Published in final edited form as:

Mach Learn Med Imaging. 2018 September ; 11046: 233–240. doi:10.1007/978-3-030-00919-9_27.

Automatic Accurate Infant Cerebellar Tissue Segmentation with Densely Connected Convolutional Network

Jiawei Chen, Han Zhang, Dong Nie, Li Wang, Gang Li, Weili Lin, and Dinggang Shen*

Department of Radiology and BRIC, University of North Carolina at Chapel Hill, USA

Abstract

The human cerebellum has been recognized as a key brain structure for motor control and cognitive function regulation. Investigation of brain functional development in the early life has recently been focusing on both cerebral and cerebellar development. Accurate segmentation of the infant cerebellum into different tissues is among the most important steps for quantitative development studies. However, this is extremely challenging due to the weak tissue contrast, extremely folded structures, and severe partial volume effect. To date, there are very few works touching infant cerebellum segmentation. We tackle this challenge by proposing a densely connected convolutional network to learn robust feature representations of different cerebellar tissues towards automatic and accurate segmentation. Specifically, we develop a novel deep neural network architecture by directly connecting all the layers to ensure maximum information flow even among distant layers in the network. This is distinct from all previous studies. Importantly, the outputs from all previous layers are passed to all subsequent layers as contextual features that can guide the segmentation. Our method achieved superior performance than other state-of-the-art methods when applied to Baby Connectome Project (BCP) data consisting of both 6- and 12-month-old infant brain images.

1 Introduction

The first year of the human life represents the most dynamic phase of postnatal brain development, with rapid brain size growth and cognitive function development. Most of previous early development studies have focused on the cerebral cortex [1], with few on the development of cerebellum. In fact, the human cerebellum has been recognized as playing equal important roles in both motor control and various cognitive function regulations as the cerebral cortex, many of which are key to daily living [2]. Accurate segmentation of the cerebellum into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) tissues is one of the most pivotal steps for quantitative analysis of early brain development, yet challenging for infant brain images. Manual segmentation of cerebellar tissues from 3D magnetic resonance (MR) images is extremely difficult and laborious, and often prone to biases or errors. It has been shown that, even manually segmented by experts, the results could still be less reproducible with unneglectable inter- and/or intra-operator variability. A more realistic and highly desired solution is computer-aided automatic cerebellum segmentation based on MR images; however, it is still challenging to date. Cerebellar cortex

*Corresponding Author: dgshen@med.unc.edu.

has far more complex structures with highly folded geometric patterns compared to the cerebral cortex, e.g., small cerebellar foliations (<0.5 mm) and comparable dense neurons. Moreover, due to ongoing maturation and largely immature myelination, the cerebellum has weak, even “isointense” contrast in MR images at 6–12 months of age. Furthermore, the small size of the cerebellum makes the partial volume effect more severe than that in the cerebral cortex; this becomes even worse for the already small-sized infant brain. Taking a 12-month-old cerebellum for example (Fig. 1, left), accurate WM, GM and CSF delineation in the blurred and isointense cerebellar cortex is among the most difficult medical image analysis problems [3]. To date, there are few studies on such topic.

