Automatic segmentation in fetal ultrasound images based on improved U-net

Yujie Yang*, Pinli Yang, Bo Zhang
School of Computer Science, Sichuan University, Chengdu, 610000, China
hxcszhangbo@163.com
*2017141463180@stu.scu.edu.cn

Abstract. As an effective way of routine prenatal diagnosis, ultrasound (US) imaging has been widely used in clinical practice. Biosignatures obtained from fetal segmentation contribute to fetal development and health monitoring. However, the artifacts, speckle noises, quality of imaging equipment and other factors make the segmentation of fetal US images extremely challenging. In this paper, aiming to improve the depth of the model, as well as to avoid the vanishing gradient problem and exploding gradient problem, we propose Residual U-net and ASPP U-net based on U-net, which further improves the accuracy of segmentation without increasing the depth of the model. The results of our experiments show that the network proposed in the paper can effectively improve the segmentation accuracy in fetal US images.

1. Introduction
The main purpose of prenatal diagnosis is to provide families with information about foreseeable anomalies and therapeutic alternatives for any anomalies detected [1]. Due to its painlessness and instantaneity, ultrasound (US) imaging has become a common approach in prenatal diagnosis [2]. It is well accepted that biosignatures obtained from fetal segmentation contribute to fetal development and health monitoring. However, the quality of US images is greatly affected by the imaging equipment and the expertise of doctors. At the same time, artifacts and speckle noises bring about the fuzzy boundary of the region of interest (ROI) in US images, causing the low contrast with the background area [3]. All these factors make the segmentation of fetal US images more challenging. There have been multiple models for image segmentation based on machine learning, such as active shape model (ASM), constrained local model (CLM) and active appearance model (AAM) [12, 13]. With the development of deep learning, lately, U-net is one of the most well-known architectures for medical image segmentation due to its simple structure and has achieved remarkable successes [4]. Nonetheless, as the result of fewer convolution layers, U-net is insufficient to extract the high-level features from input images adequately. In this paper, we propose Residual U-net and ASPP U-net inspired by U-net to address this challenging problem.

The contributions of our work are as follows.1) To improve the depth of model and effectively alleviate the vanishing gradient problem and exploding gradient problem, two new residual blocks are applied in Residual U-net as the basic components of contracting path and expansive path respectively. 2) To further improve the performance of segmentation without increasing the depth of the model, we introduce a new structure named as atrous spatial pyramid pooling (ASPP) based on Residual U-net which can extract multi-scale features from the input images by atrous convolutions with different
sampling rates [5]. Moreover, we propose two new inception-like blocks as the basic components of contracting path and expansive path for ASPP U-net.

2. Methodology

2.1. Residual U-net

The network architecture of Residual U-net is illustrated in Fig. 1, which employs the U-net as backbone [7] and takes the US slices as input and outputs the segmented images. It consists of a contracting path (left side) and an expansive path (right side). To be specific, the contracting path consists of consecutive block1s which we will discuss below and each of them is followed by a 2x2 max-pooling operation. At each max-pooling step, the number of feature channels will be doubled. In the expansive path, feature maps correspondingly copied from the contracting path is concatenated with the upsampled feature maps. After that, the concatenated feature maps are passed into block2s followed by a 2x2 up-convolution which halves the number of feature channels. At the final layer, a convolution operation with stride 1x1 is utilized to the final output of block2s and the segmented images are generated.

![Fig.1 Network architecture of Residual U-net](image)

As shown in Figure. 2, compared with the residual block in ResNet-101 [6], the block2 utilizes group convolution [14] and group normalization (GN) [8] instead of conventional convolution and batch normalization (BN), which aim to not only reduce the number of parameters but also solve the problem of segmentation accuracy degradation caused by the small batch size. Additionally, the main innovation of the block1 is that it adds two 3x3 convolution layers in comparison with block2.
2.2. ASPP U-net

2.2.1. ASPP

Atrous Spatial Pyramid Pooling (ASPP) is a basic unit of DeepLab which can capture objects as well as image context at multiple scales [5]. Atrous convolution allows us to enlarge the receptive field without the loss of spatial resolution of feature maps and control the receptive field explicitly [5]. As can be seen from Figure 3, aiming to compute the feature maps more effectively, ASPP proposed in this paper consists of multiple parallel atrous convolution layers with different sampling rates (1, 6, 12, 18 from left to right). In ASPP, the input feature maps are fed into 4 parallel atrous convolution layers shown below. Then, the feature maps extracted from the 4 atrous convolution layers are fused to generate the final result.

2.2.2. Basic modules of ASPP U-net

The inception architecture is firstly proposed in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14), whose main hallmark is the improved utilization of the computing resources inside the network [9]. In this paper, inspired by the success of inception architecture and ASPP, we utilized two inception-like blocks (block3 and block4) instead of residual blocks (block1 and block2) in Residual U-net, which further improve the ability to extract features from multiple scales. As shown in Fig. 4, the input feature maps are divided into 4 parts and each part is processed in separate branches. Besides, compared with the residual block in Residual U-net, inception-like blocks contain 3 parallel atrous convolution layers with different sampling rates.
3. Experiments and results

3.1. Dataset and data pre-processing
In this paper, we conducted our experiments on three datasets acquired from West China Hospital, Sichuan University, including fetal abdominal circumference (AC), fetal femur length (FL) and fetal crown-rump length (CRL). Concretely, 2081 US images are selected as training set, 450 as validation set, 727 as test set. Especially, the gray-scale values of pixels in ground-truth are set to 0 or 255 according to the target area manually annotated by the experts. Additionally, in order to decrease the risk of overfitting and improve the robustness, we enlarged the training set with data augmentation implemented by flipping, rotation and shift.

