ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 13, No. 2, 2018, *pp*.083-088

A Multi-Scale Fully Convolutional Networks Model for Brain MRI Segmentation

Zhihui Cao, Yuhang Qin, Yunjie Chen*

[1]*School of math and statistics, Nanjing University of Information Science & Technology, Nanjing 210044, China*
(Received October 01 2017, accepted January 15 2018)

Abstract。Accurate segmentation for brain magnetic resonance (MR) images is of great significance to quantitative analysis of brain image. However, traditional segmentation methods suffer from the problems existing in brain images such as noise, weak edges and intensity inhomogeneity (also named as bias field). Convolutional neural networks based methods have been used to segment images; however, it is still hard to find accurate results for brain MR images. In order to obtain accurate segmentation results, a multi-scale fully convolution networks model (MSFCN) is proposed in this paper. First, we use padding convolutions in conv-layer to preserve the resolution of feature maps. So we can obtain segmentation results with the same resolution as inputs. Then, different sized filters are utilized in the same conv-layer, after that, the outputs of these filters are concatenated together and fed to the next layer, which makes the model learn features from different scales. Both experimental results and statistic results show that the proposed model can obtain more accurate results.

Keywords: Convolutional neural networks; Fully convolutional networks; magnetic resonance image; multi-scale.

1. Introduction

Brain disease is one of the main diseases that threaten human health, so it is of important sense to use brain imaging to help us diagnose the brain disease [1-3]. Compared with other medical images, brain MR images are easier to be used for diagnosis of brain disease for their high contrast among different soft tissues and high spatial resolution [4]. Segmenting brain tissues accurately, including white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF), plays an important role in both clinical practice and medical study. However, some imaging artifacts such as noise, intensity inhomogeneity and weak edges, which drives scholars to find and propose more robust and more accurate approaches, hinder most segmentation methods.

Quite a lot of clustering algorithms have been proposed for brain MR image segmentation. Fuzzy C-means (FCM) algorithm, first introduced by Dunn [5], is one of the most widely used ones. FCM assumes that a pixel of an image belongs to different classes at different degree, corresponding to brain MR images' fuzziness. The FCM fails to segment images with noise, low contrast and bias field by only using intensity information. In order to improve the robustness of FCM, many scholars proposed modified models based on FCM by adding spatial information into it and obtained a certain improvement [6, 7]. However, they failed to solve the problem of low contrast.

In the last several years, deep learning (DL) [8], especially convolutional neural networks (CNN), has outperformed the state of the art in computer vision tasks. Since the breakthrough by Krizhevsky et al. [9], even larger and deeper networks have been trained [10, 11].

The traditional use of CNN is on classification tasks, where the output we want is just a class label for the input image. However, in image segmentation tasks, the desired output should be an image with each pixel labeled. Hence, Ciresan et al. [12] trained a network (DNN) to predict a class label of a pixel by labeling a patch in a square window centered on the pixel itself. They succeeded to apply CNN to image segmentation and won the EM segmentation challenge at ISBR 2012. Nevertheless, there are some shortcomings in this model. First, it requires a great lot of calculation and room, for example, if we want to segment an image with a size of 512×512, we have to classify 262144 patches with the network. Secondly, there is a lot of redundancy because of the overlapping patches of adjacent pixels. Thirdly, there is a tension between local information and global information. Small patches guarantee the localization accuracy but use little context, while large patches may reduce the localization accuracy. These cause the network inefficient.

Long et al. [19] first trained an end-to-end learning network for semantic segmentation called fully convolutional networks (FCN), which outputs dense prediction from arbitrary-sized inputs. Moreover, both

training and predicting are performed whole-image-at-a-time. In-network convolutional layers extract features and upsampling layers enable pixelwise prediction. Compared with patch-wise training networks [12], the FCN model is more uncomplicated and works more efficient with no pre- or post-processing.

However, there are millions of parameters to be trained in the FCN model, which requires a large amount of training data and a lot of training time, but it is unrealistic in biomedical tasks. Hence, Ronneberger et al. [20] modified and extended the FCN model such that it works with few training images and obtains more precise results, and this model is called U-Net for its U-shaped architecture. U-Net upsamples at stride 2 and combines shallow layers with upsampled outputs, thus it learns better and the segmentation results are more realistic with sharper edges. But U-Net underperforms when it comes to segmenting details, such as CSF in brain MR images.

