Retinal Vessel Segmentation by Inception-like Convolutional Neural Networks

Hadi Niknam Shirvan[a], Reza Askari Moghadam[b] and Kurosh Madani[2]
[1]*Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran*
[2]*LISSI Lab, Senart-FB Institute of Technology, University Paris Est-Creteil (UPEC), Lieusaint, France*

Keywords: Deep Learning, Inception-like CNN, Retinal Image Processing, Medical Application.

Abstract: Deep learning architectures have been proposed in some neural networks like convolutional neural networks (CNN), recurrent neural networks and deep belief neural networks. Among them, CNNs have been applied in image processing tasks frequently. An important section in intelligent image processing is medical image processing which provides intelligent tools and software for medical applications. Analysis of blood vessels in retinal images would help the physicians to detect some retina diseases like glaucoma or even diabetes. In this paper a new neural network structure is proposed which can process the retinal images and detect vessels apart from retinal background. This neural network consists of convolutional layers, concatenate layers and transpose convolutional layers. The results for DRIVE dataset show acceptable performance regarding to accuracy, recall and F-measure criteria.

1 INTRODUCTION

Nowadays, modern methods based on deep learning methods have been widely used in various sciences and have solved a lot of challenges and problems. One of the essential applications of deep neural networks is in medical image processing in order to diagnose various diseases. Analysis of blood vessels in retinal images is done to diagnose eye diseases. Before the advent of computer vision and deep learning methods, this operation was done manually, which was time-consuming (Soomro, 2019). However, in recent years several methods have been developed to detect blood vessels in retinal images, which have high speed and high accuracy benefits, and these methods can become helpful in this field. In the earlier ways, some methods were based on image processing techniques with using different filters and math calculations on images (Staal, 2004), and some other methods were based on simple neural networks (Zhang, 2015). Other methods e.g. fuzzy c-means (Tolias and Panas,1998; Kande, 2010) and decision tree (Fraz, 2012) were proposed for segmentation blood vessels in retinal images, but the presented methods were not very accurate and were not able to detect all the blood vessels in the image.

With advent new processing hardware and providing large volumes of datasets, deep learning networks have made significant progress in medical image processing and disease detection and replaced traditional methods (LeCun, 2015). In deep learning networks, a large number of layers and neurons perform learning tasks, and by using large amounts of training data, the trained model will be highly accurate. AlexNet (Krizhevsky, 2012) was one of the earlier proposed networks in the field of deep neural networks. In the Architecture of this network, there are some layers called convolutional layers, and the operation of extracting image features are done by these layers. In the convolutional neural networks, the lower layers extract low-level features of the image such as horizontal or vertical line detection, and upper layers of the network extract the high-level features of the image. In the following years, new structures of deep learning networks were introduced, e.g., VGGNet (Simonyan & Zisserman, 2014), GoogLeNet (Szegedy, 2015) and ResNet (He, 2016), each with its architecture and features. These networks are very accurate in image processing and object recognition applications. In the Google Network, some individual layers are used, which are called the Inspection layer, in which the convolve

[a] https://orcid.org/0000-0003-4745-6956
[b] https://orcid.org/0000-0001-8394-7256

Shirvan, H., Moghadam, R. and Madani, K.
Retinal Vessel Segmentation by Inception-like Convolutional Neural Networks.
DOI: 10.5220/0009638100530058
In *Proceedings of the 1st International Conference on Deep Learning Theory and Applications (DeLTA 2020)*, pages 53-58
ISBN: 978-989-758-441-1

operation is done in different sizes and parallel, and the results of operations concatenate to each other at the end of Inspection layer. This approach let the network to learn best weights and automatically select the more useful features. But this operation needs more computational cost for training this layer and because of this, At the beginning of this network, several pooling layers reduce the size of the input image, and as a result, the weight of the network decreases, which makes the network very fast to train.

With the advances in deep neural networks, these networks can be used for various applications such as object detection, face recognition, cancer detection. The techniques of learning in neural networks are divided into two categories of supervised learning and unsupervised learning. Supervised learning uses input data with their ground truth to makes the behavior of the network more similar to the target label. The method of supervised learning is more accurate for image processing, but preparing data with precise and appropriate ground truths is one of the main challenges of this method. In retinal images, the detection of blood vessels is done by segmentation of input images pixels. In the ground truth image of training data, the blood vessels and the other parts are separated. After the learning process, the trained model would be able to separate the blood vessel and other parts in each input picture. Many networks have been proposed for segmentation operations on retinal images that are highly accurate in the detection of blood vessels in retinal images.

