EE4476/IM4476

Image Processing

Project – Vessel Segmentation Report

Submitted by: Fan Xiaofeng
Matriculation Number: U1520572L

School of Electrical and Electronic Engineering

2018

Table of Contents

Chapter 1 Introduction

1.1 Background

Retinal vessel segmentation is an important part of computer-aided diagnosis of retinal diseases, like arteriosclerosis, vein occlusions, and diabetic retinopathy. A reliable assessment for these diseases can be achieved by regularly performing accurate measurement of the vessel width, tortuosity and proliferation. If abnormal signs are detected at early stage, timely treatment can be advised to perform on patients. Vessel segmentation based on computer vision and image processing provides an efficient and economic benefit tool for retinal image analysis.

Due to its importance, large amount of research work has been done in the past decades to develop effective automatic method of automatic vessel segmentation/ detection from retinal image. Earlier efforts focused on applying various image processing techniques while the state-of-the-arts adopt machine learning approach [1].

1.2 Related Work

1.2.1 Convolutional Neural Network

With recent advancement of computing power, machine learning methods especially deep learning have been dominating many applications. Convolutional Neural Network (CNN) is primarily employed for imaging tasks such as object detection and segmentation. The convolution operations with backpropagations are found to be able to automatically extract image features. The U-Net [2] employed various training techniques such as patch convolution, skipping connections to map an image tile to the desired segmentation map using very few training images.

1.2.2 Generative Adversarial Network

In the literature of unsupervised representation learning, Generative Adversarial Network (GAN) shines in recent years. It has been shown incredible performance on generating realistic images. The idea is to set up two convolutional neural networks, one known as generator G and other known as discriminator D. G starts with a vector of random value and passes the vector through a set of deconvolution operation to form an image. Then the image generated from G together with a real image were fed into D. The output of D is a binary digit to indicate whether the input image was from G or real dataset. The two networks are trained simultaneously (the gradient from D is also propagated back to G) to minimize loss for G while maximize accuracy for D. After reaching the equilibrium, the G is able to generate realistic images that are difficult for D to tell if they are fake or real. Inspired by this sprit, many successors of GAN were introduced to image-to-image translation such as CycleGAN [3]. Inspired by the spirit of U-Net [2] and GANs, [4] addressed the Vessel Segmentation task as an image translation process with the use of GANs.

1.3 Objectives

This project aims to apply various image processing techniques together with the state-of-the-art machine learning approach, specifically Generative Adversarial Network, to perform automatic vessel segmentation from retinal images. This project implements the model proposed by [5] with some slight modifications. The dataset being used by this project is the DRIVE [4] which consists of total 40 retinal images with manual segmentation from domain experts.

Chapter 2 Method

2.1 Preprocessing

2.1.1 Centre Cropping

To fully utilize the power of CNN and GAN, the spatial resolution of input image becomes an important factor to consider due to the processes of continuous down-sampling and up-sampling. It was found that to control the input dimension to be power of 2 would help avoid unnecessary paddings which affected the segmentation performance.

As a result, the DRIVE images originally with resolution of 565×584 were cropped from the center to 512×512. The groundtruth labels were also cropped to 512×512 in order to compute the segmentation accuracy. This cropping also made the measurements of accuracy slightly fair since we narrowed down the region of interest to mainly cover the retinal components.

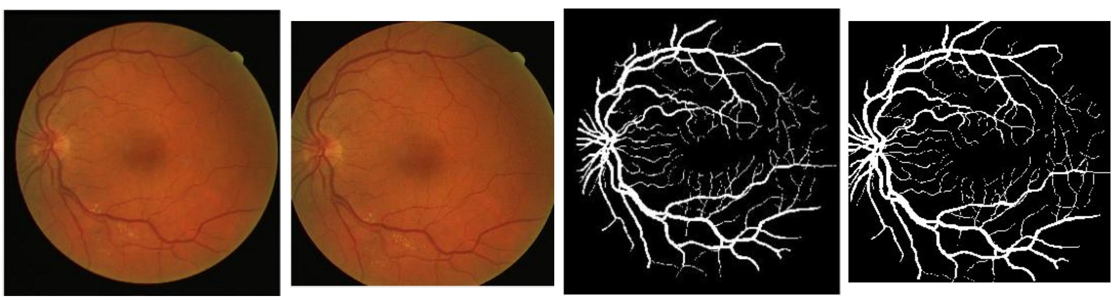

Fig1. Image Centre Cropping. From left to right: retinal image (565×584), cropped retinal image (512×512), groundtruth label (565×584), cropped label (512×512).

