Improved U-Net Model for Nerve Segmentation

Houlong Zhao and Nongliang Sun(✉)

College of Electronics, Communication and Physics,
Shandong University of Science and Technology, Qingdao 266590, China
zhaohoulong@outlook.com, nl-jackson@vip.163.com

Abstract. Noticeable gains in computer vision have been made as a result of the large-scale datasets and deep convolutional neural networks (CNNs). CNNs have been used in a wide variety of tasks, for instance, recognition, detection, and segmentation. Recently, due to the open medical images datasets, CNNs have been used in Computer Aided Detection (CADe) to help doctors diagnose lesion. In this work, we present an end-to-end method based on CNNs for automatical segmentation from medical images. The proposed network architecture is similar to U-Net, which consists of a contracting path and an expansive path. However, we take advantage of inception modules and batch normalization instead of ordinary convolutional layers, which reduce the quantity of parameters and accelerate training without loss of accuracy. In addition, we confirm Dice coefficient as loss function rather than binary cross entropy. We use this model to segment nerve from ultrasound images and achieve a better performance.

Keywords: Nerve · Segmentation · Ultrasound Images · Inception
Dice coefficient

1 Introduction

Since 2012 krizhevsky et al. [16] won the ImageNet competition [6], deep convolutional neural networks [17] have been the mainstream to create the state of the art in computer vision and pattern recognition, in which there have been a series of breakthroughs among large lots of tasks, for example to classification [11,26,27], object-detection [7,21], and segmentation [23]. A series of success is credited on the ability of CNNs to self-educate from the raw input without manual intervention.

Benefited from the large open datasets about medical-images, computer aided detection (CADe) making use of deep learning has become a reality. Up to now, CADe has been used in clinical environments for over 40 years [1], but it usually can't replace the doctor or become the role in diagnosing. However, ResNet designed by He et al. [11] has 3.57% top-5 error on the ImageNet test set excelling human beings, which provides a possibility that machine can substitute for the doctors in some tasks, for instance, detecting cancer metastases [22], and diabetic retinopathy [9] etc.

Y. Zhao et al. (Eds.): ICIG 2017, Part II, LNCS 10667, pp. 496–504, 2017.
https://doi.org/10.1007/978-3-319-71589-6_43

In this paper, we specifically consider of the problem of automatic segmentation for nerves: given an ultrasound image of neck, some with nerve, we want to fully automatically end-to-end segment the nerves. Recently, most approaches rely on deep convolutional neural networks of medical image segmentation have achieved great success, in view of which, we tested some foundational methods based on CNNs and attempted to raise our own architecture.

The proposed architecture is inspired by U-Net [4] that has a good performance in biomedical image segmentation but with too many parameters. In order to improve the performance, we adopted inception modules and batch normalization. Through some efforts we have done, the model has fewer parameters and less time for training. In addition, we confirmed Dice coefficient as loss function, which compares the pixel-wise agreement with a predicted segmentation and ground truth.

2 Related Work

Early medical image segmentation methods mostly based on statistical shape, gray level and texture collected in [12]. Recently, level set [18], graph cut [19] have been employed as approaches in biomedical image segmentation. However, these methods are not widely used owing to their speed and complex operation.

Through the rapid development in recent years, deep Convolutional Neural Networks (CNNs) have been exploited to improve the ability of machine to achieve state-of-art goals [10,11,28] in computer vision. More stirring, CNN seems to be widely-used, which prompted us to employ it to automatically segment nerves from ultrasound images.

Semantic segmentation methods based on convolutional neural networks got a big development. [23] is the first model to use fully convolutional networks that produce corresponding output as same size as input and skip architecture that combines semantic information from a deep layer with appearance information from a shallow layer, SegNet [2] uses encoder-decoder structures restore the feature maps from higher layers with spatial information from lower layers. DeepLab [3] combines the responses at the final CNN layer with a fully connected Conditional Random Field (CRF) to improve localization accuracy. PSPNet [28] proposes a pyramid scene parsing network to combine local and global features.

As for medical image segmentation, U-Net [25] can be trained end-to-end from very few medical images, [24] proposes V-Net to 3D image segmentation. [5] uses cascaded fully convolutional neural networks for liver and tumor segmentation, which firstly trains an FCN to segment the liver as input for the second FCN, secondly uses the second FCN solely segment tumor.

3 Dataset

The dataset we used to evaluate our model is provided by Kaggle[1], which contains a collection of nerves called the Brachial Plexus in ultrasound images. The

[1] https://www.kaggle.com.

training data is a set of 5635 images (580 × 420 pixels) and its corresponding masks where nerve (white) and background (black) have been manually annotated (Fig. 1). And the testing set with 5508 ultrasound images is publicly available, but its segmentation maps are kept secret. The main purpose of accurately segmenting the nerve structures in ultrasound images is to help doctors to effectively insert a patients pain management catheter that mitigate pain at the source in order to decrease the intake of narcotic drug and speed up patient recovery.