One of the earlier proposed architecture for image segmentation applications is Fully Connected Network (FCN) (Long, 2015). All layers that are used in the architecture of this network are convolutional, and there is no fully connected layer. This architecture makes the network independent of the input image size.

U-Net (Ronneberger, 2015) is another network for segmentation operations in medical images. In this network, encoder and decoder operations are done on the images. In the encoder part, the features of the image are extracted, and the size of the input image is reduced. After the encoder layers, the decoder layers are replaced, which reconstructs the image by concatenating the lower layers.

Some new methods (Guo, 2019) are proposed which their structures are inspired by other famous networks like VGG-Net, Res-Net and U-Net. Although there are some changes in the architectures of layers in these networks. Deep Retinal Images Understanding (DRIU) is name of a structure that is proposed for segmentation both blood vessels and optical disc in retinal images. The structure of DRIU is based on VGG-Net but more in-depth (Maninis, 2016). In some frameworks (Soomro, 2019), more in-depth examples of encoder-decoder architectures are used for the segmentation of retinal images. In these architectures, the polling layers are replaced by stride in the convolutional layers.

In this research a new approach for segmentation of blood vessels in retinal images is used which called Inception-like CNN. This structure first was proposed for saliency detection applications in our previous research. In the architecture of this network there are some layers which are based on inception layers in GoogLe-Net (Misaghi, 2018; Misaghi, 2018).

2 THE PROPOSED METHOD

In the structure of our network, there are five Inception-like layers. These layers are based on the idea of the inception layer but partly different. In the Inception-like layer, three convolutional layers are used separately, as shown in Figure 1. The main idea

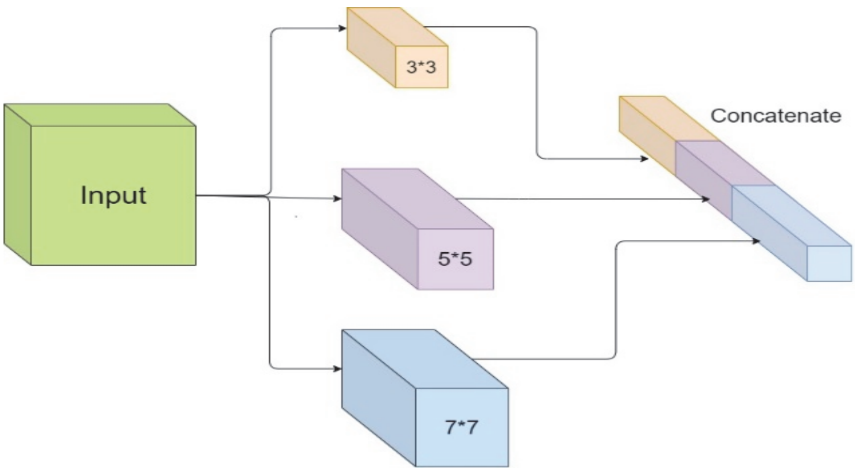

Figure 1: Inception-like layer.

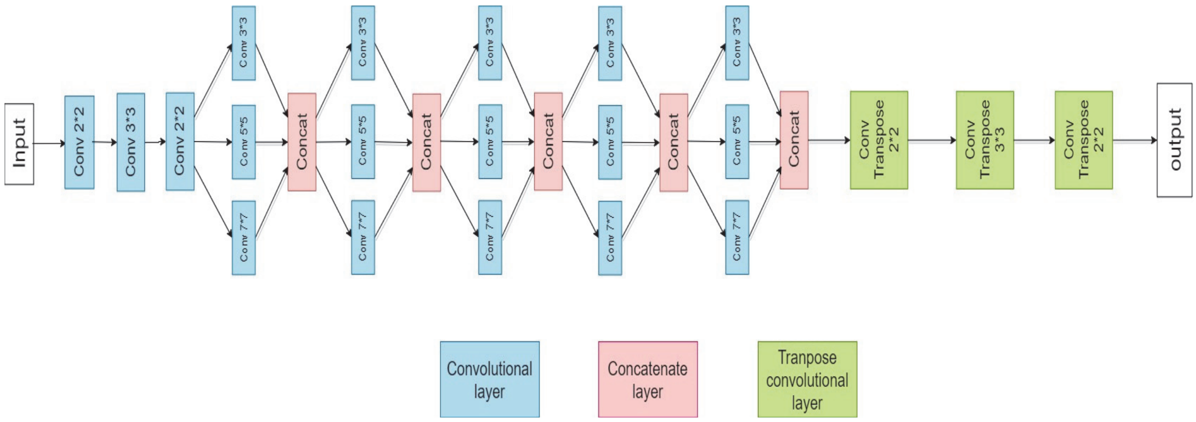

Figure 2: Architecture of proposed network.

