Retinel Vessel Segmentation using U-Net and GANs

Eric Megrabov Armin Jamshidi Sanika Patange

A12177906 A13624356 A53313444

University of California, San Diego

Abstract

Medical image segmentation is an extremely useful task in the realm of diagnosis of diseases, especially in the case of illnesses related to retinal vessels. In this project, our objective is to determine the best technique for retinal vessel segmentation to provide ophthalmologists with samples where they can detect vision-related diseases such as diabetic retinopathy, atherosclerosis and hypertension. Early detection of such conditions can lead to the prevention of more serious issues like blindness. By using fundoscopic images, the segmentation outputs were obtained from a U-Net model built by us and compared with the segmented retinal vessel maps produced by using an existing Generative Adversarial Network, which we also explore via hyperparameter tuning. Our methods achieved a 0.849 ROC, 0.799 PR, and a dice coefficient of 0.770 on the DRIVE dataset, and a 0.792 ROC, 0.759 PR and a dice coefficient of 0.701 on the STARE dataset. We compared our baseline U-Net with the state-of-the-art Image-GAN model.

1. INTRODUCTION

Analysis of retinal vessels is the study of delicate arteries and veins within the eye. Some morphological features like diameter, branching angle, length, and tortuosity can be used for the diagnosis, treatment and evaluation of various ophthalmologic and cardiovascular diseases. Some of the diseases include diabetes, hypertension, arteriosclerosis, retinopathy of prematurity and retinal vessel occlusion [12]. If these diseases are not diagnosed in time, it can eventually lead to issues such as blindness. Hence, the segmentation of retina vessels becomes very crucial since it facilitates early detection and diagnosis of such diseases.

In order for a segmentation technique to be considered acceptable, it needs to be able to detect and create segmentations for variable vessel sizes as well as preserve any abnormalities seen in the original fundoscopic image. Manual segmentation is tedious, time-intensive, and error prone, especially during the analysis of large and complex datasets. There are many scenarios in which two doctors label the same image differently, which also introduces inconsistencies depending on which medical professional was chosen. On the other hand, automated segmentation is much faster, quantitatively more accurate, and saves the doctors a lot of time by allowing the patients to be treated faster and more efficiently. It also creates a more universally accepted strategy of labeling of vessels, thus creating consistency. Therefore, our paper focuses on these computer-assisted diagnostic techniques that help in the early detection and timely treatment of these diseases.

Segmentation plays a crucial role in medical imaging because it is the initial step in clinical diagnosis. In this project, we focus on the various approaches involved in retinal vessel segmentation and compare the obtained results. The inputs given to the compared network models, U-Net and Image-GAN, are the fundoscopic images, which are photographs of the rear of an eye, also known as the fundus. These images are obtained from the DRIVE and STARE datasets.

In our project, we implemented the U-Net architecture for retinal vessel segmentation and compared it with the existing state-of-the-art method Image-GAN from [1]. The main advantage of Image-GAN is that it yields crisper results through the use of its full-image objective function, while U-Net results are blurrier. Both models have been explained in the sections that follow.

2. RELATED WORK

There are many previous literatures that emphasize the extraction blood vessels from retina images. They can be classified as unsupervised and supervised methods.

2.1 Unsupervised methods

The unsupervised methods mainly deal with enhancement of vessels combined with segmentation. Popular techniques are: model-based locally adaptive thresholding, Frangi's filter methods, matched filter responses and vessel tracking as mentioned in [5].

In [10], the authors implement an approach that makes use of the Gabor wavelet and Multiscale Line Detector for segmenting blood vessels. Also, an algorithm based on Entropy filtering and curvature evaluation is used in [11] that calculates segmentations in relation to central reflex vessels.

2.2 Supervised models

The supervised methods for retinal vessel segmentation use datasets to train the classifier to categorize the pixel into a vessel or non-vessel. These methods can be support vector machine-based methods or neural network-based methods.

Jianqing Gao et al [5] uses a random walk algorithm that uses centerlines to segment the vessels. The random walk algorithms are robust to weak object boundaries and this method is compared to other existing retinal vessel segmentation methods.

Xiao et al. [9] ran a Bayesian classifier on the retina image and used it to extract vessels. In addition, they also differentiated the vessels from the background using an energy function. Yuan Xue et al in [3] designed an end-to-end adversarial network called SegAN that had a new multiscale loss for semantic segmentation which also proved that SegAN outperformed U-Net.

3. DATASETS

We have used two datasets: DRIVE and STARE. Both datasets contain segmented labels of retina blood vessels. The images are RGB images, but for our UNET, we feed them in as grayscale as it seems to improve runtime performance.

