

Comparison of image processing techniques on Retinal Blood Vessel Segmentation

Guided By:

Dr. P.M. Pradhan

Submitted By:

Amit Manchanda

14116013

Anshul Jain

14116016


Abstract

Automatic segmentation of blood vessels in fundus images is of great importance as eye diseases as well as some systemic diseases cause observable pathologic modifications. It is a binary classification problem: for each pixel we consider two possible classes (vessel or non-vessel). Many algorithms, both unsupervised and supervised, have been proposed for this purpose in the past. We used a GPU implementation of max-pooling convolutional neural networks to segment blood vessels. We tested our CNN model on various inputs such as pixel intensity, FFT coefficients and two dimensional Gabor wavelet responses taken at multiple scales and rotations. We test our method on publicly available DRIVE dataset and our results demonstrate the effectiveness of different image processing techniques combined with neural network approach. On DRIVE dataset, we achieved an accuracy of 94.37% and area under ROC curve of 0.97022 by providing real and imaginary frequency components along with the patch to CNN.

Introduction

Optic fundus assessment has been widely used by the medical community for diagnosing vascular and nonvascular pathology. Inspection of the retinal vasculature may reveal hypertension, diabetes, arteriosclerosis, cardiovascular disease, and stroke. Diabetic retinopathy is a major cause of adult blindness due to changes in blood vessel structure and distribution such as new vessel growth (proliferative diabetic retinopathy) and requires laborious analysis from a specialist. Early recognition of changes to the blood vessel patterns can prevent major vision loss as early intervention becomes possible.

The retinal vessels are connected and create a binary treelike structure but some background features may also have similar attributes to vessels. Several morphological features of retinal veins and arteries (e.g. diameter, length, branching angle, tortuosity) have diagnostic significance so can be used in monitoring the disease progression, treatment, and evaluation of various cardiovascular and ophthalmologic diseases.



Because of the manual blood vessel segmentation is a time-consuming and repetitious task which requires training and skill, automatic segmentation of retinal vessels is the initial step in the development of a computer-assisted diagnostic system for ophthalmic disorders. Conventional supervised methods are usually based on two phases: feature extraction and classification. Finding the best set of features (which minimizes segmentation error) is a difficult task as choice of features significantly affects segmentation. Recent works use convolutional neural networks (CNNs) to segment images so feature extraction itself is learned from data and not designed manually. These approaches obtain state-of-the-art results in many applications. Today, large-scale DL-trained NNs successfully tackle generic object recognition tasks with thousands of object classes, a feat that many considered unthinkable just ten years ago.

In this project, we performed various preprocessing techniques to enhance the features of blood vessels like frequency and wavelet transformation along with Neural Network for the problem of detecting blood vessels in fundus imagery, a medical imaging task that has significant diagnostic relevance and was subject to many studies in the past.

Database

The **DRIVE** (Digital Retinal Images for Vessel Extraction) is a publicly available database, consisting of a total of 40 color fundus photographs. The photographs were obtained from a diabetic retinopathy screening program in the Netherlands. The screening population consisted of 453 subjects between 31 and 86 years of age. Each image has been JPEG compressed, which is common practice in screening programs. The images were acquired using a Canon CR5 non-mydratic 3-CCD camera with a 45 ° field of view (FOV). Each image is captured using 8 bits per color plane at 768 × 584 pixels. The FOV of each image is circular with a diameter of approximately 540 pixels. The set of 40 images was divided into a test and training set both containing 20 images. Three observers manually segmented a number of images. Fig. 1 presents an exemplary training image and the corresponding manual segmentation.