Recently, deep convolutional neural networks (CNNs) [4] have achieved great successes in medical image segmentation [5, 6]. A U-shaped convolutional network [7], shorted for U-Net, was proposed for biomedical image segmentation, with a merit of preserving the resolution of the original images. The main spirit of the U-Net is that several expansive paths are built following a fully convolutional network, where respective deconvolution layers are employed to allow skipping the connections between low- and high-level features. Milletari *et al.* proposed a 3D fully convolutional network (FCN) [6] for volumetric image segmentation, but it suffers from a gradient vanishing problem that could undermine the model accuracy. To solve this issue, Chen *et al.* proposed a deep voxel-wise residual network [8], where more representative features are generated by integrating multimodal and multi-level contextual information to better guide tissue segmentation. Varying in topological structure, all these networks have similar short paths that connect only the previous and the later layers, making the input patches impossible to be fully convolved. To tackle this problem, a densely connected CNN (DenseNet) was proposed, where all layers are inter-connected directly and each layer receives extra inputs from all preceding layers and transits the feature representations to all subsequent layers [9]. The DenseNet has been extended for volumetric cardiac image segmentation in [10]. Inspired by the U-Net [7] and DenseNet [9], we propose a 3D, densely connected, U-shaped CNN to learn robust features for infant cerebellum segmentation. Specifically, in the contracting path, dense blocks are connected by the transition layers with a pooling operator to extract the context features. Our proposed network extends the densely connected convolution networks with an expansive path, along which deconvolution layers are placed to enforce a precise localization with the help from the context feature maps. The experimental results on the public Baby Connectome Project (BCP) dataset show a significant improvement by our proposed method in infant cerebellum segmentation compared to other state-of-the-art methods.

2 Method

2.1 Dataset and Preprocessing

We demonstrate the feasibility and superiority of our method with the MR images from 10 infants at the age of 12 months and 10 infants at the age of 6 months, selected from the BCP dataset. BCP was recently initiated for promoting the understanding of early brain development with state-of-the-art imaging sequences and protocols. The data were acquired by two 3T Siemens Prisma MRI scanners with the same settings at two different sites, Center for Magnetic Resonance Research (CMRR) at University of Minnesota and

Biomedical Research Imaging Center (BRIC) at University of North Carolina at Chapel Hill. T1-weighted MR images were acquired with TR = 2400 ms, TE = 2.24 ms, flip angle = 8°, and voxel resolution = 0.8×0.8×0.8 mm³.

Skull stripping and intensity inhomogeneity correction were performed to preprocess all the images. Ground-truth tissue segmentation results are important for model training and validation; however, manual labeling from the scratch is unrealistic. Therefore, we employed a widely adopted segmentation toolbox, LINKS [11], to obtain reasonable initial segmentation results, which were then manually edited to correct possible segmentation and geometric errors.

2.2 Network Architecture Design

Fig. 2 shows the flowchart of the proposed method. The network consists of 6 dense blocks, each of which can be defined as a nonlinear mapping function:

$$\mathbf{x}_l = H_l([\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{l-1}]) \quad (1)$$

where $[\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{l-1}]$ means the feature maps of all preceding layers (0,1,...,l-1) and transmits their own feature maps to all subsequent layers. It aims to improve the information flow between each pair of layers. The mapping function $H_l(\cdot)$ is the composition of three subsequent operations: batch normalization (BN), rectified linear unit (ReLU, a nonlinear activation operator), and 3×3×3 convolution. A feature map $\mathbf{x}_l$ can thus be obtained via the composite function $H_l([\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{l-1}])$.

These blocks are connected via transition layers into a U-shaped network with a contracting path and an expansive path. Along the contracting path, each transition layer between two dense blocks contains a batch normalization operator, a ReLU and an average pooling operator. The average pooling with a 2×2×2 volumetric kernel and a stride of 2 aims to reduce the resolution of the input feature map while increasing its receptive field. In contrast to max-pooling, for average pooling operator, the switches that map the outputs of a pooling layer to the corresponding inputs are no longer required in back-propagation, which can reduce memory consumption in the training stage.

On the other hand, in the expansive path, the average-pooling operator in a transition layer is replaced by a deconvolution operator to restore the resolution of the input feature maps. In addition, the deconvolution operator also allows one to assemble the context information from a lower resolution feature map to a higher resolution one. Furthermore, to ensure accurate segmentation along the image boundary, the deconvolutional feature map is combined along the expansive path with its mirrored image in the contracting path as the inputs of subsequent convolution layer. At the final layer, the convolution with a 1×1×1 kernel is employed to produce four feature maps (i.e., WM, GM, CSF and background) and they are then converted into probability maps by the following voxel-wise soft-max:

$$p_k(\mathbf{x}) = \exp(f_k(\mathbf{x})) / \sum_{i=1}^C \exp(f_i(\mathbf{x})) \quad (2)$$

where $\mathbf{x}$ is the position of a voxel, C is the total number of different tissues, and $f_k(\mathbf{x})$ is the intensity of the k -th feature map at voxel $\mathbf{x}$.

