



Towards Accurate Segmentation of Retinal Vessels and the Optic Disc in Fundoscopic Images with Generative Adversarial Networks

Jaemin Son¹ · Sang Jun Park² · Kyu-Hwan Jung¹

Published online: 5 October 2018

© Society for Imaging Informatics in Medicine 2018

Abstract

Automatic segmentation of the retinal vasculature and the optic disc is a crucial task for accurate geometric analysis and reliable automated diagnosis. In recent years, Convolutional Neural Networks (CNN) have shown outstanding performance compared to the conventional approaches in the segmentation tasks. In this paper, we experimentally measure the performance gain for Generative Adversarial Networks (GAN) framework when applied to the segmentation tasks. We show that GAN achieves statistically significant improvement in area under the receiver operating characteristic (AU-ROC) and area under the precision and recall curve (AU-PR) on two public datasets (DRIVE, STARE) by segmenting fine vessels. Also, we found a model that surpassed the current state-of-the-art method by 0.2 – 1.0% in AU-ROC and 0.8 – 1.2% in AU-PR and 0.5 – 0.7% in dice coefficient. In contrast, significant improvements were not observed in the optic disc segmentation task on DRIONS-DB, RIM-ONE (r3) and Drishti-GS datasets in AU-ROC and AU-PR.

Keywords Retinal vessel segmentation · Optic disc segmentation · Convolutional neural network · Generative adversarial networks

Abbreviations

Convolutional Neural Networks (CNN)
Generative Adversarial Networks (GAN)

Introduction

Analysis of anatomical structures in fundoscopic images such as retinal vessels and the optic disc provides rich information about the condition of the eyes and even helps estimate general health status. From abnormality in vascular structures, ophthalmologists can detect early signs of increased systemic vascular burden from hypertension

and diabetes mellitus as well as vision threatening retinal vascular diseases such as retinal vein occlusion and retinal artery occlusion. Also, evaluation of the optic disc is essential for diagnosing glaucoma and for monitoring patients with glaucoma.

Specifically, the geometric parameters for the retinal vascular structure, such as diameter, curvature tortuosity, and branching angles, are reported to be highly correlated with coronary artery disease [1] and diabetes mellitus [2], and even cognitive ability [3, 4], dementia [5], and body mass [6] in older population. Furthermore, it is widely accepted that the diameters of the retinal vessel can even predict the risk of cardiovascular diseases in the future [7, 8].

In general, the geometric parameters are estimated with computer software to guarantee systemic and coherent estimations. Computer software such as VAMPIRE [9] and Singapore “I” Vessel Assessment automatically segments retinal vessels and the optic disc from fundoscopic images from which various measures regarding vascular structure are computed. Needless to say, the precise segmentation, as a preliminary step, is imperative to accurate and reliable estimations of the geometric parameters.

Furthermore, the precise localization of retinal landmarks including retinal vessels and the optic disc is an essential prerequisite step for automatic diagnosis systems for

The first two authors are co-first authors.

✉ Kyu-Hwan Jung
khwan.jung@vuno.co

Jaemin Son
woalsdnd@vuno.com

Sang Jun Park
sangjunpark@snu.ac.kr

¹ VUNO Inc., 6F, 507, Gangnam-daero, Seocho-gu, Seoul, Republic of Korea

² Department of Ophthalmology, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

fundoscopic images. For instance, when automatically detecting anomalies in the retinal vasculature, reliable localization of retinal vasculature should precede the process of decision making for reliable results so that the decision should only be made based on vascular structures.

For many years, automatic segmentation methods have been extensively developed. Many early attempts tackled the segmentation problems from a signal processing perspective based on empirical evidence about visual characteristics for each structure. In the case of vessel segmentation, many approaches are designed to detect lines in image patches from the observation that vessels stretch linearly in small scale [10–15]. Even though these methods tend to be intuitive to humans, however, they would not perform as expected when heuristics do not apply. Also, the fact that the optic disc is brighter than the surroundings is well exploited in heuristic methods [16, 17]; however, the performance was not reliable. With the advances in machine learning, more improved results were obtained with extracted features using non-linear classifiers such as SVM [18], Bayesian classifier [19] and boosted decision trees [20]. Still, the performance was not commensurate to human experts.

In recent years, Convolutional Neural Networks (CNN) have shown outstanding performance in various computer vision tasks including classification [21–23], detection [24, 25], and segmentation [26–28]. Several studies have already demonstrated that CNN could achieve the unprecedented performance in segmenting retinal vessels [29–32] and the optic discs [29], even surpassing the ability of human observers in several publicly available datasets. However, segmented vessels with CNN are rather blurry around minuscule and faint branches where distinction is unclear between vessel walls and retinal background. This

is because convolution operations basically respond to local structures and the conventional objective functions only rely on pixel-wise comparison between the output and gold standard. It would be desirable if natural constraints on anatomical structures are actively accommodated in to gain more realistic segmentations.

In fact, the segmentation tasks can be considered as an image translation task where a probability map for segmentation is generated from an input fundoscopic image. If the outputs are constrained to resemble the human observer's annotation, more realistic and accurate probability maps can be obtained. In this paper, we experiment the effect of GAN framework when applied to the segmentation of retinal vasculature and the optic disc. Experimental results suggest that GAN helps improve the performance in vessel segmentation by segmenting fine vessels. In contrast, significant improvement was not observed in segmenting the optic disc. We suspect that the optic disc is a noticeable blob that CNN can readily segment with high precision.

Methods

Overall Framework

Generative Adversarial Networks (GAN) are a framework that enables us to create and utilize realistic outputs as the gold standard [33]. GAN consists of two sub-networks: a discriminator and a generator. While the discriminator tries to distinguish the gold standard from outputs from the generator, the generator tries to generate outputs which are as realistic as possible so that the discriminator cannot differentiate from the gold standard.

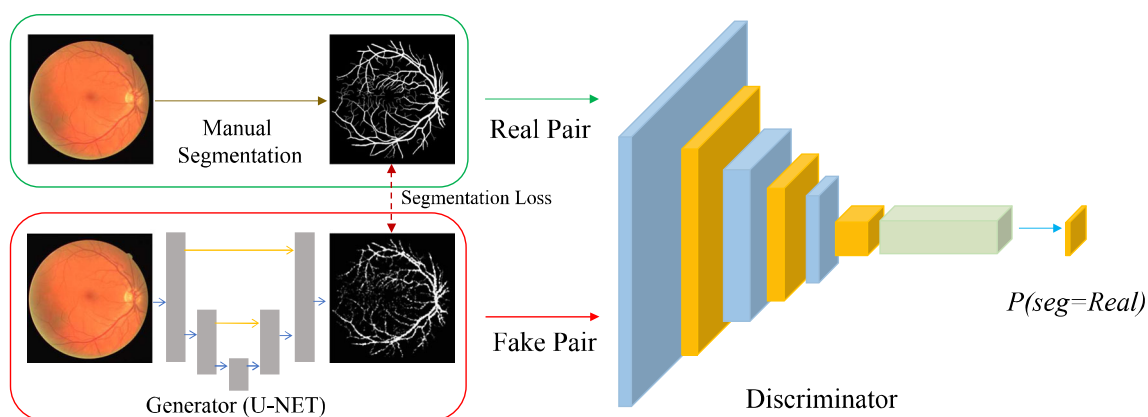


Fig. 1 RetinaGAN framework in the case of vessel segmentation. A fundus image constitutes a *real pair* with a manual vessel segmentation, and a *fake pair* with a vessel segmentation generated from the

generator. A discriminator is trained to output a probability map for the vessel segmentation being real. Varying sizes of the probability map were experimented from pixel level to image level

In principle, our framework is similar to conditional-GAN [34], which is suitable for segmentation tasks¹. In our case, the generator is given a fundoscopic image and generates probability maps of retinal vessels and the optic disc that have the same size as the input image. Values in the probability maps range from 0 to 1, indicating the probability of being a pixel of an anatomical structure. Then, the discriminator takes a pair of a fundoscopic image and the probability map and determines whether the probability map is the gold standard from the human annotator or output of the generator. The objective of the generator is to output a probability map that the discriminator cannot discern from the gold standard. In such a way, the generator is constrained to produce more realistic outputs. The overall framework, which we call *RetinaGAN*, is depicted in Fig. 1 for the case of vessel segmentation. As the generator network, we utilized U-Net style network whose details are provided in Appendix A. For the discriminator network, 4 networks with different receptive fields for the decision were experimented (Pixel GAN, Patch GAN-1, Patch GAN-2, Image GAN). Details of discriminators are described in Appendix B. We applied the identical networks to the optic disc segmentation. Also, detailed explanation about objective function for our GAN framework is described in Appendix C.

