**ISSN** 1746-7659, England, UK
**Journal of Information and Computing Science**
Vol. 14, No. 2, 2019, *pp*.156-160

# Residual and Dense Connection Combine Fully Convolutional Network for Infant Brain MRI Segmentation

Yuhang Qin[1], Mao Cai[1]

[1]School of math and statistics, Nanjing University of Information Science & Technology, Nanjing 210044, China
(Received January 03, 2019, accepted March 16, 2019)

**Abstract:** In the development of medical image segmentation, the application of convolutional neural networks has begun a profound revolution. The deep learning model is famous for excellent flexibility, efficiency and accuracy. The U-Net model is the beginning of task in the segmentation of medical images, which includes the basic operations of convolution, maxpooling, deconvolution, and concatenation. However, the U-Net model is disable to perform well on many types of data sets, because the model can't solve the exact segmentation of the details. We proposed Residual and Dense Fully Convolutional Network (RDFCN) that consist of Residual Connection Block and Dense Connection Block, which makes up for the shortcomings of U-Net. The dataset we used for training and testing comes from iSeg-2017 challenge (http://iseg2017.web.unc.edu). This dataset is comprised of infant(between 6 and 9 months of age) brain MR images. After the testing, our model outperforms the U-Net and some of its improved models in evaluation of WM, GM and CSF.

**Keywords:** medical image segmentation, infant brain MR images, Convolutional network, Residual Connection Block, Dense Connection Block

## 1. Introduction

With the development of medical imaging technology, the segmentation of medical images is becoming more and more important. More and more segmentation methods have been proposed to effectively solve various complex segmentation tasks, some of which are based on traditional statistical or partial differential equations. In recent years, with the development of computer technology, more and more methods based on deep learning have become mainstream.

Ronneberger et al.[1] created a deep network model of a U-shaped structure, namely U-Net. In the case of a small amount of data, the method can still obtain extremely high accuracy. In addition, this model also has strong noise immunity. However, the model also has some shortfalls: 1) Some details are easily lost in the segmentation result; 2) as the number of network layers deepens, the gradient will disappear and training will become difficult.

Zhou Z et al. improved the skip connection of U-Net model[2], namely the nested U-Net. They connect all the layers and enable network to learn different levels of features. Badrinarayanan V proposed the SegNet[3], which introduces the pooling indices instead of deconvolution in upsampling layers and reduces memory consumption during the training. Zhao H et al. created the PSPNet[4]. A pyramid pool block is introduced to fully obtain global information. Chen L C et al. proposed the DeepLab[5], which applies methods of atrous convolution and atrous spatial pyramid pooling(ASPP). They combine the final output of the deep network with the fully connected conditional random field(CRF) to make the location of the object boundary more precise.

II. RELATED WORK

Brain diseases affect human health and capture tens of thousands of lives every year. How to accurately analyze and effectively treat brain diseases has always been a hot topic in the medical field. The segmentation of biomedical images is an excellent means of assisted diagnosis and treatment. Many current methods demonstrate excellent performance in adult brain MR image segmentation tasks. Infant brain MR images have more serious intensity inhomogeneities, low contrast and weak boundary problems, whose segmentation challenge is greater. Traditional methods are disabled to accomplish such a segmentation task, so we need to create a deep learning model.

In the development process of deep learning, the deepening of the network layer usually leads to the disappearance of the gradient during the training process, the convergence becomes slower, and the classification accuracy is even worse. The proposal of ResNet has made this problem effectively solved[6]. The inception of a building block makes the network model focus on learning the residuals between input and output. Huang G et al. proposed DenseNet. This model creates a better feature extraction method: Dense block. Unlike ResNet, each layer in the block concatenates all of the previous feature maps as its additional input, which enhances feature reuse.