2.3 Network Training

The T1-weighted cerebellar MR intensity images of the training subjects (see Section 3 for cross validation parameters) and the corresponding cerebellar tissue segmentation ground truth are employed to train the network. The image of each subject is divided into patches with a size of $32 \times 32 \times 32$. Along the contracting path, the number of kernels is doubled at each transition layer, starting from 64 to 256. Along the expansive path, such a number is halved at each transition layer. In addition, the convolution layer is performed on the zero-padding maps to make the size of outputs identical to that of the inputs. Finally, these kernel parameters are optimized via a stochastic gradient descent algorithm implemented by Caffe [12], where the initial learning rate is 0.05 and is automatically decreased by 10% after each epoch.

3 Experimental Results

In the experiment, the training data is only derived from the subjects at the age of 12 months, where eight subjects are used for training, one subject for validation, and the remaining one for testing. For each subject, 2,000 patches are randomly selected; thus, for eight training subjects, we have a total of 16,000 patches for training, while, for one validation subject, we have 2,000 patches to monitor the overfitting issue. Then, the learning rate will decrease by 10% every 6,000 iterations. The performance is evaluated using Dice Ratio (DR), measuring the proportion of overlap voxels between the automatic segmentation map I_S and the ground truth map I_G :

$$DR = \frac{2|I_S \cap I_G|}{|I_S| + |I_G|} \quad (3)$$

3.1 Performance on 12-month-old Subjects

To validate our proposed method, we compare it with two state-of-the-art methods: LINKS [11] and U-Net [7] in the task of cerebellum segmentation for 12-month-old infants.

We first make qualitative comparisons with the results from LINKS and U-Net. A typical set of results is shown in Fig. 3. For the results from U-Net, there are several topological errors in the WM structure, while our proposed method obtains clearer and more reasonable WM structure. By comparing the 3D surface renderings of the GMWM interface among all the methods, we can see more “handles” (indicating segmentation errors) in the GM-WM interface structures generated by U-Net. To emphasize the differences among these methods, we zoom in a typical region at the cerebellar lobule VI. It can be observed that the

competing methods led to quite confusing boundary between WM and GM in this region, while our proposed method obtained a clearer and more reasonable WM structure.

For quantitative comparisons, we employ leave-one-out validation in the comparison experiment and evaluate the performance by the means and standard deviations (std) of the DRs among all the three methods. Our method achieved significantly higher DR for all three tissue types than other two methods (Table 1).

3.3 Performance on 6-month-old Infants

To test the generalization ability of our method, the model trained on ten 12-month-old subjects are directly applied to ten 6-month-old subjects. Of note, cerebellum segmentation on this age group is more challenging. For example, the T1-weighted MR image in Fig. 4(a) reaches the peak of isotropic intensity, for which even manual segmentation can be quite challenging. While other competing methods were significantly affected (see much worse result in Fig. 4(b, c)), our proposed method could still reliably and consistently generate decent tissue segmentation results, with much improved GM-WM surface reconstruction results. Such results indicate that densely connected convolutional network is capable to learn robust feature representations and provide comprehensive contextual guidance information for tissue classification.

4 Conclusions

In this paper, we developed a U-shaped densely connected convolutional network for infant cerebellar tissue segmentation. We directly connected all the layers with each other to ensure maximum information flow among even distant layers in the network, which provides comprehensive contextual guidance information for tissue classification in this challenging task. Compared to U-Net, our proposed network consists of dense blocks, thus avoiding gradient vanishing problem. Experimental results show that our method outperforms other state-of-the-art methods, even in the extremely challenging case of 6-month-old infant cerebellar tissue segmentation.

