Published in final edited form as:

Intell Med. 2024 May ; 4(2): 65–74. doi:10.1016/j.imed.2023.05.002.

An end-to-end infant brain parcellation pipeline

Limei Wang,
Yue Sun,
Weili Lin,
Gang Li,
Li Wang[*]

Department of Radiology and Biomedical Research Imaging Center, UNC-Chapel Hill, New Caledonia, 27599, USA

Abstract

Objective—Accurate infant brain parcellation is crucial for understanding early brain development; however, it is challenging due to the inherent low tissue contrast, high noise, and severe partial volume effects in infant magnetic resonance images (MRIs). The aim of this study was to develop an end-to-end pipeline that enabled accurate parcellation of infant brain MRIs.

Methods—We proposed an end-to-end pipeline that employs a two-stage global-to-local approach for accurate parcellation of infant brain MRIs. Specifically, in the global regions of interest (ROIs) localization stage, a combination of transformer and convolution operations was employed to capture both global spatial features and fine texture features, enabling an approximate localization of the ROIs across the whole brain. In the local ROIs refinement stage, leveraging the position priors from the first stage along with the raw MRIs, the boundaries o the ROIs are refined for a more accurate parcellation.

Results—We utilized the Dice ratio to evaluate the accuracy of parcellation results. Results on 263 subjects from National Database for Autism Research (NDAR), Baby Connectome Project (BCP) and Cross-site datasets demonstrated the better accuracy and robustness of our method than other competing methods.

Conclusion—Our end-to-end pipeline may be capable of accurately parcellating 6-month-old infant brain MRIs.

[*]Corresponding author: Li Wang, Department of Radiology and Biomedical Research Imaging Center, UNC-Chapel Hill, New Caledonia, 27599, USA (li_wang@med.unc.edu).

Author contributions

Limei Wang: Conceptualization, Data curation, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. **Yue Sun:** Validation, Writing – review & editing. **Weili Lin:** Resources, Funding acquisition. **Gang Li:** Resources, Funding acquisition. **Li Wang:** Resources, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Conflicts of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Keywords

Infant brain parcellation; Isointense magnetic resonance image; Convolutional neural networks; Transformer

1. Introduction

The parcellation of infant brain [1–2] MRIs is crucial for quantifying early brain development and analyzing brain structural and functional networks [3–4]. Over the years, numerous parcellation approaches have been proposed, such as atlas-based methods [5–8], clustering-based methods [9–11], graph-based methods [12–13], statistical methods [14–16], and surface-based methods [17–19]. Nevertheless, many of these methods fail to produce accurate results for infant brain MRIs. Recently, researchers have developed advanced, fine-grained brain parcellation techniques based on convolutional neural networks (CNNs). For instance, Fang et al. [20] and Tang et al. [21] successively proposed multi-atlas guided CNNs for brain image labeling. Hong et al. [22] introduced a spatially localized atlas network tile (SLANT) method to reduce computational demands. Coupe et al. [23] proposed an ensemble learning approach, named as AssemblyNet, for coarse-to-fine parcellation.

Although the aforementioned methods have demonstrated improved performance, they suffer from four significant limitations. First, most existing parcellation approaches were developed for adult brains and are inapplicable for infant brain MRIs with low tissue contrast, blurred boundaries, and dynamic appearance changes. Figure 1 illustrates the most challenging 6-month-old infant brains [24–25], with extremely low contrast between white matter (WM) and gray matter (GM), and blurred boundaries between different regions in the cortical regions [26–29]. In this work, we will focus on such challenging 6-month-old infant brain MRIs parcellation task [27]. Second, most current infant brain parcellation methods follow a step-by-step manner. As shown in Table 1, Infant FreeSurfer [24] and iBEAT [25] are the only methods that can process 6-month-old infant MRIs, and both involve several steps, such as image preprocessing, skull stripping, cerebellum removal, tissue segmentation, and parcellation (Figure 2). A failure in any of these steps can result in a flawed and unreliable parcellation outcome. Third, the CNNs-based architecture is often limited by the local receptive field, making it difficult to perceive the spatial distribution of the whole brain. Finally, deep-learning-based methods are susceptible to cross-site issues due to variations in scanners between sites, resulting in inconsistent performance across different sites [27].

