Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 00 (2018) 000–000

Procedia
Computer Science

www.elsevier.com/locate/procedia

International Conference on Data Science (ICDS 2018)

Retina Blood Vessel Segmentation Using A U-Net Based Convolutional Neural Network

Wang Xiancheng[a,b]*, Li Wei[a],Miao Bingyi[c], Jing He[d,e], Zhangwei Jiang[f], Wen Xu[e], Zhenyan Ji[g], Gu Hong[h], Shen Zhaomeng[h]

[a] State Key Laboratory of Fluid Power Transmission and Control, Zhejiang University, Hangzhou 310027, China
[b]Ningbo Institute of Technology, Zhejiang University, Ningbo 315100, China
[c]Harbin Engineering University, Harbin, 150001, China
[d]Institute of Information Technology, Nanjing University of Finance and Economics, China
[e]Swinburne Data Science Research Institute, Swinburne University of Technology, Melbourne, Australia
[f]Baidu Inc, Beijing, China
[g]School of Software Engineering, Beijing Jiaotong University, Beijing, China
[h]Department of Ophthalmology, Ningbo Lihuili Hospital

Abstract

This paper applies deep learning techniques to the retinal blood vessels segmentations based on spectral fundus images. It presents a network and training strategy that relies on the data augmentation to use the available annotated samples more efficiently. Thus, the shape, size, and arteriovenous crossing types can be used to get the evidence about the numerous eye diseases. In addition, we apply deep learning based on U-Net convolutional network for real patients' fundus images. As a result of this, we achieve high performance and its results are much better than the manual way of a skilled ophthalmologist.

Selection and/or peer-review under responsibility of the organizers of ICDS 2018

Keywords: U-Net convolutional neural network; deep learning; image segmentation; blood vessels segmentation

1. Introduction

Various eye diseases can be diagnosed through the characterization of the retinal blood vessels [1]. The characterization can be extracted by using proper imaging techniques and data analysis methods [2]. It was

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: wxc@nit.zju.edu.cn.

commonly accepted that successful training of deep networks requires many thousand annotated training samples [3-5]. In this paper, we present a network and training strategy that relies on the data augmentation to use the available annotated samples more efficiently. The architecture is based U-Net convolutional neural network which won the ISBI cell tracking challenge 2015 in these categories [6]. Apart the imaging techniques, proper technologies have to be chosen for the data analysis in order to automate the diagnostics process. In case of eye examination, one of the important tasks is the retinal image segmentation. In the last two decades, a lot of algorithms for the retinal image segmentation have been proposed [7-9]. Most of them concentrate on the blood vessel segmentation, since proper characterization of the blood vessels plays significant role in different eye diseases diagnostics. The majority of the proposed algorithms are based on classical machine learning schemes where handcrafted feature extraction methods are used and trainable classifiers utilize handcrafted features to get the segmented image. The main limitation consists in the feature engineering which is a difficult, expensive and time-consuming process. Recent developments in artificial neural networks (ANNs) and deep learning provide an effective way for feature learning [10]. In the last decade, there have been a lot of publications about deep learning applications in the field of machine vision [11] and medical image analysis [12]. Recently numerous papers describing deep learning based approaches for the colorful retinal image segmentation [13-14] have been published, whereas there have been the significant number of developments in the field of hyperspectral image segmentation [15-19]. To the best of the author's knowledge, this paper is the first work which studies application of the deep learning techniques for the retinal blood vessels segmentation using spectral fundus images. Thus, the shape, size, and arteriovenous crossing types can be used to get the evidence about the numerous eye diseases. In this paper, we apply deep learning based on U-Net convolutional network for real patients' fundus images. As a result of this, we achieve high performance and its results are much better than the manual way of a skilled ophthalmologist.

This paper is organized as follows. Section 1 gives the introduction for the background, section 2 presents the proposed approach. The Experiments on the patients' fundus images were displayed in section 3. Sections 4 concludes this paper. The first introduction section is organized as follows. Subsection 1.1 gives the background of eye structure and fundus image and utilization of the blood vessel characteristics for the eye diseases diagnostics. Subsection 1.2 describes the development and brief theoretical background of convolutional neural network, and a brief analysis of existing algorithms based on the deep learning and neural networks for the retinal image segmentation. Subsection 1.3 provides the description of U-Net convolutional neural network.