3.1 DRIVE - digital retinal images for vessel extraction

The dataset [6] were captured with digital retinal cameras from diabetic subjects, containing 20 training and 20 testing images. Each image was originally 768 by 584 pixels. It also contains a single manual segmentation for training images and two manual segmentations for each test case. One is treated as the gold standard while the other can be used to determine model output performance to. In addition, a mask image has also been provided to indicate the region of interest.

3.2 STARE - structured analysis of the retina

This dataset [8] is similar to the Drive dataset with a slightly different imaging system, with 10 train/test images. The project from which the images came was initiated at University of California, San Diego. The ophthalmologist-annotated data is originally 700 by 605. Once again, this was resized for the U-Net for the same purposes as the DRIVE network.

3.3 FEATURES

For the U-Net, the encoder helps in extracting the coarse features with a low resolution whereas the decoder(due to the skip connections) helps to get the fine details in the image and brings the image back to the original high resolution. The combination of coarse and fine features help in semantic segmentation. As far as the GAN is concerned, it learns the underlying structure of the given data. Features are extracted via convolutional layers and feature maps and are weighed in terms of importance indirectly via the loss function. With the generator and discriminator, they use an adversarial approach to search for the features corresponding to a specific image as they attempt to improve each other.

4. METHODS

4.1 U-Net

U-Net is designed to perform semantic segmentation, a classification of every pixel in an image. Our implementation of U-net's architecture can be broken down into an encoder and a decoder network. The encoder reduces the image's dimensionality by convolution and max pooling and the decoder uses the filter maps generated to then increase the image dimensionality back to the original by upsampling. The unique element of this architecture is the skip networks between the encoder and decoder. A skip connection allows information to jump from layers of the encoder to layers in the decoder, skipping over several layers of convolution and pooling layers. The spatial information that was lost in some of those layers can be used in layers of the decoder which allows for better upsampling of the images. Spatial information plays a significant role in biomedical imaging; therefore U-net outperforms other encoder-decoder networks.

Loss function - The loss function we used for out U-net Model is Binary cross-entropy, a very stable loss function for problems with only two classes (vessel or not a vessel).

Binary cross-entropy is defined as –

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i * \log(p(y_i)) + (1 - y_i) * \log(1 - p(y_i))$$

Using our dataset as an example, a retina vessel would indicate a 1 in label y_i and anything else would be a 0. These values are multiplied by their respective log probabilities and summed up. This value is then normalized by dividing by the number of points N. The negative sign is because the values of log are native for numbers between 0 and 1. This loss functions works very well for our dataset and stabilizes after about 1000 iterations.

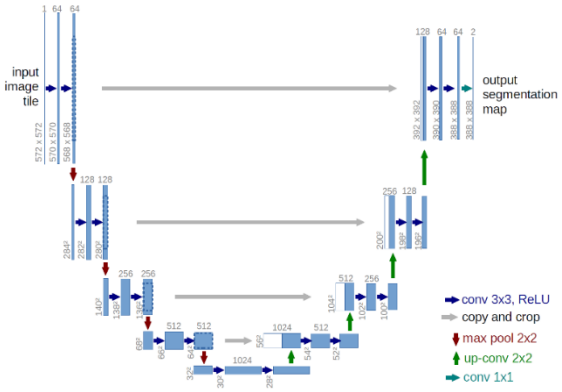


Figure 2: U-Net Architecture

4.2 GAN – Generative Adversarial Network

In our model, we used the Image GAN from [1] to produce the segmented vessel images of the retina. In this case, the generator, whose architecture is designed to perform segmentation, generates the map images of the segmented vessels from the input fundoscopic images. The discriminator distinguishes images produced by the generator from the true data distribution. Our discriminator compares the generated image with the ground truth images which are human-annotated vessel maps performed by medical professionals. This process continues until the generator produces a segmented map that is close to the ground truth and the discriminator is not able to distinguish between real and generated images. The architecture of the generator has some similarities to U-Net because of the presence

of skip connections. The main difference between Image-GAN and other types of GANs is that it uses the comparison of the entire output image during the calculation of the loss rather than just using pixels or patches of pixels.

Objective Function - In the loss equation below, the generator is represented as G and the discriminator is represented as D. x is a fundoscopic image which maps to a vessel segmentation map called y. D(x,y) creates a set of binary classifications per pixel in which an output of 1 indicates a match between the generated map and the ground truth and an output of 0 indicates a mismatch. This output for the loss is the size of the images being compared for the case of Image-GAN.
The loss function for the GAN is as seen below.

