Multi-scale U-SegNet architecture with cascaded dilated convolutions for brain MRI Segmentation

챠이트라 다야난다, 이범식
Chaitra Dayananda, Bumshik Lee

(16452) 광주광역시 남구 필문대로 309, 조선대학교 IT 융합대학 정보통신공학과
chaitrad@chosun.kr, bslee@chosun.ac.kr (교신저자)

요 약

Automatic segmentation of brain tissues such as WM, GM, and CSF from brain MRI scans is helpful for the diagnosis of many neurological disorders. Accurate segmentation of these brain structures is a very challenging task due to low tissue contrast, bias filed, and partial volume effects. With the aim to improve brain MRI segmentation accuracy, we propose an end-to-end convolutional based U-SegNet architecture designed with multi-scale kernels, which includes cascaded dilated convolutions for the task of brain MRI segmentation. The multi-scale convolution kernels are designed to extract abundant semantic features and capture context information at different scales. Further, the cascaded dilated convolution scheme helps to alleviate the vanishing gradient problem in the proposed model. Experimental outcomes indicate that the proposed architecture is superior to the traditional deep-learning methods such as Segnet, U-net, and U-Segnet and achieves high performance with an average DSC of 93% and 86% of JI value for brain MRI segmentation.

1. Introduction

Many traditional image processing approaches are involved in the segmentation of medical images, such as accurate model-based segmentation [1], random forest classification [2], level set, and sparse decomposition [3]. Recently, most of the researchers propose deep learning into medical image segmentation [4-6] since the deep-learning through its convolutional structures possess the unique capability of self-learning from a considerable volume of data. Among many of the deep learning models, U-Net [6-9], an end-to-end full convolutional neural network, is considered as the promising architecture for medical image segmentation. The U-Net architecture through skip connections obtains both coarse level and fine level details at the deconvolutional layers [7]. U-Net network is employed to perform different tasks such as to segment organs/tissues or lesions in the brain, left ventricle, prostate [8-9]. U-Net involves an encoder and decoder, where the encoder is integrated with the design of pooling operation and decoder with interpolation, which will impact the segmentation performance. SegNet [12], a similar encoder-decoder based network, uses pooling indices for upsampling the feature maps. Many researchers have revised the primary U-net by including some modules or by modifying some sub-structures [10-12] to enhance segmentation accuracy. Although the U-net with its skip connections provides better accuracy, it is prone to generate more learnable upsampling parameters, thus making the U-net model slower to train than

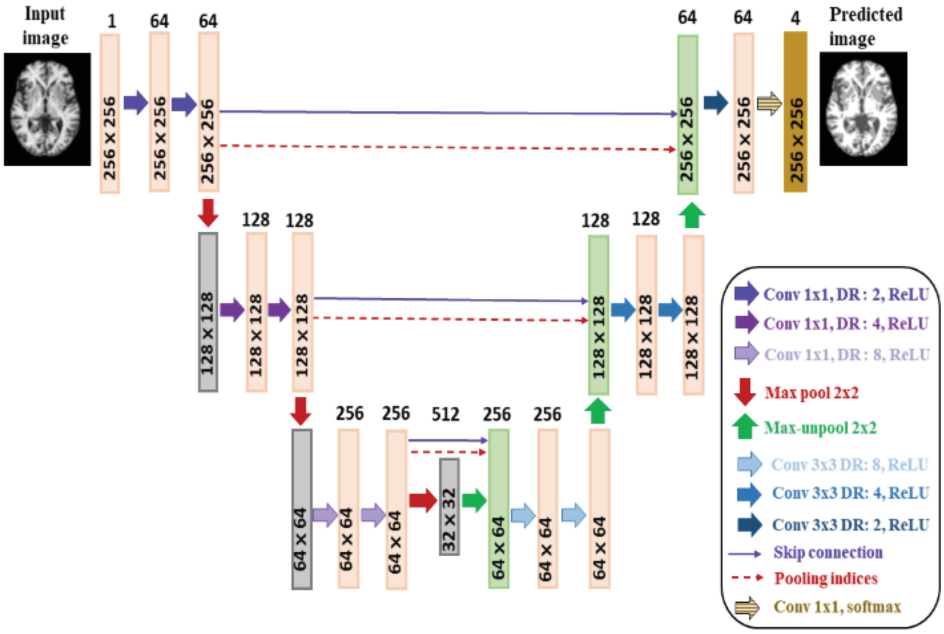

Figure 1: Schematic representation of the proposed method. The values inside and at the top of each block represent the dimension of the feature maps and number of convolution filters, respectively.

