Published in final edited form as:

Med Image Comput Assist Interv. 2019 October ; 11766: 484–491.

doi:10.1007/978-3-030-32248-9_54.

Hierarchical Parcellation of the Cerebellum

Shuo Han¹, Aaron Carass^{2,3}, Jerry L. Prince^{2,3}

¹Department of Biomedical Engineering, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

²Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA

³Department of Computer Science, The Johns Hopkins University, Baltimore, MD 21218, USA

Abstract

Parcellation of the cerebellum in an MR image has been used to study regional associations with both motion and cognitive functions. Despite the fact that the division of the cerebellum is defined hierarchically—i.e., the cerebellum can be divided into lobes and the lobes can be further divided into lobules—previous automatic methods to parcellate the cerebellum do not utilize this information. In this work, we propose a method based on convolutional neural networks (CNNs) to explicitly incorporate the hierarchical organization of the cerebellum. The network is constructed in a tree structure with each node representing a cerebellar region and having child nodes that further subdivide the region into finer substructures. Thus, our CNN is aware of the hierarchical organization of the cerebellum. Furthermore, by selecting tree nodes to represent the hierarchical properties of a given training sample, our network can be trained with heterogeneous training data that are labeled to different hierarchical depths. The proposed method was compared with a state-of-the-art cerebellum parcellation network. Our approach shows promising results as a first parcellation method to take the cerebellar hierarchical organization into consideration.

1 Introduction

The cerebellum of the brain plays an important role in both motor and cognitive functions [2]. Parcellating the cerebellum into its subregions from a structural magnetic resonance (MR) image can be used to study the topological mapping of its functions and characterize its morphological differences between groups [7,10]. Like the cerebrum, the anatomical organization of the human cerebellum is hierarchically defined [3]. It can be first divided into the corpus medullare and the gray matter, the gray matter can then be divided into the anterior, superior posterior, inferior posterior, and flocculonodular lobes. The lobes themselves can be further divided into lobules, primarily as I through X but with further subdivisions possible—VIIB, for example. Previous automatic methods to parcellate the cerebellum are mainly based on multi-atlas segmentation [4,9,12] and supervised machine learning [6,12]. However, unlike manual delineation protocols [1], none of these automatic methods explicitly utilize the hierarchical organization of the cerebellum.

Convolutional neural networks (CNNs) can achieve state-of-the-art accuracy for semantic segmentation (parcellating the cerebellum, for example [6]). Recently, Liang et al. [8] incorporated semantic hierarchy concepts to construct a tree-structured CNN to achieve improved performance for segmentation. Based on their work, we explicitly built a cerebellar hierarchical organization into our 3D CNN. Our 3D network is comprised of a feature generator and a predictor. At each voxel, our network generates features that are used to perform the corresponding hierarchical classification. The predictor is implemented within a tree data structure. Each node of the tree detects a cerebellar region with child nodes subdividing it into finer subregions. For example, the first node in the hierarchy tree differentiates the cerebellum from the background; the cerebellum is then broken down into the corpus medullare and the gray matter. See Fig. 1 for the complete hierarchy. We note that different data sets label the cerebellum using different hierarchies [3]. As in [8], these different data sets can be easily used simultaneously to train the network by selecting different subsets of the tree nodes. Despite the differences in the hierarchies of the training data, they can contribute to the training of the nodes that the hierarchies have in common. The performance of our network was compared with a state-of-the-art method [6], and it shows promising results.

2 Methods

2.1 Cerebellar Hierarchical Organization

We have constructed a cerebellar hierarchy, shown in Fig. 1, which is compatible with the Adult and Pediatric cohorts in [3]. In this tree structure, a node represents a cerebellar region and its child nodes correspond to its subdivisions. At Level 1, an image is divided into the cerebellum and the background. Level 2 subdivides the cerebellum into the corpus medullare (CM) and the gray matter (GM). In Level 3, the gray matter is divided into the anterior (AL), superior posterior (SPL), and inferior posterior lobes (IPL)—our IPL also incorporates the flocculonodular lobe, similar to the manual delineation protocol of Bogovic et al. [1]. At Level 4, lobes are broken down into left, right, and vermal. Finally, the hemispherical lobes are divided into lobules, with some lobules being grouped together to be compatible with the Pediatric data [3], such as VIIIA and VIIIB grouped as VIII. We note that the AL is handled differently in both cohorts; the Adult data divides the AL into left and right, whereas the Pediatric data also has the vermis. The child nodes of AL are swapped out when working with the two different data sets.