Datasets

Vessel Segmentation

Our methods were evaluated on 2 public datasets, namely DRIVE [14] and STARE [35]. DRIVE dataset consists of fundoscopic images with the corresponding FOV masks and manual segmentation of the vessels. A total of 40 sets of images are split into 20/20 sets for training and test sets. In the training set, a single manual segmentation was provided for each fundoscopic image. In the test set, each fundoscopic image was given two manual segmentation annotated by two independent human observers who were trained by an experienced ophthalmologist. For training and testing, we regarded the first annotator's vessel segmentation as the gold standard and compared the results with the segmentation of the second annotator.

The STARE dataset includes 20 sets of fundoscopic images and vessel segmentation from two independent annotators, *Adam Hoover* (AH) and *Valentina Kouznetsova* (VK). We regarded the annotation of AH as the gold standard and used that of VK as a reference point for a

human observer. Since masks for Field of View (FOV) are not given in the STARE dataset, we generated them. In our experiments, we retained the first 10 images for training and tested on the rest as was done in the previous literature [29] to compare the performance.

Optic Disc Segmentation

Our methods were tested on 3 datasets - DRIONS-DB [36], RIM-ONE [37] and Drishti-GS [38]. DRIONS-DB dataset consists of 110 sets of fundoscopic images and two manual annotations of the optic disc. Annotations were given by coordinates of vertices for polygons in an orderly manner. In the experiment, the first 100 images were used as training/validation sets and the last 10 images were evaluated for testing the models. The results in the test set were compared with DRIU [29].

RIM-ONE (r3) dataset contains 159 fundoscopic images and two manual segmentations for each fundoscopic image. Images were cropped and magnified around the optic disc as the dataset was dedicated to the segmentation of the optic disc and optic cup. We split the data into 145/14 for training and test sets respectively and the results in the test set were compared with DRIU [29].

Drishti-GS dataset consists of 101 macula-centered fundus images which are divided into 50 training and 51 test images. High-resolution images with a width of 2896 and height of 1944 were collected in Indian hospital and annotated by Indian experts. A mask of the optic disc was retrieved by averaging masks of 4 ophthalmologists with varying experience (3, 5, 9, 20 years).

Experimental Design

In our experiments, we evaluated the performance of GAN with 4 discriminators and a U-Net architecture after training from the scratch. Throughout all experiments, we maintained the same experimental settings, including utilization of uniform hyper-parameters during training and application of identical procedures for image preprocessing, to eliminate external variability that can potentially confound the performance evaluation. Details of experimental settings are described in Appendix D.

As metric, area under the receiver operating characteristic (AU-ROC) and area under the precision and recall curve (AU-PR) were compared between the models. The curves were drawn by shifting the threshold for vessels and the optic disc from 0 to 1. In order to measure statistical significance in the improvement of AU-ROC and AU-PR, we ran experiments 5 times on the same setting and estimated the p -value via Kolmogorov-Smirnov test which was widely

¹Our source code is available at <https://bitbucket.org/woalsndnd/retinagan>

Table 1 Comparison of models with different discriminators on two datasets with respect to area under the receiver operating characteristic (AU-ROC) and area under the precision and recall curve (AU-PR)

Model	DRIVE		STARE		
	AU-ROC	AU-PR	AU-ROC	AU-PR	
U-Net (No discriminator)	0.9671 ± 0.0028	0.8898 ± 0.0016	0.9736 ± 0.0038	0.9085 ± 0.0028	
RetinaGAN	Pixel (1×1)	0.9723 ± 0.0005	0.8927 ± 0.0009	0.9142 ± 0.0021	
	Patch-1 (8×8)	0.9730 ± 0.0004	0.8951 ± 0.0008	0.9123 ± 0.0019	
	Patch-2 (64×64)	0.9766 ± 0.0029	0.9045 ± 0.0075	0.9832 ± 0.0019	0.9194 ± 0.0011
	Image (640×640)	0.9800 ± 0.0010	0.9134 ± 0.0011	0.9858 ± 0.0016	0.9197 ± 0.0045

Digits are rounded to the 4th decimal point and shown in the form of $\mu \pm \sigma$ where σ is the empirical standard deviation. Receptive Field size of the discriminator is shown in the parenthesis

used for testing whether two distributions differ with a small number of samples. We fixed the training set and test set to exclude any external perturbations induced by the shuffling of datasets.

We compared the best result of our models with other techniques regarding AU-ROC, AU-PR and dice coefficient or F1 measure. To compare dice coefficient, best operating point that yielded highest value was selected for all methods as in the previous literature [29]. For fair measurement, only pixels inside the Field Of View (FOV) were counted when computing the measures.

Results

Vessel Segmentation

Statistics of AU-ROC and AU-PR are shown in Table 1 for different GAN architectures with distinct discriminators. U-Net, which is not equipped with a discriminator,

demonstrated inferior performance to other GAN architectures with substantial confidence regarding both AU-ROC and AU-PR on both datasets (p value < 0.05). This suggests that GAN meaningfully contributes to the improvement of segmentation quality. Also, it is noticeable that a deeper and more complex discriminator yields better measures. Image GAN outperformed pixel GAN and patch GAN-1 statistically significantly (p value < 0.05) in both measures on both datasets. Image GAN surpassed patch GAN-2 in both measures only on DRIVE dataset (p value < 0.05) and no statistically significant improvement was observed on STARE dataset. This observation is consistent to the claim that a competent discriminator is key to successful training with GAN [33, 39].

As obvious in Table 1, AU-ROC and AU-PR of Image GAN do not seem overwhelmingly higher than those of U-Net. However, the gap in performance measures between U-Net architecture and GAN architectures mainly comes from segmentation of fine vessels. Exemplar comparison is shown in Fig. 2 between U-Net and Image GAN. As shown

