Retinal Vessel Segmentation of Dehazed Fundus Images

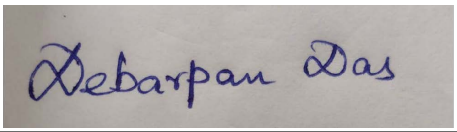
CMPUT 610 : Project Write-up

Debarpan Das

Declaration of Originality

I, Debarpan Das, declare that this project titled, Retinal Vessel Segmentation of Dehazed Fundus Images and the work presented in it are my own. I confirm that:

- This work was fully done by me as my course project for CMPUT 610 course at University of Alberta.
- Wherever I have consulted any published work or code written by others has always clearly acknowledged them.
- I have acknowledged all main sources of information.

Signed: 

Date:

Abstract

Retinal vessel analysis is very fundamental and crucial for the diagnosing various kinds of disorders. While using modern cameras for retinal image acquisition, there often occurs some reflections from the retinal surface that causes luminosity and contrast variations. These irregularities in illumination might introduce distortions in the acquired image thereby distorting the anatomical structures. So, the illumination correction proposed by Savelli et. al [1] which draws a connection between shadows and haze is used to reduce these illumination defects in this project. Since the performance of this method has only been tested on older retinal vessel segmentation algorithms, we try to use the results obtained from this method and apply it to train modern vessel segmentation networks and compare their performances.

Goal and motivation

Retinal images are used extensively in diagnosing and treating disorders related to heart as well as opthalmologic diseases [2]. A characteristic symptom of these disorders is that they affect the anatomical features of blood vessels. That's why the vessel structure of these retinal images are very useful for detecting and estimating the severity of such disorders [3]. However manually delineating these vessel structures is a very expensive task both in terms of time as well as skill. So, in recent years, a number of methods for retinal vessel segmentation have been introduced by various researchers which include both supervised as well as unsupervised methods.

Most of these methods are highly sensitive to the data-set with which they are trained and any illumination errors in the training images might have degrading effects in their performance. During image acquisition, non-uniformity in the image illumination is often observed. These arise from inter reflections and shading artifacts due to the spherical geometry of the eye.

This issue was addressed by Savelli et. al [1] where they proposed that removing haze from an image in it's inverted intensity domain is equivalent to correcting the non uniform illumination in the original intensity domain.

This method is meant to be a pre-processing step for retinal vessel segmentation.The authors had evaluated the proposed method by using two segmentation algorithms, one mathematical morphology based unsupervised method and the other CNN based supervised method inspired from the work of Liskowski and Krawiec [4]. However both these methods are quite old and many new methods have been published in the recent years that outperform these methods. We have used three commonly used frameworks namely U-Net, DU-Net and

LadderNet for segmenting the retinal vessel structure and have compared their results with and without using the illumination correction.

Inputs and specifications

The dataset used for this project is the DRIVE public database [5] of retinal images. The images for this dataset were acquired from a series of diabetic retinopathy test conducted in the Netherlands. Each image is of 768x584 pixels in size and consists of 8 bits per color plane. The dataset consists of 40 images : 20 training and 20 testing. In this project we have divided the training set into equal halves of 10 samples each and used one half for training and the other one for evaluating the performance.

Assumptions and constraints

The illumination correction method is based on the assumption that shadows, which causes the overall image to tend towards darker intensities are the inverse phenomenon of haze, which causes the overall image to tend towards the purest white color. Although dehazing is an operation which is supposed to be applicable only to natural images (since it is fundamentally based on atmospheric illumination), an equivalence between dehazing of the fundus images in their inverted intensity domain and correcting the non-uniformity in illumination in the original fundus images has been proposed, which makes it possible to theoretically apply dehazing to medical images. The proposed theory can also be understood by the observation that haze and shadows, both being low-frequency phenomena, presence of haze makes white as the dominant color and

the presence of shadows makes the darker intensities in the images dominant.