Fig. 1. The left image is the raw ultrasound image containing nerve structure, the middle image is the corresponding mask manually annotated by experts, and the right image is the ultrasound image overlaid with ground truth segmentation map (red border). (Color figure online)

4 Improved U-Net Model

The network architecture is illustrated in Fig. 2. There are two paths in this architecture, similar to U-Net [25], which are a contracting path and an expansive path. In this network, we combine the two paths with inception module [20,27] and batch normalization [14]. The contracting path (Fig. 2 left) is a normal convolutional neural network for recognition, it involves 3 basic convolutional units (Fig. 4) and 4 inception modules (Fig. 3), besides these, there are 5 max pooling operations with stride of 2 for downsampling. Between the contracting path and expansive path, there also have an inception module. In expansive path (Fig. 2 right), there are 5 upsample layers followed basic convolutional units or inception modules, generally symmetrical to the contracting path. At the final layer, a 1×1 convolution and a sigmoid activation function are used to output 2 class segmentation images which have the same size as inputs. We also used skip architectures [23], which concatenate features of shallower layers from contracting path with the features of deeper layers from expansive path. We concatenate the features in the inception modules with the 4 convolutional paths.

Each inception module has 4 paths, which act together on one input and are concatenated as an output following the practice in [27]. Before every expansive 3×3 and 5×5 convolution, 1×1 convolution layer is used to reduce computation. In total, the inception module has 6 convolutional layers and an average pooling layer. Besides the final layer, all of the convolutional operations are followed by a batch normalization and a rectified linear unit (ReLU), which make

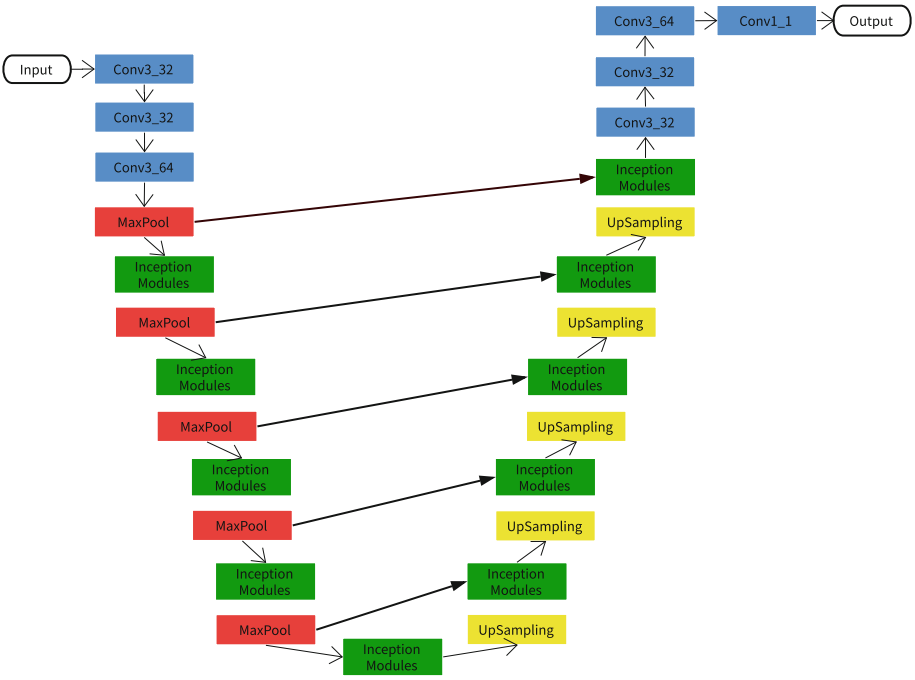

Fig. 2. The proposed architecture. The blue and green box represent the basic convolutional units and the inception modules respectively. And what the red and yellow box signify are downsampling (max pooling) and upsampling. Simultaneously, there are 5 skip architectures. (Color figure online)

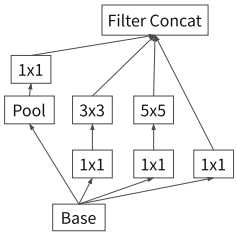

Fig. 3. Inception modules where four convolutional paths act on one input and connect to one output with average pooling as pooling layer.

up a basic convolutional unit (Fig. 4). We employed sample linear interpolation as the upsample operation to make the size of output segmentation maps to reach the size of the input image rather than using deconvolutional layers as in [23], which demand supernumerary computation.