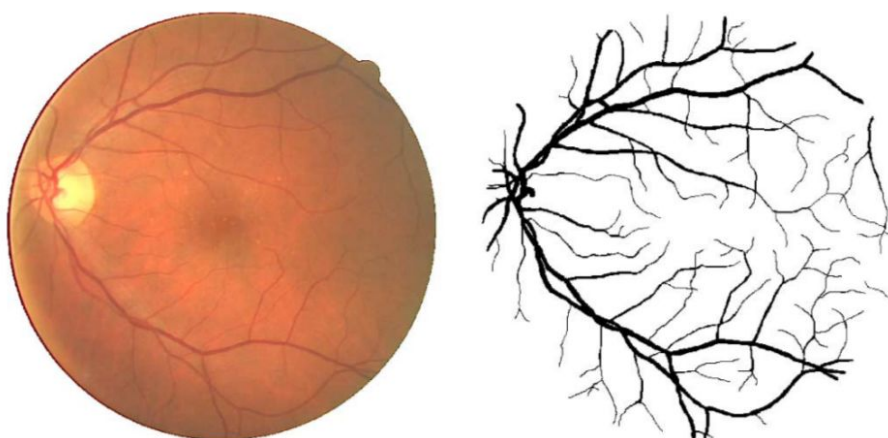


Fig. 1. A training image from the DRIVE database (left) and the corresponding manual segmentation (right).

Related Works

A large number of algorithms and techniques have been published relating to the segmentation of retinal blood vessels. These developments have been documented and described in a number of review papers (Bühler et al., 2004; Faust et al., 2012; Felkel et al., 2001; Fraz et al., 2012, 2012; Kirbas and Quek, 2004; Winder et al., 2009).

A common categorization of algorithms for segmentation of vessel-like structures in medical images (Kirbas and Quek, 2004) includes image driven techniques (such as edge-based and region-based approaches), pattern recognition techniques, model-based approaches, tracking-based approaches and neural network based approaches. Similarly approach has been used by Fraz et al. (2012).

Many articles in which supervised methods are used have been published to date. The most prevalent approach in these articles has been matched filtering. The performance of algorithms based on supervised classification is better in general than on unsupervised. Almost all articles using supervised methods report AUCs of approximately 0.95. However, these methods do not work very well on the images with non uniform illumination as they produce false detection in some images on the border of the optic disc, hemorrhages and other types of pathologies that present strong contrast. Some results show that Gabor Wavelets are very useful in retinal image analysis.

The recent research ^[1] proved that using Deep Learning approach with multiple Convolutional Layers followed by Max Pooling has shown state of the art results reporting AUCs of approximately 0.9790

Methods

This section explains the main component that we have used in our approach: convolutional neural network.

A. Convolutional Neural Network

A convolutional neural network (CNNs) consists of several weighted input features and one output, performing convolution of input features with weights and transforming the outcome with some form of nonlinearity. The units are arranged in rectangular layers (grids), and their locations in a layer correspond to pixels in an input image. Apart from the spatial characteristics of CNN, other features like local connectivity, parameter sharing and pooling of hidden units.

1) Local Connectivity: The unit receives data from its receptive field (RF) only, a small rectangle of image pixels (in the first layer) or units in the previous layer (for subsequent layers). Local connectivity substantially reduces the number of weights in comparison to the fully-connected conventional networks.

2) Parameter Sharing: It consists of sharing weights across units in the same layer. When the units in a given layer share the same vector of weights, they form a feature map, with each

of them calculating the same local feature, in different part of the image. This reduces the number of parameters even further and makes the extracted features equivariant.

3) Pooling: (subsampling) consists in aggregating outputs of multiple units by other means than convolution. In the popular method consists of Max-Pooling and Average Pooling. Like local connectivity, pooling reduces the resolution w.r.t. previous layer and provides for translational invariance.

A typical CNN architecture consists of several convolutional feature maps intertwined with max-pooling layers, finalized with at least one fully-connected layer that ‘funnels’ the excitations into output neurons, each corresponding to one decision class.

Data Generation

The raw fundus images are subject to preprocessing comprising image processing, extraction of patches and optional transformation of image.

