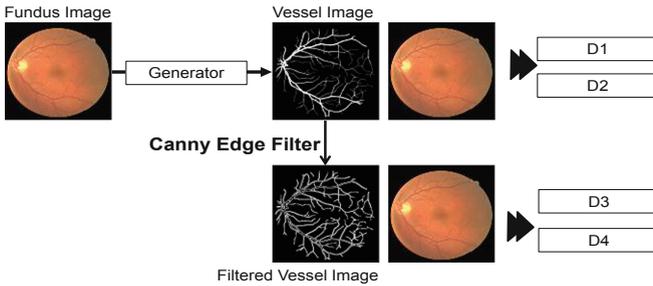
$$\min_{\theta_G} \max_{\theta_D} \mathcal{L}(D, G) = \mathbb{E}_{x \sim p_{data}(x)} \log(D(x)) + \mathbb{E}_{z \sim p_{data}(z)} \log(1 - D(G(z))) \quad (1)$$

where  $\theta_D$  and  $\theta_G$  refer to the parameters for the discriminator  $D$  and generator  $G$  respectively,  $x$ : source data,  $z$ : noise inputs,  $p_{data}$ : probability distribution of data. The generator must be able to generate an output such that the discriminator is unable to differentiate between that output and the ground truth. We modify the objective function [7] to include the loss function for the a given discriminators  $D_i$ ,  $\lambda$  is used to weight the segmentation loss  $\mathcal{L}_{seg}$ :

$$\min_{\theta_G} \max_{\theta_{D_i}} \mathcal{L}(D_i, G) + \lambda \mathcal{L}_{seg}(G) \quad (2)$$

The segmentation loss function  $\mathcal{L}_{seg}$  uses the binary cross entropy loss, which compares the generator output  $G(x)$  with the ground truth (segmented vasculature)  $y$  and  $x$  is the source data (input fundus images):

$$\mathcal{L}_{seg} = \mathbb{E}_{x,y \sim p_{data}(x)} \{-y \log G(x) - (1 - y) \log(\{1 - G(x)\})\} \quad (3)$$



**Fig. 2.** Workflow for training the multiple discriminator GAN (MuGAN). The discriminators D1 and D3 have the same architecture. Similarly, D2 and D4 have the same architecture. D1 and D2 (D3 and D4) have different effective receptive field sizes.

Figure 2 illustrates the proposed workflow. The discriminators D1 (D3) and D2 (D4) have different effective receptive field sizes. We vary effective receptive field size by changing the dilation rate of the convolutional layers [9]. The first set of discriminators (D1 and D2) was trained to distinguish between the ground truth and the trained segmentations. The second group of discriminators (D3 and D4) was trained to distinguish

between edge enhanced ground truth and edge enhanced segmented vessels. The ground truth vessel images were processed with a Canny edge filter before training with the default settings (lower bound for thresholding: 10% of maximum pixel value of input; upper bound of maximum pixel value: 20%;  $\sigma$  of Gaussian: 1.0), instead of tuning the parameters to suit the training data. This was done because the default settings are likely to have the best performance across a wide range of image types. The generated vessel images were filtered before training the discriminator. Edge enhancement was not applied during training, eliminating the need for a differentiable edge detection method.

### 2.2 Architecture

Figure 3 describes the MuGAN architecture. The basic unit of the generator and discriminator architecture is the convolutional block, which comprises of a Conv2D  $3 \times 3$ ,  $n$  ( $n$  is the depth: 32, 64, 128), a batch normalization layer, an activation function and a max pooling layer of kernel size  $3 \times 3$ . The Rectified Linear Unit (ReLU) activation function was used for all convolution layers, with the exception of the last layer (Conv2D  $1 \times 1$ , 1) in the generator, which uses a sigmoid activation function. The generator is based on the UNET architecture [15] for its capability to resolve high level and low level features. The discriminators have relatively shallow architectures (3 convolutional layers) to minimize computational requirements. For the discriminator, skip connections are used to pass the output from each convolutional block to be subsequently concatenated and pooled using a global max pooling layer. The discriminator D2 has the same architecture as D1, except that D2 has a dilation rate of 2 for the convolutional block. D1 and D3 both do not utilize dilated convolutional filters (dilation rate = 0). D3 has the same architecture as D1 while D4 has the same architecture as D2.