Recently, transformer architectures have gained attention for their ability to capture global spatial information [35–37]. Inspired by this advance and to address the above limitations, we propose an “end-to-end” pipeline for infant brain MRIs parcellation, which employs a global-to-local strategy with two stages: global regions of interest (ROIs) localization and local ROIs refinement. In the first stage, we use a combination of transformer and convolution operations to roughly locate ROIs throughout the brain. In the second stage, we refine the boundaries of the ROIs using position priors generated in the first stage and raw MRIs, resulting in a more precise parcellation. Notably, our pipeline requires minimal preprocessing (i.e., N4 inhomogeneity correction [38]), as shown in Figure 2, reducing potential errors from preprocessing steps. We evaluated the effectiveness of our method using 263 subjects from four 6-month-old infant MRI datasets and compared it with other competing methods.

The paper is structured as follows. Section 2 provides an overview of the datasets used and introduces the proposed pipeline. In Section 3, we present the results from experiments on different 6-month-old datasets. Finally, Section 4 offers a discussion of the results.

2. Methods

2.1. Dataset acquisition and image preprocessing

In this research, T1w and T2w MRIs were obtained from multiple datasets, as detailed in Table 2 along with the respective imaging parameters. All imaging protocols and studies were approved by the Institutional Review Boards (IRBs) at each clinical site, and written informed consent was obtained from all parents or legal guardians of the infants. At the time of scanning, all infants were approximately 6 months old.

NDAR dataset: 330 subjects were randomly selected from the NDAR [39] and were acquired with both T1w and T2w MR images, using a Siemens scanner equipped with a 12-channel head coil. During the scan, the infants were asleep and their heads were fixed in a vacuum device. According to the annotation protocols for the cerebrum (http://www.neuromorphometrics.com/ParcellationProtocol-2010-0405.PDF) and cerebellum [40], each subject was manually divided into 146 ROIs, with 129 ROIs for the cerebrum and 17 ROIs for the cerebellum.

BCP dataset: 83 unlabeled subjects with T1w and T2w MR images were from the BCP [41]. All images were acquired on Siemens scanners while infants were naturally sleeping, wearing ear protection and a head vacuum-fixation device.

Cross-site dataset: 10 unlabeled subjects with T1w and T2w MR images were collected from two clinical sites with distinct imaging protocols and scanners. Five subjects were imaged using a GE scanner, while the other five were imaged using a Philips scanner.

For the image preprocessing, two basic steps were carried out: uniform resampling to a resolution of 1.0 × 1.0 × 1.0 mm^3 and correction of intensity inhomogeneity [38]. 160 subjects from the NDAR dataset were utilized for training. The remaining subjects from multiple sites were used for testing, including 170 subjects from the NDAR, 83 subjects from the BCP, and 10 subjects from the Cross-site.

2.2. Methods

Our end-to-end pipeline is following a global-to-local strategy with two stages: (1) global ROIs localization, and (2) local ROIs refinement, as depicted in Figure 3. The first stage utilizes a combination of transformer and convolution operations to extract both global spatial features and local texture features comprehensively. This allows for the rough localization of ROIs within the whole brain space. In the second stage, the convolution operation is used to further refine the boundaries of the ROIs by leveraging the global personalized position priors obtained in the first stage, in conjunction with the raw MRIs. This refinement results in a more accurate identification of the ROIs.