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

The expected value of log(D(x,y)) indicates the log probability that x is real while the second term (1-D(x,G(x))) indicates the log probability that the image generated by the generator is fake. The optimal G is shown below.

$$G^* = \arg\min_G \left[\max_D \mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(x,y)] + \mathbb{E}_{x\sim p_{data}(x)}[\log(1 - D(x,G(x)))]\right]. \quad (2)$$

In order to improve the classification, we create a penalty for the measured distance between the output of the generator and the ground truth using binary cross entropy.

$$L_{SEG}(G) = \mathbb{E}_{x,y\sim p_{data}(x,y)} - y \cdot \log G(x) - (1 - y) \cdot \log(1 - G(x)). \quad (3)$$

Combining the optimal G* with the segmentation loss, we obtain:

$$G^* = \arg\min_G \left[\max_D L_{GAN}(G,D)\right] + \lambda L_{SEG}(G) \quad (4)$$

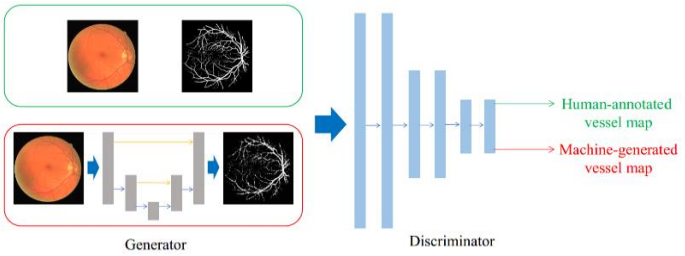


Figure 3: GAN for Retinal Vessel Segmentation

5. EXPERIMENTAL RESULTS

We compared the performance of our models on both the datasets and observed that our baseline model U-Net shows comparable but inferior performance to Image-GAN indicating that the GAN framework improves the quality of segmentation. Both models were trained using the adam optimizer. U-Net was run for 2000 iterations while the GAN was run for 50,000 iterations.

5.1 U-NET Results

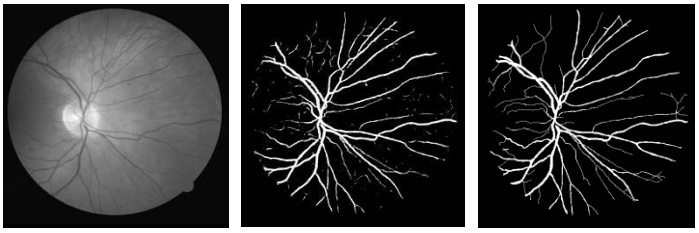


Figure 4**: The grayscale fundoscopic image (left), predicted vessel map (middle), and ground truth (right) (DRIVE**)

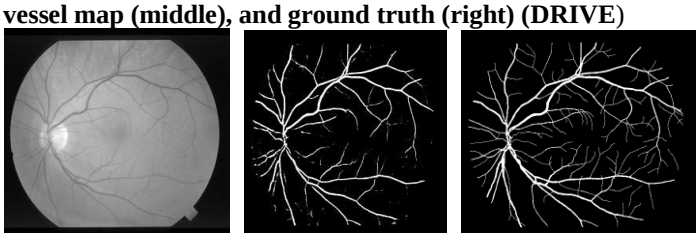


Figure 5: **The grayscale fundoscopic image (left), predicted vessel map (middle), and ground truth (right) (STARE**)

5.2 Image-GAN Results

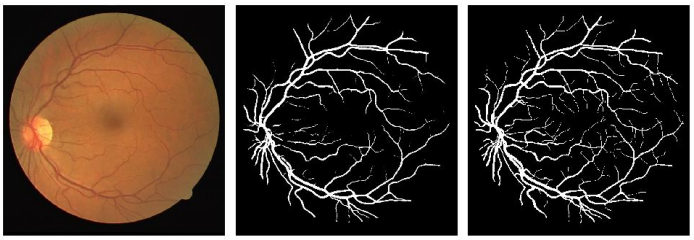


Figure 6: **The input fundoscopic image, the predicted vessel map and the corresponding ground truth image (DRIVE)**

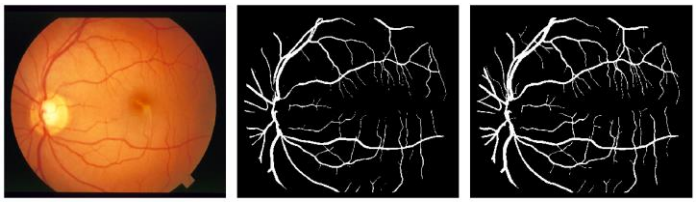


Figure 7: **The input fundoscopic image, the predicted vessel map and the corresponding ground truth image (STARE)**