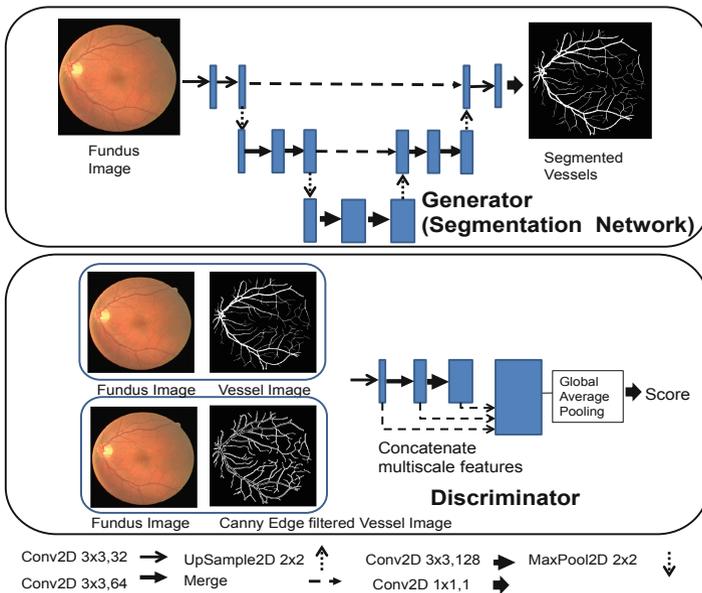


Fig. 3. Architecture of the generator and discriminator networks.

### 3 Experiments

#### 3.1 Dataset

We use the publicly available DRIVE [13] and STARE [14] dataset. The DRIVE dataset contains 40 images from diabetic patients ( $584 \times 565$  pixels,  $45^\circ$  field of view FOV). The STARE dataset contains 20 images ( $605 \times 700$  pixels,  $35^\circ$  FOV). The DRIVE dataset is evenly split for training and testing. We performed leave one out cross validation (LOOCV) for the STARE dataset. Training is performed with 19 images and testing is conducted on the ‘left out’ image. 20 iterations of these train-test cycles are performed to evaluate performance for all images. The fundus images and ground truth segmented vessel images are padded to  $640 \times 640$  pixels (DRIVE) or  $720 \times 720$  pixels (STARE). Data augmentation was performed by flipping and rotating the images, generating 266 images (DRIVE) and 252 images (STARE).

#### 3.2 Training

We implement our approach in Python 2.7 using the Keras framework. We used the Adam optimizer, initial learning rate  $2 \times 10^{-4}$  and trained for 20 epochs. It took approximately 2–3 h to train the proposed MuGAN model on the DRIVE dataset and each fold of the STARE dataset (Intel(R) Xeon(R) W-2145 CPU, NVIDIA Titan Xp GPU, Keras version 2.2.4).

#### 3.3 Evaluation Criteria

We employ the following evaluation criteria: area under the Receiver Operator Characteristic (AUC) curve, accuracy, the skeletal similarity metrics (CAL and rAccuracy) [10]. CAL comprises of 3 parameters, C: measure of fragmentation extent (ratio of the number of connected components) between the ground truth and the output, A: measure of overlap between the ground truth and the output, L: similarity in total lengths of ground truth and the output. These 3 parameters are multiplied to give a single score between 0 and 1. rAccuracy is defined as

$$rAcc = \frac{\text{Skeletal Similarity} \times \text{Pixels}_{\text{vessels}} + \text{Pixels correctly classified as background}}{\text{Pixels}_{\text{total}}} \quad (4)$$

where skeletal similarity is the weighted sum of curve similarity and thickness consistency for segmented vessels relative to the ground truth. Yan et al. [10] suggest that their proposed metric rAccuracy gives equal importance to both thick and thin vessels and we have therefore adopted their metrics for evaluation.

## 4 Results

**Comparisons with Existing Methods:** We compare our approach with DeepVessel [11], which is a deep learning approach combined with conditional random fields and

M2U-Net [12], an efficient deep learning approach based on the UNET architecture [15]. To further ascertain the effectiveness of including multiscale inputs and edge information, we train the GAN networks ( $\text{GAN}_{\text{single}}$  and  $\text{MuGAN}_{\text{noedge}}$ ). Both networks do not use the edge enhanced vessel images for training ( $\text{GAN}_{\text{single}}$  comprises of the generator and discriminator D1;  $\text{MuGAN}_{\text{noedge}}$  comprises of the generator, discriminators D1 and D2). Tables 1 and 2 gives the accuracy, CAL, rAccuracy and AUC for the DRIVE and STARE datasets.