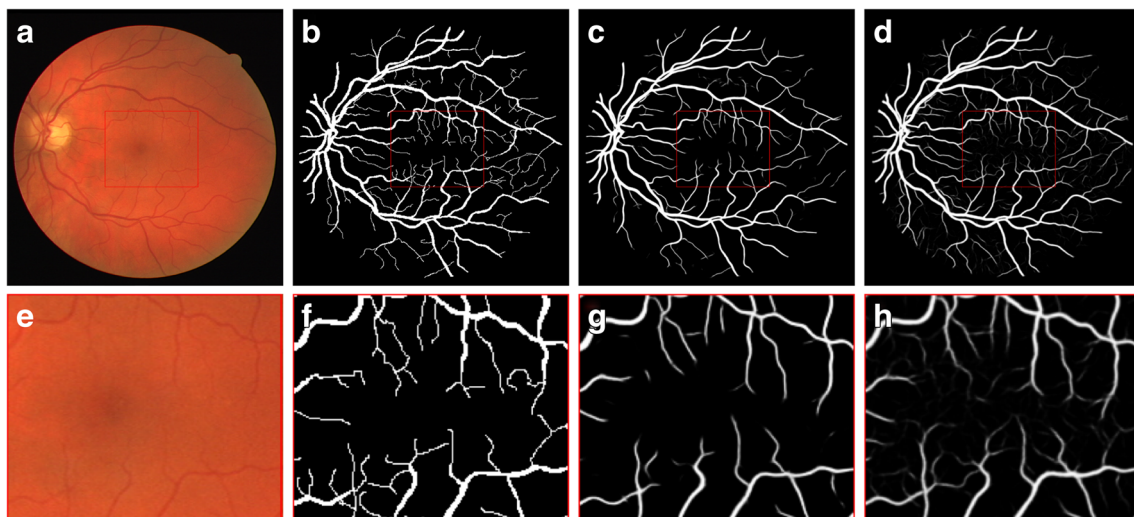


Fig. 2 Fundus image (a, e), gold standard (b, f), output of U-Net (c, g), and output of RetinaGAN-Image (d, h). Central area of images at the top row is zoomed up at the bottom row

in the figure, fine vessels are more preserved in the result of Image GAN than that of U-Net. Basically, detecting thin vessels is difficult unlike thick main branches since colors differ merely around them. Under such constraints, U-Net is only penalized for missing fine vessels by segmentation loss which is minuscule due to the small number of pixels in fine vessels. However, GAN is also penalized for the tendency to miss fine vessels included in the gold standard. Therefore, GAN architecture captures fine vessels better than U-Net.

Figure 3 shows ROC and PR curves for RetinaGAN (Image) and other existing methods (DRIU [29], N^4 -Fields [40], HED [31], Kernel Boost [41], CRFs [42], and Wavelets [19]) around an operating regime with high F1 score. As shown in the figure, Image GAN draws ROC and PR curves above other methods in the operating regime on both datasets. The margin between curves of Image GAN

and DRIU seems inconspicuous; however, the difference is stark in probability maps. This is because AUROC and AUPR can still be high even though probability values are unrealistic.

Figure 4 compares vessel images of the gold standard with results of current best method (DRIU) and our method. Probability maps are converted into binary masks by Otsu threshold [43] and original fundoscopic images are multiplied element-wise with the binary masks. As shown in the figure, vessels extracted by our method are more concordant to the gold standard. DRIU tends to over-segment fine vessels and boundaries between vessels and the background.

Also, main trunks are over-segmented as well preventing precise measurement of vascular caliber which is crucial for quantitative studies in ophthalmology. Therefore, our

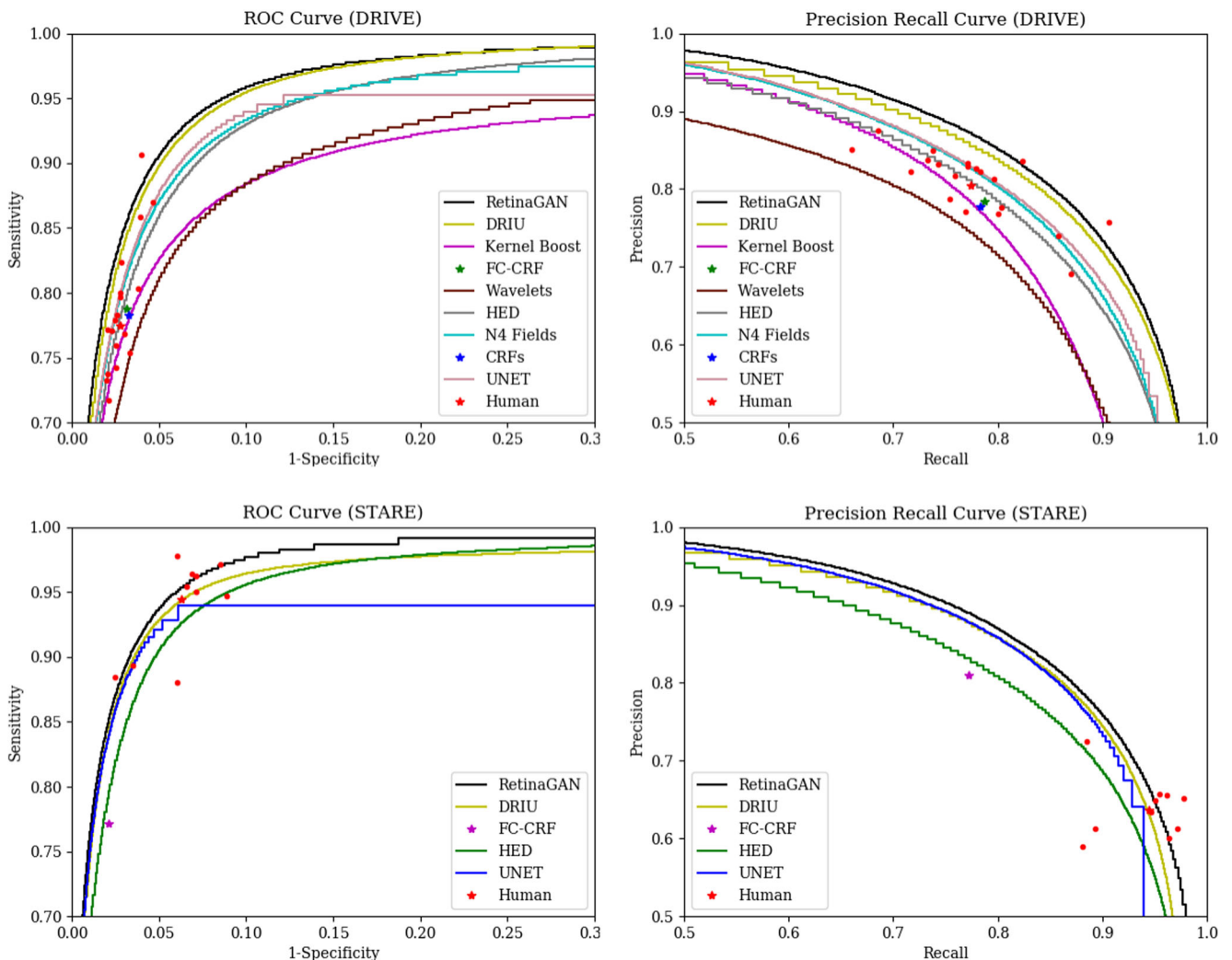
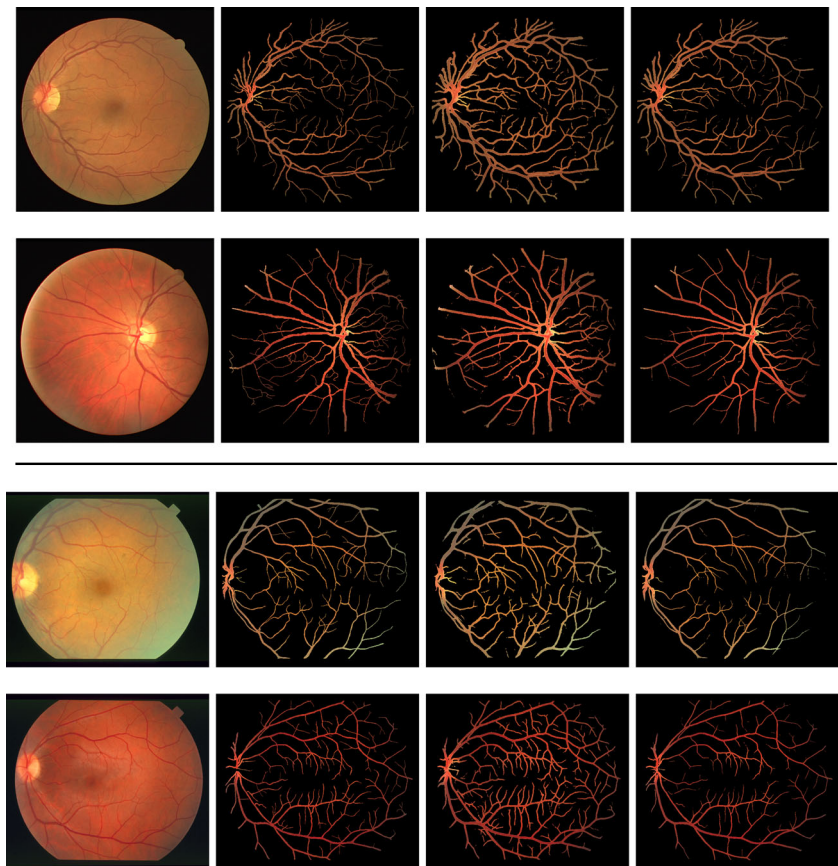