There are problems with the U-Net model: loss of details, gradient dispersion, and difficult improvement in accuracy. Therefore, we create a model called Residual and Dense Fully Convolutional Network(RDFCN) inspired by the idea of ResNet and DenseNet. We introduce Residual Connection Block and Dense Connection Block to optimize the model. Our main contributions are: 1) the Residual Connection Block helps the train speed up the convergence and avoid the disappearance of the gradient. 2) the Dense Connection Block enhances the reusability of features.3) our model achieved excellent performance on the test set

## 2. RDFCN

### *A. RDFCN model*

We propose an optimized U-Net model in this paper. Residual connections and dense connections are modularized for applying in full convolutional neural networks. The network is divided into an encoding path and a decoding path. Fig.2 clearly demonstrates the overall structure of our model. We will describe this residual connection block in detail in Section B. We apply the dense connections to the original feature fusion(concatenation), which will be elaborated in Section C.

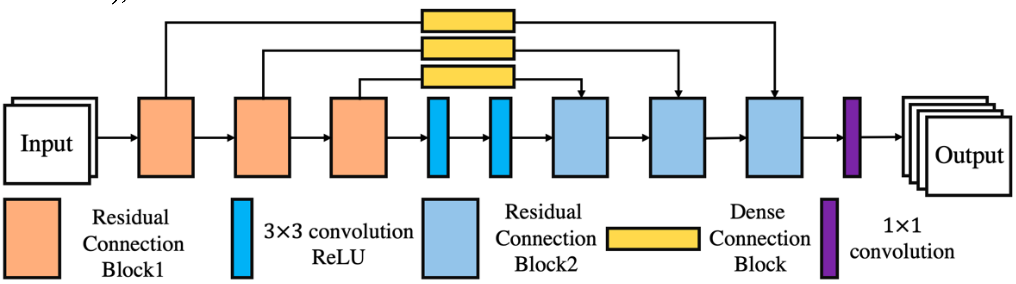


**Fig. 2.** The entire process is shown in the RDFCN structure map. The input to the model is a two-channel image. Residual Connection Block and Dense Connection Block are important parts of the structure, the details of them are shown in Fig. 3 and Fig. 4, respectively.

### *B. Residual Connection block*

The residual connection block in the structure is divided into two cases: residual connection block1 and residual connection block2. This design can avoid gradient dispersion and speed up the convergence of the model during the training process. We respectively represent the calculation results of the two blocks:

$$x_{l+1} = g_p([\mathcal{F}(f_{c3}(x_l)) + f_{c3}(x_l)]) \tag{1}$$

$$x_{l+1} = F[f_{c3}(g_d(x_l) \circ y_l)] + f_{c3}(g_d(x_l) \circ y_l) \tag{2}$$

Equation (1) and equation (2) represent the results of block1 and block2, respectively. where $x_l$ and $x_{l+1}$ are input and output of the $l$-th layer. The $g_p(\cdot)$ denotes the Maxpooling with pool-size 2. The $f_{c3}(\cdot)$ denotes the convolution function followed by ReLU, whose kernel is $3 \times 3$. The $\mathcal{F}(\cdot)$ represents the residual function with three convolutions followed by ReLU, the three kernels of which are $1 \times 1$, $3 \times 3$ and $1 \times 1$. The $g_d(\cdot)$ denotes deconvolution function with stride 2. $y_l$ is the corresponding result of Dense Connection Block. The $\circ$ is the concatenation.

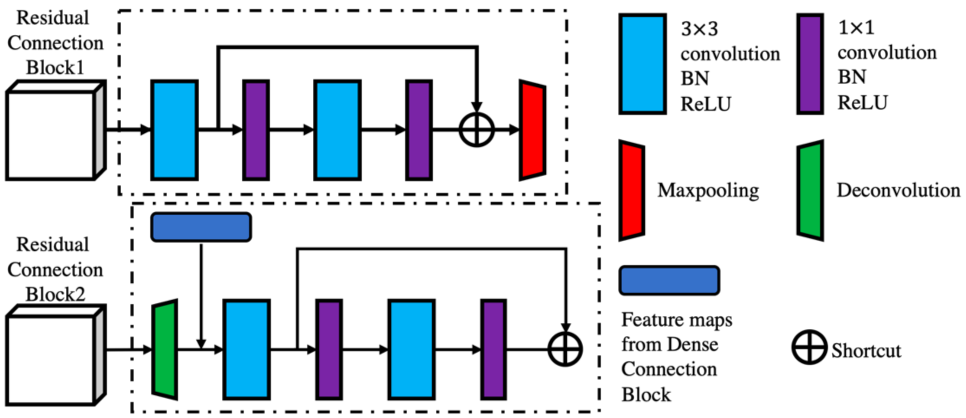


**Fig. 3.** The residual connection block introduces a residual connection, the upper half is the residual connection block1, and the lower half is the residual connection block 2.