Based on the analysis above, we can find gray value based methods are sensitive to outliers. Although some modified models reduce the influence of noise and outliers to some extent, most improvements are at the price of increasing parameters and complexity of the model. FCN and U-Net are efficient and find better results, while details and narrow bands in images are lost. To address these drawbacks, we propose a multi-scale FCN model (MSFCN) in this paper. Different-scaled filters are added into the model to obtain more accurate results, where large-scaled filters see more context and small-scaled ones keep details and narrow bands. We have compared proposed model with other state-of-the-art segmentation models to show that our model can obtain more precise results.

2. Backgrounds

2.1 Deep convolutional networks (DNN) in segmentation

Considering the excellent performance of deep learning in classification task, many scholars try to take advantage of it for image segmentation. Patchwise training was used in many approaches, in which each pixel is labeled with the class of its enclosing region. For example, Ciresan et al. [12-18] succeeded to apply CNN to segmentation task, but the poor efficiency made it impossible for their model to be used in clinical medicine.

By contrast, FCN [19] is more efficient. Long et al. reinterpreted classification networks as fully convolutional and add upsampling layers, which decreased parameters and complexity a lot and enabled pixelwise prediction. However, there are some drawbacks in the FCN model. First, the FCN model takes some typical CNN models such as AlexNet, VGGNet, etc as its contracting part, which requires a large amount of training data. Secondly, the model has to be trained three times (FCN-32s, FCN-16s, FCN-8s). Thirdly, the result of FCN is too smooth and the details are lost.

Based on FCN, Ronneberger et al. [20] built a more elegant architecture called U-Net. This network uses successive convolutional layers followed by a maxpooling layer in contracting part, and in expansive part, the process is inverted. Further, feature maps with the same resolution from both parts are concatenated together followed by successive convolutional layers and two 1×1 filters are used in the final to obtain the segmentation results. But due to the unpadded convolution, the resolution of results is lower than that of inputs. Moreover, so many times of convolutions make the details lost in high layers, leading to erroneous results at these pixels.

3. Proposed Model

We replace some 3×3 kernels with 1×1 and 5×5 ones based on U-Net without changing the depth of the network, which enable the network to extract features from different scales at the same time. In the following, an overview of our model is given.

3.1 Motivation

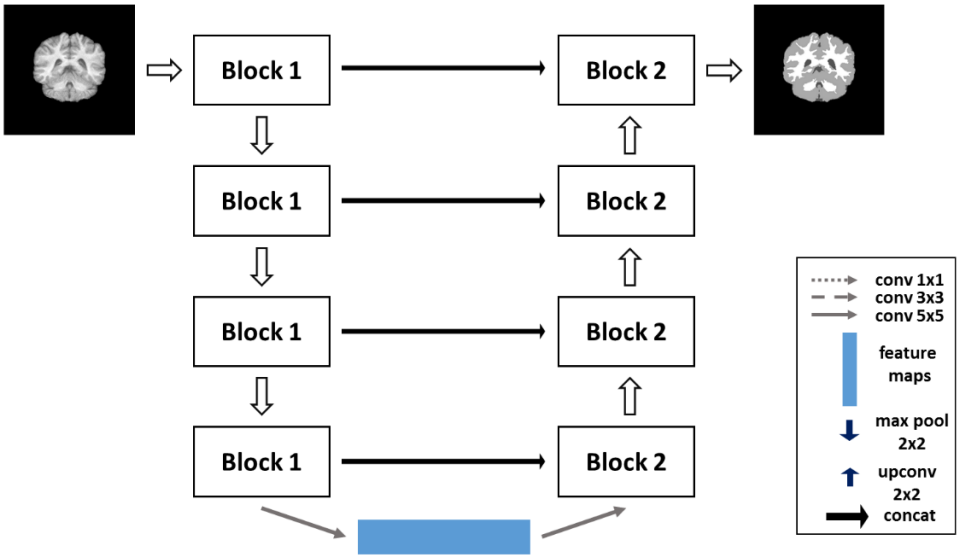
Szegedy et al. [10] trained GoogLeNet model in 2014 and won the classification task of Imagenet Large Scale Visual Recognition Challenge 2014(ILSVRC2014). They proposed an inception module, where 1×1, 3×3, 5×5 filters and max-pooling are used at the same time. This work improved their final accuracy and reduced the number of parameters quite a lot.