Fig. 3 Receiver Operating Characteristic (ROC) curve and Precision and Recall (PR) curve for various methods on the DRIVE dataset (top) and STARE dataset (bottom). Red dots mark performance of the human annotator in individual images

Fig. 4 (From the leftmost column to the rightmost column) fundoscopic images, gold standard, current state-of-the-art technique (DRIU [29]) and RetinaGAN on DRIVE (top half) and STARE (bottom half) dataset



method expresses the probability of vessels in more sensible ways than DRIU and is more suitable to tasks that require probability maps to describe uncertainty.

Table 2 summarizes AU-ROC, AU-PR and dice coefficient. We retrieved output images of other methods from previously published literature [29] and computed measures

from the images. Our method demonstrated superior AU-ROC, AU-PR and dice coefficient to any other existing methods. Also, our method surpassed the human annotator's ability on both DRIVE and STARE datasets.

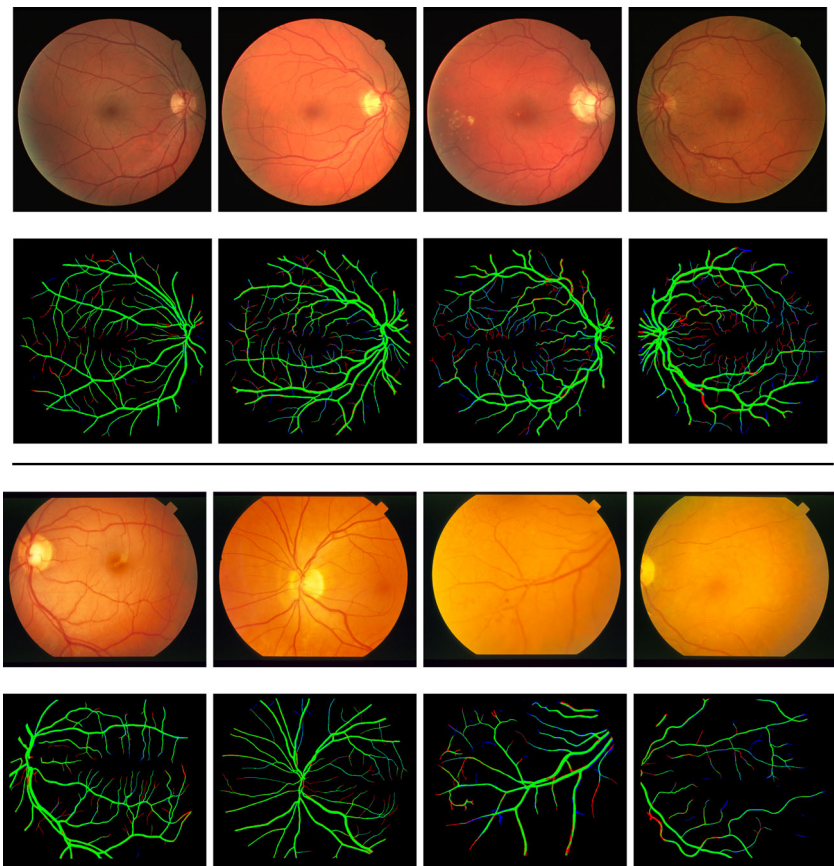
Figure 5 shows results of RetinaGAN with the best and worst dice coefficients. The threshold was chosen to

Table 2 Comparison of different methods on two datasets with respect to area under the receiver operating characteristic (AU-ROC), area under the precision and recall curve (AU-PR) and dice coefficient

Method	DRIVE			STARE		
	AU-ROC	AU-PR	Dice	AU-ROC	AU-PR	Dice
CRF [42]	–	–	0.7797	–	–	–
FC-CRF [44]	–	–	0.7858	–	–	0.7902
Kernel Boost [41]	0.9307	0.8464	0.7795	–	–	–
HED [31]	0.9696	0.8773	0.7935	0.9764	0.8888	0.8057
Wavelets [19]	0.9436	0.8149	0.7600	0.9694	0.8433	0.7756
N^4 -Fields [40]	0.9686	0.8851	0.8018	–	–	–
DRIU [29]	0.9793	0.9064	0.8207	0.9772	0.9101	0.8323
Human	–	–	0.7889	–	–	0.7605
U-Net	0.9718	0.8904	0.8039	0.9759	0.9093	0.8307
RetinaGAN (Image)	<i>0.9810</i>	<i>0.9145</i>	<i>0.8275</i>	<i>0.9873</i>	<i>0.9226</i>	<i>0.8378</i>

Digits are rounded to 4th decimal point

Fig. 5 Fundoscopic images and segmentation results with best dice coefficient (2 columns on the left) and worst dice coefficient (2 columns on the right) on DRIVE (top half) and STARE (bottom half) dataset. Green region represents correct segmentation while red and blue denote false negative and false positive respectively



maximize the average dice coefficient for all images. In general, thick and conspicuous trunks were not missed. When thick trunks were missed, the trunks were hard to discern from the retinal wall in the fundoscopic images. On the other hand, extremely thin vessels that span only 1 pixel in width are often missed (red) and sometimes additionally detected (blue). Drawing extremely thin vessels demands understanding of vasculatures as invisible vessels

need to be extrapolated according to surrounding vascular structures based on anatomical knowledge that vessels are connected. We suspect that our model does not possess such a reasoning processes and ended up missing them. Other techniques would be necessary to reduce false negatives. When it comes to false positives, on the other hand, it is desirable to detect those that the human annotator may miss due to intra-observer variance. Results suggest that visible

Table 3 Comparison of different discriminators with respect to area under the receiver operating characteristic (AU-ROC) and area under the precision and recall curve (AU-PR) on DRIONS-DB, RIM-ONE (r3) and Drishti-GS

Method	DRIONS-DB		RIM-ONE		Drishti-GS	
	AU-ROC	AU-PR	AU-ROC	AU-PR	AU-ROC	AU-PR
U-Net (No discriminator)	0.9996 ± 0.0001	0.9886 ± 0.0010	0.9995 ± 0.0001	0.9913 ± 0.0020	0.9997 ± 0.0000	0.9957 ± 0.0017
RetinaGAN Pixel (1 × 1)	0.9995 ± 0.0001	0.9876 ± 0.0029	0.9995 ± 0.0000	0.9915 ± 0.0007	0.9990 ± 0.0012	0.9929 ± 0.0028
Patch-1 (8 × 8)	0.9996 ± 0.0001	0.9880 ± 0.0015	0.9995 ± 0.0001	0.9910 ± 0.0011	0.9991 ± 0.0010	0.9937 ± 0.0029
Patch-2 (64 × 64)	0.9995 ± 0.0001	0.9868 ± 0.0018	0.9995 ± 0.0000	0.9915 ± 0.0002	0.9994 ± 0.0006	0.9891 ± 0.0095
Image (640 × 640)	0.9990 ± 0.0004	0.9769 ± 0.0075	0.9995 ± 0.0001	0.9909 ± 0.0013	0.9989 ± 0.0004	0.9719 ± 0.0112