***C. Dense Connection Block***

The design of this block is derived from the densenet[2]. In this block, dense connections are used between layers, which improves feature reuse and effectively suppresses overfitting. The calculation result of the block can be expressed as follows:

$$y_{l+1} = f_1{}'(y_l \circ f_1'(y_l) \circ f_2'(y_l) \circ f_3'(y_l) \circ \ldots \circ f_n'(y_l)) \quad (3)$$

where $y_l$ and $y_{l+1}$ represent input and output of block, respectively. The $f_n'(y_l)$ represents the result of the convolution of the $y_l$ after $n$ times, and each time the convolution is followed by BN[7] and ReLU. The $\circ$ denotes the concatenation.

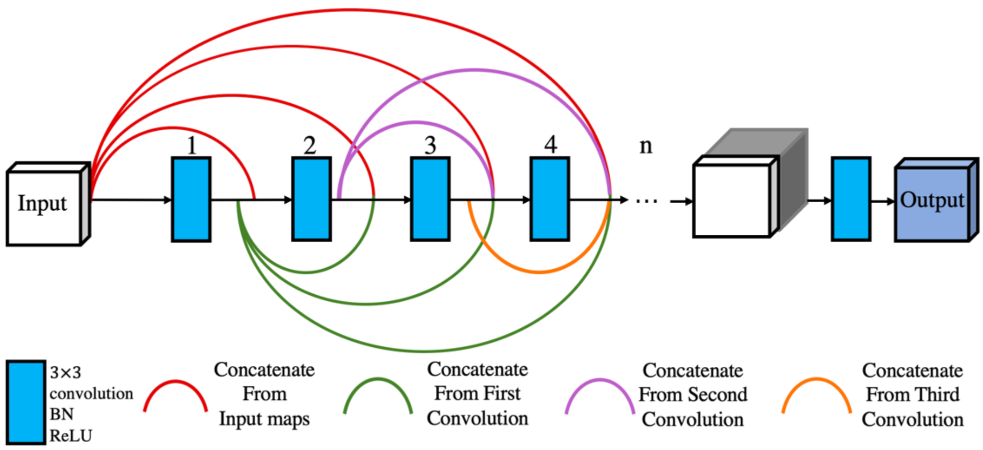


**Fig. 4.** The input of the dense connection block is derived from the output of the residual connection block before each downsampling. An n-layer block has a growth rate of n-1. The input of each layer concatenates the output of all previous layer.

***D. Model Training***

The data sets we chose for training and testing were from iSeg-2017 challenge (http://iseg2017.web.unc.edu), and the average age of these babies was 6 months[8] without any pathology. We concatenate the T1w images and the corresponding T2w images in the dataset as an input. We used the Adam optimizer in the model training. The batch size was set to 10 and the learning rate was set 0.0001. Furthermore, we adopt dropout and set the dropout[9] rate is 0.3 after the convolution and residual multiscale block. Our experimental code is based on Keras, and model is trained on a NVIDIA GeForce GTX 1080 Ti GPU(11GB).

## 3. Experiments

### *A. Evaluation criterions*

The Jaccard similarity(Js values)[10] is a metric, which can effectively help us evaluate the experimental results of the RDFCN and comparison methods. We express this metric: $Js(S_1, S_2) = (S_1 \cap S_2)/(S_1 \cup S_2)$, where $S_1$ represents the segmentation result and $S_2$ represents the ground truth.

In Table 1, our model has the highest Js values on CSF, GM and WM, indicating the improvement of RDFCN's performance in segmentation results.