- A. Extraction of patches :** The ground-truth data provided in the manual segmentations frames the blood vessel detection as a binary classification problem. As in many other studies, in our approach the decision on the class of a particular pixel is based on a $(m \times m)$ patch centered at that pixel. The patches are extracted from a preprocessed grayscale image and they form the input fed into a neural network. In this project, we use $m=27$, so an example is a vector of length $1 \times 27 \times 27$. We consider only patches that completely fit in the circular active area in fundus images, called field of view (FOV) in the following (Fig.1). For DRIVE, we use the original FOVs provided in the description.
- B. Image Preprocessing :** Deep learning architectures can effectively learn from raw image data. However, they tend to perform better on appropriately preprocessed images. In the following, we describe the preprocessing applied in this project.
 - a. RGB to gray:** All the images in dataset are converted to grayscale images as color channels do not provide any additional information in detection of vessels. All further processes are applied on grayscale images.
 - b. Normalization and Standardization:** In order to help the learning process to abstract from fluctuations among images such as brightness and contrast and focus on vessel detection, we perform local (per-patch) brightness and contrast normalization. Every image is normalized by subtracting the mean and dividing by the standard deviation of its elements. The images are further standardized by converting the whole range of values to 0-255. The normalization process is also later applied on the extracted patches.
 - c. CLAHE (Contrast-Limiting Adaptive Histogram Equalization):** Adaptive histogram equalization (AHE) is a technique used to improve contrast in images. It differs from ordinary histogram equalization in the respect that the adaptive method computes several histograms, each corresponding to a distinct section of the image. The image is divided into small blocks called "tiles" . Then each of these blocks are histogram equalized as usual. So in a

small area, histogram would confine to a small region (unless there is noise). If noise is there, it will be amplified. To avoid this, contrast limiting is applied. If any histogram bin is above the specified contrast limit, those pixels are clipped and distributed uniformly to other bins before applying histogram equalization. After equalization, to remove artifacts in tile borders, bilinear interpolation is applied. The tile size used in our project is 8x8.

- C. Fourier Transform :** The DFT is the sampled Fourier Transform containing a set of samples in frequency domain which is large enough to fully describe the spatial domain image. The number of frequencies corresponds to the number of pixels in the spatial domain image, *i.e.* the image in the spatial and Fourier domain are of the same size. For a square image of size $N \times N$, the two-dimensional DFT is given by:

$$F(k, l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i, j) e^{-i2\pi(\frac{ki}{N} + \frac{lj}{N})}$$

The real and imaginary frequency components at a location are given as input to the CNN model. The complexity involved with DFT is $O(n^2)$, so a faster algorithm called Fast Fourier Transform(FFT) is used which has a time complexity of $O(n \log n)$.

- D. 2D Gabor Wavelet Transform :** The Fourier transform gives poor localization of a signal in time domain, hence Discrete wavelet transform is used to get localization in time as well frequency domain. We chose the 2-D Gabor wavelet for the purposes of this work, due to its directional selectiveness capability of detecting oriented features and fine tuning to specific frequencies. This latter property is especially important in filtering out the background noise of the fundus images. The two dimensional mother Gabor wavelet :

$$\psi(\vec{x}) = \frac{1}{\sigma^2} e^{-\frac{\vec{x}^2}{2\sigma^2}} e^{i \vec{e}_h^T \vec{x}}$$

is a complex-valued filter that is divided into two overlaying parts. The first part is a Gaussian with standard deviation σ , which makes the Gabor wavelet localized both in spatial and frequency domain. The second part is the complex-valued even wave $\exp(i\vec{e}_h^T \vec{x})$ with spatial frequency $\vec{e}_h = (1, 0)^T$ pointing along the horizontal axis. The mother Gabor wavelet in frequency domain :

$$\check{\psi}(\vec{\omega}) = e^{-\frac{\sigma^2(\vec{\omega} - \vec{e}_h)^2}{2}}$$

Out of this mother Gabor wavelet, the continuous Gabor wavelet family in spatial domain:

$$\begin{aligned}\psi_{k,\vartheta,\vec{t}}(\vec{x}) &= k^2 \psi(k Q(\vartheta)^T (\vec{x} - \vec{t})) \\ &= \frac{k^2}{\sigma^2} e^{-\frac{k^2(\vec{x}-\vec{t})^2}{2\sigma^2}} e^{i \vec{e}_h^T (k Q(\vartheta)^T (\vec{x}-\vec{t}))}\end{aligned}$$

$$\vec{k} = \begin{pmatrix} k_h \\ k_v \end{pmatrix} = k Q(\vartheta) \vec{e}_h = \begin{pmatrix} k \cos(\vartheta) \\ k \sin(\vartheta) \end{pmatrix}$$

Here $Q(v)$ is the rotation matrix of the wavelet and v is the rotation angle. For each patch, we compute the 2D gabor wavelet transform and give the real and imaginary components of the coefficient matrix to our model as input.

