Research on Assistant Diagnosis of Fundus Retinopathy Based on Deep Learning

Haoyu Li, Ziwen Yuan, Kun Zhang
Northeastern University at Qinhuangdao, Qinhuangdao, Hebei, 066004, China

Abstract. Macular edema has three types of lesions: REA, PED and SRF. Early detection of edema areas can play a key role in the treatment of diseases. Neural network is a powerful tool for image processing in medical field. Deep learning automatically finds features that are ideal for “AI+Medical Imaging” diagnostics. This paper mainly proposes a new method of neural network that includes Object recognition and classification to improve the accuracy and speed of detection of macular edema area. The method is offered to be evaluated and compared to the traditional network method. The results indicate that the method applied reducing the over segmentation effect and getting a more accurate result of the option than traditional network method.

1. Introduction
This paper mainly studies the segmentation and detection of three regions of macular edema (PED, SRF, REA). It adopts the current advanced U-net neural network structure. On the basis of U-net, it uses many short connections and long connections in the horizontal connection of network structure. The network structure cooperates with deep supervision, which makes the deep network with huge parameters greatly reduce the parameters in an acceptable range of accuracy, so it is easy to train. The experimental results show that the structure of U-net neural network based on deep learning algorithm can segment and recognize the edema lesion area of fundus in medical images. The overall framework of target recognition is shown in Figure 1. This research has important practical application value for the recognition of edema lesion area of fundus.

2. Background
With the application of electronic equipment becoming more and more popular in social production and people's life, digital image has become an indispensable information medium, generating huge amounts of image data every moment [1]. Image segmentation is an important and difficult task in image analysis technology. The goal of segmentation is to divide an image into a set of disjoint regions object recognition and classification relies excessively on the result of the segmentation process [2]. At the same time, it is becoming more and more important to recognize the target accurately in the image. We not only pay attention to the simple classification of images, but also hope to accurately obtain the interested objects and their locations in images. Target detection is an important prerequisite for many high-level visual processing and analysis tasks. In the area of lesion detection of medical images, target detection has great practical value and application prospects [3].

Since 2006, deep learning has emerged as a branch of machine learning. Deep learning refers to a multi-layer neural network that can continue to implement extremely complex patterns [4]. It is difficult
for other algorithms to model the middle representation of data by using hidden layer between input layer and output layer. Compared with traditional machine learning, the main advantage of neural network lies in its ability to surpass almost all other machine learning algorithms, which makes deep learning have great potential in medical image analysis [5]. Macular edema is an eye disease that can cause vision loss and affect normal life in severe cases. Early detection of edema symptoms can play a key role in the treatment of diseases. The disease has three lesions: retinal pigment epithelial detachment (PED), subretinal edema (SRF) and retinal edema area (REA) [6]. At present, in the treatment of fundus edema, the rate of misdiagnosis is very high. In this paper, a new assistant method is proposed, which uses neural network intelligent system to assist doctor's diagnosis, to reduce the rate of misdiagnosis.

3. U-net and u-net++

3.1. Convolutional Networks for Biomedical Image Segmentation(U-net)

U-net is improved on the basis of FCNs, which solves the problem that selecting different patches in the segmented image will lead to great uncertainty of the results [7]. Unet architecture (figure 1) is a contracting path and an expanding path. Contracting path increases the resolution of the output as the depth increases, while expanding path gradually restores details and image accuracy with skip-connection from contracting path. Unlike the FCNs network, U-Net has a large number of feature Channels in contracting path, so the network can propagate context information to higher resolution layers, so that it can segment large images and medical images with better results. Expanding path is a typical convolutional network architecture. Each expanding path doubles the number of feature channels. Each step of expanding path uses deconvolution first, and each deconvolution reduces the number of feature Channels by half, doubling the size of feature map. After deconvolution, the result of deconvolution is joined with the feature map of corresponding steps in expanding path through skip-connection, which can extract spatial information better.

![Figure 1. UNet Architecture.](image)

3.2. A Nested U-Net Architecture for Medical Image Segmentation(U-net++)

U-net++ adds deconvolution and re-designed skip pathways to each layer on the basis of u-net, and forms a network topology through long and short links [8] (figure 2). Through multi-level U-net sub-network, each node is composed of a convolution layer and adjacent pooling layer. We can extract
features from different perspectives, which makes the network more robust. Another advantage of U-net++ is deep supervision. In each graph, $X_{i,1}$, $X_{i,2}$, $X_{i,3}$ is followed by a $1 \times 1$ convolution core, which is equivalent to the output of U-NET to supervise each level, or to segment branch, so as to ensure that the information of each layer is preserved. Compared with U-net, although U-net++ increases the parameters by 20%, it also improves the accuracy greatly.

In order to solve this problem, we use four layers of U-net++ Architecture with a parameter of 2.98M, which has achieved a good improvement in performance and accuracy.

![Figure 2. UNet++ consists of an encoder and decoder that are connected through a series of nested dense convolutional blocks.](image)

4. Experiments

4.1. Datasets

As shown in figure 3, we use the real data set provided by AI CHALLENGE, including 100 groups of OCT about retinal sections. We used 70 groups of pictures as training set, 15 groups as verification set and 15 groups as test set. The size of each picture is 1285121024 pixels. Some images of the data set are shown in the figure. OCT images include three types of edematous lesions: retinal pigment epithelial detachment (PED), subretinal edema (SRF) and retinal edema area (REA). In order to achieve the best effect of the network, we superimpose every three pictures in the data set, then input them into the network and normalize them. At the same time, the image is reversed horizontally to increase the size of the data set.
4.2. Optimization and Improvement

In the aspect of image enhancement (preprocessing), for the two-dimensional image hierarchical file given by the data set, we get three of his front and back images as three channels, which are combined into RGB images to make greater use of spatial information. We tested the composite loss function on the basis of the original, compared with the original single loss function, we got a good improvement.