2.1.2 Data Augmentation

To improve the performance of any deep learning model, it is commonly agreed that to feed the model with more training data is beneficial. Data augmentation techniques including horizontal flipping and vertical flipping were introduced to increase the training data size.

2.1.3 Image Random Enhancing

Randomness was introduced to the training images to make the model more generalized to input variance. It also helps in augmenting the training data size. This process was achieved by adding random values to the image's brightness, contrast, sharpness.

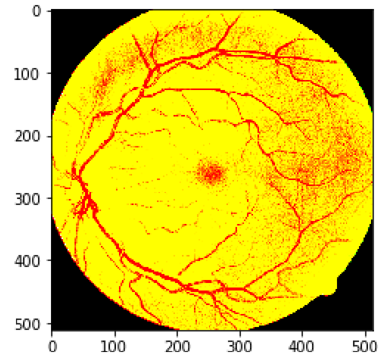

Fig2 Image After Random Enhancing

2.1.4 Data Normalisation

It is commonly agreed that to normalise the input data would be helpful for deep learning training. Each colour channel of the images was normalized to z-score. The same normalisation should be applied to test images before feeding into the model for inference.

2.2 GAN Training

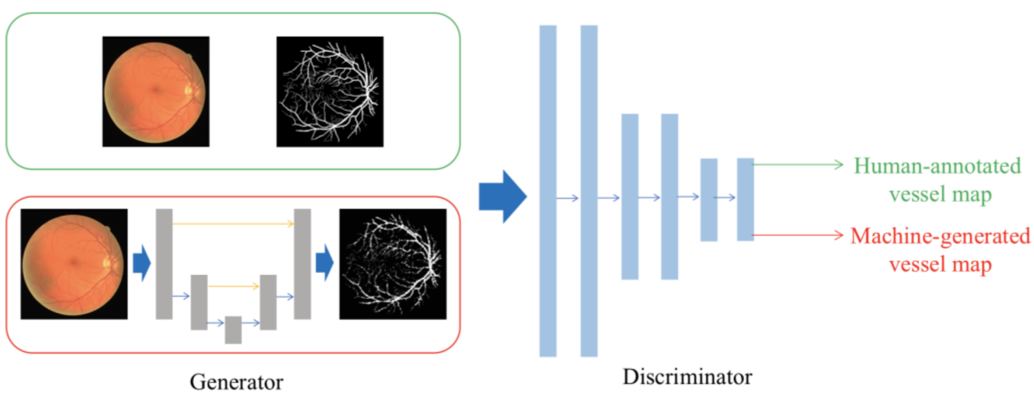

Fig3. The framework proposed by [5]

As discussed in 1.3, this project implements the generative adversarial network proposed by [5]. The generator G performs the mapping from the retinal image x to a vessel map y using U-Net [3]. The discriminator takes a pair of $\{x, y\}$ and outputs a binary digit 0 or 1 to classify whether the vessel map y is from G or the groundtruth label. The loss function used for D is binary crossentropy. The overall objective function described by the paper

$$G^{*} = argmin_{G}[max_{D}\mathrm{E}_{x,y \sim p_{data}(x,y)}[\log D(x,y)] + \mathrm{E}_{x \sim p_{data}(x)}[\log(1 - D(x, G(x)))] \quad [5]$$

2.3 Postprocessing

2.3.1 Thresholding

As discussed in section 2.2, after training of the GAN reached certain level of equilibrium, the Generator can be used as model for segmentation. The output of Generator is an array of float number ranged in (0,1) indicating the probability being a vessel pixel. The output array was multiplied by 255 to obtain a grayscale image. Various thresholds were then tested to filter out the pixels with low confidence. For the image assigned to this project, it was found that a threshold of 90 (roughly 35% confidential level for the trained model) can lead to the highest accuracy. Any pixel with intensity lower than 85 was set to zero and others were set to 255. Various morphological operations were tested on the output obtained from 2.3.1. Unfortunately, those operation did not improve the accuracy.

2.3.2 Filtering – Morphological Closing

Various morphological operations were tested on the output obtained from 2.3.1. The reason was the image became sparse due to thresholding. Some pixels became 0 although human cannot notice. However, this sparsity may affect the performance reported by accuracy since the accuracy was measured at pixel-to-pixel level. Thus, the output after thresholding was then morphologically closed using a structure element of square of 3×3.