Digits are rounded to the 4th decimal point and shown in the format of $\mu \pm \sigma$ where σ is the empirical standard deviation. Receptive Field size of the discriminator is shown in the parenthesis

Table 4 Comparison of U-Net, DRIU and a human annotator with respect to boundary error and dice coefficient on DRIONS-DB, RIM-ONE (r3) and Drishti-GS

Method	DRIONS-DB		RIM-ONE		Drishti-GS	
	Boundary Error	Dice	Boundary Error	Dice		
Human	1.6111 ± 0.5875	0.9636	3.5530 ± 1.511	0.9453	–	–
DRIU	1.2933 ± 0.3017	0.9672	2.3321 ± 1.1483	0.9548	–	–
U-Net	1.1466 ± 0.3351	0.9688	2.3443 ± 0.9479	0.9562	1.4348 ± 0.6048	0.9678
RetinaGAN (Pixel)	1.2461 ± 0.3930	0.9685	2.2290 ± 1.2608	0.9546	1.3961 ± 0.7257	0.9674

Digits are rounded to the 4th decimal point and shown in the format of $\mu \pm \sigma$ where σ is empirical standard deviation

vessels are still missed while more invisible vessels are segmented by the annotator.

Optic Disc Segmentation

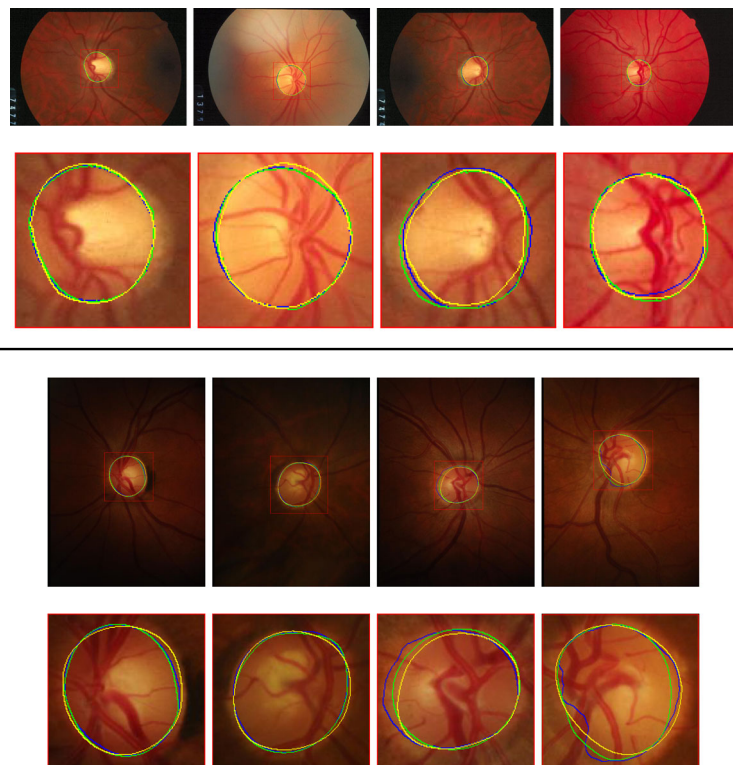
Statistics of AU-ROC and AU-PR are shown in Table 3 for various GAN architectures with different discriminators. It turned out that there were no statistically significant improvements in AU-ROC and AU-PR with GAN architectures on both datasets (p value > 0.2 for all pairs). In fact, U-Net already achieves very high AU-ROC and AU-PR. Unlike the case of vessels, segmentation of the optic discs can be reliably done without GAN architecture since the optic disc is a salient blob-like structure that U-Net is especially good at segmenting.

In Table 4, performance of the RetinaGAN (Pixel) is compared with U-Net, the current state-of-the-art method (DRIU [29]) and the 2nd annotator regarding dice coefficient and boundary error. Boundary error, or average distance between boundaries, is formally described as

$$\frac{1}{|B_{out}|} \sum_{o \in B_{out}} \min_{g \in B_{gt}} \|o - g\|_2 \quad (1)$$

where B_{out} and B_{gt} are sets of pixels constituting the boundary of the output and ground truth. Operating point with best F1 score is found to compare with results of the current state-of-the-art method (DRIU) [29]. As shown in the table, all CNN models (DRIU, U-Net and RetinaGAN) showed similar performance on both datasets which is comparable to or slightly better than an independent human

Fig. 6 Original fundoscopic images and detected boundary of the optic disc with smallest boundary error (2 columns on the left) and largest boundary error (2 columns on the right) on DRIONS-DB (top half) and RIM-ONE (r3) (bottom half) dataset. Curves are color-coded by type of annotation (Green : Ground Truth, Yellow : Human, Blue : U-Net)



expert. Superiority of one model over others could not be clearly seen. Dice coefficients differ only in the 4th decimal place, the quantity of which is interpreted as unnoticeable difference to human eyes. Also, boundary errors do not significantly vary between the models.

Figure 6 shows results of a U-Net with best and worst boundary error. Overall, U-Net robustly segments the optic disc. Worst cases on RIM-ONE (r3) dataset involve a tumultuous boundary that humans would not draw. This is mainly because humans would draw disc-like shapes even when the contour is rough. By contrast, U-Net solely depends on color information to detect the optic disc which results in a contour that is not smooth. Even though it seems that GAN framework can forge smoother segmentation, we could not witness any substantial differences in the worst cases. We suspect that the ability to draw smooth contours in the absence of clear visual cues is unattainable with GAN framework as is the case for segmentation of invisible vessels that require extrapolation.

Discussion

In this study, we introduced Generative Adversarial Networks (GAN) framework [33] to segmentation of retinal vessels and the optic disc to leverage the capability of Convolutional Neural Networks (CNN). The rationale behind our approach is that the GAN loss that penalizes dissimilarity from the gold standard would encourage the segmentation results to be more realistic and thus, improve the segmentation quality [33, 34, 39].

In the experiments, the GAN framework was formulated with a U-Net [27] style generator and several discriminators with different discriminatory capabilities. Statistical significance in the improvement of AU-ROC and AU-PR was investigated among independent GAN frameworks and U-Net. The choice of AU-ROC and AU-PR as performance measures is motivated for the comparison in the entire operating regime following the suit of previously published literature [19, 29, 31, 40, 41].

Experimental results in vessel segmentations suggest that the GAN framework meaningfully boosts AU-ROC and AU-PR compared to U-Net as shown in Table 1. GAN frameworks with any discriminator outperformed U-Net statistically significantly (p value < 0.05). The amount of improvement seems insignificant in nominal digits, however, improvement in AU-ROC and AU-PR mainly derived from localization of fine and thin vessels at terminal branches that are sometimes nearly indistinguishable with the naked eyes (Fig. 2). Since fine and thin vessels constitute only a minor fraction among pixels of vessels, the improvement was minimal, though they were captured. However, the ability to localize fine and thin retinal

vessels has clinical meaning in certain cases such as when estimating visual acuity of patients with diabetic macular edema [45].

The improvements originate from additional loss in the GAN architecture. Unlike U-Net that penalizes under-segmentation only, the GAN architecture additionally penalizes overlooked fine vessels with the discriminator. Therefore, the GAN architecture could detect fine and thin vessels better than U-Net. We believe that the GAN framework can also be applied to other modalities. Especially, GAN framework is expected to help segment microvasculature in Optical Coherence Tomography Angiography [46].