3.2. Implementation details
We implemented our networks with PyTorch and trained them on the NVIDIA GeForce GTX 1080Ti with 11GB memory. Specifically, we introduced kaiming initialization [10] to initialize the parameters and stochastic gradient descent (SGD) [11] optimization with a high momentum (0.99) to optimize the networks. The proposed networks were trained 100 epochs with an initial learning rate of 1.0e-9 and the batch size was experimentally set to 4. Furthermore, to solve the problem of overfitting, the weight decay was set to 0.0005. Notably, we set different numbers of iterations for different datasets of fetal US images (40000 for AC and CRL, 20000 for FL).

3.3. Evaluation criteria
In this paper, in order to evaluate and analyze the experimental results quantitively, we adopt three evaluation metrics, including Dice, PM and CR, which are widely utilized in the evaluation of image segmentation networks and can be expressed as Equation 1-3.

\[
\text{Dice}(GT, SEG) = \frac{2 \times |GT \cap SEG|}{|GT| + |SEG|} \tag{1}
\]
\[ PM = \frac{TP_s}{GT} \times 100\% \]  \hspace{1cm} (2)

\[ CR = \frac{TP_s - 0.5 \times FP_s}{GT} \times 100\% \]  \hspace{1cm} (3)

where ground truth (GT) represents the target areas manually segmented by experts and SEG represents the target areas segmented by networks proposed in this paper. TPs denotes the areas which match the ground truth, while FPs denotes the areas segmented falsely. It is obvious that the Dice, PM and CR are directly proportional to the accuracy of segmentation.

### 3.4. Experiment results

Aiming to reflect the advantages of our networks intuitively, we visualized the outputs of the three networks on the three datasets, which are shown in Fig. 5. As observed, there are some subtle visible connected regions segmented falsely in the segmentation results of U-net while Residual U-net and ASPP U-net perform better, which may due to the increased number of convolution layers and the introduced residual blocks and ASPP.

![Fig. 5 Qualitative comparison of different models on AC, FL and CRL](image)

As can be seen from Table 1, obviously, in comparison of U-net, networks proposed in this paper can effectively improve the segmentation accuracy of fetal US images. For instance, on the AC dataset, the Residual U-net performs better with a great improvement from 0.8985 to 0.9241 in Dice compared with U-net, while ASPP U-net further increases to 0.9412 by the usage of parallel atrous convolution layers with different sampling rates. In conclusion, the methods proposed in this paper are beneficial to the segmentation performance improvement.

| Dataset | Network      | Dice   | PM      | CR    |
|---------|--------------|--------|---------|-------|
| AC      | U-net        | 0.8985 | 0.8416  | 0.8351|
|         | Residual U-net| 0.9241 | 0.9243  | 0.8709|
|         | ASPP U-net   | 0.9412 | 0.9254  | 0.9011|
| FL      | U-net        | 0.7790 | 0.6739  | 0.6494|
|         | Residual U-net| 0.7963 | 0.7008  | 0.6730|
|         | ASPP U-net   | 0.8092 | 0.7123  | 0.6902|
| CRL     | U-net        | 0.9169 | 0.8648  | 0.8520|
|         | Residual U-net| 0.9203 | 0.8656  | 0.8568|
|         | ASPP U-net   | 0.9180 | 0.8742  | 0.8568|
4. Conclusion

In the prenatal diagnosis, automatic and accurate segmentation of fetal US images is still challenging owing to a variety of interference factors. In this paper, to further improve the segmentation accuracy in fetal US images, we proposed Residual U-net and ASPP U-net inspired by ASPP and U-net. Different from U-net, we introduced two new residual blocks as the basic component of contracting path and expansive path in Residual U-net. Additionally, in ASPP U-net, we utilized two new Inception-like blocks instead of residual blocks in Residual U-net, which improves the ability to extract features from different scales. The effectiveness of our networks is extensively verified by conducting comparison experiments on the datasets. In the near future, we will further optimize our models and further improve the segmentation performance so that our model can be widely applied in clinical practice.

References

[1] César Martín Martínez, Darnell A, Escofet C, et al. Fetal magnetic resonance imaging[J]. The Ultrasound Review of Obstetrics and Gynecology, 2004.
[2] Rueda S, Fathima S, Knight C L, et al. Evaluation and Comparison of Current Fetal Ultrasound Image Segmentation Methods for Biometric Measurements: A Grand Challenge[J]. IEEE TRANSACTIONS ON MEDICAL IMAGING MI, 2014.
[3] Xu L, Liu M, Shen Z, et al. DW-Net: A Cascaded Convolutional Neural Network for apical four-chamber view segmentation in fetal echocardiography[J]. Computerized Medical Imaging and Graphics, 2019, 80:101690.
[4] Chen L, Bentley P, Mori K, et al. DRINet for Medical Image Segmentation[J]. IEEE Transactions on Medical Imaging, 2018:1-1.
[5] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4):834-848.
[6] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2015.
[7] 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, Cham, 2015.
[8] Wu Y, He K. Group Normalization[J]. International Journal of Computer Vision, 2018.
[9] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
[10] He K, Zhang X, Ren S, et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification[J]. 2015.
[11] Bottou L. Stochastic gradient descent tricks[M]//Neural networks: Tricks of the trade. Springer, Berlin, Heidelberg, 2012: 421-436.
[12] Cootes T F, Taylor C J. Active shape models—‘smart snakes’[M]//BMVC92. Springer, London, 1992: 266-275.
[13] Cootes T F, Edwards G J, Taylor C J. Active appearance models[J]. IEEE Transactions on pattern analysis and machine intelligence, 2001, 23(6): 681-685.
[14] Zhang T, Qi G J, Xiao B, et al. Interleaved Group Convolutions for Deep Neural Networks[J]. 2017.