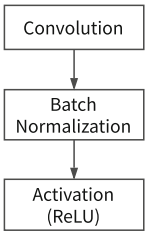

Fig. 4. The basic convolutional unit we used, where the convolutional layer followed a batch normalization and a ReLU.

5 Training

The input images and their corresponding masks are resized to 64×64 so as to reduce the computational cost. We adopt batch normalization [14] after each convolution and before activation. We use Adam [15] with a mini-batch size of 30. The learning rate is 0.001 and the models are trained for up to 78 epochs with Keras [4]. We use β_1 of 0.9 and β_2 of 0.999 following [15]. We set the initial random weights using Xavier normal initializer [8]. We have not used dropout [13] and any other approach for regularization.

In medical image segmentation, the interested region is more essential than background. If we use binary cross entropy to predict the probability of each pixel to belong to foreground or to background, all the pixels in input images are equally important. However, the interested region (white) just accounts for a smaller proportion of the area than background (black) as shown in Fig. 1, as a result, the interested region will be often missed. Binary cross entropy defined as

$$L = -\frac{1}{n}\sum_{i=1}^{n}[t_i \; log(o_i) + (1 - t_i) \; log(1 - o_i)] \tag{1}$$

where the sums run over the n pixels, i denotes the pixel position, t_i is the ground truth value, and the o_i is the predicted pixel value. In this paper, we employed Dice coefficient as loss function followed the practice in [24]:

$$L = -\frac{2\sum_{i=1}^{n} o_i t_i}{\sum_{i=1}^{n} o_i^2 + \sum_{i=1}^{n} t_i^2} \tag{2}$$

where n, i, o_i, t_i denote same meaning as binary cross entropy.

6 Experiments and Analysis

The model is trained end-to-end on the 5635 training ultrasound images, and tested on the 5508 testing images. An evaluated score can be obtained by submitting the predicted segmentation maps to the Kaggle's sever. The result is evaluated on the mean Dice coefficient:

$$L = \frac{2|X \cap Y|}{|X| + |Y|} \tag{3}$$

where X is the set of predicted pixel values and Y is the set of ground truth values, correspondingly $|X|$ and $|Y|$ are the numbers of elements in them. We trained and evaluated our method and U-Net [25] (Fig. 5), which inspired our idea and was used widely in biomedical image segmentation. Same as our model, the U-Net we tested in this paper adopt Dice coefficient as the loss function and use basic convolutional units (Fig. 4) to replaced original convolutional layers without any dropout and other regularization.

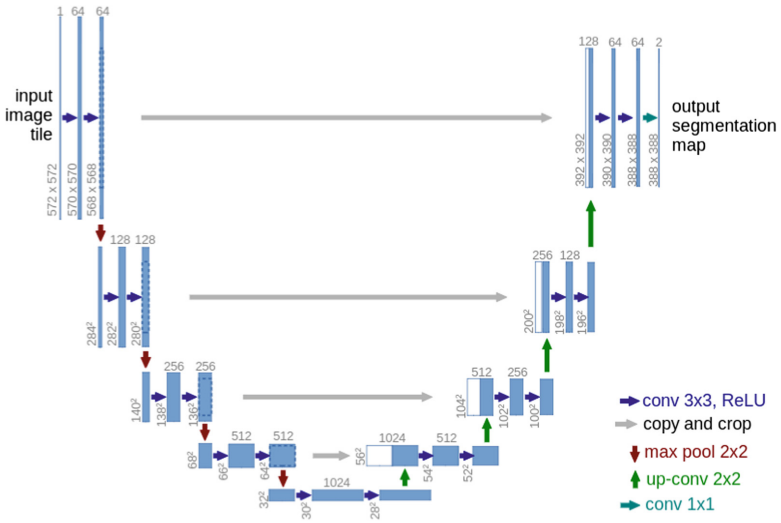

Fig. 5. U-Net [25] used to compare with our model, which totally has 23 convolutional layers. In our work, batch normalization is used after each convolution and before activation instead of the original model.

In Fig. 6 we compare Dice coefficient of our model with U-Net [25] during the training procedure. We have observed that our model reached 0.34 after first epoch and 0.80 after 32th epoch, faster than U-Net, which just reached 0.05 after first epoch and 0.80 after 50th epoch. The much faster convergence proves that the inception modules we adopted can accelerate training procedure.

The results in Table 1 show that our model achieves a score of 0.653, roughly equals to the score of 0.658 from U-Net. However, our model has fewer parameters, only 16% of the parameters of U-Net. The reason for parameters' reduction is that the 1×1 convolution does not care the correlation of information in same feature map and leads to dimension reduction. Figure 7 shows a testing result using our model.