Using composite loss functions can combine the advantages of various loss functions and reduce the influence of boundary fluctuations which have a sensitive loss. Compared with BCE only, it has made great progress in convergence speed and accuracy. We tested several loss functions such as BCE, WCE+Dice, WCE+Dice+BCE, and finally concluded that UNET++ achieved the highest accuracy when using WCE+ELDice+BCE as loss, which was higher than using a single loss.

Table 1. Accuracy and dice under different loss functions.

| Model   | Multi-Task | Params | Loss           | Val_Dice | Val_Auc  | Test_Dice | Test_Auc |
|---------|------------|--------|----------------|----------|----------|-----------|----------|
| UNet    | No         | 2.47 M | WCE+Dice       | 0.772    | -        | 0.683     | -        |
| UNet    | Yes        | 2.47 M | WCE+Dice+BCE   | 0.785 (+0.013) | 0.985 (+0.014) | - | - |
| UNet++  | Yes        | 2.95 M | WCE+Dice       | 0.781 (+0.009) | 0.985 (+0.014) | 0.701 (+0.018) | 0.968 (+0.064) |
| UNet++  | Yes        | 2.95 M | WCE+Dice+BCE   | 0.784 (+0.012) | 0.986 (+0.015) | - | - |
| UNet++  | Yes        | 2.95 M | WCE+ELDice+BCE | 0.799 (+0.027) | 0.989 (+0.018) | 0.736 (+0.053) | 0.983 (+0.079) |

4.3. Advantages and disadvantages
Assistant diagnosis can extract deeper environmental information better, and has higher accuracy in the segmentation of lesion areas.

Training network needs certain medical data set, which requires large amount of data and high calculation cost.

5. Conclusions
For the problems presented in this paper, as shown in Table 1, using U-net++ with compound loss function can improve the accuracy to 98.3%, 8.74% compared with RESNET (90.4%) and 1.55% compared with U-net ++network (96.8%) using BCE loss only. The results are quite remarkable.

Table 2. Progress of ResNet18, U-net and U-net++ under composite loss function.

| Model       | Multi-Task | Param | Loss     | Val_Dice | Val_Auc | Test_Dice | Test_Auc |
|-------------|------------|-------|----------|----------|---------|-----------|----------|
| ResNet18(pre) | No         | 11.18 | BCE      | 0.971    | -       | -         | 0.904    |
| UNet        | No         | 2.47  | WCE+Dice | 0.772    | -       | 0.683     | -        |
| UNet        | Yes        | 2.47  | WCE+Dice+BCE | 0.785 (+0.013) | 0.985 (+0.014) | -         | -        |
| UNet++      | Yes        | 2.95  | BCE      | 0.776 (+0.004) | 0.979 (+0.008) | 0.699 (+0.016) | 0.947 (+0.043) |
| UNet++      | Yes        | 2.95  | WCE+Dice | 0.781 (+0.009) | 0.985 (+0.014) | 0.701 (+0.018) | 0.968 (+0.064) |
| UNet++      | Yes        | 2.95  | WCE+Dice+BCE | 0.784 (+0.012) | 0.986 (+0.015) | -         | -        |
| UNet++      | Yes        | 2.95  | WCE+ELDice+BCE | 0.799 (+0.027) | 0.989 (+0.018) | 0.736 (+0.053) | 0.983 (+0.079) |

The dice of single model on the test set is 0.736. The dice of fusion model on the test set is 0.744 and the detection AUC is 0.986. In addition, the memory of inference stage is 7.3G when the set batch is 8 and the inference time is 9.5s per patient.

References
[1] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman. The pascal visual object classes (VOC) challenge[J]. Int. J. Comput. Vis., 2010.
[2] Ronneberger O, Fischer P, Brox T . U-Net: convolutional networks for biomedical image segmentation[C]International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015.
[3] Zhou Z, Siddiquee M M R, Tajbakhsh N, et al. UNet++: A nested u-net architecture for medical image segmentation[J]. 2018.
[4] Zhu Jiangbing, Ke Xin, Liu Chang, Yang Jinkui, He Jianguo. Diagnostic system for diabetic retinopathy based on computer vision[J]. Journal of Capital Medical University. 2015(06): 28-32.
[5] Lu Peilin, Zou Fang, Li Dan, Zhu Zhenhua, Zhang Wei, Wang Wei. Development and application
of diagnostic imaging system for diabetic retinopathy based on remote consultation[J]. China Digital Medicine. 2015(07): 68-72.

[6] Chen Xiang. Automated detection algorithm for exudate in diabetic retinopathy images[D]. Harbin Institute of Technology, 2012.

[7] Tian Man, Liu Wei. Progress in diagnosis and treatment of diabetic retinopathy[J]. Journal of Liaoning University of Traditional Chinese Medicine. 2018(11): 122-128.

[8] Pang Hao, Wang Wei. Deep learning model for the detection of diabetic retinopathy[J]. Journal of Software. 2017(11): 89.