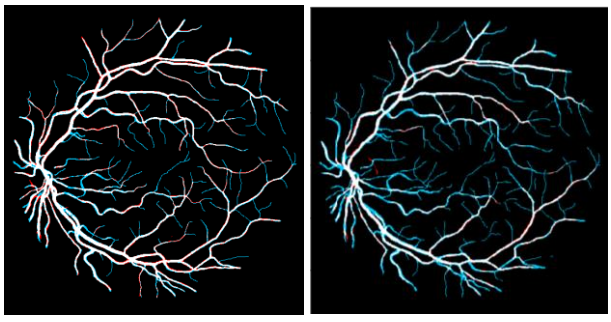
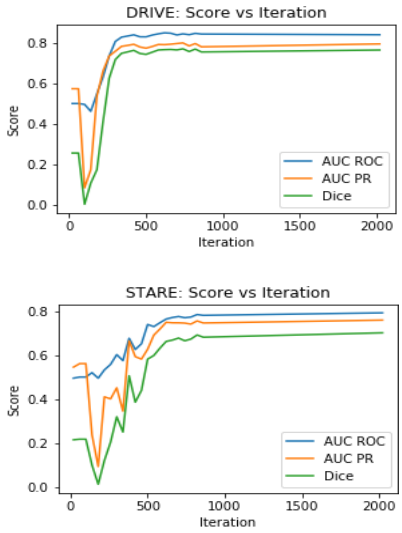


Figure 8: **DRIVE predictions overlaid with ground truth. White: true-positive, red: false-positive, blue: false negative. Left is Image-GAN prediction, right is U-Net prediction.**

Examining the results of the U-Net, we see that it does in fact yield blurrier photos; the retinal vessels are not as bright or sharp as in the ground truth nor as in the Image-GAN. In addition, both the GAN and U-Net have trouble replicating delicate vessels, but the

U-Net is far worse; in figure 8, the difference image shows how the U-Net does not detect many vessels that the GAN does, as seen by the blue lines. However, both can replicate the original ground truth image with high accuracy despite losing some detail. Overall, the GAN is preferable since it yields sharper results. Furthermore, we also evaluated our models with Area Under Curve for Receiver Operating Characteristic (ROC AUC), Area Under Curve for Precision and Recall Curve (PR AUC) and dice coefficients.

5.3 Curves for U-Net



5.4 Curves for Image -GAN

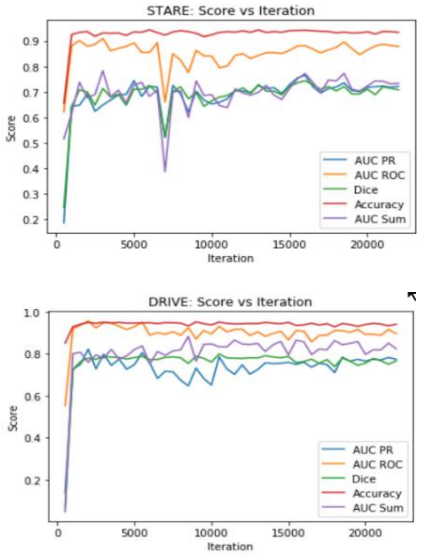


Table 1: **Comparison of models in ROC, PR and Dice (DRIVE)**

		DRIVE	
Model	ROC	PR	DICE
U-Net	0.849646	0.799663	0.770534
Image-GAN	**0.956901**	**0.822135**	**0.799085**

Table 2: **Comparison of models in ROC, PR and Dice (STARE)**

		STARE	
Model	ROC	PR	DICE
U-Net	0.792442	0.759346	0.701426
Image-GAN	**0.909608**	**0.765179**	**0.743937**

By qualitative inspection, we see that the GAN is able to better predict small vessels within the fundoscopic image. In addition, the predictions made by the U-Net are darker in color and less prominent. Quantitatively, as seen in Tables 1 and 2, the Image-GAN outperforms the U-Net in all three scores: ROC, PR, and Dice. Therefore, there is great value in transitioning from a classical CNN model to a GAN for the purpose of retinal image segmentation.

6. EXTENSION: Transfer Learning

After we completed our project, we decided to see if either model would generalize to a dataset in which we are provided with aerial images of Massachusetts Roads. This dataset consists of a training set with 1108 images, which are satellite views of a scene [13]. We decided to run the scenes through our segmentation models in order to extract main roads.