Algorithm expression and methodologies

Pre-processing the fundus images

Dehazing

The following algorithm was proposed in [1]: If *I(x)* is the captured image (graylevel/ RGB) corresponding to pixel *x*, *i(x)* is the illumination conditions of the scene and *r(x)* is the true reflectance of the object, we can write the illumination-reflectance model of image formation as :

$$I(x) = i(x)r(x) \tag{1}$$

Assuming the the image are normalized, we can rewrite equation 1 in the inverted intensity domain as

$$1 - I(x) = 1 - i(x)r(x) \tag{2}$$

which can be rearranged as

$$1 - I(x) = (1 - r(x))i(x) + 1 - i(x) \tag{3}$$

By denoting the captured intensity as $\tilde{I}(x) = 1 - I(x)$ and true reflectance as $\tilde{r}(x) = 1 - r(x)$ in the inverted intensity domain, equation 3 can be written as :

$$\tilde{I}(x) = i(x)\tilde{r}(x) + 1 - i(x) \tag{4}$$

which closely resembles the atmospheric illumination model given by

$$I(x) = t(x)R(x) + A - At(x) \tag{5}$$

where *R(x)* : radiance in the haze-free scene, *t(x)* : transmission of light in the atmosphere and A is the most prominent color of the atmosphere (purest white). So by assuming $A \approx (1, 1, 1)$ [21] the haze model can be written as

$$I(x) = t(x)R(x) + 1 - t(x) \tag{6}$$

which is like the dual form of equation 4 which is obtained by exchanging $\tilde{r}(x)$ with *R(x)* (reflectance with radiance) and *i(x)* with *t(x)* (illumination with transmission). This dualism can also be understood by observing that haze and shadows, both being low-frequency phenomena, presence of haze makes white as the dominant color and the presence of shadows makes the darker intensities in the images dominant. The value of *t(x)* is estimated by the Dark Channel Prior [22] method implemented by [6] as follows :

$$t(x) \approx 1 - \omega \frac{\tilde{I}_{dark}(x)}{A} \tag{7}$$

Here ω is a constant whose value is taken as 0.9 and $\tilde{I}_{dark}(x)$ is the dark channel of the inverted image in a neighbourhood $\Omega(x)$ as follows :

$$\tilde{I}_{dark}(x) = \min_{c \in \{R,G,B\}} \left(\min_{z \in \Omega(x)} \tilde{I}(z) \right) \tag{8}$$

where the neighbourhood $\Omega(x)$ is of side $d = 20$ pixels.

So, we can summarize the entire algorithm as : given an image affected by

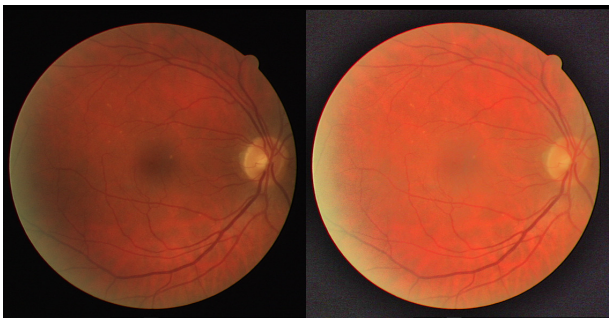


Figure 1: Original vs illumination corrected versions of sample image 30 from the training set of DRIVE dataset

shadows/illumination problems, it's intensities are inverted. In the inverted intensity domain, haze removal is applied to the image for which the transmission $t(\tilde{x})$ is estimated by the Dark Channel Prior method. The intensities are again inverted to correct the non-uniform illumination problem.

As we can see in Figure 1, the original image from the dataset is highly corrupted by non-uniform illumination, whereas the illumination corrected image is much better in terms of clarity and perception and it is much easier the blood vessels from it.

Other pre-processing methods

Once we have two datasets of dehazed images and non-dehazed original images, we firstly convert them from their RGB scale to grayscale and thereby subject them to intensity normalization over their respective datasets. This generalizes the datasets to much an extent. This is followed by the application of Contrast Limited Adaptive Histogram Equalization (CLAHE) which is a modification of the Adaptive Histogram Equalization method. Lastly we apply the Gamma Adjustment which corrects the non-linear luminance effects of the image.

After the pre-processing stage is applied, patches of size 48x48x3 are cropped

from both the dehazed and non-dehazed set of fundus images along with their corresponding masks. These patches are used to extensively train the frameworks of U-Net, DU-Net and LadderNet respectively.