2.2.1. Global ROIs localization stage—

The global ROIs localization network is designed to explore the spatial information of ROIs throughout the whole brain space. The network is an extension of the 3D UNet architecture [42] and incorporates transformer and convolution operations in the encoder to integrate multi-scale spatial information, as shown in Figure 4. Each layer in the encoder of the network consists of a transformer module (TM) and a convolution module (CM). The TM captures global spatial information by representing sequences, while the CM extracts local texture details through local receptive fields. These features are concatenated and utilized in subsequent operations to achieve a comprehensive extraction of spatial information. To generate precise position priors for each ROI in the whole brain space, the global ROIs localization network is operated on whole-brain images that contain complete spatial information. The network then outputs a corresponding probability map for each ROI. To balance performance and GPU memory, the resolution of the whole-brain images is downsampled from 1 × 1 × 1 mm^3 to 2 × 2 × 2 $mm^{3.}$ The loss function of the global ROIs localization network combines Soft Dice loss and Cross-Entropy loss [43]. Table 3 provides the specific configurations of TM and CM, which will be further discussed in the following sections.

<u>Transformer Module.:</u> Vision Transformer (ViT) [35–37] has successfully adapted the transformer architecture from the field of natural language processing to computer vision. It captures global spatial information by inputting an embedded one-dimensional sequence, allowing for a comprehensive understanding of the spatial relationships between different regions of an image. Given an input volume $x \in R^{(H \times W \times D \times C)}$ with a size of $(H,\ W,\ D)$ and C channels, we first divide the image into flattened non-overlapping patches $x_v \in R^{\left(N \times \left(P^3 \cdot C\right)\right)}$, where $(P,\ P,\ P)$ denotes the size of patch and $N = \left(H \times W \times D\right)/P^3$ is the length of the sequence. Then, these patches x_v are mapped into d-dimensional embedding space using a trainable linear projection. In order to encode patch spatial information, 1D learnable positional embedding $E_{pos} \in R^{(N \times d)}$ is added to the patch embeddings $E \in R^{\left(\left(P^3 \cdot C\right) \times d\right)}$ as follows:

$$z_0 = \left[x_v^1 E; x_v^2 E; \ldots; x_v^N E\right] + E_{pos} \tag{1}$$

After the embedding layer, we perform a layer normalization operation $Norm()$ on the sequence of embeddings, and then employ a multi-head self-attention (MSA) mechanism [44–45] to attend to different parts of the sequence with varying weights, which can be written as:

$$z_i^{'} = MSA\left(Norm\left(z_{(i-1)}\right)\right) + z_{(i-1)}, \quad i = 1 \ldots L \tag{2}$$

where i denotes the identifier of the transformer block and L denotes the number of transformer operations. Specifically, MSA learns the mapping between the query (q) and the key (k) in a sequence $z \in R^{N \times d}$ to measure the attention weights A as following [46–47]:

$$A = Softmax\left(\left(qk^{T}\right)/\sqrt{d}\right) \tag{3}$$

where d is the dimension of embedding space used to maintain qk^{T} as a constant value across different space scales. According to the computed attention weights A, the output of MSA can be expressed as Av, where v is the value representations in a sequence $z \in R^{N \times d}$.

The output of MSA adds its input sequence of embeddings z_{i-1} as an intermediate result $z_{i}^{'}$ to participate in the subsequent progression. The $z_{i}^{'}$ is normalized first by layers and then by performed multilayer perceptron (MLP), which can be defined as:

$$z_{i} = MLP\left(Norm\left(z_{i}^{'}\right)\right) + z_{i}^{'}, \quad i = 1 \ldots L \tag{4}$$

The MLP contains two linear layers with GELU activation functions, and the summation z_{i} of its input and output is the final feature output of the transformer block.

To reconcile the difference between the embedding and convolution space, we transform the feature tensor of the transformer from the embedding space to the convolution space. The reshaped transformer features are then combined with the convolutional features from the same layer, resulting in the final output of the feature extraction.

Convolution Module.: Inspired by the proven success of UNet for a multitude of medical image segmentation tasks, the design of our Convolution Module is based on a dual convolution structure. Both convolution operations have a uniform implementation, employing 3 × 3 × 3 kernels with zero padding, followed by an instance normalization (IN) layer and a LeakyRelu activation function.

Additionally, in the training stage, we split the training set into two subsets of equal size and train two separate global ROIs localization subnetworks. The cross-tested probability maps generated by these subnetworks are then used in the training of local ROIs refinement network.