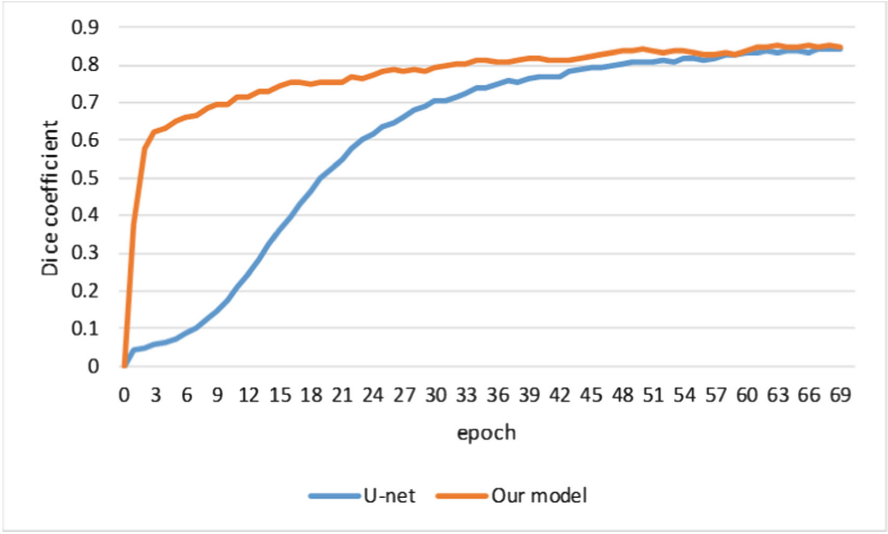

Fig. 6. Training on 5635 ultrasound images using U-Net (orange) and our model (blue). The lines denote training Dice coefficient. (Color figure online)

Table 1. The results from the test set. The first column shows the models we have tested, the second column their Dice coefficient and the third the total number of these models

Model	Dice coeff.	Total param.
Our model	0.653	5M
U-Net	0.658	31M

Fig. 7. One result of the test data, the left image is the testing ultrasound image containing nerve structure, the middle image is the predicted segmentation map, and the right image is the ultrasound image overlaid with segmentation map (green border). (Color figure online)

7 Conclusion

We present an approach based on convolutional neural networks, which achieves a good performance on ultrasound images segmentation and possesses fewer parameters thanks to inception modules. We adopted an efficacious loss function, Dice coefficient between the predicted segmentation maps and ground truth. Our model has satisfactory training time of 54 h on a Intel CORE i7 vPro and 16 GB of memory. Future work will aim at different biomedical segmentation applications and train our model over GPUs.

Acknowledgements. This work was supported by Leading talent development program of Shandong University of Science and Technology.

References

1. Computer-aided diagnosis. https://en.wikipedia.org/wiki/Computer-aided_diagnosis
2. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561 (2015)
3. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. arXiv preprint arXiv:1606.00915 (2016)
4. Chollet, F., et al.: Keras (2015). https://github.com/fchollet/keras
5. Christ, P.F., Ettlinger, F., Grün, F., Elshaera, M.E.A., Lipkova, J., Schlecht, S., Ahmaddy, F., Tatavarty, S., Bickel, M., Bilic, P., et al.: Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. arXiv preprint arXiv:1702.05970 (2017)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 248–255. IEEE (2009)
7. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)
8. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: AISTATS, vol. 9, pp. 249–256 (2010)
9. Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA **316**(22), 2402–2410 (2016)
10. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. arXiv preprint arXiv:1703.06870 (2017)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
12. Heimann, T., Van Ginneken, B., Styner, M.A., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., et al.: Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Trans. Med. Imaging **28**(8), 1251–1265 (2009)

13. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 (2012)
14. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
15. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
16. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
17. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
18. Li, C., Wang, X., Eberl, S., Fulham, M., Yin, Y., Chen, J., Feng, D.D.: A likelihood and local constraint level set model for liver tumor segmentation from CT volumes. IEEE Trans. Biomed. Eng. **60**(10), 2967–2977 (2013)
19. Li, G., Chen, X., Shi, F., Zhu, W., Tian, J., Xiang, D.: Automatic liver segmentation based on shape constraints and deformable graph cut in CT images. IEEE Trans. Image Process. **24**(12), 5315–5329 (2015)
20. Lin, M., Chen, Q., Yan, S.: Network in network. arXiv preprint arXiv:1312.4400 (2013)
21. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: SSD: single shot MultiBox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
22. Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G.E., Kohlberger, T., Boyko, A., Venugopalan, S., Timofeev, A., Nelson, P.Q., Corrado, G.S., et al.: Detecting cancer metastases on gigapixel pathology images. arXiv preprint arXiv:1703.02442 (2017)
23. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
24. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 4th International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
25. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
27. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
28. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. arXiv preprint arXiv:1612.01105 (2016)