U-Net

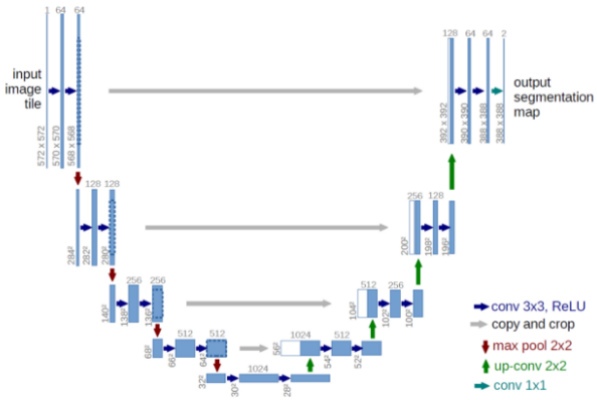


Figure 2: U-Net architecture, the blue boxes represent the various feature maps and the arrows represent the different operations (image source : [7]

The U-Net architecture was proposed by Ronneberger et. al [7] and it was meant to be applied specifically for semantic segmentation of biomedical images. It consists of an encoder (for learning the feature context) and a decoder (for localising the features).

We used orobix's [8] implementation of U-Net wherein we used 10 images from the DRIVE dataset along with their ground truths for training the network. About 100,000 patches were cropped from these samples which were fed to the network as input. A batch size of 32 has been used and the number of epochs was fixed at 150. It took about 3:05 hours to train the network with the non-dehazed dataset and the dehazed dataset took about 3:08 hours on a P100 GPU on the graham cluster of Compute Canada.

LadderNet

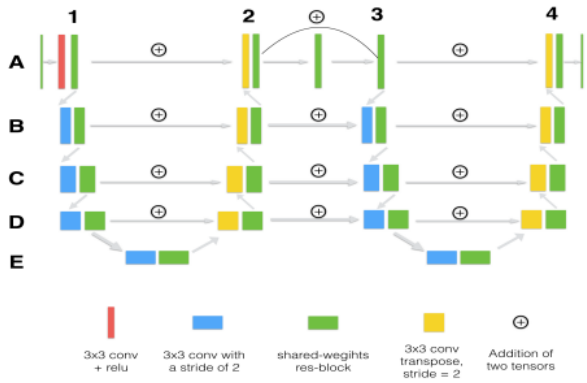


Figure 3: LadderNet architecture (image source : [9])

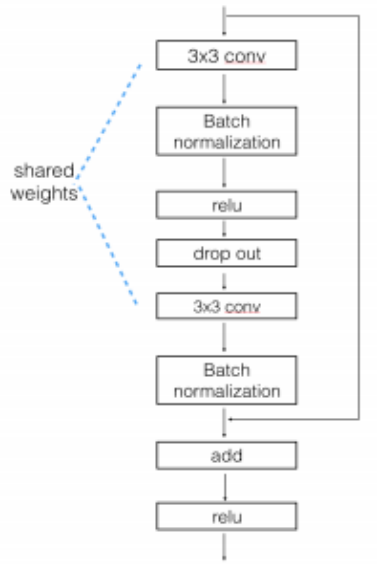


Figure 4: Shared weights residual block (image source : [9])

LadderNet can be viewd as multiple U-Nets linked together. There are skip connections between layers at every spatial scale which provide multiple pathways or data-flow. This helps it to learn much more complex features. So consequently, the number of parameters should also be more. In order to combat this problem, shared weight residual blocks (Figure 4) have been used

in which the weights of the convolutional layers belonging to the same block are the same.

About 190,000 patches are cropped from 10 samples which are fed to the network as input. A batch size of 1024 is used and the number of epochs is fixed at 150. It takes about 7:53 hours to train the network with the non-dehazed dataset and the dehazed dataset took about 7:23 hours on a P100 GPU on the graham cluster of Compute Canada.

DU-Net

This convolutional encoder and decoder of this model consist of deformable convolutional blocks which efficiently learn the various features of the retinal images via local-adaptive kernels. Each block has a convolution offset which is analogous to the kernel concept.

About 200,000 patches are cropped from 10 samples which are fed to the network as input. A batch size of 60 is used and the number of epochs is fixed at 100. It takes about 19 hours to train the network with the non-dehazed dataset and the dehazed dataset took about 18:57 hours on a P100 GPU on the graham cluster of Compute Canada.