1.1. Eye Structure and fundus images

The eye is an organ of sight which typically has a spherical form and located in an orbital cavity. The human eye has a complicated structure which is presented in Fig. 1a. Usually three layers of the eyeball are distinguished: the outer fibrous layer, the middle vascular layer, and the inner nervous tissue layer [20] shown in Figure 1.

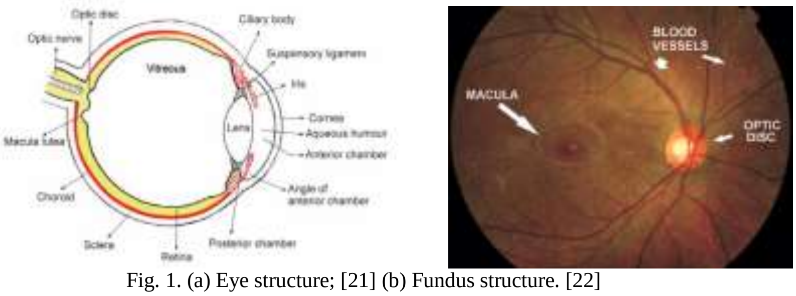


Fig. 1. (a) Eye structure; [21] (b) Fundus structure. [22]

Eyes diseases include macular, hypertensive retinopathy, diabetic retinopathy and etc. Most of the retinal diseases are usually detected by identifying the size, shape and widen of vessels [23] in the manual way. Thus it will be helpful for diagnosis if we can get vessel diameter automatically.

1.2. Deep neural networks

Over the past few years major computer vision research efforts have concentrated on convolutional neural networks, commonly referred to ConvNets or CNNs. These research works have produced a better performance on a wide range of classification (e.g [24-25]) and regression (e.g [26-28]) tasks.

A typical neural network architecture is made of an input layer, x, an output layer, y, and a stack of multiple hidden layers, h, where each layer consists of multiple cells or units, as depicted in Figure 2. Usually, each hidden unit, h_j, receives input from all units at the previous layer and is defined as a weighted combination of the inputs followed by a nonlinearity according to

$$h_j = F\left(b_j + \sum_i w_{ij} x_i\right) \quad (1)$$

where, w_{ij}, are the weights controlling the strength of the connections between the input units and the hidden unit, b_j is a small bias of the hidden unit and $\boldsymbol{F}(\cdot)$ is a certain saturating nonlinearity such as the sigmoid function. Deep neural networks can be seen as a modern day instantiation of Rosenblatt's perceptron [29] and multilayer perceptron [30]. For example, the figure as Fig. 2 was reproduced from [31].

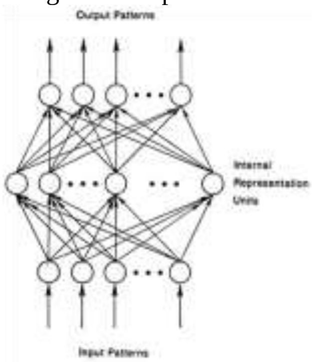


Fig. 2. Illustration of a typical Neural Network architecture [31].

Although, neural network models have been around for many years (*i.e.* since the 1960's) they were not heavily used until more recently. There were a number of reasons for this delay. For example, a major contribution that allowed for a progress in the field of deep neural networks is layer-wise unsupervised pretraining, using Restricted Boltzman Machine (RBM) [32]. Restricted Boltzman Machines can be seen as two-layer neural networks where, in their restricted form, only feedforward connections are allowed. These are some shortcomings which limit the application of neural network models.