in this approach is to extract more features by using different size convolution on input images. It is desired to know some relations between regions of the retinal image and their surroundings to learn whether that region is a blood vessel or not. The size of three convolutional layers is 3*3, 5*5, and 7*7. And the RELU activation function is used at the end of each convolution layer, which is the most commonly used activation function in the CNNs. RELU function compares the input value with zero and returns the maximum value between them. After the activation function, the results of three convolution layers are concatenated in depth. With the advantage that different size convolve operation is done, more data is extracted in each input image, and this operation improves the performance of the network.

As shown in Figure 2, in the architecture of this network, there are other layers except the Inception-like layer. The input image is passed, and its size is reduced by three down sampling layers. Same as the structure of other convolutional neural networks with passing each layer, the depth of layers increases. In these layers, we actually try to teach the neural network to focus on the fewer activation points than all of it, because to reduce resolution of the feature map which helps to reduce time and memory while training. After extracting data and features of the image in Inception-like layers, three transpose convolution layers up samples the feature map to rescaling it to the desired size. Transpose convolution layers operate despite convolution layers, which means in 3*3 kernel, they map from 1 input pixel to 3x3 pixels instead of mapping from 3x3 input pixels to 1 output.

p sampling and down sampling layers use the RELU activation function, the same as convolutional layers in the Inception-like layer. In the last layer of the network, the sigmoid activation function is used to bounds the output map to a grayscale image. In the output map, the pixels are divided into two regions that show there are blood vessels or not.

3 TRAINING PROCEDURE

In the application of blood vessel segmentation in retinal images, there are several datasets. The mostly used datasets are DRIVE (Staal, 2004) and Stare (Hoover, 2000). We use DRIVE dataset for training our network. It consists of 40 colour images and is divided into training and test, each containing 20 retinal images. For each image, there are a manual segmentation ground truth and a binary mask. Figure 3 shows an example of DRIVE dataset with its ground truth and binary mask. In the training procedure, the whole 20 training images divided into 10000 sub-images, and these sub-images stick to each other randomly, to build the input map. This operation is done because of shortage dataset and the trained model by this method would be more powerful in segmentation operation. Using more sub-images for training takes more time for training process, thus choosing an efficient number for sub-images is so important. Ninety percent of training images are used for train and other 10 percent for the validation set.

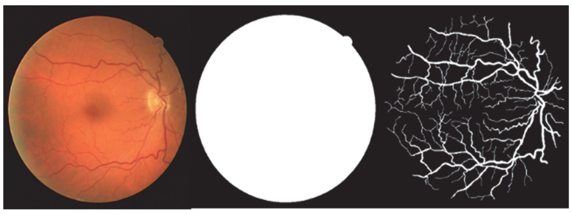

Figure 3: An example of DRIVE dataset. Original image- Binary mask- Ground truth.

