Ultrasonic Image Segmentation Method based on Improved Fully Convolution Network

Hai Ye[1, a], Kaiping Feng[2, b], Lihong Luo[2, c] and Hongning Xie[3, d]

[1]School of Computer Science, Guangdong University of Technology, Guangzhou Guangdong 510006, China;

[2] School of Art and Design, Guangdong University of Technology, Guangzhou Guangdong 510090, China;

[3] Department of Ultrasonic Medicine, Fetal Medical Center, First Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, 510080, China

[a] yehai_work@163.com, [b] fengkp@gdut.edu.cn, [c]453072195@qq.com, [d] hongning_x@126.com

Abstract. The clinical fetal brain ultrasound image contains a lot of noise, which is impact the classification or recognition results in deep learning tasks. Therefore, proposed an improved fully convolutional network which can automatically eliminate noise and extract effective area in fetal brain ultrasound images. By adding dilated convolution, U-net obtained a larger receptive field. The original U-net neural network was compared with the method by adding dilated convolution one. The experiment of test data shows that the improved segmentation model was effective and robust, and performed better in precise rate, recall rate, f1-score and DICE coefficient. This improved method can be wildly used in other medical image segmentation tasks.

Keywords: Computer vision; Deep learning; Ultrasound image; Image segmentation.

1. Introduction

In recent years, fetal ultrasound image processing has had some published work. Yu Z, Chen H et al. proposed a method to identify standard fetal ultrasound section based on deep learning [1, 2], which is used to assist doctors as a tool for clinical diagnosis. Baumgartner C F et al. implemented real-time detection of standard section using deep learning method [3]. As an important step in clinical diagnosis, standard section detection has been increasingly studied and achieved in the field of deep learning. At present, how to use deep learning methods to diagnose fetal diseases will become the focus and difficulty of the next research.

There is a large amount of noise in the fetal ultrasound image, which is not conducive to the classification and diagnosis of abnormal diseases. Different from general image classification, the essence of medical image recognition is fine-grained image classification, which requires segmentation of lesion areas with high similarity. At present, there are few related researches on the segmentation of such fine-grained images of fetal brain regions. In this paper, an image segmentation method based on full convolution is proposed to segment the skull halo, and the cavity convolution is added for improvement. This segmentation method can also be easily extended to other medical image segmentation fields as a means to eliminate noise and extract target areas.

2. Method

2.1 Data Set and Image Preprocessing

In the experiment, cranial rings of each ultrasound image were artificially labeled by fitting an elliptical region, and the original image size, the coordinates of the central point of the fitting ellipse, the length of the long and short axes, and the Angle were saved in an XML file. For each ultrasonic image, a mask image can be generated through the corresponding XML file. The mask image is a binary image with the same size as the original image. The pixel value inside the skull halo area is 255, while the pixel value in the non-skull halo area is 0. Each original image has a corresponding mask image as the real tag of model training.

Before getting into the model, all the original image from the color image into single channel of three-channel gray image, and for each image for a z - score of standardized operation, namely each image minus all picture means and divided by their standard, there will be data distribution difference image into a unified level, to ensure the comparability of the data set, to improve the stability of the network. The data set standardized by z-score will be resized to 256 by 256 as the input of the model. Finally, all data are divided into training set and test set.

2.2 U-net Model

U-net [4] is an improved model of full convolutional network (FCN) [5], and its network model structure is shown in Fig 1. It takes a 256 by 256 gray image as input and a 256 by 256 feature image as output. The whole U-net architecture can be divided into two parts: the feature extraction process with down sampling and the feature expansion process with up sampling. In the process of down sampling, each feature map passing through two convolutions which has kernel size of 3 x 3, stride and padding were set as 1, after the convolution kernels with a Rectified Linear Unit (Relu) as the activation function, then after a 2 x 2 down sampling. In the down sampling process, the number and size of each set of feature images will be doubled and the original size will be halved after the above operation. In the process of up sampling, the feature map with size of 8 x 8 are up sampled by transposed convolution [6], moreover after each up-sampling, the of feature maps will be concatenate to corresponding feature maps that has same size, to mixed information before sampling again, after repeated five times end up with a single channel image with an original image size, the output of U-net model can be obtained.