Also, it is confirmed in Table 1 that a discriminator with more discriminatory capabilities lead to higher AU-ROC and AU-PR which is consistent with previous literature [33, 39]. The best model in our experiments (Image GAN) surpassed the current state-of-the-art method (DRIU [29]) let alone an independent human annotator in the measures of AU-ROC, AU-PR and dice coefficient on DRIVE and STARE dataset as shown in Fig. 3 and Table 2. It has already been reported that CNN-based models [29] easily outperformed any other methods that depend on heuristics [10–17] and hand-crafted features [18–20]. Up to our knowledge, our method achieves the state-of-the-art performance on the two public datasets in the measures that are widely accepted in retinal vessel segmentation.

However, as can be seen in Figs. 2 and 5, our method still under-segments fine vessels that the human annotator can identify with the help of anatomical knowledge that the vessels are connected and do not appear out of nowhere. Also, the knowledge that retinal vessels always branch in acute angle helps detect vessels more precisely. We believe that post-processing of the segmentation results based on such anatomical knowledge about retinal vascular structure would yield more reliable vessel segmentation. Region growing algorithm [13, 15] that expands the region from several seed points or morphological dilation and erosion would help connect scattered pieces of vessels.

Unlike vessels, segmentation of the optic disc would not witness statistically meaningful improvement in AU-ROC and AU-PR through the introduction of the GAN framework. As obvious in Table 3, the optic disc is reliably segmented with U-Net presenting high AU-ROC and AU-PR on both datasets. The optic disc is a noticeable blob that U-Net can readily segment with high precision. Still, our U-Net architecture outperformed an independent human annotator and showed commensurate or slightly superior performance compared with the current state-of-the-art method [29] with respect to boundary error and dice coefficient (Table 4).

However, our U-Net style segmentation network sometimes failed to clearly segment the optic disc in an oval shape (Fig. 6). As was the case with vessel segmentation,

we believe that post-processing can complement limitation of our method with the anatomical knowledge that the optic disc is generally an oval shape and a convex hull. Therefore, transforming the detected optic disc into a convex hull [47] would result in a more reliable and interpretable segmentation result.

Conclusion

We have demonstrated that the introduction of GAN framework to retinal vessel segmentation improves segmentation quality by detecting fine and thin vessels better than the U-Net. The best model that we have discovered also surpassed the current state-of-the-art performance. When it comes to segmentation of the optic disc, however, the improvements were not statistically significant. We suspect that the optic disc can be easily segmented with U-Net due to its conspicuity. For future work, we look forward to combining anatomical knowledge about vascular structure and the optic disc to further advance the segmentation quality.

Funding Information This study was supported by the Research Grant for Intelligence Information Service Expansion Project, which is funded by National IT Industry Promotion Agency (NIPA-C0202-17-1045) and the Small Grant for Exploratory Research of the National Research Foundation of Korea (NRF), which is funded by the Ministry of Science, ICT, and Future Planning (NRF-2015R1D1A1A02062194). The sponsors or funding organizations had no role in the design or conduct of this research.

Appendix A: Network of a Generator

For the generator, we follow the main structure of U-Net [27] where initial convolutional feature maps are depth-concatenated to layers that are upsampled from the bottleneck layer. U-Net is a fully convolutional network [28]

which consists only of convolutional layers and depth concatenation is known to be crucial to segmentation tasks as the initial feature maps maintain low-level features such as edges and blobs that can be properly exploited for accurate segmentation. We modified U-Net to maintain width and height of feature maps during convolution operations by padding zeros at the edges of each feature map (zero-padding). Though concerns could be cast on the boundary effect with zero-padding, we observed that our U-Net style network does not suffer from false positives at the boundary. Also, the number of feature maps for each layer is shrunk to avoid overfitting and accelerate training and inference. Details of the generator network are shown in Table 5. We used ReLU for the activation function [48] and batch-normalization [49] before the activation.

Appendix B: Network of Discriminators

There is variability in the choice of discriminators by the size of pixels on which the decision is made. We explored several models for the discriminators with different output size as done in the previous work [50]. In the atomic level, a discriminator can determine the authenticity pixel-wise (Pixel GAN) while the judgment can also be made in the image level (Image GAN). Between the extremes, it is also possible to set the receptive field to a $K \times K$ patch where the decision can be given in the patch level (Patch GAN). We investigated Pixel GAN, Image GAN, and Patch GAN with two intermediary patch sizes (8×8 , 64×64). Details of the network configuration for discriminators are given in Table 6.

Appendix C: Objective Function

Let the generator G be a mapping from a fundoscopic image x to a probability map y , or $G : x \mapsto y$.

Table 5 Details of the generator network

Block	Input	Conv 1-4	Conv 5	Conv 6-9	Output
Output Size	(640, 640, 3)	$\left(\frac{640}{2^{n-1}}, \frac{640}{2^{n-1}}, 32 * 2^{n-1}\right)$	$\left(\frac{640}{2^{n-1}}, \frac{640}{2^{n-1}}, 32 * 2^{n-1}\right)$	$\left(\frac{640}{2^{9-n}}, \frac{640}{2^{9-n}}, 32 * 2^{9-n}\right)$	(640, 640, 1)
Upsample	-	-	-	Conv $n - 1$	-
Depth Concat	-	-	-	Conv $10 - n$	-
Operations	-	$\left\{ \begin{array}{l} 3 \times 3\text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2$ 2 × 2 maxpool	$\left\{ \begin{array}{l} 3 \times 3\text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2$	$\left\{ \begin{array}{l} 3 \times 3\text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2$	$\left\{ \begin{array}{l} 1 \times 1\text{conv} \\ \text{sigmoid} \end{array} \right\} \times 1$

Output size represents (width, height, depth). n refers to index of convolution block from 1 to 9. In conv 6-9, lower layer and upsampled layer are first concatenated along depth and is applied operations designated. Output depth for convolution operations is set equal to output depth of the block

Table 6 Details of the discriminator networks with difference size of receptive fields

Block	Pixel GAN		Patch GAN-1		Patch GAN-2		Image GAN	
	Operation	Output size	Operation	Output size	Operation	Output size	Operation	Output size
conv 1	$\left\{ \begin{array}{l} 3 \times 3 \text{conv} \\ \text{Leaky ReLU} \end{array} \right\} \times 3$	(640,640,32)	$\left\{ \begin{array}{l} 3 \times 3 \text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2^*$ $2 \times 2 \text{maxpool}$	(160, 160, 32)	$\left\{ \begin{array}{l} 3 \times 3 \text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2^*$ $2 \times 2 \text{maxpool}$	(160, 160, 32)	$\left\{ \begin{array}{l} 3 \times 3 \text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2^*$ $2 \times 2 \text{maxpool}$	(160, 160, 32)
conv 2	-	-	$\left\{ \begin{array}{l} 3 \times 3 \text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2$ $2 \times 2 \text{maxpool}$	(80, 80, 64)	$\left\{ \begin{array}{l} 3 \times 3 \text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2^*$ $2 \times 2 \text{maxpool}$	(40, 40, 64)	$\left\{ \begin{array}{l} 3 \times 3 \text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2^*$ $2 \times 2 \text{maxpool}$	(40, 40, 64)
conv 3	-	-	$\left\{ \begin{array}{l} 3 \times 3 \text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2$ $2 \times 2 \text{maxpool}$	(80, 80, 128)	$\left\{ \begin{array}{l} 3 \times 3 \text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2$ $2 \times 2 \text{maxpool}$	(20, 20, 128)	$\left\{ \begin{array}{l} 3 \times 3 \text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2$ $2 \times 2 \text{maxpool}$	(20, 20, 128)
conv 4	-	-	-	-	$\left\{ \begin{array}{l} 3 \times 3 \text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2$ $2 \times 2 \text{maxpool}$	(10, 10, 256)	$\left\{ \begin{array}{l} 3 \times 3 \text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2$ $2 \times 2 \text{maxpool}$	(10, 10, 256)
conv 5	-	-	-	-	$\left\{ \begin{array}{l} 3 \times 3 \text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2$ $2 \times 2 \text{maxpool}$	(10, 10, 512)	$\left\{ \begin{array}{l} 3 \times 3 \text{conv} \\ \text{batch-norm} \\ \text{ReLU} \end{array} \right\} \times 2$ $2 \times 2 \text{maxpool}$	(10, 10, 512)
output	$\left\{ \begin{array}{l} 1 \times 1 \text{conv} \\ \text{sigmoid} \end{array} \right\} \times 1$	(640, 640, 1)	$\left\{ \begin{array}{l} 1 \times 1 \text{conv} \\ \text{sigmoid} \end{array} \right\} \times 1$	(80, 80, 1)	$\left\{ \begin{array}{l} 1 \times 1 \text{conv} \\ \text{sigmoid} \end{array} \right\} \times 1$	(10, 10, 1)	$\left\{ \begin{array}{l} \text{GAP} \\ 1 \times 1 \text{conv} \\ \text{sigmoid} \end{array} \right\} \times 1$	(1, 1, 1)