Segnet [12]. Moreover, Segnet fails to achieve multi-scale learning as effective as U-net. Hence, a new combination model with the complementary strengths in the two models was explored, which is called U-Segnet [13]. U-Segnet, use Segnet [12] as the base architecture with skip connections at the selected deconvolution layer. The pooling indices aid in faster convergence of the model and skip connections provide multi-scale information for better performance [13]. Although the U-SegNet gives better segmentation performance, the up-sampled outputs are still blurry and smooth, and the network fails to capture the details in certain regions of the image.

A general approach to promote the performance of a fully convolutional network (FCN) is to deepen the network by increasing its convolutional layers [14]. This approach results in a significant rise in the model parameters compared to the accuracy enhancement, which leads to more memory and computational requirements by the network. It shows that this deepening method should be implemented on devices with excellent hardware. Also, numerous pooling operations result in low-resolution feature maps. Therefore, this scheme is not suitable for medical image segmentation. To avoid the above problems, we designed a novel cascaded dilated convolution-based U-SegNet architecture with multi-scale convolution kernels. The architectural representation of the proposed method is as shown in **Fig. 1**.

2. Proposed method

Our proposed method is a two-dimensional convolutional network, which accepts original MR images with a dimension of 256×256 as input along with their corresponding segmented maps used as ground truth and predicts segmented tissue maps as output. As shown in **Fig. 1**, the proposed method contains an encoder and decoder. The encoder consists of 3 layers, where each layer is composed of a set of 1×1 convolutions with ReLU activations followed by max-pooling with a stride of 2×2. The convolutions in each encoder layer consist of different dilation rates forming cascaded dilated convolutions. The 1×1 convolution in the first encoder layer is set with dilation rate 2, second encoder layer with dilation rate 4, and third layer with dilation rate as 8. The encoder layer takes the input from the previous encoding layer and outputs a feature map to the upcoming encoder layer. This feature map contains the information or features that represent the input. Each encoder layer consists of a corresponding decoder layer. The decoder is composed of 3×3 convolution with ReLU activation, followed by the max-unpool operation. Similar to the encoder, the first, second, and third decoding layer consists of convolutions with dilation rates 2, 4, and 8, respectively. The decoder aims to accept the low-resolution features, learned by the encoder, and unpools these features onto the higher resolution pixel space.

Table 1 Performance comparisons between the proposed and the traditional methods such as U-net, SegNet, and U-SegNet (DSC: Dice Similarity Coefficient, JI: Jaccard Index, MSE: Mean Square Error)

Models	WM		GM		CSF		MSE	Computation	# Learnable
	DSC (%)	JI (%)	DSC (%)	JI (%)	DSC (%)	JI (%)		Time(5 epochs)	parameters
SegNet[12]	87.56	77.18	84.93	72.20	80.67	67.85	0.015	2.6 hours	3873860
U-Net[6]	92.18	85.50	90.32	82.35	89.53	81.53	0.008	3.4 hours	4832324
U-Segnet[13]	92.86	87.07	91.15	84.77	90.85	85.28	0.007	2.9 hours	4279172
Proposed Method	**93.98**	86.63	**92.97**	84.22	**91.65**	87.91	**0.006**	1.8 hours	**2214276**