The results show that our method performs better than other methods, particularly for CAL and rAccuracy.  $\text{MuGAN}_{\text{noedge}}$  shows mixed performance relative to  $\text{GAN}_{\text{single}}$ , performing better for the STARE dataset than the DRIVE dataset.  $\text{GAN}_{\text{single}}$  shows comparable performance with M2U-Net [12] on DRIVE. M2U-Net [12] was not trained on STARE and therefore not evaluated for STARE. DV [11] is also comparable with  $\text{GAN}_{\text{single}}$  on STARE, but also performs less well than MuGAN. MuGAN has higher accuracy compared to  $\text{MuGAN}_{\text{noedge}}$  and  $\text{GAN}_{\text{single}}$ . These results suggest that edge information, combined with multiscale discriminators, help improve segmentation accuracy. Though the edge information used in this study emphasizes the boundaries at the expense of eroding the center of the vessels, the results suggest that segmentation accuracy is not worsened due to the loss of information.

**Table 1.** Averaged performance metrics for DRIVE test images (n = 20)

	DV [11]	M2UNET [12]	$\text{GAN}_{\text{single}}$	$\text{MuGAN}_{\text{noedge}}$	MuGAN
Accuracy	0.949	0.951	0.953	0.953	<b>0.955</b>
CAL	0.687	0.812	0.811	0.786	<b>0.829</b>
rAccuracy	0.898	0.931	0.939	0.932	<b>0.944</b>
AUC	–	0.971	0.976	0.975	<b>0.978</b>

**Table 2.** Performance metrics for STARE images (n = 20, Leave-one-out cross validation)

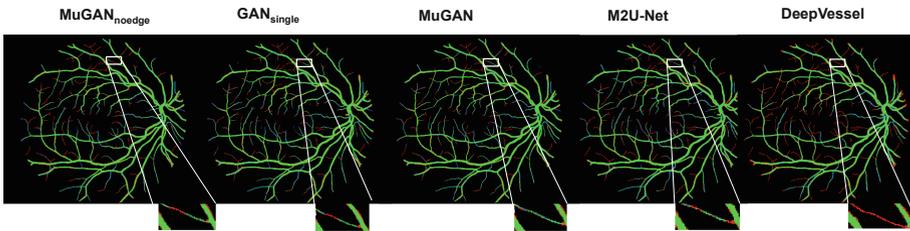
	DV [11]	$\text{GAN}_{\text{single}}$	$\text{MuGAN}_{\text{noedge}}$	MuGAN
Accuracy	0.958	0.958	0.959	<b>0.960</b>
CAL	0.713	0.715	0.737	<b>0.746</b>
rAccuracy	0.942	0.950	0.956	<b>0.960</b>
AUC	–	0.974	0.978	<b>0.981</b>

**Cross Training:** We also perform cross training between the STARE and DRIVE datasets. Table 3 shows the cross training performance. Overall, performance across the different approaches is similar. MuGAN AUC fell from 0.981 to 0.966 and 0.978 to 0.953 for the STARE and DRIVE datasets.  $\text{MuGAN}_{\text{noedge}}$  AUC fell from 0.978 to 0.968 and 0.975 to 0.956 for the STARE and DRIVE datasets. One possible explanation for the results is that STARE contains more pathological images than DRIVE, resulting in lower performance when the model trained on DRIVE is implemented on the STARE dataset. This is also supported by the greater difference in CAL scores (DRIVE: 0.829 to 0.641, STARE: 0.746 to 0.761) for MuGAN. Interestingly,  $\text{MuGAN}_{\text{noedge}}$  performs better than  $\text{GAN}_{\text{single}}$  for both datasets. We suggest that the

multiple discriminators reduce the effects of overfitting. Figure 4 also shows the qualitative improvements with our approach, with MuGAN detecting the thin vasculature absent in the outputs from other methods.