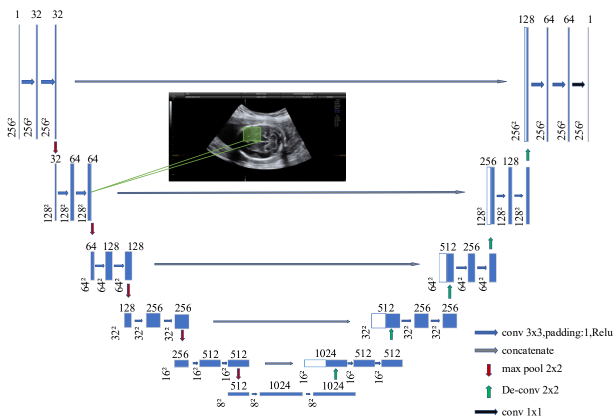
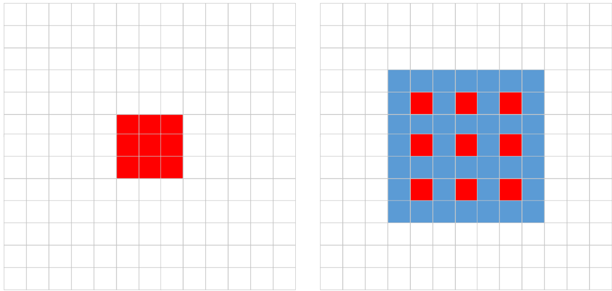


Fig. 1 U-net structure

Compared with the soft-max activation function and the cross-entropy loss function commonly used in the network output layer, the sigmoid activation function was used in this experiment to normalize the characteristic graph to 0 to 1, and the Dice coefficient was used as the loss function after the sigmoid activation function.

2.3 Dilated Convolution

There is an obvious defect in most DCNN that sampling is a process that cannot be learned. After sampling (up or down) the feature graph, the original internal data structure and spatial hierarchy information of the feature map will be lost, which means that the pixels in a group of small areas cannot be reconstructed by the reverse sampling operation. Intuitively speaking, the information lost by the U-net model in the image pooling operation for several times during the down-sampling process, and it cannot be found back in the up-sampling process, and the segmentation of the skull ring area requires not only the local image features such as texture, edge and color, but also the information of the overall brain structure and the position relations between different organs. In order to extract the global information, we improved the original of the U - net by adding dilated convolution [7], Dilated convolution can effectively improve reception field without reduce the image size, which can avoid information loss in pooling operation.



(a) Conv 3X3 dilate=1 (b) Conv 3X3 dilate=2

Fig. 2 Receptive field of dilated convolution

The illustration of dilated convolution is shown in Fig. 2. Dilated convolution separates continuous pixels in the convolution to expand the sensing field. In Fig. 2 (a), the size of the convolution kernel is 3x3, and the dilated rate is 1, the dilated convolution (Conv 3x3, dilation=1) is same as the common convolution operation, and the receptive field is 3x3. In Fig. 2 (b), the dilated convolution with an expansion rate of 2 only convolved the red pixel. The actual size of the convolution kernel is still 3x3, but the receptive field has increased to 7x7 like the blue region. In other words, the feature map with dilated convolution can be "seen" more than the non- dilated layer. For the initial convolutional layer that needs to deal with large-size feature maps, dilated convolution can effectively ignore some details and pay more attention to the acquisition of overall structure information.

2.4 Training and Predict Method

The mini-batch Stochastic Gradient Decent (SGD) algorithm with a batch size of 64 and Adam optimizer with an initial learning rate of 0.0001, beta1, beta2 and 0.999, respectively, were used to optimize model parameters in the training. In the process of experiment, the epoch was set as 80.