The Experiments

Network Architecture

We used the model in fig.2 for the classification purpose. The input image patches after suitable processing are passed through a single convolutional layer with a 0.2 dropout with units equipped with 3x3 RFs. Max-pooling is performed over a 2x2 pixel window, with stride 2. ReLU activation function is used in 2D convolutional layer. Thanks to their non-saturating form, ReLUs greatly accelerate the convergence of stochastic gradient descent compared to the squeezing nonlinearities (sigmoid and tanh). Also, ReLUs can be implemented by simply thresholding a matrix of activations at zero and so execute much faster. The output of max pooling is flattened out and given to a single fully connected layer having softmax activation.

Training is carried out by stochastic gradient descent with batch updates and momentum. We used adaptive learning rate technique to reduce the learning rate if no considerable change in validation loss.

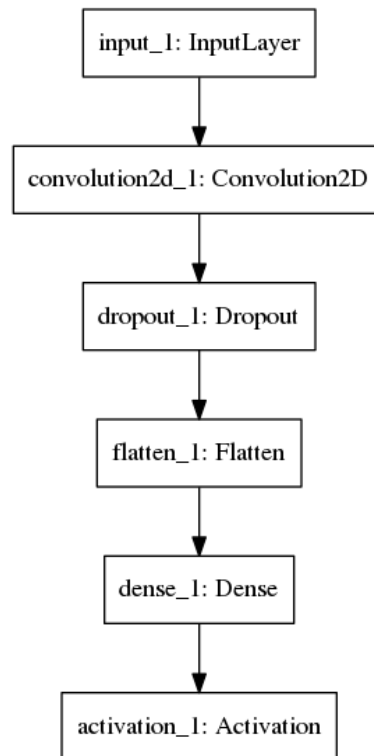


Fig. 2: Architecture of evaluated CNN model.

CNN with CLAHE:

In this experiment, we used only the preprocessing techniques such as normalization and grayscale conversion along with CLAHE. Rest other experiments involving transformed images did not use CLAHE.

2D FFT : The frequency domain equivalent of patches were evaluated using Fast Fourier Transformation. The output is a matrix of size same as patch but with complex values. There were multiple experiments involving fourier coefficients if a patch given as inputs to the neural network to compare the importance of different components.

1. Only Real components
2. Real and Imaginary components
3. Real and imaginary components along with original patch.

2D Gabor Wavelet Transform : Experiments similar to the 2D FFT case were also performed with 2D Gabor DWT but with different number of scalings and rotation angles. These experiments were performed with the reduced dataset having only 100000 patches

in the training set as compared to 180,000 patches in FFT and CLAHE experiment. This is due to the increase in data size after computing DWT. The output of wavelet transform is also complex, so experiments with only real components were performed.

1. 2 scaling factors and 2 angle of rotations
2. 2 scaling factors and 4 angles of rotations
3. 4 scaling factors and 2 angles of rotations

Results

All the experiments were tested with the testing images provided in the dataset. The metrics of result taken are accuracy and area under the ROC(Receiver Operating Characteristics) curve. The comparative study of different experiments has been plotted in ROC curves taking False positive rate(FPR) along x axis and True Positive Rate(TPR) along y axis.

$$TPR = \frac{\text{True Positive}}{\text{True Positives} + \text{False Negatives}}$$

$$FPR = \frac{\text{False Positive}}{\text{False Positives} + \text{True Negatives}}$$