**Table 1**: Comparison of mean Js values of CSF, GM, WM% in three methods

| | CSF(%) | GM(%) | WM(%) |
|---|---|---|---|
| U-Net[1] | 87.96 | 80.70 | 72.24 |
| RDFCN(our) | **88.25** | **81.49** | **74.58** |

We also calculate the accuracy(Acc) and variance(Var). The accuracy here removes the black background, which we define as: $Acc(S_1', S_2') = (S_1' \cap S_2')/S_2'$, where $S_1'$ indicates the segmentation result after the removed background, and $S_2'$ indicates the true value after the removed background. The variance to be calculated is the Js values. In Table 2, We find our model to have the highest accuracy and the lowest variance, which proves that our model is more robust.

**Table 2**: Methods of accuracy(%) and variances($10^{-4}$) of Js values

| | Acc | Var(Acc) | Var(CSF) | Var(GM) | Var(WM) |
|---|---|---|---|---|---|
| U-Net[1] | 90.01 | 6.21 | 19.65 | 16.76 | 137.23 |
| RDFCN(our) | **90.49** | **4.84** | **18.21** | **14.11** | **116.79** |

***B. Experiment Results***

We randomly selected a segmentation result from the experiment. In Fig. 5, the U-Net results show the fracture of the white matter belt, and our method can perform the segmentation well.

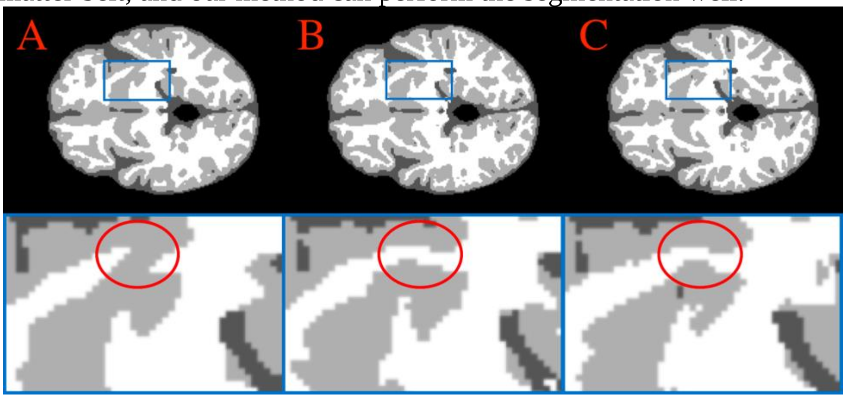


**Fig. 5.** A and B represent one of the segmentation results of U-Net and RDFCN, respectively, and C represents the label.

## 4. DISCUSSION AND CONCLUSION

In our model, we created a new residual connection module to replace the traditional convolution, which makes the model alleviate the problem of gradient dispersion and become easier to train. We also added dense connection blocks in the concatenation to enhance feature reuse, which allows the process of feature fusion to gather more information. These changes contribute to the improvement of network performance. Our model performed better than the other models in the experiment, and it was very obvious in the segmentation details.

## References

[1] 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.

[2] Zhou Z, Siddiquee M M R, Tajbakhsh N, et al. UNet++: A Nested U-Net Architecture for Medical Image Segmentation[J]. 2018.

[3] Badrinarayanan V , Kendall A , Cipolla R . SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017:1-1.

[4] Zhao H , Shi J , Qi X , et al. Pyramid Scene Parsing Network[J]. 2016.

[5] Chen L C , Papandreou G , Kokkinos I , et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs[J]. Computer Science, 2014(4):357-361.

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in CVPR, 2016, pp. 770–778.

[7] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in ICML, 2015, pp. 448–456.

[8] L. Wang, et al., "Longitudinally Guided Level Sets for Consistent Tissue Segmentation of Neonates," Human Brain Mapping, vol. 34, pp. 956-972, Apr 2013.

[9] N. Srivastava, et al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," The Journal of Machine Learning Research, vol. 15, pp. 1929-1958, 2014.

[10] 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.