2.2 Network Architectures

Our 3D network is comprised of a feature generator and a predictor. The feature generator is modified from the state-of-the-art cerebellum parcellation network in [6] to generate 32 features for each voxel and its output layers are replaced by our predictor network. As in [8], the predictor is constructed using a tree structure corresponding to the cerebellar hierarchical organization shown in Fig. 1 and each tree node uses a projection convolution to convert the input features into a single-channel image for binary classification of the corresponding region to distinguish it from the remaining labels. Note that the classification in each node is performed separately and does not compete with its sibling nodes. The whole prediction is then done recursively. Two variations of the predictor are reported for comparison purposes.

The first one, *Identity Predictor*, simply takes the same 32 features from the feature generator to perform classification at all nodes in the tree. The second one, *Dense Predictor*, has a similar architecture to [8]. Each node has an additional encoding block to generate two features from its input and only the two features are used by the following projection convolution for the binary classification. Its child nodes then take the concatenation of all ancestors' newly generated two-features with the features from the feature generator as the input. Their architectures are illustrated in Fig. 2A and B. To compare their performance with the state-of-the-art network in [6], a third predictor, *Multi-head Predictor*, is used. For use with multiple data sets, this network simply uses different projection convolutions for each data set to directly classify all the presented labels (Fig. 2C).

2.3 Dynamic Selection of Predictor Nodes

The predictor tree is the union of all the possible manual delineation hierarchies used by the different data sets. Therefore, only a subset of the nodes in the predictor are selected for a particular sample during training. In other words, the training loss is only calculated over the available hierarchy concepts of this sample and only the parameters of the corresponding nodes are updated in back-propagation. For example, if the data set only labels the cerebellar lobes rather than the lobules, although the predictor has classifiers for detecting the lobules, their parameters would not be updated. As a result, the proposed method can efficiently incorporate data sets that are delineated with different manual protocols. Similarly, during inference, subsets of the predictor nodes can be chosen in order to produce the parcellation of different hierarchy levels, which provides flexibility to study different levels of cerebellar anatomical structure.

2.4 Training and Inference

To generate the truth binary image for each node of the predictor during training, the label values of a node are recursively unioned from the label values of its child nodes. Since different data sets define different hierarchies, different predictor nodes must be selected. As a result, back-propagation of the training would be inefficient if images from different data sets are combined into a single mini-batch. Therefore, for a fixed batch size, we select—at random—our training images from the same data set. Although these data sets have different hierarchies, they always share the shallower portions of the hierarchy and differ only in the deeper levels. For example, both data sets separate the cerebellum into CM and GM but one can lack the subdivision of a certain lobule. Consequently, the parameters of the deeper nodes of the predictor tree are updated less frequently compared to the shallower nodes. Therefore, we use different learning rates for different nodes according to the occurrences of the corresponding regions in the training data. Suppose the learning rate of the feature generator is 0.002; then for a node that only presents in half of the training images, its learning rate is set to 0.004. We train our network (the feature generator as well as the predictor) from scratch using the Dice loss I ,

$$I = 1 - \frac{1}{N} \sum_i \frac{2 \sum_j^M \text{sigmoid}(p_{ij}) q_{ij} + \epsilon}{\sum_j^M \text{sigmoid}(p_{ij}) + \sum_j^M q_{ij} + \epsilon}, \quad (1)$$

where N is the number of selected nodes from the predictor tree, M is the number of voxels, p_{ij} is the network output for voxel j at node i , q_{ij} is the truth for voxel j at node i , and $\epsilon = 1 \times 10^{-8}$ prevents division by zero.