2.2.2. Local ROIs refinement stage—

To achieve precise differentiation of the boundaries of localized ROIs, we employ a local ROIs refinement model that operates in the original resolution space, leveraging the probability map of each ROI from the first stage and the raw MRIs. Firstly, we upsample the probability maps of ROIs to the original resolution ($1 \times 1 \times 1$ mm^{3}) and concatenate them with the raw MRIs. Then, this concatenated data is used to train the local ROIs refinement network, which produces an optimized parcellation result. For this model, we use the 3D UNet architecture [42], renowned for its ability to preserve fine details through consecutive convolution operations. During training, we employ the Cross-Entropy loss function.

2.2.3. Implementation details—The proposed global ROIs localization network and the local ROIs refinement network were implemented on a single Tesla V100-SXM2 GPU (16GB) utilizing the PyTorch framework. For the global ROIs localization network, the ViT-B16 architecture from MONAI (https://doi.org/10.5281/zenodo.4323059) was adopted and optimized using the AdamW optimizer. The learning rate was set at 0.0001, with a warmup cosine annealing decay. For the local ROIs refinement network, the AdamW optimizer was used with a learning rate of 0.01, and a warmup cosine annealing decay was applied [48–49].

2.2.4. Evaluation metrics—The parcellation results were quantitatively evaluated by the $Dice$ ratio, including Dice_{BR}, which reflects the average parcellation performance of all brain regions, and $Dice_{BR}$, which represents parcellation performance weighted by the proportion of each region in the whole brain volume. The $Dice_{BR}$ ratio was defined as:

$$Dice_{BR}(G_i, P_i) = \frac{2 \times G_i P_i}{G_i + P_i} \tag{5}$$

where G_i and P_i denote the number of voxels in ith brain region from the ground truth and prediction result, respectively. The $Dice_{WB}$ ratio was calculated by:

$$\begin{aligned} Dice_{WB}(G, P) &= \sum_{i=1}^{N} \frac{G_i}{G} \times \frac{2 \times G_i P_i}{G_i + P_i} \\ &= \sum_{i=1}^{N} \frac{G_i}{G} \times Dice_{BR}(G_i, P_i) \end{aligned} \tag{6}$$

where G and P denotes the number of voxels of the whole brain from ground truth and prediction result, and N is the number of brain regions.

3. Results

To thoroughly evaluate the performance of our pipeline on infant brain parcellation, we validated it on a total of 263 subjects aged 6 months from four datasets. These included 170 subjects acquired by Siemens scanners from the NDAR dataset, 83 subjects acquired by Siemens scanners from the BCP dataset, 5 subjects acquired by a Philips scanner, and 5 subjects acquired by a GE scanner.

Competing Methods.

We compared the performance of our pipeline with three state-of-the-art brain parcellation methods, including Infant FreeSurfer [24], SLANT [22], and AssemblyNet [23]. Infant FreeSurfer is a well-established infant brain image processing tool, while AssemblyNet and SLANT have demonstrated effective performance for parcellation of lifespan datasets (1–90 years old) and children's dataset, respectively. Since the SLANT and AssemblyNet methods can only be applied to T1w MRIs, all parcellation methods in the comparison were conducted on T1w MRIs to ensure a fair comparison. To ensure consistency in the ROIs of the parcellation results for comparison, we merged certain labels from our results to match those of Infant FreeSurfer, SLANT, and AssemblyNet.

3.1. Quantitative evaluation on the NDAR dataset

To evaluate the parcellation performance of our pipeline, we tested it on 170 six-month-old subjects from the NDAR dataset. The whole brain was divided into 129 functional regions for the cerebrum and 17 functional regions for the cerebellum. The average $Dice_{BR}$ ratios of 170 subjects across 146 ROIs in the cerebrum and cerebellum are presented in Figure 5. Our pipeline demonstrates favorable performance in almost all brain regions, with over 80% accuracy in 101 of 129 cerebral regions and 14 of 17 cerebellar regions.