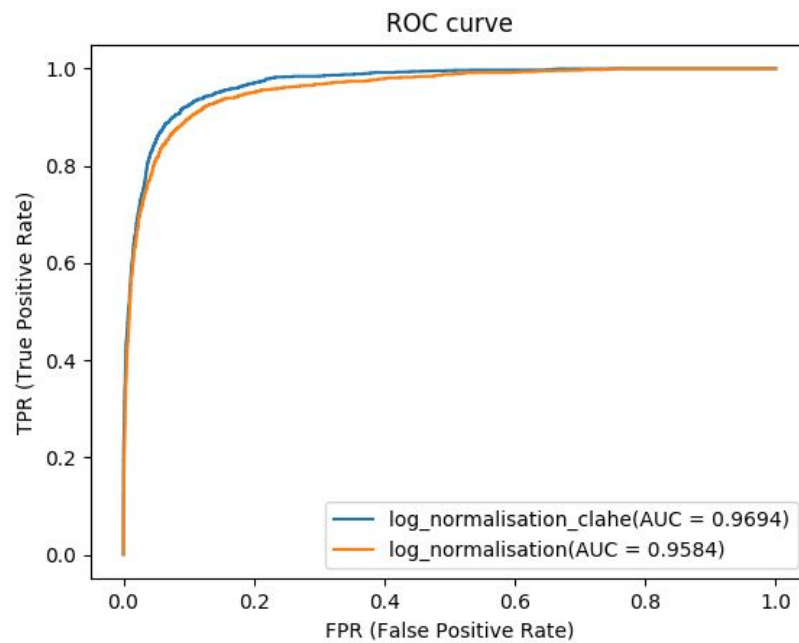


Fig.3: ROC curve of neural network with CLAHE and without CLAHE

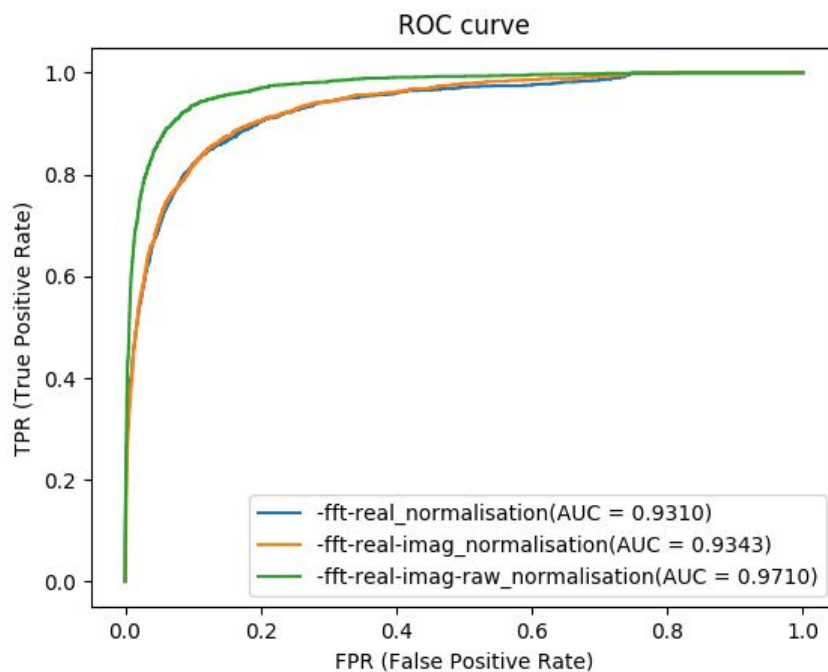


Fig.4: ROC curves of experiments with FFT.