3. Experiment

3.1 Segmentation Result

In this paper, Dice coefficient, precision rate, recall rate and their harmonic average f1-score were used to evaluate the performance of the original model and the model with the addition of dilated convolution. In other words, the segmentation results of the two models were quantitatively analyzed for a given test data set.

The original U-net completed by training and the U-net using dilated convolution were tested with 2000 pieces of test data, and the above evaluation criteria were calculated respectively, and the average value of each evaluation criterion was taken. The final quantitative analysis and comparison results are shown in Table 1.

Table 1. Comparing of experiment result

Method	Precision	Recall	F1-score	DICE
Original convolution	0.883	0.985	0.931	0.930
Dilated convolution	0.979	0.909	0.943	0.941

As can be seen from Table 1, the improved segmentation method after dilated convolution has a higher precision rate than the original method, which indicates that the improved method will segment more precisely and exclude redundant information out of the segmentation region as much as possible to eliminate more noises. However, the decrease of recall rate also means that the model may over-segment some effective areas, such as the skull area of skull halo in this paper. In this experiment, the oval area was expanded as far as possible to the peripheral area of the skull halo during labeling, so as to facilitate the subsequent identification of diseases such as abnormal head type. Low recall rate had little negative impact on the diagnosis of brain diseases. On the whole, the f1-score and Dice coefficients increased by 1.2% and 1.1% respectively. The experiment proved that the addition of dilated convolution had a good effect on the u-net segmentation results.

3.2 Segmentation Result

The experimental segmentation results are shown in Fig. 3. The left image is the original image, and the blue oval area in the right image is the ellipse fitted by the segmentation result. The fitting ellipse area is made into an external rectangular box, and the segmentation image is extracted to obtain the desired target segmentation image.

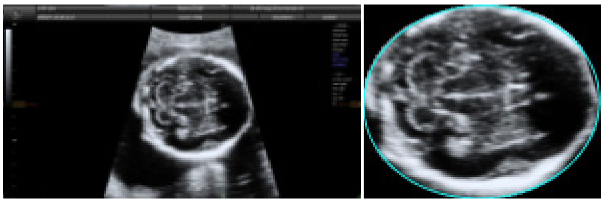


Fig. 3 Segmentation result

4. Conclusion

Proposes a U-net with dilated convolution segmentation model for automatic extraction, By using image standardization method with the ultrasonic image data set standardization, we train a model using labeled ultrasonic images. After testing the data set of 2000 fetal brain segmentation test, the u-net network model with dilated convolution improved the accuracy by 9.6% compared with the original network model, the recall rate decreased by 7.6%, the f1-score increased by 1.2%, and the Dice coefficient increased by 1.1%, which was better than the original u-net network. There was no significant difference in the time consumption, and the u-net network model could effectively improve the segmentation accuracy. This method can be easily extended to other medical image segmentation tasks, providing a reliable and effective method for removing medical image noise and improving detection and classification accuracy.

References

[1]. Yu Z, Ni D, Chen S, et al. Fetal facial standard plane recognition via very deep convolutional networks[C]// Engineering in Medicine & Biology Society. IEEE, 2016.

[2]. Chen H, Wu L, Dou Q, et al. Ultrasound Standard Plane Detection Using a Composite Neural Network Framework[J]. IEEE Transactions on Cybernetics, 2017:1-11.

[3]. Baumgartner C F, Kamnitsas K, Matthew J, et al. SonoNet: Real-Time Detection and Localisation of Fetal Standard Scan Planes in Freehand Ultrasound[J]. IEEE Transactions on Medical Imaging, 2017:1-1.

[4]. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]// International Conference on Medical Image Computing & Computer-assisted Intervention. 2015.

[5]. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4):640-651.

[6]. Zeiler M D, Krishnan D, Taylor G W, et al. Deconvolutional networks[C]// Computer Vision & Pattern Recognition. 2010.

[7]. Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions[C]// International Conference on Representation Learning. 2016.