We compared our pipeline with three competing methods (Infant FreeSurfer, SLANT, and AssemblyNet) by merging certain ROIs from our parcellation. Figure 6 shows the parcellation results obtained by our method and the other three competing methods on a randomly selected subject from the NDAR dataset, with the upper three rows corresponding to the cerebrum and the bottom three rows corresponding to the cerebellum. Table 4 presents the average quantitative results of our method and the competing methods for $Dice_{BR}$ and $Dice_{WB}$ on 170 subjects from the NDAR dataset in both the cerebrum and cerebellum. Visual inspection of the cerebrum parcellation results in Figure 6 reveals that our pipeline demonstrates excellent performance in brain parcellation for 6-month-old infants, with higher overall consistency with ground truth. Specifically, the parcellation results of Infant FreeSurfer, SLANT, and AssemblyNet showed a significant portion of missing WM compared to the ground truth, while our pipeline obtained a more complete and reasonable WM, as shown in the third row of Figure 6. In terms of GM extraction, our method preserved better brain sulci and gyri structures compared to the other three competing methods, demonstrating the advantage of our pipeline in cortical parcellation, as shown in the second row of Figure 6. The quantitative analysis in Table 4 further demonstrates the superiority of our pipeline over the other competing methods in infant brain parcellation, with higher $Dice_{BR}$ and $Dice_{WB}$ ratios.

A more detailed comparison of the parcellation results for the cerebellum, as shown in the lower three rows of Figure 6, further confirms the superiority of our method in white matter (WM) extraction compared to the competing methods. This advantage is also reflected in the higher $Dice_{BR}/Dice_{WB}$ ratios achieved by our method.

3.2. Qualitative evaluation on the BCP dataset

Figure 7 displays a visual comparison between our pipeline and three other competing methods on a randomly selected subject from the BCP dataset. The top three rows depict the cerebrum, and the bottom three rows depict the cerebellum. While ground truth parcellation is unavailable, we can still assess performance through visual inspection. Our pipeline extracts more reasonable WM structures that are more consistent with the T1w MRIs, as shown in the third and sixth rows, while the other competing methods fail to extract WM due to the low tissue contrast in the T1w MRIs. Furthermore, our method outperforms the other methods in dividing the cortical area, as evidenced in the second and fifth rows, producing more reasonable cortical folding.

3.3. Qualitative evaluation on the cross-site data with different scanners

To evaluate the performance of our parcellation pipeline on imaging data from different scanners, we tested it on 5 Philips scanner imaging data and 5 GE scanner imaging data. Figure 8 shows the parcellation results for the cerebrum and cerebellum acquired from each respective scanner. Through visual inspection, our pipeline effectively divided the brain into functional/anatomical regions with correct sulcus and gyrus structures. Furthermore, the parcellation results for the WM in the fourth column are structurally sound and consistent with the T1w MRIs. Overall, our pipeline demonstrates robustness to imaging differences caused by different scanners and produces satisfactory parcellation results in both the cerebrum and cerebellum.

4. Discussion

In the global ROIs localization stage, our pipeline utilizes a combination of transformer module (TM) and convolution module (CM) to reinforce an accurate understanding of the spatial position of ROIs in the whole brain space. For simplicity, we refer to the network combining TM and CM as TCNet. To quantitatively evaluate the performance improvement attributable to TCNet, we compared its parcellation results with those of UNet [42] and UNETR [43] methods on the NDAR dataset. UNet only uses convolution operations to capture features, while UN-ETR only extracts features based on transformer operations. Specifically, TCNet, UNet, and UNETR all utilize the whole-brain image-based learning approach and were trained on 160 downsampled subjects from NDAR, each with a size of 96 × 96 × 96 voxels (2 × 2 × 2 mm^3). We assessed their parcellation performance using 170 testing subjects from the NDAR dataset. Table 5 presents the average $Dice_{BR}$ and $Dice_{WB}$ ratios of all brain regions for all 170 subjects, and Figure 9 displays the average $Dice_{BR}$ ratios for all subjects corresponding to 146 brain regions. As shown in Figure 9, UNETR correctly located all brain regions, while UNet failed in some regions, demonstrating the positive effect of transformers on acquiring position information in the whole brain space. Furthermore, for the regions that were correctly located, UNet achieved a higher $Dice_{BR}$ ratio, indicating that detailed information acquired through convolution operations plays a significant role in the quality of the parcellation results. In contrast, TCNet, which leverages both TM and CM, not only correctly located all brain regions but also outperformed both UNETR and UNet in terms of $Dice_{BR}$ and $Dice_{WB}$. These results demonstrate that the combination of TM and CM, as proposed, can indeed enhance the performance of parcellation in infant brain.