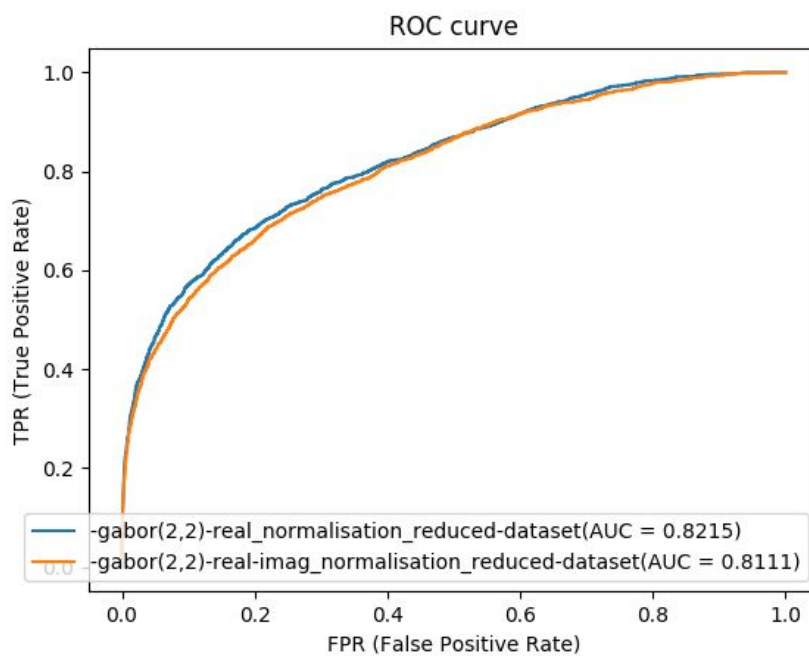


Fig.5: ROC curves of Gabor with a) Real components(in blue) , b)Real and imaginary components (in orange)

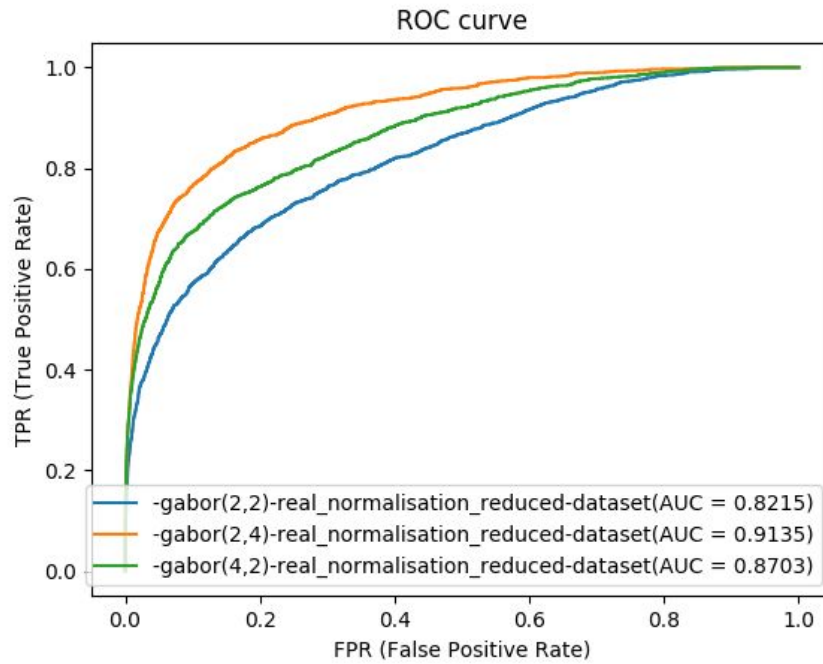


Fig.6: ROC curves of Gabor wavelet transform with different scalings and rotation angles