Repetitive building blocks with asterisk superscript on the number of repetition start with convolution with 2×2 stride. Output depth in convolution operations is set equal to output depth of the block

Then, the discriminator D maps a pair of $\{x, y\}$ to binary classification $\{0, 1\}^N$ where 0 represents machine-generated y and 1 denotes human-annotated y and N is the number of decisions. Note that $N = 1$ for Image GAN and $N = W \times H$ for Pixel GAN with image size of $W \times H$.

Then, the objective function of GAN for the segmentation problem can be formulated as

$$L_{GAN}(G, D) = \mathbb{E}_{x, y \sim p_{data}(x, y)}[\log D(x, y)] + \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D(x, G(x)))]. \quad (2)$$

Note that G takes the input of an image, thus, analogous to conditional GAN [34], but there is no randomness involved in G . Then, the GAN framework solves the optimization problem of

$$G^* = \arg \min_G \left[\max_D \mathbb{E}_{x, y \sim p_{data}(x, y)}[\log D(x, y)] + \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D(x, G(x)))] \right]. \quad (3)$$

For training the discriminator D to make correct judgment, $D(x, y)$ needs to be maximized while $D(x, G(x))$ should be minimized. On the other hand, the generator should prevent the discriminator from making the correct judgment by producing outputs that are indiscernible to real data. Since the ultimate goal is to obtain realistic outputs from the generator, the objective function is defined as the minimax of the objective.

In fact, the segmentation task can also utilize gold standard images by adding a loss function that penalizes distance from the gold standard such as binary cross entropy

$$L_{SEG}(G) = \mathbb{E}_{x, y \sim p_{data}(x, y)} [-y \cdot \log G(x) - (1 - y) \cdot \log(1 - G(x))]. \quad (4)$$

By summing up both the GAN objective and the segmentation loss, we can formulate the objective function as

$$G^* = \arg \min_G \left[\max_D L_{GAN}(G, D) \right] + \lambda L_{SEG}(G) \quad (5)$$

where λ balances the two objective functions.

Appendix D: Experimental Details

D.1 Image Preprocessing

In the public datasets that are used in our experiments, the number of data is not sufficient for successful training. To overcome the limitation, we augmented the data to retrieve new images that are similar but slightly different to the originals. First, each image is rotated with interval of 3 degrees and flipped horizontally yielding additional 239 augmented images from one original image. We empirically found that 3 degree was sufficient for the

rotational augmentation. Finally, photographic information is perturbed by the following formula,

$$I'_{xyd} = (1 - \alpha)I_{xy}^{gray} + \alpha I_{xyd}, \quad I_{xy}^{gray} = 0.299I_{xyr} + 0.587I_{xyg} + 0.114I_{xyb} \quad (6)$$

where I_{xyd} is a pixel value in channel $d \in \{r, g, b\}$ at coordinate (x, y) in an image I and α is a coefficient that balances a color-deprived image (gray image) and the original image. In our experiment, α is randomly sampled from $[0.8, 1.2]$. When α is above 1, color in each channel intensifies while subtracting intensity in gray scale. From multiple experiments, we found that these augmentation methods can result in a reliable segmentation for both vessels and the optic discs.

When feeding fundoscopic input images to the network, we computed z-score by subtracting the mean and dividing by standard deviation per channel for each individual image. Formally, the following equation is applied to each pixel,

$$z_{xyd} = \frac{I_{xyd} - \mu_d}{\sigma_d}, \quad \mu_d = \frac{1}{WH} \sum_{x, y} I_{xyd} \quad \sigma_d = \frac{1}{WH} \sum_{x, y} (I_{xyd} - \mu_d)^2 \quad (7)$$

where I_{xyd} denotes a pixel value in channel $d \in \{r, g, b\}$ at (x, y) in an image $I \in \mathbb{R}^{W \times H}$. Since the mean and standard deviation in every image are set to 0 and 1 respectively, the sheer difference in brightness and contrast between the two images is eliminated. Therefore, any meaningful features will not be extracted based upon brightness or contrast during training.

D.2 Validation and Training

After the augmentation of original images, 5% of images are reserved for validation set and the remaining 95% of images are used for training. The models with the least generator loss on the validation set are chosen for comparison. We ran 10 rounds of training in which the discriminator and the generator are trained for 240 epochs alternatively. Total of 2400 epochs during training was enough to observe convergence in the generator loss. We emphasize that training both networks alternatively with sufficient iterations can lead to stable learning in which the discriminator perfectly classifies at first and the generator completely fakes the discriminator by letting the discriminator become gradually weaker.

D.3 Hyper-parameters

As an optimizer during training, we used Adam [51] with fixed learning rate of $2e-4$ and first moment coefficient (β_1) of 0.5, and second moment coefficient (β_2) of 0.999. We fixed the trade-off coefficient in Eq. 5 to 10 ($\lambda = 10$).