Our proposed pipeline consists of two stages: first, locating the position of ROIs in the whole brain space, and second, fine-tuning the boundary details. To demonstrate the advantage of this two-stage approach over a one-stage approach, we conducted two groups of experiments on the NDAR dataset using TCNet. One group used a patch-wise strategy

(TCNet), and the other group adopted our global-to-local framework (f-TCNet). TCNet was trained on 160 NDAR training subjects in the original space by building patches of size 64 × 64 × 64 voxels (1 × 1 × 1 mm^3). For f-TCNet, the global ROIs localization network was first trained on 160 downsampled whole-brain images from NDAR with a size of 96 × 96 × 96 voxels (2 × 2 × 2 mm^3), and personalized probability maps were generated corresponding to each brain region. Using these probability maps generated in the first stage, together with raw MRIs, f-TCNet then trained the local ROIs refinement stage by cropping patches. We used 170 subjects from the NDAR dataset to evaluate the parcellation performance of TCNet and f-TCNet. Figure 10 provides the parcellation results obtained by TCNet and f-TCNet on one randomly selected image from the NDAR dataset and Table 6 presents the average $Dice_{BR}$ and $Dice_{WB}$ ratios of all subjects. The results show that f-TCNet provides more precise parcellation results that are consistent with the ground truth. This is because the patch-wise strategy fails to consider the spatial position of ROIs in the whole brain space, leading to inaccurate or missing ROIs, particularly cortical area. In contrast, the f-TCNet method can successfully locate each ROI’s position in the whole brain space and further achieve more accurate parcellation results, with the most significant performance improvement in the cortical region, as evidenced by higher $Dice_{BR}$.

In this section, we thoroughly evaluated the effectiveness of incorporating multimodal images (i.e., T1w and T2w MRIs) on our pipeline’s performance by conducting experiments on 170 subjects from the NDAR dataset. Table 7 presents the average quantitative results of our pipeline on T1w+T2w MRIs and T1w MRIs. The results show that our method achieves slightly higher parcellation accuracy on multimodal images than on T1w MRI alone, in both the cerebrum and cerebellum. Moreover, the variance of the parcellation results obtained with multimodal images is smaller, indicating that the incorporation of T1w and T2w MRIs enhances parcellation performance.

We have developed an end-to-end pipeline for parcellating infant brains using raw MRIs. Our pipeline consists of two stages: a global ROIs localization stage and a local ROIs refinement stage. In the global ROIs localization stage, we utilized a combination of transformer and convolution operations to integrate global spatial features and local texture features. The latter stage focused on improving boundary details using convolution. Our pipeline was tested on 263 subjects aged 6 months from various sites, and the results showed that our pipeline outperformed other published pipelines and was robust to the cross-site issue.

However, there are still some limitations to our pipeline. Firstly, in the global ROIs localization stage, we divided the TM and CM components into two independent paths, which could be further optimized in future research. Secondly, a unified framework that integrates global position information and local boundary details from whole-brain images would lead to better parcellation results. However, this approach requires careful consideration of computational cost and parcellation accuracy.

Acknowledgments

A portion of the data used in this study was obtained from the National Database for Autism Research (NDAR), which is a collaborative informatics network established by the NIH to enhance autism research through providing access to a national resource. The findings presented in this paper represent the views of the authors and may not reflect the opinions or perspectives of the NIH or the original contributors to NDAR.