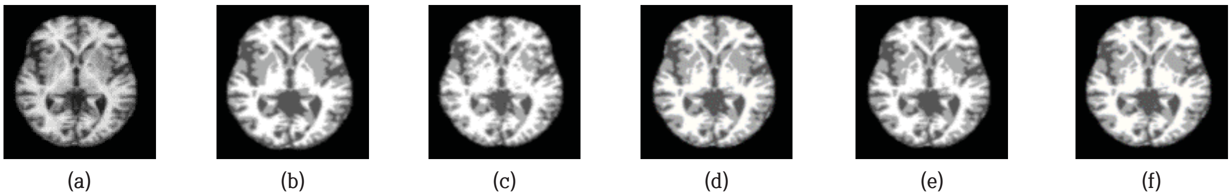

Figure 2: Segmentation results obtained for our proposed method: (a) Original input image, (b) ground truth segmentation map, (c) predicted segmentation map for SegNet, (d) predicted segmentation map for U-Net (e) predicted segmentation map for U-SegNet, (f) predicted segmentation map for the proposed method.

As a result, a dense pixel classification can be achieved. The unpooling operation captures class-specific structures by following the original pooling indices retained while encoding. As a result, it efficiently restores the complete structure of an image in finer resolutions. In addition, skip connections introduced from encoder to decoder captures both coarse level and fine level details while unpooling. The final decoder layer feature maps are given to the softmax classifier to predict the brain tissue classes (GM, WM, CSF). Further, the proposed architecture designed with both 1×1 and 3×3 kernels forms a multi-scale convolution model and achieve multi-scale information of brain tissues. The 1×1 extracts highly local features and helps to capture smaller, complex information in the image. Whereas 3×3 spreads across the image extract more generic features and capture the basic image components. Hence proposed network in combination with both 1×1 and 3×3 convolution achieves better accuracy with fewer network parameters. Further, the dilated convolutions enlarge the receptive field without resolution or coverage loss and combine context information to improve the segmentation accuracy.

3. Experimental results

Open-access series of imaging studies (OASIS) [15] dataset is used to analyze our proposed method. The dataset consists of 416 subjects with their T1- weighted MRI scans. To show the effectiveness of our experiments with less training data, we randomly chose only the first 15 subjects. We used 10 subjects for training the network, and the remaining 5 subjects were used for testing the model. Axial plane images were used in our experiments. For each subject, axial scans consist of 176 slices in total. "Stochastic Gradient Descent (SGD)" is used for model training with a learning rate of 0.001and a high momentum rate of 0.99.

The segmentation results of the proposed method and the traditional methods such as SegNet, U-net, and U-SegNet presented in **Fig. 2**. shows that the proposed method presents subjectively well-segmented results. **Table 1** shows the performance comparisons between the proposed method, and conventional SegNet [12], U-net [6], and U-SegNet [13]. To objectively evaluate the performances of the methods, the Dice Similarity Coefficient (DSC) [16] and Jaccard Index (JI) [17] were used. The DSC and JI scores between ground truth and predicted segmentation map expressed as (1) and (2)

$$DSC = \frac{2TP}{2TP+FP+FN} \quad (1)$$

$$JI = \frac{DSC}{2-DSC} \quad (2)$$

where FP, FN, and TP are, false positives and false negative, true positives, respectively. We also further used mean square error (MSE) to evaluate the segmentation performance, which is an average square difference between the original S and predicted S' values and computed as (3),

$$\text{MSE} = \frac{1}{RC}\sum_{i=1}^{R}\sum_{j=1}^{C}(S-S') \quad (3)$$

where R and C are the width and height of the image, respectively. As mentioned in Table 1, the accuracy and the loss for the proposed deep learning network for brain MRI segmentation achieved 93% DSC accuracy with a substantially lower MSE value of 0.006 on average than other methods. It indicates that the proposed method can predict more similar values to the original ones. The proposed method generates 0.2 million learnable parameters, which constitute a 50% reduction in parameter requirement, consuming less computation time than conventional methods.