For inference, only the outputs of the leaf nodes are used. These outputs are concatenated channel-wise and softmax is applied to convert them into a label probability map. The label of the channel with the largest probability is assigned to the voxel in the final classification.

3 Experiments and Results

3.1 Data

Magnetization-prepared rapid acquisition with gradient echo (MPRAGE) images from the Adult and Pediatric data sets [3] were used to train and test the proposed method. N4 [11] was applied to correct the inhomogeneity and the Pediatric data set was rigidly registered to the 1 mm isotropic ICBM 2009c template [5] in MNI space. The images were then resized to $192 \times 256 \times 192$ by zero-padding. Three spinocerebellar ataxia subtype 6 (SCA6) subjects and two healthy controls from the Adult data set and ten random subjects from the Pediatric data set were selected as the testing data. The remaining twenty images (ten from each data set) were used as the training data.

3.2 Training and Testing

To train our network, the training images were cropped to a size of $128 \times 96 \times 96$ around the manual delineation of the cerebellum. The whole cropped-out region was then used as the input to the proposed networks. For the testing images, instead of using a cerebellum mask in MNI space to crop around the cerebellum, we used another cerebellum-locating network modified from the U-Net in [6] with less parameters to more accurately put the cerebellum at the center of the cropped-out region. This network takes the whole 3D image as the input and outputs a binary prediction of the cerebellum. The testing images were then cropped to $128 \times 96 \times 96$ around the largest connected component of this network output. The Adam optimizer was used with the learning rate of the feature generator equal to 0.002 and other parameters were default. The batch size was 2. No data augmentation was used for comparing these networks.

We first trained the networks only on the Pediatric data set but tested on both data sets. All three networks were trained for 600 epochs. The Dice coefficients between the network outputs and the manual delineation were evaluated (Table 1). For the Pediatric data set, all five hierarchy levels were evaluated. For the Adult data set, since its delineation protocol is different from the Pediatric data set's and thus there is no truth available for the last two levels, only the first three levels were evaluated.

We then trained the networks on both data sets. The Multi-head Predictor and the Identity Predictor were trained for 300 epochs. The Dense Predictor was trained for 500 epochs. The Dice coefficients between the network outputs and the manual delineation are shown in Table 1. Two-sided paired Wilcoxon tests were performed between the Multi-head Predictor and the Dense Predictor for each label and for each data set. The Adult data set did not show any statistical differences. For the Pediatric data set, two regions, the corpus medullare and

the right lobule Crus II/VIIIB, were statistically improved ($p < 0.05$) for the Dense Predictor, and one region, vermis inferior posterior, was statistically better ($p < 0.05$) for the Multi-head Predictor for the Pediatric data set. Other regions were comparable. The visual comparison between these three predictors is further shown in Fig. 3.

For the Pediatric data set, the Dice coefficients of the double-data-set training are not always better than those of the single-data-set training, despite the fact that it had more training data; for the Adult data set, the Dice coefficients of the double-data-set training are better than those of the single-data-set training in each available level.

4 Discussion and Conclusions

In this work, a tree-structured network was used to explicitly incorporate the cerebellar hierarchical organization into parcellation and also efficiently take different data sets simultaneously as training data. To improve the performance, an additional loss function could be used during the training to further encourage the agreement between the output of a node and the outputs of its child nodes. Instead of using binary classification separately, the sibling nodes could be trained to learn to compete with each other via multi-label classification. During inference, the prediction of a node could be more explicitly involved in the prediction of its child nodes, for example, with conditional probability. Although the proposed method was only comparable to a state-of-the-art network to parcellate the cerebellum, it shows promising results as a first step to explicitly take the anatomical hierarchy into the design of the cerebellum parcellation algorithm.