Convolutional networks (ConvNets) are a special type of neural network that are especially well adapted to computer vision applications because of their ability to hierarchically abstract representations with local operations. There are two key design ideas driving the success of convolutional architectures in computer vision. First, ConvNets take advantage of the 2D structure of images and the fact that pixels within a neighbourhood which are usually highly correlated. Further, ConvNet architectures rely on feature sharing and each channel (or output feature map) is thereby generated from convolution with the same filter at all locations. This important characteristic of ConvNets leads to an architecture that relies on far fewer parameters compared to standard Neural Networks. Second, ConvNets also introduce a pooling step that provides a degree of translation invariance making the architecture less affected by small variations in position. Notably, pooling also allows the network to gradually see larger portions of the input. The increase in receptive field size (coupled with a decrease in the input's resolution) allows the network to represent more abstract characteristics of the input as the network's depth increase. For example, for the task of object recognition, it is advocated that ConvNets layers start by focusing on edges to parts of the object to finally cover the entire object at higher layers in the hierarchy. For example, Fig.3 was reproduced from [33].

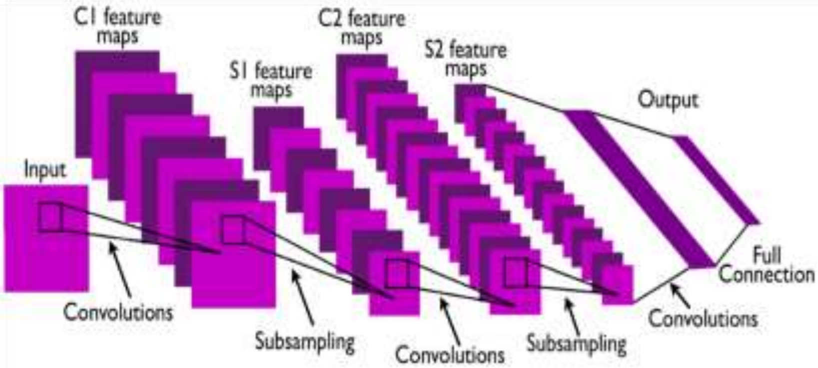


Fig. 3. Illustration of the structure of a standard Convolutional Network [33].

1.3. Brief introduction of U-Net convolutional networks

The U-Net architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. The U-Net architecture can be trained end-to-end from very few images and outperforms most of the methods on the ISBI challenge for segmentation of neuronal structures in electron microscopic stacks.

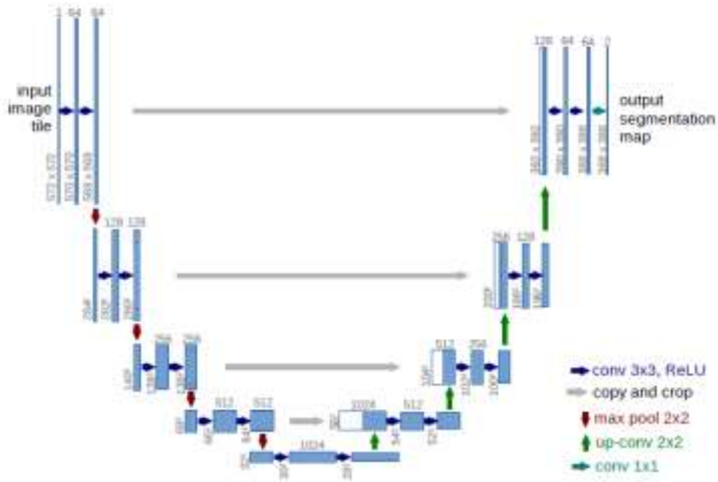


Figure 4. U-net architecture [6]

U-net architecture is shown in Figure 4 (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on the top of the box. The *x*-*y*-size is provided at the lower left edge of the box. White boxes represent copied feature maps and the arrows denote the different operations.

2. The proposed approach

This section is organized as follows. Subsection 2.1 gives the improved U-Net architecture. Subsection 2.2 describes the results of testing and predicting the standard database STARE by using the proposed approach.

2.1. Network architecture

The architecture of CNN used for the blood vessel segmentation of the fundus images is presented in Figure 5. It was derived from the U-Net network presented in Figure 4. The U-Net exhibits the encoder-decoder architecture where the decoder gradually recovers it. As a result, it produces a pixel-wise probability map instead of classifying an input image as a whole. The U-Net in opposition to other CNN architectures does not require a huge amount of training samples and can be effectively trained with only a few images. This was also in the case of the dataset considered in this study.