Chapter 3 Results and Discussion

3.1 Experimental Results

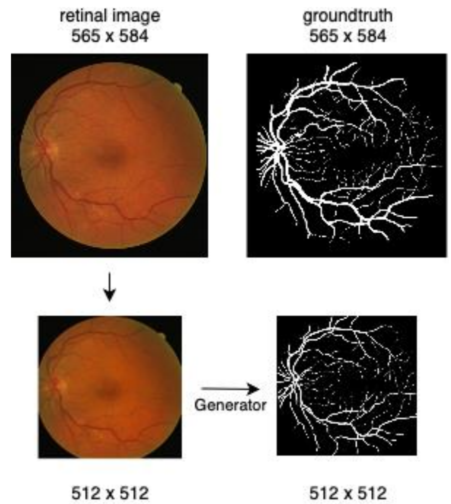

Fig4 Segmentation Results

The DRIVE dataset contains 40 pairs of retinal images and groundtruth labels. The first 20 images are for testing purpose and the last 20 images are training. The training of this project followed this data split and the image assigned to this project is 03_test.tif which falls into the test split. Thus, the model was prevented from seeing the test image during training. There won't be overfitting issue as long as the segmentation performs well on the test images.

The segmentation map of 03_test.tif is shown in Fig4. After performing thresholding, the pixels are converted to 255 in order to compute accuracy with the groundtruth label from human experts. As discussed in 2.1.1, the groundtruth was also cropped from the centre to 512×512. By setting a threshold of 85 (33% confidential level), the segmentation accuracy achieved is 92.7322%. After morphological closing using a structure element of square of 3×3, the accuracy changed to 92.8322%.

3.2 Discussion

The model was trained on one Nvidia GTX 1080 for 1.5 hours. As demonstrated in Fig5, the generator started with segmenting the eyeball edge. After iterating roughly 100 epochs,

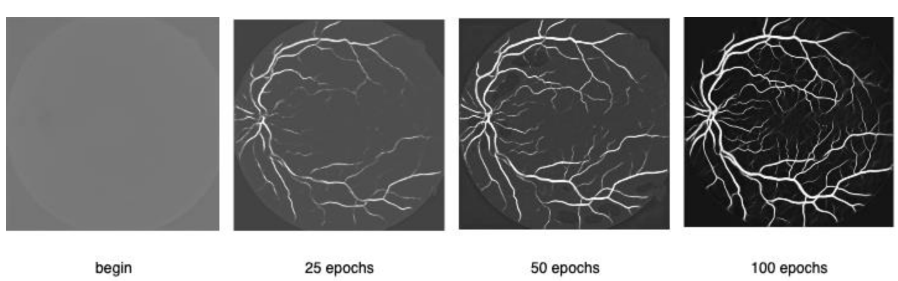

Fig5 Generator outputs during training

the generator began to perform quite precise segmentation which made it difficult for discriminator to tell.

From human's point of view, the output from the generator is quite "realistic" with respect to the groundtruth used for training thanks to the design of generative adversarial network. Since the generator is well trained to generate vessel mappings which are very similar to what domain experts would do, it is now safe to use the generator as the model for automatic segmentation. Fig6 illustrates the segmentation error according to the algorithm given by the lecturer. The majority of the error is from the thresholding process. Because of the generator's raw output is ranged in [0,1], the thresholding filtered out some correct pixels. Thus, a morphological closing was adapted as discussed in 2.3.2 with the hope to fill in some of those holes. Unfortunately, the accuracy did not improve significantly. The possible reason is the choice of structure element may affect the closing results. This project stopped from further exploring with the consideration of time.

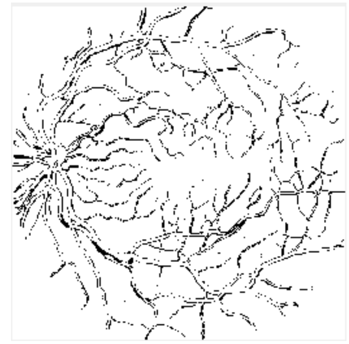

Fig6 Segmentation Error Map after morphological operation. (Output from MatLab code given by the lecturer)

Chapter 4 Summary

In this project, we further enhanced the understanding of various image processing techniques and applied them to the automatic vessel segmentation task and achieved 92.8322% accuracy. Furthermore, the state-of-the-art approach of machine learning for computer vision was also integrated. The implementation of generative adversarial network also opened the door for mainly interesting applications where image-to-image mapping is desired. After this project, we are now equipped with the knowledge and skills to solve more challenging computer vision problems by combining traditional image processing techniques with the state-of-the-art machine learning approaches.

Reference

[1] EE4476 Project Notes.

[2] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015.

[3] J.-Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," in *ICCV*, 2017.

[4] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 501-509, 2004.

[5] J. Son, S. J. Park and K.-H. Jung, "Retinal Vessel Segmentation in Fundoscopic Images with Generative Adversarial Networks," in *arXiv preprint arXiv:1706.09318*, 2017.