References

1. Bogovic JA, et al.: Approaching expert results using a hierarchical cerebellum parcellation protocol for multiple inexpert human raters. *NeuroImage* 64, 616–629 (2013) [PubMed: 22975160]
2. Buckner RL: The cerebellum and cognitive function: 25 years of insight from anatomy and neuroimaging. *Neuron* 80(3), 807–815 (2013) [PubMed: 24183029]
3. Carass A, et al.: Comparing fully automated state-of-the-art cerebellum parcellation from magnetic resonance images. *NeuroImage* 183, 150–172 (2018) [PubMed: 30099076]
4. Diedrichsen J, Balsters JH, Flavell J, Cussans E, Ramnani N: A probabilistic MR atlas of the human cerebellum. *NeuroImage* 46(1), 39–46 (2009) [PubMed: 19457380]
5. Fonov VS, Evans AC, McKinstry RC, Almlí C, Collins D: Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage* 47, S102 (2009)
6. Han S, He Y, Carass A, Ying SH, Prince JL: Cerebellum parcellation with convolutional neural networks. In: *Medical Imaging 2019: Image Processing*. vol. 10949, p. 109490K International Society for Optics and Photonics (2019)
7. Kansal K, et al.: Structural cerebellar correlates of cognitive and motor dysfunctions in cerebellar degeneration. *Brain* 140(3), 707–720 (2016)
8. Liang X, Zhou H, Xing E: Dynamic-structured semantic propagation network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 752–761 (2018)
9. Romero JE, et al.: CERES: a new cerebellum lobule segmentation method. *NeuroImage* 147, 916–924 (2017) [PubMed: 27833012]
10. Steele CJ, Chakravarty MM: Gray-matter structural variability in the human cerebellum: lobule-specific differences across sex and hemisphere. *NeuroImage* 170, 164–173 (2018) [PubMed: 28461060]
11. Tustison NJ, et al.: N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* 29(6), 1310–1320 (2010) [PubMed: 20378467]

12. Yang Z, et al.: Automated cerebellar lobule segmentation with application to cerebellar structural analysis in cerebellar disease. *NeuroImage* 127, 435–444 (2016) [PubMed: 26408861]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1. Cerebellar hierarchical organization. The color in a label image is marked on the side of each region. Example parcellation of (A) the Adult and the (B) Pediatric cohort is also shown at the bottom. AL: anterior lobe, BG: background, CE: cerebellum, CM: corpus medullare, GM: gray matter, IPL: inferior posterior lobe, and SPL: superior posterior lobe.

Fig. 2.

Predictor architectures: (A) the Identity Predictor, (B) the Dense Predictor, and (C) the Multi-head Predictor. Only part of the tree structure of (A) and (B) is shown.

Fig. 3.

Example parcellation of two testing images from the Adult and the Pediatric data sets, respectively. From top to bottom: Level 1, Level 4, and Level 5 hierarchies. (A) the image, (B) the truth parcellation, (C) the results of the Multi-head Predictor, (D) the results of the Identity Predictor and (E) the results of the Dense Predictor.

Table 1.

Dice coefficients of the single-data-set and double-data-set training. The Dice coefficients are averaged across all labels and all subjects. The largest Dice coefficients among the three methods are highlighted.

			Level 1	Level 2	Level 3	Level 4	Level 5
Single	Adult	Multi-head Identity Dense	0.8374	0.8435	0.7952	NA	NA
			0.8613	0.8587	0.8138	NA	NA
			0.8585	0.8550	0.8024	NA	NA
	Pediatric	Multi-head Identity Dense	0.9617 0.9649 0.9632	0.9314 0.9229 0.9246	0.9033 0.9013 0.8996	0.8669 0.8668 0.8653	0.8465 0.8471 0.8454
Double	Adult	Multi-head Identity Dense	0.9617	0.9246	0.8980	0.8826	0.7951
			0.9540	0.9231	0.8897	0.8679	0.7790
			0.9603	0.9266	0.8988	0.8802	0.7847
	Pediatric	Multi-head Identity Dense	0.9621 0.9637 0.9631	0.9217 0.9265 0.9246	0.8984 0.9013 0.8988	0.8648 0.8667 0.8651	0.8456 0.8491 0.8470