Considering the improvement by the inception module, we proposed MSFCN model for segmenting brain MR images. Different from classification task, a class label should be assigned to each pixel in segmentation task, which means not only the global information and large regions matter, but small regions, details and the pixel itself are also essential. Thus, we increase the proportions of 1×1 and 3×3 filters and

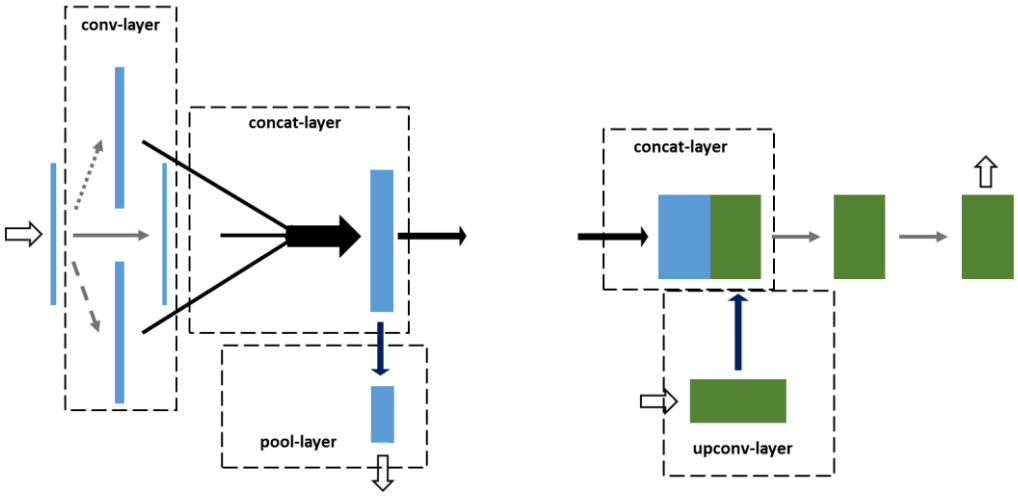
reduce that of 5×5 ones. Traditional methods underperform when segmenting details and narrow bands, while our strategy to utilize multi-scale filters is more flexible.

3.2 Model structure

Our model is made up of two parts, the contracting part and the expansive part. Each part consists of several blocks (see Fig. 1 (b) and (c)), and our network model structure is illustrated in Fig. 1 (a) (we show our model with blocks just to illustrate it clearly). It concatenates outputs from different scaled filters, and concatenates maps from shallow layers and high layers as well. In contracting part, every "max-pooling" halves the width of maps and the following "convolution" doubles the number of feature channels, while it is inverted in expansive part. We use padded convolutions to find segmentation results with same resolution as inputs. At the final layer, we use four 1×1 convolution filters to map the outputs, making the network segment brain MR images into four classes (WM, GM, CSF and background). The detailed parameters of our model is reported in Table 1. The expansive part is same as U-Net so we don't list it here.



(a)



(b) (c)

Fig. 1 Illustration of MSFCN model (a) MSFCN architecture and a description of some icons. The number of feature maps in each block is different, but the structures are the same (b) Block 1 architecture. (c) Block 2 architecture

3.3 Model training

We fed the network with 2000 brain MR images and corresponding segmentation results to train it. In the training process, we used the Adam [21] optimizer, batch size 5 and learning rate $\lambda = 0.0001$, and the network was trained for 30 epochs. For a given image sized by 256×256, the network outputs a probability matrix sized by $256 \times 256 \times 4$. The probability is calculated by the softmax which is defined as

$$p_c(I_{ij}) = e^{O_c(I_{ij})} / \sum_{c'}^{C} e^{O_{c'}(I_{ij})} \quad (1)$$

where $p_c(I_{ij})$ is the probability that the pixel I_{ij} belongs to class c, and $O_c(I_i)$ is the activation in feature channel c at position (i, j) in the last layer (the value at (i, j, c) in the probability matrix). In addition, the cross entropy loss function was used in the network as the energy function and is shown as follow:

$$E = \sum_{I_{ij} \in I} \sum_{c=1}^{C} p_c'(I_{ij}) \log\left(p_c(I_{ij})\right)^{-1} \quad (2)$$

where $p_c'(I_{ij})$ is the true distribution. That means the loss function is computed by adding up pixelwise softmax over the whole final map with the cross entropy.

Our network is based on Keras and trained on a NVIDIA GeForce GTX 1050Ti GPU (4GB).