Funding

This research was partially funded by National Institutes of Health (Grant Nos. MH117943, MH109773, MH116225, and MH123202). Additionally, the work leverages approaches developed through an National Institutes of Health (Grant No. 1U01MH110274) and the efforts of the Baby Connectome Project Consortium at UNC/UMN.

Data and code availability statement

The raw image data and associated information for the datasets utilized in this study are publicly accessible through the following links: NDAR (https://nda.nih.gov/edit_collection.html?id=19/) and BCP (https://nda.nih.gov/edit_collection.html?id=2848/). The cross-site dataset includes ten in-house subjects that can be obtained upon request. Our source code and model are available (https://github.com/DBC-Lab/An-End-to-end-Infant-Brain-Parcellation-Pipeline.git).

References

[1]. Eickhoff SB, Yeo B, Genon S. Imaging-based parcellations of the human brain. Nat Rev Neurosci 2018;19(11):672–86. doi:10.1038/s41583-018-0071-7. [PubMed: 30305712]

[2]. Knickmeyer RC, Gouttard S, Kang C, et al. A structural MRI study of human brain development from birth to 2 years. NeuroSci 2008;28(47):12176–82. doi:10.1523/JNEUROSCI.3479-08.2008.

[3]. Wang L, Gao Y, Shi F, et al. learning-based multi-source integration framework for segmentation of infant brain images. Neuroimage 2015;108:160–72. doi:10.1016/j.neuroimage.2014.12.042. [PubMed: 25541188]

[4]. Wang L, Shi F, Gao Y, et al. Integration of sparse multi-modality representation and anatomical constraint for isointense infant brain MR image segmentation. Neuroimage 2014;89:152–64. doi:10.1016/j.neuroimage.2013.11.040. [PubMed: 24291615]

[5]. Power JD, Cohen AL, Nelson SM, et al. Functional network organization of the human brain. Neuron 2011;72(4):665–78. doi:10.1016/j.neuron.2011.09.006. [PubMed: 22099467]

[6]. Craddock RC, James GA, Holtzheimer PE 3rd, et al. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum Brain Mapp 2012;33(8):1914–28. doi:10.1002/hbm.21333. [PubMed: 21769991]

[7]. Iglesias JE, Sabuncu MR. Multi-atlas segmentation of biomedical images: a survey. Med Image Anal 2015;24(1):205–19. doi:10.1016/j.media.2015.06.012. [PubMed: 26201875]

[8]. Artaechevarria X, Munoz-Barrutia A, Ortiz-de Solorzano C. Combination strategies in multi-atlas image segmentation: application to brain mr data. IEEE Trans Med Imaging 2009;28(8):1266–77. doi:10.1109/TMI.2009.2014372. [PubMed: 19228554]

[9]. Nanetti L, Cerliani L, Gazzola V, et al. Group analyses of connectivity-based cortical parcellation using repeated k-means clustering. Neuroimage 2009;47(4):1666–77. doi:10.1016/j.neuroimage.2009.06.014. [PubMed: 19524682]

[10]. Luo Z, Zeng LL, Qin J, et al. Functional parcellation of human brain precuneus using density-based clustering. Cereb Cortex 2020;30(1):269–82. doi:10.1093/cercor/bhz086. [PubMed: 31044223]

[11]. Dillon K, Wang YP. Resolution-based spectral clustering for brain parcellation using functional MRI. J Neurosci Methods 2020;335:108628. doi:10.1016/j.jneumeth.2020.108628. [PubMed: 32035090]

[12]. Gopinath K, Desrosiers C, Lombaert H. Graph convolutions on spectral embeddings for cortical surface parcellation. Med Image Anal 2019;54:297–305. doi:10.1016/j.media.2019.03.012. [PubMed: 30974398]

[13]. Shen X, Papademetris X, Constable RT. Graph-theory based parcellation of functional subunits in the brain from resting-state fMRI data. Neuroimage 2010;50(3):1027–35. doi:10.1016/j.neuroimage.2009.12.119. [PubMed: 20060479]