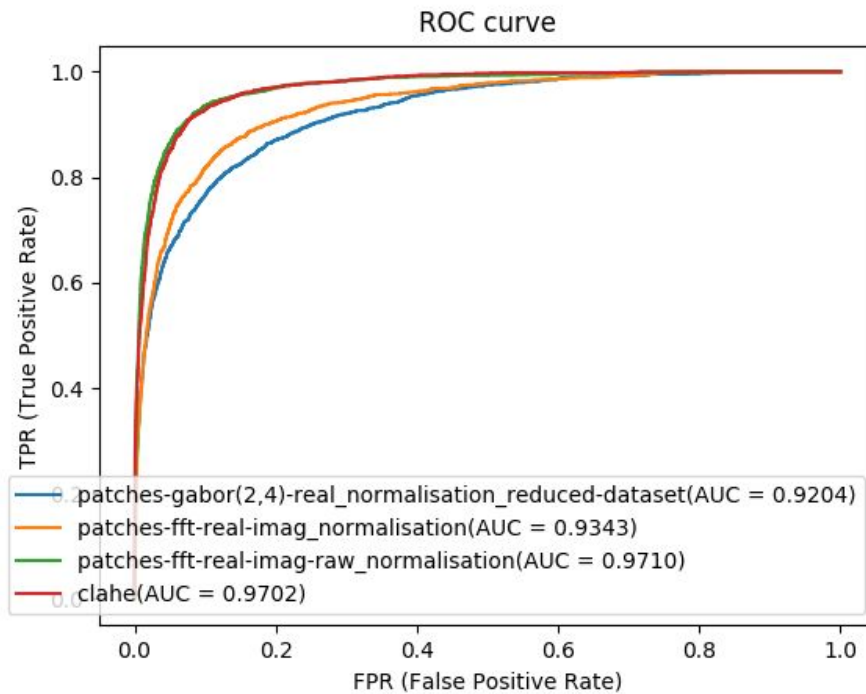


Fig.7: ROC curves of 3 best results obtained using 3 different techniques(based on ROC and AUC).

Predicted Outputs of test image 1 along with the corresponding ground truth is shown in following figures.

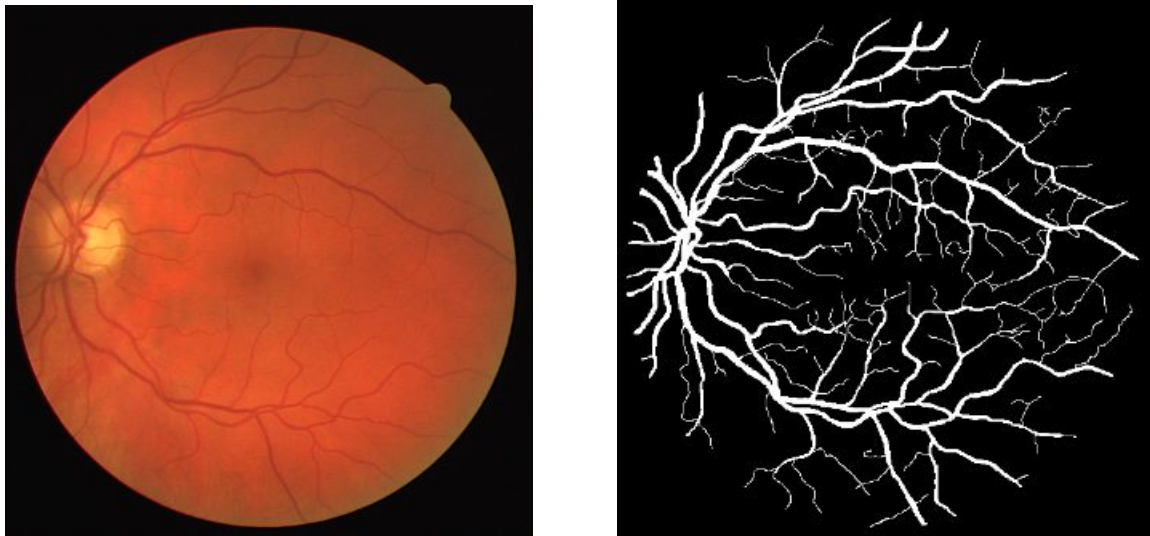


Fig.8: Original test Image 1(on left) and corresponding ground truth image(on right).