Table 1: The architecture and parameters of MSFCN

Part	Layer	Input	Kernel Size	OutSize
Contracting Part	Conv1	Brain MRI	1*1	256*256*32
			3*3	256*256*32
			5*5	256*256*10
	Concat1	Conv1	-	256*256*74
	Pool1	Concat1	2*2	128*128*74
	Conv2	Pool1	1*1	128*128*64
			3*3	128*128*64
			5*5	128*128*20
	Concat2	Conv2	-	128*128*148
	Pool2	Concat2	2*2	64*64*148
	Conv3	Pool2	1*1	64*64*128
			3*3	64*64*128
			5*5	64*64*40
	Concat3	Conv3	-	64*64*296
	Pool3	Concat3	2*2	32*32*296
	Conv4	Pool3	1*1	32*32*256
			3*3	32*32*256
			5*5	32*32*80
	Pool4	Conv4	2*2	16*16*592
	Conv5	Pool4	3*3	16*16*512
Expansive Part	Conv5	Conv5	3*3	16*16*512
	…			

4. Experiment Results

In this section, we experimentally evaluate our proposed model in a set of clinical brain MR images, which is generated from Internet Brain Segmentation Repository (IBSR). We also evaluate U-Net for comparison. The U-Net and our model are both trained with 2000 images and corresponding ground truth for 30 epochs, and tested on 200 images. Furthermore, the two models are both trained for ten times to show the robustness.

First, we compare the result of our model with that of other methods in order to show that our model performs better in segmenting details and narrow bands. Fig. 4(a) shows the original clinical brain image and Fig. 4(b) is the standard segmentation result of it. Fig. 4(c) is the segmentation result of U-Net and Fig. 4(d) is ours. Fig. 4(e-h) show the details of Fig. 4(a-d). It can be seen from the results that our model segment the details and narrow bands better than U-Net and our model is more flexible.

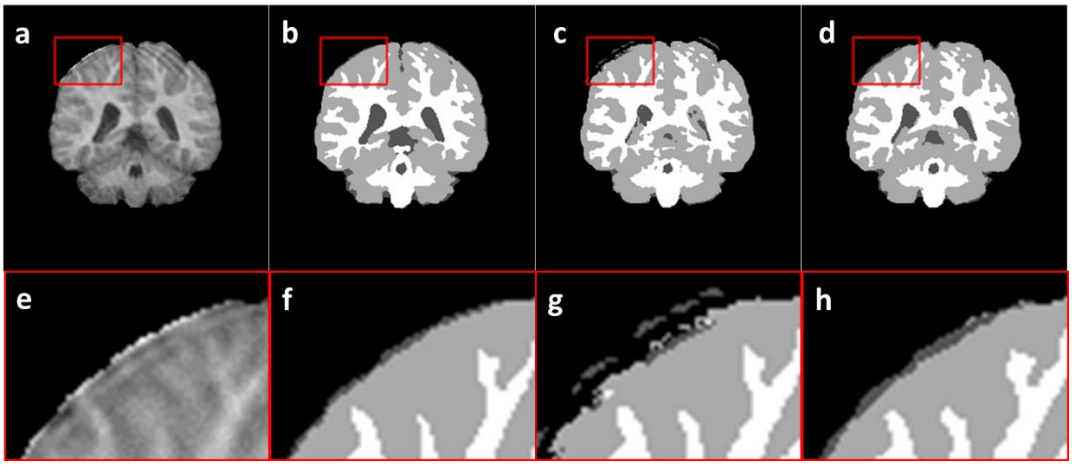


Fig. 2 (a) clinical brain MR image (b) ground truth (c) U-Net (d) proposed model (e-h) details in red boxes of (a-d)

Then, in order to quantitatively analyze the effects of various segmentation models, we use Jaccard similarity (Js values)[22] as a metric to evaluate their performance, which is define as a ratio between the intersection and union of two sets S_1 and S_2:

$$Js(S_1, S_2) = (S_1 \cap S_2)/(S_1 \cup S_2) \quad (3)$$

where S_1 is the segmentation result and S_2 is the ground truth. A more accurate segmentation result should have a higher Js value. We applied above methods on 200 clinical brain MR images from IBSR to show the robustness and accuracy of our proposed model. The Js values on CSF, GM and WM are listed in Table 2 with means computed over 200 results predicted by ten trained models. From Table 1 we can find that our proposed model has higher means of Js values on all brain tissues, which indicates that our model has more accurate segmentation results.

Table 2 ： Mean Js values of segmentation results on CSF, GM, WM (%)

	CSF	GM	WM
U-Net[20]	31.79	84.25	83.04
MSFCN	**39.05**	**86.60**	**85.80**

In Table 3, we list some other criterions for comparison, including Acc, Var, training time and running time, where Acc (accuracy) is a value computed by $(S_1 \cap S_2)/S_2$ over the whole map and Var is the variance of Js values. From Table 2, we can find that our model has lower variance which proves that our model is more robust.