Compared to the original architecture, some important modifications were introduced in the CNN used in this work. First, the network was downscaled. Particularly, the depth of the network was reduced by removing two (out of five) levels of pooling/upsampling operations with the corresponding convolution. Additionally, the number of feature vectors at each level was halved. As a result, the number of filters varies from 32 at the input to 128 in the lowest resolution. The downscaling was performed since shallower architecture allowed to obtain equivalent results as the original U-Net, but the training became easier and its time was significantly reduced. The final number of layers and their configuration were selected via experimentation and were balanced between the training time and the accuracy of network. Additionally, dropout layers were introduced between the convolutional layers to improve the training performance.

2.2. Training and Prediction

The performance of this neural network is tested on the DRIVE database, and it achieves the best score in terms of area under the ROC curve in comparison to the other methods published so far. Also on the STARE datasets, this method reports one of the best performances.

Before training, the 20 images of the DRIVE training datasets are pre-processed with the following transformations:

- Gray-scale conversion
- Standardization
- Contrast-limited adaptive histogram equalization (CLAHE)
- Gamma adjustment

The proposed U-Net based network was trained using patches of a size 32*32 pixels. This patch size provided the best accuracy of pixel classification by CNN. Each patch is obtained by randomly selecting its center inside the full image. Also the patches partially or completely outside the Field Of View (FOV) are selected, in this way the neural network learns how to discriminate the FOV border from blood vessels.

A set of 190,000 patches is obtained by randomly extracting 9,500 patches in each of the 20 DRIVE training images. Although the patches overlap, i.e. different patches may contain same part of the original images, no further data augmentation is performed. The first 90% of the dataset is used for training (171,000 patches), while the last 10% is used for validation (19,000 patches).

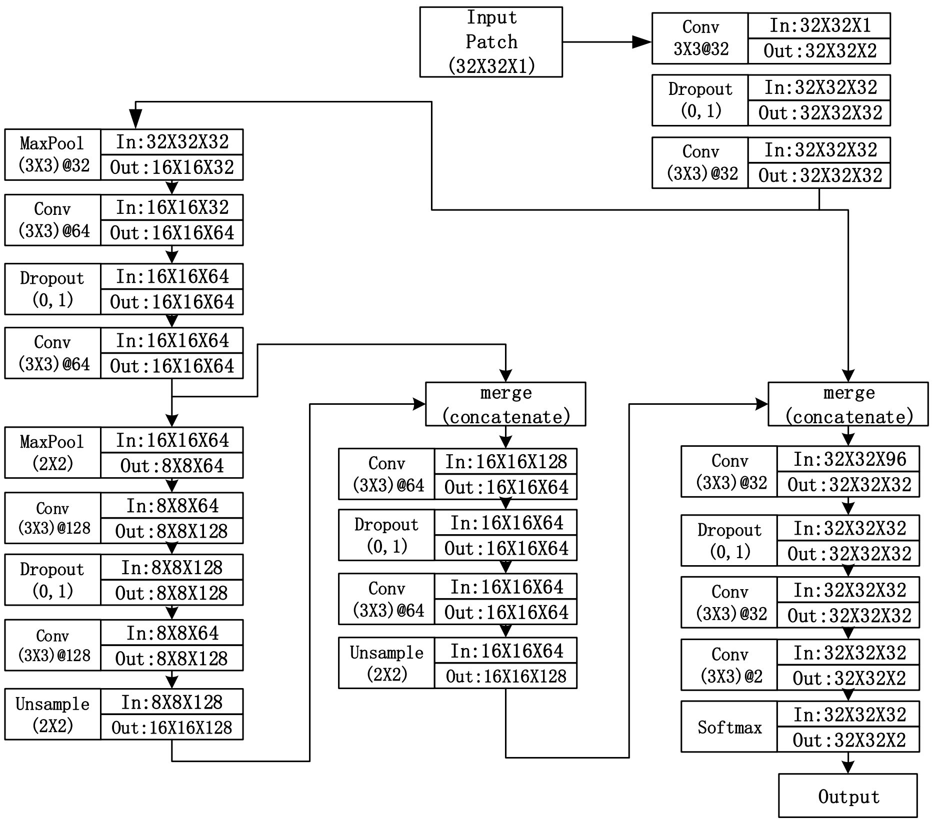


Fig.5 The architecture of the proposed CNN network.