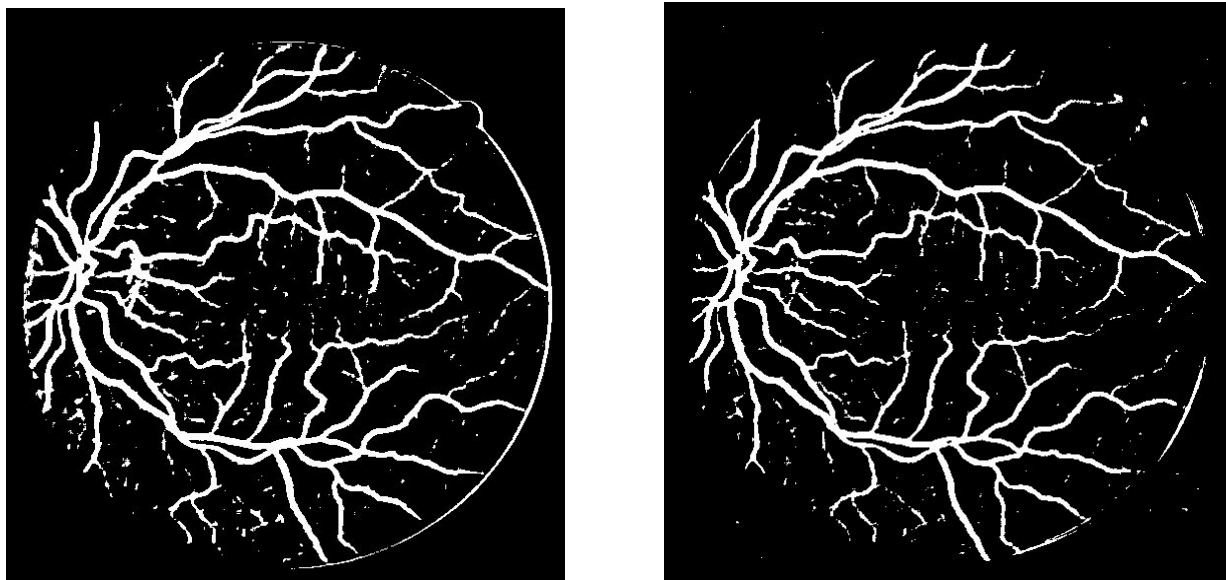


Fig.9: The predicted output using CLAHE equalization (on left) and predicted output using real and imaginary components of FFT along with original image patches (on right).

Conclusion


This study brings evidence that neural networks are a viable methodology for medical imaging. We find this encouraging, in particular given the entirely supervised character of the neural approach, which learns from raw pixel data and does not rely on any prior domain knowledge on vessel structure. While learning, a network autonomously extracts low-level features that are invariant to small geometric variations, then gradually transforms and combines them into higher-order features. In this way, the raw image is transformed into a more abstract – and a priori unknown representation that fosters effective vessel segmentation. The features learned at multiple levels of abstraction are then automatically composed into a complex function that maps an input patch to its label.

Inclusion of features obtained by frequency transformation and Gabor wavelet transformation or after Adaptive Histogram Equalization along with original image makes the learning better and faster. Providing frequency components along with the image has made the network learn better as compared to directly providing the raw pixels.

The Gabor wavelet transformation also showed promising results but due to less computational memory while training we could not explore the full power of the transformation.

References

1. Liskowski et al., "Segmenting Retinal Blood Vessels with Deep Neural Networks", IEEE Transactions on Medical Imaging, vol. PP, no. 99, pp. 1-1, 2016.
2. M. Fraz et al., "Blood vessel segmentation methodologies in retinal images—A survey," Comput. Methods Prog. Biomed., vol. 108, no. 1, pp. 407–433, Oct. 2012.
3. A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images", Computer Sci. Dept., Univ. Toronto, Tech. Rep., 2009, vol. 1, p. 7.
4. Soares et al., "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," Medical Imaging, IEEE Transactions on, vol. 25, no. 9, pp. 1214–1222, 2006.

- 
5. Martina Melinscak , Pavle Prentasic and Sven Loncaric, "Retinal Vessel Segmentation Using Deep Neural Networks", IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 35, NO. 11, November 2016.
 6. Manuel Guenther. "Statistical Gabor graph based techniques for the detection, recognition, classification and visualization of human faces", PhD thesis, Technical University of Ilmenau, June 2011.
 7. A. Anjos et al., "Bob: a free signal processing and machine learning toolbox for researchers," in 20th ACM Conference on Multimedia Systems (ACMMM), Nara, Japan. ACM Press, Oct. 2012.