Acknowledgement

This work was supported by the National Institutes of Health (MH109773, MH100217, MH070890, EB006733, EB008374, EB009634, AG041721, AG042599, MH088520, MH108914, and MH107815). This work also utilizes approaches developed by an NIH grant (1U01MH110274) and the efforts of the UNC/UMN Baby Connectome Project Consortium.

References

1. Li G, Lin W, Gilmore JH, & Shen D (2015). Spatial patterns, longitudinal development, and hemispheric asymmetries of cortical thickness in infants from birth to 2 years of age. *Journal of Neuroscience*, 35(24), 9150–9162. [PubMed: 26085637]
2. Wolf U, Rapoport MJ, Schweizer TA (2009). “Evaluating the affective component of the cerebellar cognitive affective syndrome”. *J. Neuropsychiatry Clin. Neurosci* 21 (3), 245–53. [PubMed: 19776302]
3. Poretti A, Boltshauser E, & Huisman TA (2016). Pre- and postnatal neuroimaging of congenital cerebellar abnormalities. *The Cerebellum*, 15(1), 5–9. [PubMed: 26166429]

4. Long J, Shelhamer E, & Darrell T (2015). Fully convolutional networks for semantic segmentation In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3431–3440)
5. Ronneberger O, et al. (2015). “U-net: Convolutional networks for biomedical image segmentation” In International Conference on Medical Image Computing and Computer-Assisted Intervention, 234–241.
6. Milletari F, Navab N, & Ahmadi SA (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation In 3D Vision (3DV), 2016 Fourth International Conference on (pp. 565–571). IEEE
7. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, & Ronneberger O (2016). 3D U-Net: learning dense volumetric segmentation from sparse annotation In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 424–432). Springer International Publishing.
8. Chen H, Dou Q, Yu L, Qin J, & Heng PA (2017). VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage*.
9. Huang Gao, et al. (2016). “Densely connected convolutional networks.” arXiv preprint arXiv: 1608.06993.
10. Yu L, Cheng JZ, Dou Q, Yang X, Chen H, Qin J, & Heng PA (2017). Automatic 3d cardiovascular mr segmentation with densely-connected volumetric convnets In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 287–295). Springer, Cham.
11. Wang L, Gao Y, Shi F, Li G, Gilmore JH, Lin W, Shen D (2015). LINKS: learning-based multi-source IntegratioN frameworkK for Segmentation of infant brain images. *Neuroimage* 108, 160–172. [PubMed: 25541188]
12. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, & Darrell T (2014). Caffe: Convolutional architecture for fast feature embedding In Proceedings of the 22nd ACM international conference on Multimedia (pp. 675–678).

Fig. 1.
Infant cerebellar MR image with low tissue contrast, scanned at 12 months of age.

Fig. 2.
Illustration of the proposed densely connected convolutional network.

Fig. 3.

Tissue segmentation results based on a randomly selected 12-month-old infant. The first row shows a T1-weighted MR image. The second row shows the segmentation results from manual labeling (ground truth), LINKS [11], U-Net [7], and our method. The third row shows the surface rendering results of the GM-WM interface derived from respective methods.

Fig. 4.

Tissue segmentation results on a randomly selected 6-month-old infant. The first row shows a T1-weighted MR image. The second row shows the segmentation results by LINKS [11], U-Net [7], and our method. The last row shows the surface renderings results from the three methods.

Table 1.

Performance on 10 12-month-old subjects with leave-one-out cross-validation.

	WM	GM	CSF
LINKS	0.856±0.040	0.880±0.007	0.843±0.052
U-Net	0.862±0.023	0.886±0.005	0.846±0.034
Proposed	0.908±0.020	0.894±0.003	0.856±0.014