The loss function is the cross-entropy and the stochastic gradient descent is employed for optimization. The activation function after each convolutional layer is the Rectifier Linear Unit (ReLU), and a dropout of 0.2 is used between two consecutive convolutional layers shown in Figure 5.

Training is performed for 150 epochs, with a mini-batch size of 32 patches. Using a GeForce GTX TITAN GPU, the training lasts for about 20 hours.

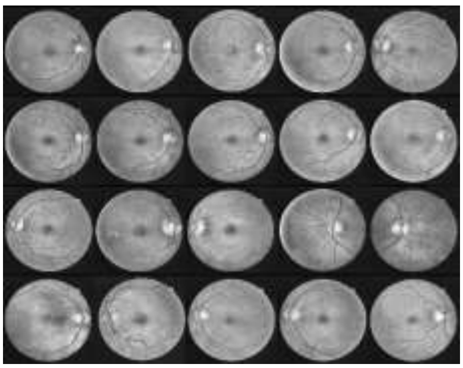


Fig. 6. Preprocessed original images.

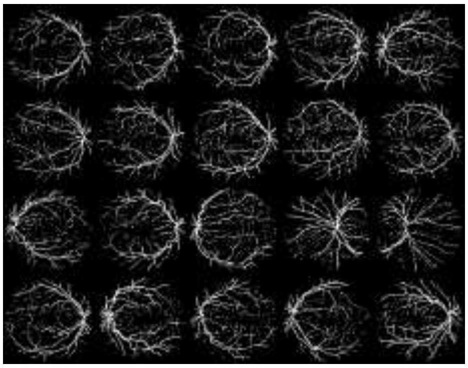


Fig.7. All predictions.

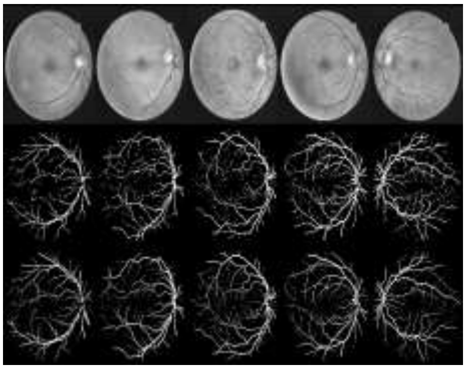


Fig.8. Original ground-truth prediction (1).

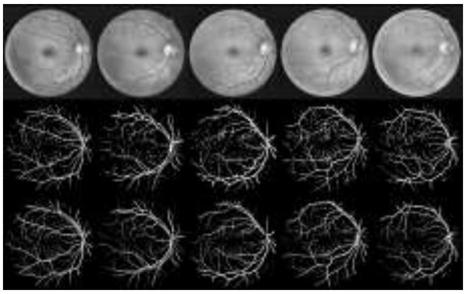


Fig.9. Original ground-truth prediction (2).

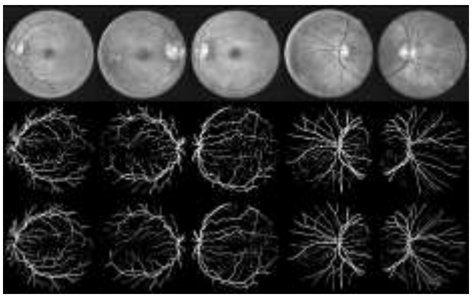


Fig.10. Original ground-truth prediction (3).

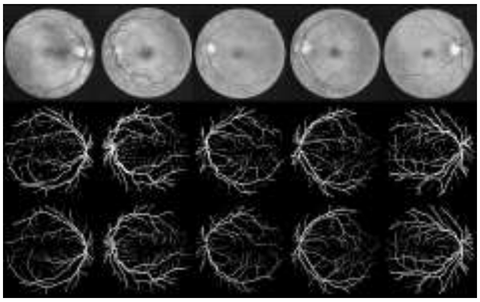


Fig.11. Original ground truth prediction (4).