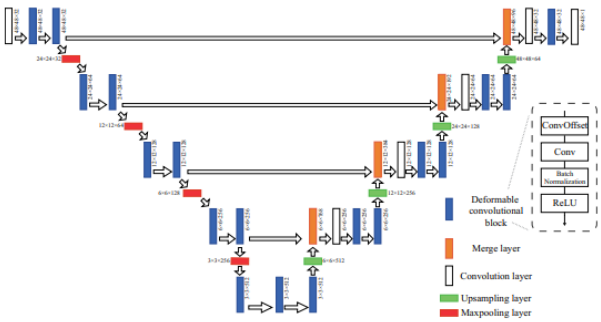


Figure 5: DU-Net architecture (image source: [10])

These models are trained separately on the dehazed dataset as well as the non-dehazed dataset which we had curated from the fundus images and their segmentation performance is evaluated based on several metrics.

Results

For the evaluation of the segmentation performance of the above mentioned models, we use commonly used performance metrics such as global accuracy, specificity, sensitivity, precision and area under ROC. The results of the U-Net and DU-Net model given in Table 1 and 3 clearly reveal that illumination correcting results in improving the performance of the network to much an extent.

On the other hand segmentation by LadderNet reveals opposite results. We see from Table 2 that the results obtained from the original dataset is outperforms the ones obtained from the corrected images. One explanation of this phenomenon might be the fact that there might be some loss in information from the haze removal method which causes the performance of the LadderNet model which has high learning capacity due to the weight shared residual blocks.

Metrics	Original	Illumination corrected
GA	0.935	0.9398
Spec	0.9625	0.9693
Sen	0.7422	0.7319
Pre	0.7367	0.7708
ROC	0.9313	0.9403

Table 1: Performance metrics of U-Net before and after correcting the non-uniform illumination in the retinal images. (GA: Global accuracy, Spec: Specificity, Sen: Sensitivity, Pre: Precision, ROC: area under ROC)

Metrics	Original	Illumination corrected
GA	0.95	0.93
Spec	0.97	0.96
Sen	0.789	0.73
Pre	0.86	0.73
ROC	0.975	0.945

Table 2: Performance metrics of LadderNet before and after correcting the non-uniform illumination in the retinal images. (GA: Global accuracy, Spec: Specificity, Sen: Sensitivity, Pre: Precision, ROC: area under ROC)

Metrics	Original	Illumination corrected
GA	0.949	0.948
Spec	0.971	0.974
Sen	0.797	0.765
Pre	0.794	0.806
ROC	0.973	0.967

Table 3: Performance metrics of DU-Net before and after correcting the non-uniform illumination in the retinal images. (GA: Global accuracy, Spec: Specificity, Sen: Sensitivity, Pre: Precision, ROC: area under ROC)

Conclusion and comments

From the comparative study we see that the effect of illumination correction has a positive effect on the performance of U-Net and DU-Net whereas a negative effect on the LadderNet model. Although the improvement is segmentation performance is good, but nowadays many modern models have been published (such as SA-Net and IterNet) which outperform the above three models with respect to the DRIVE dataset and that too without any illumination correction. However a benefit of this method is in the fact that it improves the perceptual quality of the image to much an extent which greatly useful for diagnostic and treatment purposes.

ACKNOWLEDGEMENT: This project was possible thanks to the advice and training by Professor Herbert Yang and Compute Canada's resources.

References

[1] B. Savelli, A. Bria, A. Galdran, C. Marrocco, M. Molinara, A. Campilho, and F. Tortorella, "Illumination correction by dehazing for retinal vessel segmentation," in *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, 2017, pp. 219–224.

[2] M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. Rudnicka, C. Owen, and S. Barman, "Blood vessel segmentation methodologies in retinal images – a survey," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 1, pp. 407 – 433, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169260712000843

[3] G. Leontidis, B. Al-Diri, and A. Hunter, "Summarising the retinal vascular calibres in healthy, diabetic and diabetic retinopathy eyes," *Computers in Biology and Medicine*, vol. 72, pp. 65 – 74, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010482516300609

[4] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 11, pp. 2369–2380, 2016.

[5] J. Staal, M. D. Abramoff, 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, no. 4, pp. 501–509, 2004.

[6] H. Zhang, *image_dehaze*, 2017 (accessed November 5, 2020). [Online]. Available: https://github.com/He-Zhang/image_dehaze

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.

[8] Orobix, *retina-unet*, 2018 (accessed November 6, 2020). [Online]. Available: https://github.com/orobix/retina-unet

[9] J. Zhuang, "Laddernet: Multi-path networks based on u-net for medical image segmentation," 2019.

[10] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "Dunet: A deformable network for retinal vessel segmentation," *Knowledge-Based Systems*, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705119301984