Table 3 ： Comparison of accuracy (%) and variances ($\mathbf{10^{-4}}$) of Js values

	Acc	Var(Acc)	Var(CSF)	Var(GM)	Var(WM)
U-Net[20]	89.02	4.42	**91.02**	11.18	23.30
MSFCN	**90.74**	**2.50**	92.98	**6.71**	**16.47**

We also compare training time and running time in Table 4. Although our model takes longer time in training and running, it's acceptable for clinical applications.

Table 4： Training time and running time of U-Net and MSFCN

	Training time(s/epoch)	Running time(s/img)
U-Net	**127.73**	**0.0226**
MSFCN	229.15	0.0382

5. Acknowledgement

This work was supported in part by the National Nature Science Foundation of China 61672291.

6. Conclusion

In this paper, we have proposed a multi-scale FCN model for brain MRI segmentation based on U-Net. In order to obtain more accurate results, we use different sized filters in a layer to get more features. We have shown that this model yields more accurate segmentation results and that it performs well in details, weak edges and narrow bands. In addition, we can conclude, from the experimental results, that this model has higher robustness.

7. Reference

[1] Boesen K, Rehm K, Schaper K, et al. Quantitative comparison of four brain extraction algorithms[J]. Neuroimage, 2004, 22(3):1255-1261.
[2] Le Goualher G; Argenti AM; Duyme M; Baaré WF; Hulshoff Pol HE; Boomsma DI; Zouaoui A; Barillot C; Evans AC. Statistical sulcal shape comparisons: application to the detection of genetic encoding of the central sulcus shape.[J]. Neuroimage, 2000, 11(5 Pt 1):564.
[3] Logothetis N K, Pauls J, Augath M, et al. Neurophysiological investigation of the basis of the fMRI signal[C]// Nature. 2001:150-157.
[4] Li Wang, Feng Shi, Pew-Thian Yap, et al. Longitudinally guided level sets for consistent tissue segmentation of neonates[J]. Human Brain Mapping, 2013, 34(4):956-972.
[5] Dunn J C. A fuzzy relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters[J]. Journal of Cybernetics, 1974, 3(3):32-57.
[6] Krinidis S, Chatzis V. A Robust Fuzzy Local Information C-Means Clustering Algorithm[J]. IEEE Transactions on Image Processing, 2010, 19(5):1328-1337.
[7] Zhang K, Liu Q, Song H, et al. A Variational Approach to Simultaneous Image Segmentation and Bias Correction[J]. IEEE Transactions on Cybernetics, 2017, 45(8):1426-1437.
[8] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553):436.
[9] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the Acm, 2012, 60(2):2012.
[10] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]// Computer Vision and Pattern Recognition. IEEE, 2015:1-9.
[11] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
[12] Dan C C, Giusti A, Gambardella L M, et al. Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images[J]. Advances in Neural Information Processing Systems, 2012, 25:2852--2860.
[13] Farabet C, Couprie C, Najman L, et al. Learning hierarchical features for scene labeling.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(8):1915.
[14] Ganin Y, Lempitsky V. \(N^4\), -Fields: Neural Network Nearest Neighbor Fields for Image Transforms[C]// Asian Conference on Computer Vision. Springer International Publishing, 2014:536-551.
[15] Gupta S, Girshick R, Arbeláez P, et al. Learning Rich Features from RGB-D Images for Object Detection and Segmentation[C]// European Conference on Computer Vision. Springer, Cham, 2014:345-360.
[16] Hariharan B, Arbeláez P, Girshick R, et al. Simultaneous Detection and Segmentation[J]. Lecture Notes in Computer Science, 2014, 8695:297-312.
[17] Ning F, Delhomme D, Lecun Y, et al. Toward automatic phenotyping of developing embryos from videos[J]. IEEE Transactions on Image Processing, 2005, 14(9):1360-71.
[18] Pinheiro P, Collobert R. Recurrent convolutional neural networks for scene labeling[C]//International Conference on Machine Learning. 2014: 82-90.
[19] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]// Computer Vision and Pattern Recognition. IEEE, 2015:3431-3440.
[20] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[M]// Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015. Springer International Publishing, 2015:234-241.
[21] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization[J]. Computer Science, 2014.
[22] Li C, Xu C, Anderson A W, et al. MRI Tissue Classification and Bias Field Estimation Based on Coherent Local Intensity Clustering: A Unified Energy Minimization Framework[M]// Information Processing in Medical Imaging. Springer Berlin Heidelberg, 2009:288-299.