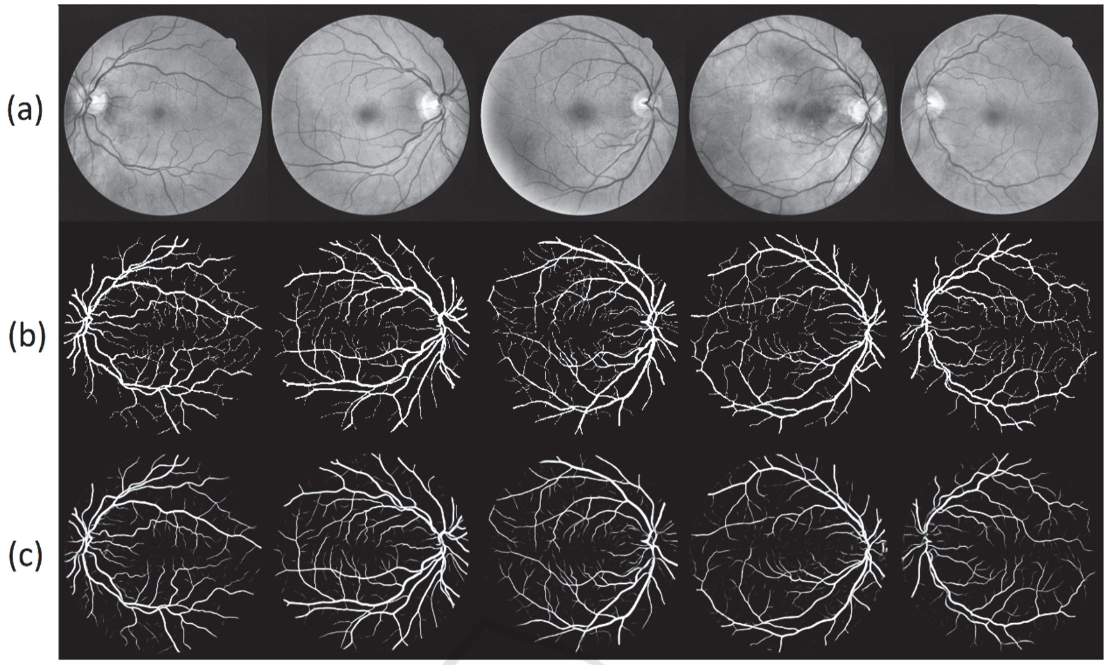

Figure 4: (a) Gray scale of original image (b) Ground truth (c) Output of network.

The training process is done with the help of the GPU service of the Google Colab framework. An Adam optimizer with exponential decay is used to update weights. Adam optimizer is used with a learning rate of 0.01. Adam optimizer has the advantage of high performance and high speed in optimization deep neural networks.

4 RESULTS AND EVALUATION

In this paper, a network for the segmentation of blood vessels in retinal images has been proposed. After training the model by 20 training images, the prepared model is tested by 20 test images of DRIVE dataset. In the output map, all pixels are divided into a vessel or non-vessel pixel. In Figure 4 a few test examples of retinal images with the corresponding ground truth and the output of the network are shown.

Evaluation parameters can be measured by comparing between output map and ground truth picture. For a vessel pixel in the output map, it would be considered as true positive (TP) and false positive (FP) if the corresponding point is defined as vessel or non-vessel, respectively. Also, true negative (TN) and false negative (FN) are defined for non-vessel pixel in output map, same as TP and FP.

Precision and recall parameters are defined in (1) and (2). These parameters can't show the quality of the result, alone, and they should present with each other. So another parameter is defined as F-measure that contains both precision and recall parameters.

$$precision : \frac{TP}{TP + FP} \quad (1)$$

$$recall = \frac{TP}{TP + FN} \quad (2)$$

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} \quad (3)$$

The accuracy of the results is another parameter that is being used for evaluating the quality of the result of deep neural networks. It defines as the ratio of all truly predicted pixels to whole pixels of the input image.

$$\text{accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

The Receiver Operating Characteristic (ROC) curve and the Area Under ROC (AUC) are two important parameters for comparing different methods of segmentation in research works. ROC is a probability curve, and AUC represents the power of the trained model in separating the pixels of the input image into a vessel or non-vessel. AUC ranges in value from 0 to 1. The results with closer AUC to one has better quality in segmentation problems. These parameters are calculated almost in all researches in this subject and by comparing the parameters with other state of arts the performance of method could be specified.

In Figure 5, the ROC curve of our results is shown, and the AUC is measured. And in Table 1, the F-measure, accuracy, and AUC are compared with other proposed methods.

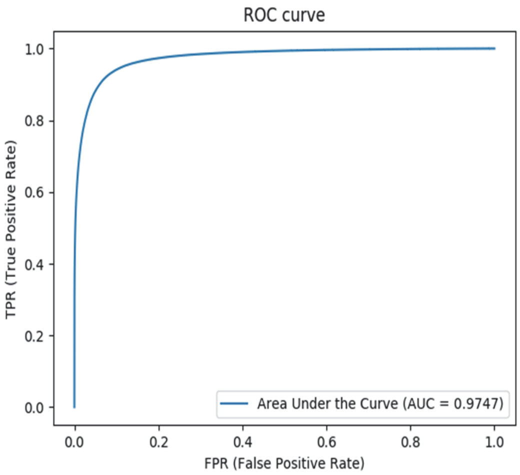

Figure 5: ROC curve for our results.

