Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2018 April ; 2018: 692–695. doi:10.1109/ISBI.2018.8363668.

A Label-fusion-aided Convolutional Neural Network for Isointense Infant Brain Tissue Segmentation

Tengfei Li¹, Fan Zhou², Ziliang Zhu², Hai Shu¹, and Hongtu Zhu^{1,2}

¹Department of Biostatistics, The University of Texas MD Anderson Cancer Center Houston, TX 77030, USA

²Department of Biostatistics, The University of North Carolina at Chapel Hill Chapel Hill, NC 27599, USA

Abstract

The extremely low tissue contrast in white matter during an infant's isointense stage (6–8 months) of brain development presents major difficulty when segmenting brain image regions for analysis. We sought to develop a label-fusion-aided deep-learning approach for automatically segmenting isointense infant brain images into white matter, gray matter and cerebrospinal fluid using T1- and T2-weighted magnetic resonance images. A key idea of our approach is to apply the fully convolutional neural network (FCNN) to individual brain regions determined by a traditional registration-based segmentation method instead of training a single model for the whole brain. This provides more refined segmentation results by capturing more region-specific features. We show that this method outperforms traditional joint label fusion and FCNN-only methods in terms of Dice coefficients using the dataset from iSEG MICCAI Grand Challenge 2017.

Keywords

Brain segmentation; fully convolutional neural network; isointense stage; joint label fusion

1. INTRODUCTION

To replace the time-consuming and labor-intensive manual segmentation procedure for classifying images of infant brains into regions of white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF), automated segmentation approaches have been extensively studied [1, 2, 3]. Different from the adult brain, autosegmentation of the infant brain is especially challenging due to the reduced tissue contrast, poor spatial resolution, severe partial volume effect, and the ongoing maturation and myelination processes [2, 3]. The first-year infant brain is usually divided into three distinct stages: (i) infantile stage (5 months), (ii) isointense stage (6–8 months), and (iii) early adult-like stage (9 months). During the isointense stage, it is particularly difficult to discriminate WM from GM in brain magnetic resonance (MR) images because of overlapping areas with low tissue contrast. The traditional multi-atlas and label fusion (MALF) method uses the registration from a group of atlases with manually segmented labels. This approach has the limitation of treating different available imaging modalities equally and is often computationally expensive [3, 4].

Deep learning techniques have emerged as powerful methods for integrating the information from multiple modalities and capturing a wide range of highly discriminative imaging features. With an increased number of parameters and intensive memory and computational requirements, the 3-dimensional (3D) fully convolutional neural network (FCNN) offers more accurate predictivity by using more complex hierarchical features for its dense layers, small kernels and deeper architectures [5]. Patch-based FCNNs usually focus on the local tissue structures instead of the regional location information, while full-3D-image networks with end-to-end predictions often require multiple graphic processing units (GPUs), adequate random-access memory (RAM) to load a sufficient number of samples, and much longer time to train. To grasp both the local structure and global spatial information with limited computational power and RAM, we developed a label-fusion-aided convolutional neural network (LFA-CNN). The key idea is to apply the FCNN onto subregions that are defined by a traditional registration-based segmentation method such as MALF rather than training a single model for the whole brain, and then combine the segmentation results from those subregions by label fusion. In this way, the proposed approach captures more region-specific features and yields more refined segmentation results. The approach was evaluated by the dataset from the iSEG MICCAI Grand Challenge 2017, which consists of 10 training and 13 testing isointense infant brain subjects.

2. METHOD FRAMEWORK

To develop a fast and reliable automatic segmentation approach that is competitive with manual segmentation, we first used an atlas-based approach to obtain a rough segmentation result, and then applied the FCNN model onto each individual sub-region derived from the initial result to generate refined segmentation results.

For each subject, we need the T1-, T2-weighted MR images and manual segmentations for training purposes. Before we started, some preprocessing steps were necessary including within-subject T1 and T2 registration, resolution standardization, skull stripping, intensity inhomogeneity correction and removal of the cerebellum and brain stem. Example preprocessed images for a single subject from the iSEG Challenge are illustrated in Fig. 1.

2.1. Multi-atlas label fusion

After preprocessing, we used atlas-based segmentation to segment the WM, GM and CSF regions. This type of methods have an advantage over the level set method and watershed transform by borrowing prior knowledge about the shape and distribution of the segmented structures from a pre-segmented reference atlas [6, 7, 8].

Assume we have N subjects. For real data, N is the number of training subjects. A target image is segmented by treating images of all the other $N-1$ subjects as atlases. First, we performed rigid, affine and then diffeomorphism registration (with mutual information as the loss function) using the Advanced Normalization Tools software to align all the T1-weighted images from atlases to the target T1-weighted image [9]. We then transferred the annotation results from those atlases to the target space and obtained $N-1$ candidate labels for the target.