From images above shown in Fig. 7 to Fig. 11, it is not hard to find that both manual and convolutional neural network perform well. But the manual vessel is required for high skills and patience while the network just needs proper architecture and parameters. In Figure 6. it shows the original images were processed by the four steps. In Figure 7, we can see all the predictions about the Figure 6. It is with very clear vision and high contrast. In Figure 8-11, it is obvious to find that there is a little difference between the manual results and neural network's predictions. More results are shown in Table 1 and Table 2.

Table 1. AUC ROC of STARE images

STARE image	AUC ROC	STARE image	AUC ROC	STARE image	AUC ROC	STARE image	AUC ROC
im0239.ppm	.9751	im0003.ppm	.9856	im0081.ppm	.9930	im0044.ppm	.9883
im0324.ppm	.9661	im0319.ppm	.9702	im0291.ppm	.9635	im0001.ppm	.9709
im0139.ppm	.9845	im0163.ppm	.9952	im0005.ppm	.9703	im0002.ppm	.9588
im0082.ppm	.9929	im0077.ppm	.9925	im0235.ppm	.9912	im0236.ppm	.9893
im0240.ppm	.9832	im0162.ppm	.9913	im0004.ppm	.9732	im0255.ppm	.9819

Table 2. AUC ROC comparison of different methods

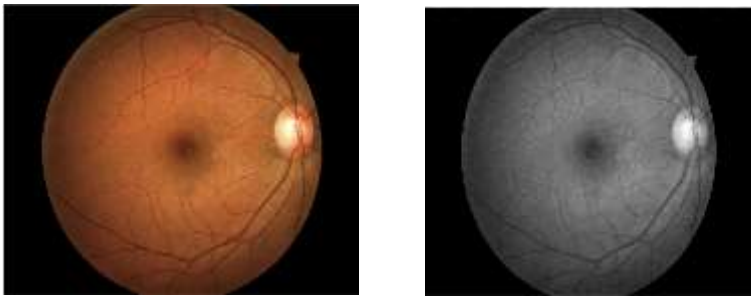
Method	AUC ROC on DRIVE
Soares et al [34]	.9614
Azzopardi et al. [35]	.9614
Osareh et al [36]	.9650
Roychowdhury et al. [37]	.9670
Fraz et al. [38]	.9747
Qiaoliang et al. [39]	.9738
Melinscak et al. [40]	.9749
Liskowski et al [41]	.9790
this method	.9790

3. The Experiments on the patients fundus images

For section 2, we can see that what we have proposed has been applied for the practical medical research or the practical medical treatment will make real sense. Therefore, it is necessary to test accuracy in real fundus photographed by a skilled ophthalmologist as a comparison.

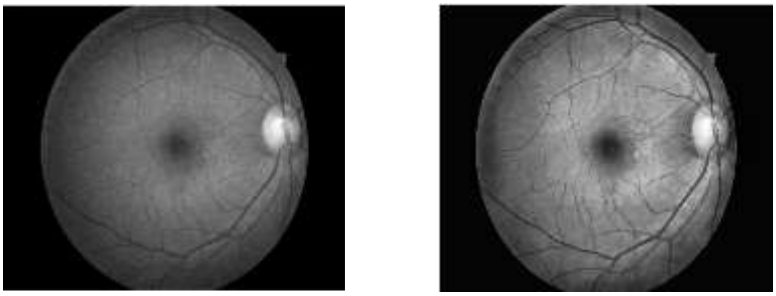
3.1. Pre-processing and vessel segmentation

We use Python3.6 to transfer RGB images to gray images shown in Fig. 12. Then Image in Fig12. (b) is processed by contrast-limited adaptive histogram equalization (CLAHE) and gamma adjustment. The results are shown in Figure 13 and Figure 14.



(a) (b)

Fig.12 (a) Original RGBimage. (b) Gray image.



(a) (b)

Fig.13 (a) Original grey image. (b) Grey image.

Fig.14 Vessel segmentation.

4. Conclusion

In this paper, a new convolutional network architecture was proposed for retinal image vessel segmentation. It achieved a better outcome in the DRIVE database and performed better than a skilled ophthalmologist. From Table 2 and comparing to the different methods for image vessel segmentation, the accuracy of the proposed method in this paper on DRIVE is 0.9790. That is to say that our method is on the top of these compared methods.