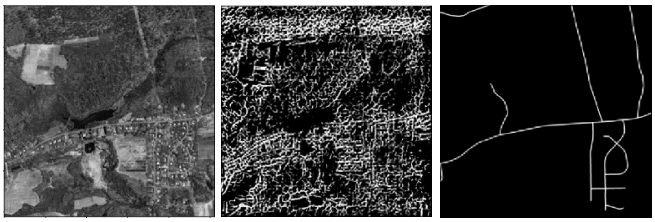


Figure 9: **The input aerial image (left), prediction (middle), ground truth (right)**

This likely failed because retinal vessel segmentation is trained to pick up very small and delicate vessels, so seeing an image with too much detail causes the model to overclassify too much of the image as vessels (i.e. roads). Much of the field area looks like noise, which becomes mistaken as vessels. Clearly, this did not transfer well, so our models are not capable of being used on non-medical imaging.

7. CONCLUSION AND FUTURE WORK

We performed retinal vessel segmentation using U-Net and compared it with the existing state-of-the-art Image GAN. U-Net yields somewhat blurry results, and both struggle to detect very thin vessels but maintain excellent overall performance. The segmented vessels maps obtained from U-Net and generated by the GAN suggest that the presence of discriminator can help segment vessels more accurately and clearly. The ROC, PR and Dice results also indicate that the Image-GAN outperformed our baseline model U-Net.

For future work, we want to find better optimizers and hyperparameters to generate more detailed and realistic segmentation maps than the state-of-the-art. We would also like to find a way to generalize to other medical datasets.

8. CONTRIBUTIONS

In this project, all three of us jointly worked on U-Net Model architecture. Armin did the pre-processing and experimental testing for U-Net. Eric and Sanika tested the four GAN models from [1] - Pixel GAN, Image GAN, Patch 1 GAN and Patch 2 GAN and finally chose Image GAN and tuned its model parameters to give better results. Then, three of us worked on the Presentation and Project Report in collaboration.

9. REFERENCES

[1] Jaemin Son, Sang Jun Park and Kyu-Hwan Jung, “Retinal Vessel Segmentation in Fundoscopic Images with Generative Adversarial Networks”, In: Computer Vision and Pattern Recognition, Jun 2017.

[2] Olaf Ronneberger, Philipp Fischer and Thomas Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, In: Computer Vision and Pattern Recognition, May 2015.

[3] Yuan Xue,Tao Xu, Han Zhang,Rodney Long,Xiaolei Huang, “SegAN: Adversarial Network with Multiscale L1 Loss for Medical Image Segmentation”, In: Computer Vision and Pattern Recognition, July 2017.

[4] Yun Jiang, Hai Zhang, Ning Tan, Li Chen, “Automatic Retinal Blood Vessel Segmentation Based on Fully Convolutional Nueral Networks”, Symmetry Journal 2019.

[5] Jianqing Gao, Guannan Chen, Wenru Lin, “An Effective Retinal Blood vessel Segmentation by Using Automatic Random Walks Bsed On Centerline Extraction”, Biomed Research International, 2020.

[6] J. J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, *Digital Retinal Image for Vessel Extraction (DRIVE) Database*, Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands, 2004.

[7] A. Hoover, V. Kouznetsova and M. Goldbaum, "Locating Blood Vessels in Retinal Images by Piece-wise Threhsold Probing of a Matched Filter Response", *IEEE Transactions on Medical Imaging*, vol. 19 no. 3, pp. 203-210, March 2000. *STARE: Structured Analysis of the Retina.*

[8] D. Siva Sundhara Raja, S. Vasuki, “Automatic Detection of Blood Vessels in Retina Images for Diabetic Retinopathy Diagnosis”, Computational and Mathematical methods in Medicine, Vol 2015.

[9] Z. Xiao, M. Adel, and S. Bourennane, “Bayesian method with spatial constraint for retinal vessel segmentation,” *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 401413, 9 pages, 2013.

[10] Syed Ayaz Ali Shah,Aamir Shahzad, Muhammad Amir Khan, Cheng-Kai Lu, “Unsupervised Method for Retinal Vessel Segmentation Based on Gabor Wavelet and Multiscale Line Detector”, IEEE Volume 7.

[11] Xiao-Xia Yin, Brian W, Jing He,Yanchun Zhang, Derek Abbott, “Unsupervised Segmentation of Blood Vesssels from Colour Retinal Fundus Images, In: International Conference on Health Information Science, 2014.

[12] Joshi, Vinayak Shivkumar. "Analysis of retinal vessel networks using quantitative descriptors of vascular morphology." PhD (Doctor of Philosophy) thesis, University of Iowa, 2012.

[13] Mnih, Volodymr. “Machine Learning for Aerial Image Labeling.” PhD (Doctor of Philosophy) thesis, University of Toronto, 2013.