The next step is to combine the candidate labels by label fusion. Define the Normalized Cross-Correlation (NCC) between registered image I_1 and target image I_2 :

$$NCC = \frac{Cov(I_1, I_2)}{\sqrt{Var(I_1)}\sqrt{Var(I_2)}} \quad (1)$$

The NCC was computed in an $r \times r$ square local neighborhood around each voxel, which characterizes the I_1 - I_2 local similarity. In this way, we calculated the N -1 NCC maps between the N -1 atlases and the target T1-weighted image. Then, for a given voxel, from the N -1 registered local label maps, we chose the one with the largest NCC at that voxel as its segmented label. This method is referred to as the NCC-based local weighted voting in [8]. The workflow is shown in Fig. 2.

2.2. Integration of label fusion and FCNN

Based on the label fusion segmentation, we developed two strategies to determine the subregions for building FCNN models separately. Strategy A: training two FCNN models, where in the first model the extracted patches are centered within the areas with $NCC < 0.85$ while in the second model they were centered within those with $NCC > 0.85$. Strategy B: training three FCNN models using patches centered within the two-voxel-dilated WM, GM and CSF areas. See Fig. 3 for illustration. From the last subsection, we had MALF segmentation for target images a1) and a2). By thresholding the NCC map, we obtained b11) the ($NCC < 0.85$) area and b12) the ($NCC > 0.85$) area; by dilating the CSF, GM and WM areas by 2 voxels in a2), we obtained the dilated CSF in b21), dilated GM in b22) and dilated WM in b23). We can obtain those 5 areas in b11), b12), b21), b22) and b23) from the target image as well as from each atlas by alternating the atlas to a target. Next, we extracted patches from the five areas of each atlas to train 5 FCNN models and predicted the target labels at their corresponding areas. More details of the FCNN structure are shown in the next section. The predicted labels are shown in c11), c12), c21), c22) and c23) for the five areas, respectively. Finally we combined the predicted probability maps of different areas by the label fusion to obtain the final segmentation results, which are shown in d1) for Strategy A and d2) for Strategy B.

3. EXPERIMENTAL STUDY

We implemented our strategies on the dataset from the iSEG MICCAI Grand Challenge 2017, which is publicly available at <http://iseg2017.web.unc.edu/>. Selected scans for training and testing were acquired at the UNC-Chapel Hill and randomly chosen from the pilot study of Baby Connectome Project. All scans were acquired on a Siemens head-only 3T scanners with a circular polarized head coil. During the scan, infants were asleep, unsedated, fitted with ear protection, and their heads were secured in a vacuum-fixation device. It was followed by the image preprocessing steps described in Section 2. Reliable manual segmentations were first generated by a publicly available software iBEAT using infants' follow-up scans and then manually edited under the guidance of an experienced radiologist [10, 11].

We used 8 subjects for training the network, 2 for validation and 13 for testing. We made comparisons among the results of the following 7 methods: the naive MALF as Method 1; FCNN without MALF as Method 2; the combination of the two FCNN models using Strategy A as Method 3; the combination of the three FCNN models using Strategy B as Method 4; and the three FCNN models separately trained from the areas b21), b22) and b23) in Fig. 3 as Methods 5–7, respectively. Here for MALF we empirically chose $r=7$; for each FCNN model as above, we randomly selected 3500 voxels as frame centers from each labeled area of both T1- and T2-weighted images to extract 25×25 3D patches for each subject, and trained the FCNN framework according to Fig. 4. In this network, patches from T1 and T2 were concatenated, processed through 8 convolutional layers with a $3 \times 3 \times 3$ kernel size and one more with a $1 \times 1 \times 1$ kernel size before the final classification layer which uses softmax to produce the probability map in a $9 \times 9 \times 9$ region. In this way it enables us to predict the labeling of the $9 \times 9 \times 9$ voxels around the same center of the input $25 \times 25 \times 25$ cube. We also tried using the kernel size 5 but found size 3 worked better on the training set. Other tuning parameters such as patch size followed the recently published literature which used similar architecture and performed good at brain subcortical segmentation [12].

All models were trained within 24 hours on a GPU server using one Nvidia K80 card, 2496 CUDA cores, and 12GB RAMs, and a prediction in Fig. 3 only costs less than five minutes for each subject. We evaluated the Dice coefficients (DC) for WM, GM and CSF on validation dataset as in [13]. Finally, the performance of MALF, FCNN only and Strategy B on test set was evaluated and compared with the highest DCs in this challenge from the top ranking team “MSL_SKKU”. See the results in Table 1.

4. CONCLUSION

From Table 1, compared with the CSF in terms of the DC, the GM and WM were more difficult to discriminate; compared with MALF, deep-learning based methods performed much better; the Strategy-B-based FCNN method outperformed the FCNN-only method. A two-sample paired t-test shows that the Dice coefficients of all the three labels by Strategy B are significantly larger than using MALF and FCNN only, both with p-values $< 10^{-7}$. Compared with the top ranking team, it is competitive on CSF segmentation. The low standard deviation on test set implies this approach is robust. A comparison of Strategy B and manual annotations is shown in Fig. 5. By integrating information from the CSF, WM, and GM-based FCNN models, Strategy B refined the boundary between GM and WM. Another advantage is that it increased the sampled patch density by concentrating on each of the three areas (CSF, GM, and WM) with 1/3 of the number of patches for the whole brain, which can save RAM. We believe that using more complex brain parcellations to train a variety of FCNN models may capture more region-specific features, but training multiple models undoubtedly increases the computational burden.