While in the practical fundus images, the results is not as good as DRIVE database. The reason is that the images photographed by ophthalmologist would have some noises. If the preprocessing to remove the noise is performed based on the raw images, we could guarantee the results could be much better.

Acknowledgement

This work is supported by National Natural Science Foundation of Innovative Research Groups Science Foundation of China (51221004), ARC DECRA and ARC Discovery projects (DE130100911, DP130101327), the NSFC funding (61332013), the International Science and Technology Cooperation Projects (No. 2016D10008, 2013DFG12810, 2013C24027), the Municipal Natural Science Foundation of Ningbo (No.2015A610119), the Guangzhou Science and Technology Project under Grants (2016201604030034), the Major Projects of Natural Science Research in Jiangsu Higher Education Institutions (NO. 14KJA520001), Jiangsu Production and Research Project (NO. BY2015010-05), the Natural Science Foundation of Zhejiang Province (NO. Y16F020002), the Ningbo Natural Science Fund (NO. 2015A610119), International Science & Technology Cooperation Projects of Ningbo (NO. 2016D10008) and the project of research and development of intelligent resource allocation and sharing platform for marine electronic information industry (NO. 2017GY116).

Reference

[1] Preethi, M., and R. Vanithamani. "Review of retinal blood vessel detection methods for automated diagnosis of Diabetic Retinopathy." International Conference on Advances in Engineering, Science and Management IEEE, 2012:262-265..

[2] Dhara, Ashis K., et al. "Methods for the detection of blood vessels in retinal fundus images and reduction of false-positive pixels around the optic nerve head." E-Health and Bioengineering Conference IEEE, 2014:1-6.

[3] Chutatape, O, L. Zheng, and S. M. Krishnan. "Retinal blood vessel detection and tracking by matched Gaussian and Kalman filters." Engineering in Medicine and Biology Society, 1998. Proceedings of the, International Conference of the IEEE IEEE, 1998:3144-3149 vol.6.

[4] Fu, Huazhu, et al. "Retinal vessel segmentation via deep learning network and fully-connected conditional random fields." IEEE, International Symposium on Biomedical Imaging IEEE, 2016:698-701.

[5] Fu, Huazhu, et al. DeepVessel: Retinal Vessel Segmentation via Deep Learning and Conditional Random Field. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. 2016:132-139.

[6] Ronneberger, Olaf, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention Springer, Cham, 2015:234-241.

[7] C. Sinthanayothin, J. F. Boyce, T. H. Williamson, H. L. Cook, E. Mensah, S. Lal, and D. Usher, “Automated detection of diabetic retinopathy on digital fundus images,” Diabetic Medicine, vol. 19, no. 2, pp. 105–112, 2002.

[8] G. Gardner, D. Keating, T. Williamson, and A. Elliott, “Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool.,” The British Journal of Ophthalmology, vol. 80, pp. 940–944, 1996.

[9] K. Malathi and R. Nedunchelian, “Detecting and classifying diabetic retinopathy in fundus retina images using artificial neural networks-based firefly clustering algorithm,” ARPN Journal of Engineering and Applied Sciences, vol. 11, no. 5 , pp. 3419–3426, 2016.

[10] Y. Bengio, “Learning deep architectures for AI,” Foundations and trends in Machine Learning, vol. 2, no. 1, pp. 1–127, 2009

[11]S. Srinivas, R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. Kruthiventi, and R. V. Babu, “A taxonomy of deep convolutional neural nets for computer vision,” arXiv preprint arXiv:1601.06615, 2016.

[12] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A.van der Laak, B. van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” arXiv preprint arXiv:1702.05747, 2017.

[13]S. Wang, Y. Yin, G. Cao, B. Wei, Y. Zheng, and G. Yang, “Hierarchical retinal blood vessel segmentation based on feature and ensemble learning,” Neurocomputing, vol. 149, pp. 708–717, 2015.

[14]M. v. Grinsven, B. v. Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sanchez, “Fast convolutional neural network training using selective data sampling: Application to haemorrhage detection in color fundus images,” IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1273 – 1284, 2016.