Table 1: Performance comparison with other proposed methods on the DRIVE dataset.

methods	accuracy	AUC	F_measure
Active Contour Model (Zhao, 2015)	0.9540	0.8620	0.7820
DRIU (Maninis, 2016)	0.9552	0.9793	0.8220
Three-stage FCN (Yan, 2018)	0.9538	0.9750	-
Modified U-net (Zhang, 2018)	0.9504	0.9799	-
Our method	**0.954**	**0.979**	**0.818**

The results of Figure 4, 5 and Table 1 show that the proposed method is powerful in segmentation task and it could be useful for diagnosing eye diseases. The accuracy of this network is acceptable, due to the result of feature extraction by several convolve operation in Inception-like layers.

5 CONCLUSIONS

The applications of artificial intelligence methods and machine learning techniques are growing drastically in many fields like medical subjects. One major intelligent tool for medical image processing is deep learning neural networks. In this paper a convolutional neural network is proposed which is able to process retina images fast and detects vessels apart from retina background. It can help the physicians to find and detect some retina diseases like glaucoma or even detect some other diseases like diabetes. The proposed CNN consists of three major parts including convolutional layers, concatenate layers and transpose convolutional layers. The features are extracted by several convolve operation in Inception-like layers. In proposed CNN accuracy is about 0.954, AUC is 0.979and F-measure value is 0.818.

ACKNOWLEDGEMENT

The authors would thank Mr. Hooman Misaghi for his helps and supports.

REFERENCES

Soomro, T. A., Afifi, A. J., Zheng, L., Soomro, S., Gao, J., Hellwich, O., & Paul, 2019, Deep learning models for retinal blood vessels segmentation: A review. *IEEE Access*, Vol.7, pp.71696-71717.

Staal, J., et al., 2004, "Ridge-based vessel segmentation in color images of the retina." *IEEE transactions on medical imaging*, Vol. 23(4), pp. 501-509.

Zhang, J., et al., 2015, Blood vessel segmentation of retinal images based on neural network. *International Conference on Image and Graphics, Springer*, pp.11-17.

Tolias, Y. A. and S. M. Panas, 1998, "A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering." *IEEE transactions on medical imaging,* Vol. 17(2) , pp. 263-273.

Kande, G. B., Subbaiah, P. V., & Savithri, T. S., 2010, Unsupervised fuzzy-based vessel segmentation in pathological digital fundus images. *Journal of medical systems*, Vol. 34(5), pp. 849-858.

Fraz, M. M., et al., 2012, "An ensemble classification-based approach applied to retinal blood vessel segmentation." *IEEE Transactions on Biomedical Engineering,* Vol. 59(9), pp. 2538-2548.

LeCun, Y., et al., 2015, "Deep learning." *Nature*, Vol. 521(7553), pp. 436-444.

Krizhevsky, A., et al., 2012, Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, pp. 1097-1105.

Simonyan, K. and A. Zisserman, 2014, "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv*, pp. 1409-1556.

Szegedy, C., et al., 2015, Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9.

He, K., et al., 2016, Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778.

Long, J., et al., 2015, Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431-3440.

Ronneberger, O., et al., 2015, U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention, Springer*, pp. 234-241.

Guo, S., et al., 2019, "BTS-DSN: Deeply supervised neural network with short connections for retinal vessel segmentation." *International journal of medical informatics*, Vol. 126, pp. 105-113.

Maninis, K.-K., et al., 2016, Deep retinal image understanding. *International conference on medical image computing and computer-assisted intervention, Springer,* pp. 140-148.

Soomro, T. A., et al., 2019, "Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation." *Expert Systems with Applications,* Vol. 134, pp. 36-52.

Misaghi, H., et al., 2018, Image Saliency Detection By Residual And Inception-like CNNs. *2018 6th RSI International Conference on Robotics and Mechatronics (IcRoM), IEEE*, pp. 94-99.

Misaghi, H., et al., 2018, Convolutional neural network for saliency detection in images. *6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), IEEE*, pp. 17-19.

Hoover, A., et al., 2000, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response." *IEEE transactions on medical imaging,* Vol. 19(3), pp. 203-210.

Zhao, Y., et al., 2015, "Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images." *IEEE transactions on medical imaging,* Vol. 34(9), pp. 1797-1807.

Maninis, K.-K., et al., 2016, Deep retinal image understanding. *International conference on medical image computing and computer-assisted intervention, Springer*, 140-148.

Yan, Z., et al., 2018, "A three-stage deep learning model for accurate retinal vessel segmentation." *IEEE journal of Biomedical and Health Informatics* Vol. 23(4), pp. 1427-1436.

Zhang, Y. and A. C. Chung, 2018, Deep supervision with additional labels for retinal vessel segmentation task. *International conference on medical image computing and computer-assisted intervention, Springer*, pp. 83-91.