5. REFERENCES

- [1]. Weisenfeld NI and Warfield SK, “Automatic Segmentation of Newborn Brain MRI”, *NeuroImage*, vol. 47, no. 2, pp. 564–572, 8 2009. [PubMed: 19409502]

- [2]. Wang L, Shi F, Gao Y, Li G, Gilmore JH, Lin W, and Anatomical D Constraint for Isointense Infant Brain MR Image Segmentation”, *NeuroImage*, vol. 89, pp. 152–164, 2014. [PubMed: 24291615]
- [3]. Wang L, Gao Y, Shi F, Li G, Gilmore JH, Lin W, and Shen D, “LINKS: Learning-based Multi-source Integration Framework for Segmentation of Infant Brain Images”, *NeuroImage*, vol. 108, pp. 160–172, 3 2015. [PubMed: 25541188]
- [4]. Wang H, Suh JW, Das SR, Pluta J, Craige C, and Yushkevich PA, “Multi-Atlas Segmentation with Joint Label Fusion”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 611–623, 3 2013. [PubMed: 22732662]
- [5]. Kamnitsas K, Ledig C, Newcombe VF, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B, “Efficient Multi-scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation”, *Medical Image Analysis*, vol. 36, pp. 61–78, 2 2017. [PubMed: 27865153]
- [6]. Sethian J, *Level Set Methods and Fast Marching Methods*, Cambridge Univ. Press, Cambridge, U.K., pp. 76–165, 1999.
- [7]. Dougherty E, *Mathematical Morphology in Image Processing*, Marcel Dekker, New York, U.S., pp. 433–482, 1992.
- [8]. Artaechevarria X, Munoz-Barrutia A, and Ortiz-de-Solorzano C, “Combination Strategies in Multi-atlas Image Segmentation: Application to Brain MR data”, *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1266–1277, 8 2009. [PubMed: 19228554]
- [9]. Tustison NJ, Cook PA, Klein A, Song G, Das SR, Duda JT, Kandel BM, van Strien N, Stone JR, Gee JC, and Avants BB, “Large-scale Evaluation of ANTs and FreeSurfer Cortical Thickness Measurements”, *NeuroImage*, vol. 99, pp. 166–179, Aug 10 2014. [PubMed: 24879923]
- [10]. Wang L, Shi F, Yap PT, Gilmore JH, Lin W, and Shen D, “4D Multi-Modality Tissue Segmentation of Serial Infant Images”, *PLoS One*, vol. 7, no. 9, pp. e44596, 2012. [PubMed: 23049751]
- [11]. Wang L, Shi F, Yap PT, Lin W, Gilmore JH, Shen D, “Longitudinally Guided Level Sets for Consistent Tissue Segmentation of Neonates”, *Human Brain Mapping*, vol. 34, no. 4, pp. 956–972, 2013. [PubMed: 22140029]
- [12]. Dolz J, Desrosiers C, and Ayed IB, (2017), “3D Fully Convolutional Networks for Subcortical Segmentation in MRI: A Large-scale Study”, *NeuroImage*, Apr 2017.
- [13]. Dice LR (1945), “Measures of the Amount of Ecologic Association between Species”, *Ecology*, vol. 26, no. 3, pp. 297–302.

Fig. 1.
T1-, T2-weighted and manual segmentation (left to right) preprocessed images from the same subject.

Fig. 2.
MALF workflow using T1-weighted MRI.

Fig. 3.
Workflow for Strategy A (top) and B (bottom).

Fig. 4.
FCNN network.

Fig. 5.
Comparison: segmentation of annotation, Strategy B, b21 (WM), b22 (GM), b23 (CSF) on the same area.

Table 1.

Comparison of Dice coefficients.

Mean Dice coefficient on validation set			
	WM	GM	CSF
MALF	0.762	0.804	0.864
FCNN only	0.881	0.900	0.949
LFA-CNN (Strat. A)	0.884	0.903	0.951
LFA-CNN (Strat. B)	0.898	0.914	0.958
LFA-CNN (Strat. B b21)	0.874	0.895	0.952
LFA-CNN (Strat. B b22)	0.891	0.908	0.951
LFA-CNN (Strat. B b23)	0.889	0.906	0.951
Mean (SD) of Dice coefficients on test set			
MALF	0.771 (0.015)	0.820 (0.011)	0.874 (0.010)
FCNN only	0.867 (0.018)	0.892 (0.010)	0.948 (0.006)
LFA-CNN (Strat. B)	0.882 (0.019)	0.904 (0.011)	0.956 (0.005)
iSEG highest	0.901 (0.016)	0.919 (0.009)	0.958 (0.005)