[15]P. Liu, H. Zhang, and K. B. Eom, “Active Deep Learning for Classification of Hyperspectral Images,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 99, 2016.

[16]W. Zhao and S. Du, “Spectral;Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach,” IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 8, 2016.
[17]A. Romero, C. Gatta, and G. Camps-Valls, “Unsupervised Deep Feature Extraction for Remote Sensing Image Classification,” IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 3, 2016.
[18]X. Ma, J. Geng, and H. Wang, “Hyperspectral image classification via contextual deep learning,” EURASIP Journal on Image and Video Processing, vol. 2015, no. 1 , pp. 20 – 28, 2015.
[19]Y. Chen, X. Zhao, and X. Jia, “Spectral-Spatial Classification of Hyperspectral Data Based on Deep Belief Network,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 6, 2015.
[20] Oyster C W. The Human Eye: Structure and Function[J]. Quarterly Review of Biology, 1999.
[21] Jogi, Basic Ophthalmology. New Delhi, India: Jaypee Brothers Medical Publishers, 2008
[22] “Figure 3 - Normal Fundus Image.” http://www.rpfightingblindness.org. uk/index.php?pageid=201&tln=aboutrp. Accessed: 2017-01-14.
[23]Cheung N, Huynh S, Wang J J, et al. Relationships of retinal vessel diameters with optic disc, macular and retinal nerve fiber layer parameters in 6-year-old children[J]. Investigative Ophthalmology & Visual Science, 2008, 49(6):2403-2408.
[24] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, 2016.
[25]D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3D convolutional networks. In ICCV, 2015.
[26] A. Dundar, J. Jin, and E. Culurciello. Convolutional clustering for unsupervised learning. In ICLR workshops, 2016.
[27] F. Liu, C. Shen, and G. Lin. Deep convolutional neural fields for depth estimation fom a single image. In CVPR, 2015.
[28] C. Zzegedy, A. Toshev, and D. Erhan. Deep neural networks for object detection. In NIPS, 2013.
[29]F. Rosenblatt. The perceptron–a perceiving and recognizing automaton. Technical Report 85-460-1, Cornell Aeronautical Laboratory, 1957.
[30] F. Rosenblatt. Principles of Neurodynamics. Spartan Books, Washington, DC, USA, 1962.
[31]N. Buduma and N. Locascio. Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms. O’Reilly Media, 2017
[32] G. E. Hinton, S. Osindero, and Y-W. Teh. A fast learning algorithm for deep belief nets. Neural Computation, 18(7):1527–1554, 2006.
[33] Y. LeCun, K. Kavukcuoglu, and C. Farabet. Convolutional networks and applications in vision. In ISCAS, 2010.
[34] 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.
[35] Azzopardi et al., “Trainable cosfire filters for vessel delineation with application to retinal images,” Medical image analysis, vol. 19, no. 1, pp. 46–57, 2015.
[36] Osareh et al., “Automatic blood vessel segmentation in color images of retina,” Iran. J. Sci. Technol. Trans. B: Engineering, vol. 33, no. B2, pp. 191–206, 2009.
[37] Roychowdhury et al., “Blood vessel segmentation of fundus images by major vessel extraction and subimage classification,” Biomedical and Health Informatics, IEEE Journal of, vol. 19, no. 3, pp. 1118–1128, 2015.
[38] Fraz et al., "An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation", IEEE Transactions on Biomedical Engineering, vol. 59, no. 9, pp. 2538-2548, 2012.
[39] Qiaoliang et al., "A Cross-Modality Learning Approach for Vessel Segmentation in Retinal Images", IEEE Transactions on Medical Imaging, vol. 35, no. 1, pp. 109-118, 2016.
[40]Melinscak et al., "Retinal vessel segmentation using deep neural networks", In Proceedings of the 10th International Conference on Computer Vision Theory and Applications (VISIGRAPP 2015), (2015), pp. 577–582.
[41] Liskowski et al., "Segmenting Retinal Blood Vessels with Deep Neural Networks", IEEE Transactions on Medical Imaging, vol. PP, no. 99, pp. 1-1, 2016.