References

- Sarah BW, Mitchell P, Liew G, Wong TY, Phan K, Thiagalingam A, Joachim N, Burlutsky G, Gopinath B: A spectrum of retinal vasculature measures and coronary artery disease. *Atherosclerosis*, 2017
- Liew G, Benitez-Aguirre P, Craig ME, Jenkins AJ, Hodgson LAB, Kifley A, Mitchell P, Wong TY, Donaghue K: Progressive retinal vasodilation in patients with type 1 diabetes: A longitudinal study of retinal vascular geometry. *retinal vascular geometry in type 1 diabetes*. *Invest Ophthalmol Vis Sci* 58(5):2503–2509, 2017
- Taylor AM, MacGillivray TJ, Henderson RD, Ilzina L, Dhillon B, Starr JM, Deary IJ: Retinal vascular fractal dimension, childhood iq, and cognitive ability in old age: The lothian birth cohort study 1936. *Plos one* 10(3):e0121119, 2015
- McGrory S, Taylor AM, Kirin M, Corley J, Pattie A, Cox SR, Dhillon B, Wardlaw JM, Doubal FN, Starr JM, et al: Retinal microvascular network geometry and cognitive abilities in community-dwelling older people: The lothian birth cohort 1936 study. *Br J Ophthalmol*, pp bjophthalmol–2016, 2016
- McGrory S, Cameron JR, Pellegrini E, Warren C, Doubal FN, Deary IJ, Dhillon B, Wardlaw JM, Trucco E, MacGillivray TJ: The application of retinal fundus camera imaging in dementia: A systematic review. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 6:91–107, 2017
- Sumukadas D, McMurdo M, Pieretti I, Ballerini L, Price R, Wilson P, Doney A, Leese G, Trucco E: Association between retinal vasculature and muscle mass in older people. *Arch Gerontol Geriatr* 61(3):425–428, 2015
- Seidelmann SB, Claggett B, Bravo PE, Gupta A, Farhad H, Klein BE, Klein R, Di Carli MF, Solomon SD: Retinal vessel calibers in predicting long-term cardiovascular outcomes: the atherosclerosis risk in communities study. *Circulation*, pp CIRCULATIONAHA–116, 2016
- McGeechan K, Liew G, Macaskill P, Irwig L, Klein R, Sharrett AR, Klein BEK, Wang JJ, Chambless LE, Wong TY: Risk prediction of coronary heart disease based on retinal vascular caliber (from the atherosclerosis risk in communities [aric] study). *Am J Cardiol* 102(1):58–63, 2008
- Perez-Rovira A, MacGillivray T, Trucco E, Chin KS, Zutis K, Lupascu C, Tegolo D, Giachetti A, Wilson PJ, Doney A, et al: Vampire: vessel assessment and measurement platform for images of the retina. In: *Engineering in medicine and biology society, EMBC, 2011 annual international conference of the IEEE*. IEEE, 2011, pp 3391–3394
- Nguyen UTV, Bhuiyan A, Park LAF, Ramamohanarao K: An effective retinal blood vessel segmentation method using multi-scale line detection. *Pattern Recog* 46(3):703–715, 2013
- Jiang X, Mojon D: Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(1):131–137, 2003
- Vlachos M, Dermatas E: Multi-scale retinal vessel segmentation using line tracking. *Comput Med Imaging Graph* 34(3):213–227, 2010
- Martinez-Perez ME, Hughes AD, Thom SA, Bharath AA, Parker KH: Segmentation of blood vessels from red-free and fluorescein retinal images. *Med Image Anal* 11(1):47–61, 2007
- Staal J, Abràmoff MD, Niemeijer M, Viergever MA, Ginneken BV: Ridge-based vessel segmentation in color images of the retina. *IEEE Trans Med Imaging* 23(4):501–509, 2004
- Mendonca AM, Campilho A: Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Trans Med Imaging* 25(9):1200–1213, 2006
- Niemeijer M, Abràmoff MD, Ginneken BV: Fast detection of the optic disc and fovea in color fundus photographs. *Med Image Anal* 13(6):859–870, 2009
- Sinthanayothin C, Boyce JF, Cook HL, Williamson TH: Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *Br J Ophthalmol* 83(8):902–910, 1999
- Marín D, Aquino A, Gegúndez-Arias ME, Bravo JM: A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Trans Med Imaging* 30(1):146–158, 2011
- Soares JVB, Leandro JGG, Cesar RM, Jelinek HF, Cree MJ: Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *IEEE Trans Med Imaging* 25(9):1214–1222, 2006
- Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG, Barman SA: An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans Biomed Eng* 59(9):2538–2548, 2012
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A: Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp 1–9
- He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp 770–778
- Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*, 2014
- Girshick R: Fast r-cnn. In: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp 1440–1448
- Ren S, He K, Girshick R, Sun J: Faster r-cnn: Towards real-time object detection with region proposal networks. In: *Advances in neural information processing systems*, 2015, pp 91–99
- Noh H, Hong S, Han B: Learning deconvolution network for semantic segmentation. In: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp 1520–1528
- Ronneberger O, Fischer P, Brox T: U-net: Convolutional networks for biomedical image segmentation. In: *International conference on medical image computing and computer-assisted intervention*. Springer, 2015, pp 234–241
- Long J, Shelhamer E, Darrell T: Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp 3431–3440
- Maninis K-K, Pont-Tuset J, Arbeláez P, Gool LV: Deep retinal image understanding. In: *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp 140–148
- Fu H, Xu Y, Lin S, Wong DWKW, Liu J: Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. In: *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp 132–139
- Xie S, Tu Z: Holistically-nested edge detection. In: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp 1395–1403

32. Melinščak M, Prentašić P, Lončarić S: Retinal vessel segmentation using deep neural networks. In: VISAPP 2015 (10th international conference on computer vision theory and applications), 2015
33. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y: Generative adversarial nets. In: Advances in neural information processing systems, 2014, pp 2672–2680
34. Mirza M, Osindero S: Conditional generative adversarial nets. arXiv:1411.1784, 2014
35. Hoover AD, Kouznetsova V, Goldbaum M: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans Med Imaging* 19(3):203–210, 2000
36. Carmona EJ, Rincón M, García-feijoó J, Martínez-de-la Casa JM: Identification of the optic nerve head with genetic algorithms. *Artif Intell Med* 43(3):243–259, 2008
37. Fumero F, Alayón S, Sanchez JL, Sigut J, Gonzalez-Hernandez M: Rim-one: An open retinal image database for optic nerve evaluation. In: 2011 24th international symposium on Computer-based medical systems (CBMS). IEEE, 2011, pp 1–6
38. Sivaswamy J, Krishnadas S, Joshi G, Jain M, Tabish AUS: Drishti-gs: Retinal image dataset for optic nerve head(oh) segmentation, 04 2014
39. Radford A, Metz L, Chintala S: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434, 2015
40. Ganin Y, Victor L: N^4 -fields: Neural network nearest neighbor fields for image transforms. In: Asian conference on computer vision. Springer, 2014, pp 536–551
41. Becker C, Rigamonti R, Lepetit V, Fua P: Supervised feature learning for curvilinear structure segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, 2013, pp 526–533
42. Orlando JI, Blaschko M: Learning fully-connected crfs for blood vessel segmentation in retinal images. In: International conference on medical image computing and computer-assisted intervention. Springer, 2014, pp 634–641
43. Otsu N: A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 9(1):62–66, 1979
44. Orlando JI, Prokofyeva E, Blaschko MB: A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. *IEEE Trans Biomed Eng* 64(1):16–27, 2017. <https://doi.org/10.1109/TBME.2016.2535311>
45. Sakata K, Funatsu H, Harino S, Noma H, Hori S: Relationship of macular microcirculation and retinal thickness with visual acuity in diabetic macular edema. *Ophthalmology* 114(11):2061–2069, 2007
46. Prentašić P, Heisler M, Mammo Z, Lee S, Merkur A, Navajas E, Beg MF, Šarunić M, Lončarić S: Segmentation of the foveal microvasculature using deep learning networks. *J Biomed Op* 21(7):075008–075008, 2016
47. Zhang Z, Liu J, Cherian NS, Sun Y, Lim JH, Wong WK, Tan NM, Lu S, Li H, Wong TY: Convex hull based neuro-retinal optic cup ellipse optimization in glaucoma diagnosis. In: Annual international conference of the IEEE Engineering in medicine and biology society, 2009. EMBC 2009. IEEE, 2009, pp 1441–1444
48. Nair V, Hinton GE: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10), 2010, pp 807–814
49. Ioffe S, Szegedy C: Batch Accelerating deep network training by reducing internal covariate shift. In: 32nd international conference on international conference on machine learning, volume 37 of ICML'15, 2015, pp 448–456. JMLR.org
50. Isola P, Zhu J-Y, Zhou T, Efros AA: Image-to-image translation with conditional adversarial networks. In: The IEEE conference on computer vision and pattern recognition (CVPR), 2017
51. Kingma D, Ba J: Adam: A method for stochastic optimization. In: Proceedings of the International Conference on Learning Representations (ICLR), 2015