**Table 3.** Cross training performance

	train (DRIVE), test (STARE)			train (STARE), test (DRIVE)		
	GAN <sub>single</sub>	MuGAN <sub>noedge</sub>	MUGAN	GAN <sub>single</sub>	MuGAN <sub>noedge</sub>	MUGAN
Accuracy	0.946	0.949	<b>0.949</b>	0.947	<b>0.950</b>	0.949
CAL	0.599	0.640	<b>0.641</b>	0.736	<b>0.764</b>	0.761
rAccuracy	0.939	0.943	<b>0.946</b>	0.918	0.924	<b>0.928</b>
AUC	0.951	<b>0.956</b>	0.953	0.964	<b>0.968</b>	0.966



**Fig. 4.** Difference images obtained from comparing the ground truth with the segmented results. Red indicates vessels incorrectly classified as background, blue indicates background wrongly classified as vessels and green indicates correctly classified vessel pixels. The rAccuracy for this subject is 94.9%, 95.8% and 95.9% for MuGAN<sub>noedge</sub>, GAN<sub>single</sub> and MuGAN respectively, while M2U-Net [12] and DeepVessel [11] achieve a rAccuracy of 94.9% and 90.5%. (Color figure online)

## 5 Conclusion

Segmentation of thin retinal vessels is one of the main challenges in retinal image analysis. We implemented a multiple discriminator GAN approach to improve multi-scale segmentation of the retina vessels, with particular focus on the thin vessels. Our proposed method has two main novel aspects: multi-input multiscale discriminators which use (1) discriminators with varying effective receptive field sizes and (2) additional input (edge filtered vessel images) to improve segmentation performance. Future work will explore custom loss functions for thin vessel segmentation and other approaches to convey edge information during training.

## References

1. Poplin, R., et al.: Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2**, 158–164 (2018)
2. Klein, R., et al.: The relation of retinal vessel caliber to the incidence and progression of diabetic retinopathy: XIX: the Wisconsin epidemiologic study of diabetic retinopathy. *Arch. Ophthalmol.* **122**, 76–83 (2004)

3. Wang, J.J., et al.: Retinal vessel diameter and cardiovascular mortality: pooled data analysis from two older populations. *Eur. Heart J.* **28**, 1984–1992 (2007)
4. Yan, Z., Yang, X., Cheng, K.T.: A three-stage deep learning model for accurate retinal vessel segmentation. *IEEE J. Biomed. Health Inform.* **23**(4) (2019). <https://ieeexplore.ieee.org/document/8476171>
5. Goodfellow, I.J., et al.: Generative Adversarial Networks. [arXiv:1406.2661](https://arxiv.org/abs/1406.2661) (2014)
6. Luc, P., Couprie, C., Chintala, S., Verbeek, J.: Semantic Segmentation using Adversarial Networks. [arXiv:1611.08408](https://arxiv.org/abs/1611.08408) (2016)
7. Son, J., Park, S.J., Jung, K.-H.: Retinal Vessel Segmentation in Fundoscopic Images with Generative Adversarial Networks. [arXiv:1706.09318](https://arxiv.org/abs/1706.09318) (2017)
8. Durugkar, I., Gemp, I., Mahadevan, S.: Generative Multi-Adversarial Networks. [arXiv:1611.01673](https://arxiv.org/abs/1611.01673) (2016)
9. Yu, F., Koltun, V.: Multi-Scale Context Aggregation by Dilated Convolutions. [arXiv:1511.07122](https://arxiv.org/abs/1511.07122) (2015)
10. Yan, Z., Yang, X., Cheng, K.: A skeletal similarity metric for quality evaluation of retinal vessel segmentation. *IEEE Trans. Med. Imaging* **37**, 1045–1057 (2018)
11. Fu, H., Xu, Y., Lin, S., Kee Wong, D.W., Liu, J.: DeepVessel: retinal vessel segmentation via deep learning and conditional random field. In: Ourselin, S., Joskowicz, L., Sabuncu, M., Unal, G., Wells, W. (eds.) *MICCAI 2016*. LNCS, vol. 9901, pp. 132–139. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46723-8\\_16](https://doi.org/10.1007/978-3-319-46723-8_16)
12. Laibacher, T., Weyde, T., Jalali, S.: M2U-Net: Effective and Efficient Retinal Vessel Segmentation for Resource-Constrained Environments. [arXiv:1811.07738](https://arxiv.org/abs/1811.07738) (2018)
13. Staal, J., Abramoff, M.D., Niemeijer, M., Viergever, M.A., Ginneken, B.V.: Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **23**, 501–509 (2004)
14. Hoover, A.D., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Med. Imaging* **19**, 203–210 (2000)
15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, William M., Frangi, Alejandro F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)