4. Conclusion

In this paper, we show that the use of 1×1 convolution in the encoder help in the dimensionality reduction reduces the computational load by reducing parameter map, and create a smaller CNN network while 3×3 convolution at the decoder side helps to retain a higher degree of accuracy. Further, the cascading of convolutions with different dilation rates prevent the spatial resolution loss of the responses. The proposed model requires 50% less number of learnable parameters compared to U-SegNet leading to reduced computation time. The proposed method gives an overall DSC score of 93%, which shows an improvement of 1.5% over the conventional U-SegNet.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (No. NRF-2019R1A4A1029769 and No. 2019R1I1A3A01058959).

References

[1] A. Mastmeyer, G. Pernelle, R. Ma, L. Barber, and T. Kapur, "Accurate model-based segmentation of gynecologic brachytherapy catheter collections in MRI-images," Med. Image Anal., vol. 42, pp. 173-188, Dec. 2017.

[2] A. Mastmeyer, D. Fortmeier, and H. Handels, "Random forest classification of large volume structures for visuo-haptic rendering in CT images," Proc. SPIE, vol. 9784, Mar. 2016, Art. No. 97842H. doi: 10.1117/ 12.2216845.

[3] Z. Zhang and J. Song, "An adaptive fuzzy level set model with local spatial information for medical image segmentation and bias correction," IEEE Access, vol. 7, pp. 27322-27338, 2019.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention—MICCAI. Cham, Switzerland: Springer, 2015, pp. 234-241.

[5] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-net based fully convolutional networks," in Proc. Annu. Conf. Med. Image Understand. Anal., Edinburgh, U.K., 2017, pp. 506-517.

[6] M. U. Dalmış, G. Litjens, K. Holland, A. Setio, R. Mann, N. Karssemeijer, and A. Gubern-Mérida, "Using deep learning to segment breast and fibroglandular tissue in MRI volumes," Med. Phys., vol. 44, no. 2, pp. 533-546, Feb. 2017.

[7] S. Charmchi, K. Punithakumar, and P. Boulanger, "Optimizing U-net to segment left ventricle from magnetic resonance imaging," in Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM), Madrid, Spain, Dec. 2018, pp. 327-332.

[8] W. Chen, Y. Zhang, J. He, Y. Qiao, Y. Chen, H. Shi, and X. Tang. "Prostate segmentation using 2D bridged U-net," 2018, arXiv: 1807.04459. [Online]. Available: https://arxiv.org/abs/1807.04459.

[9] P. F. Jaeger, S. A. A. Kohl, S. Bickelhaupt, F. Isensee, T. A. Kuder, H.-P. Schlemmer, and K. H. Maier-Hein, "Retina U-net: Embarrassingly simple exploitation of segmentation supervision for medical object detection," 2018, arXiv: 1811.08661. [Online]. Available: https://arxiv.org/abs/1811.08661

[10] M. Z. Alom, M. Hasan, C. Yakopcic, M. T. Taha, and K. V. Asari, "Recurrent residual convolutional neural network based on U-Net (R2UNet) for medical image segmentation," 2018, arXiv: 1802.06955. [Online]. Available: https://arxiv.org/abs/1802.06955

[11] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, arXiv:1804.03999. [Online]. Available: https://arxiv. org/abs/1804.03999

[12] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Trans. Pattern Anal. Mach. Intell., 2017.

[13] Kumar, Pulkit & Nagar, Pravin & Arora, Chetan & Gupta, Anubha. (2018). "U-SegNet: Fully Convolutional Neural Network based Automated Brain tissue segmentation Tool".

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Boston, MA, USA, Jun. 2015, pp. 3431-3440.

[15] H. Tracy, G. John, C. Csernansky, M. Marcus, S. Daniel, and L. Randy, "Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in Young, Middle Aged, Non demented, and Demented Older Adults". Journal of Cognitive Neuroscience, vol.19, pp.1498-1507, 2007.

[16] L. R. Dice, "Measures of the Amount of Ecologic Association between Species," Ecology, 1945

[17] P. Jaccard, "The distribution of the flora in the alpine zone," New Phytol., 1912.