[14]. Pohl KM, Bouix S, Nakamura M, et al. A hierarchical algorithm for mr brain image parcellation. IEEE Trans Med Imaging 2007;26(9):1201–12. doi:10.1109/TMI.2007.901433. [PubMed: 17896593]

[15]. Shan ZY, Yue GH, Liu JZ, et al. Automated histogram-based brain segmentation in T1-weighted three-dimensional magnetic resonance head images. Neuroimage 2002;17(3):1587–98. doi:10.1006/nimg.2002.1287. [PubMed: 12414297]

[16]. Arslan S, Ktena SI, Makropoulos A, et al. Human brain mapping: a systematic comparison of parcellation methods for the human cerebral cortex. Neuroimage 2018;170:5–30. doi:10.1016/j.neuroimage.2017.04.014. [PubMed: 28412442]

[17]. Makris N, Kaiser J, Haselgrove C, et al. Human cerebral cortex: a system for the integration of volume-and surface-based representations. Neuroimage 2006;33(1):139–53. doi:10.1016/j.neuroimage.2006.04.220. [PubMed: 16920366]

[18]. Adamson CL, Alexander B, Ball G, et al. Parcellation of the neonatal cortex using surface-based melbourne children’s regional infant brain atlases (M-CRIB-S). Sci Rep 2020;10(1):1–11. doi:10.1038/s41598-020-61326-2. [PubMed: 31913322]

[19]. Cole M, Murray K, St-Onge E, et al. Surface-based connectivity integration: an atlas-free approach to jointly study functional and structural connectivity. Hum Brain Mapp 2021;42(11):3481–99. doi:10.1002/hbm.25447. [PubMed: 33956380]

[20]. Fang L, Zhang L, Nie D, et al. Automatic brain labeling via multi-atlas guided fully convolutional networks. Med Image Anal 2019;51:157–68. doi:10.1016/j.media.2018.10.012. [PubMed: 30447544]

[21]. Tang Z, Liu X, Li Y, et al. Multi-atlas brain parcellation using squeeze-and-excitation fully convolutional networks. IEEE Trans Image Process 2020;29:6864–72.

[22]. Huo Y, Xu Z, Xiong Y, et al. 3D whole brain segmentation using spatially localized atlas network tiles. Neuroimage 2019;194:105–19. doi:10.1016/j.neuroimage.2019.03.041. [PubMed: 30910724]

[23]. Coupé P, Mansencal B, Clément M, et al. AssemblyNet: a large ensemble of CNNs for 3D whole brain MRI segmentation. Neuroimage 2020;219:117026. doi:10.1016/j.neuroimage.2020.117026. [PubMed: 32522665]

[24]. Zöllei L, Iglesias JE, Ou Y, et al. Infant FreeSurfer: an automated segmentation and surface extraction pipeline for T1-weighted neuroimaging data of infants 0–2 years. Neuroimage 2020;218:116946. doi:10.1016/j.neuroimage.2020.116946. [PubMed: 32442637]

[25]. Dai Y, Shi F, Wang L, et al. iBEAT: a toolbox for infant brain magnetic resonance image processing. Neuroinformatics 2013;11(2):211–25. doi:10.1007/s12021-012-9164-z. [PubMed: 23055044]

[26]. Wang L, Nie D, Li G, et al. Benchmark on automatic six-month-old infant brain segmentation algorithms: the iSeg-2017 challenge. IEEE Trans Med Imaging 2019;38(9):2219–30. doi:10.1109/TMI.2019.2901712.

[27]. Sun Y, Gao K, Wu Z, et al. Multi-site infant brain segmentation algorithms: the iSeg-2019 challenge. IEEE Trans Med Imaging 2021;40(5):1363–76. doi:10.1109/TMI.2021.3055428. [PubMed: 33507867]

[28]. Wang L, Gao Y, Li G, et al. Latest: local adaptive and sequential training for tissue segmentation of isointense infant brain MR images. In: Medical computer vision and Bayesian and graphical models for biomedical imaging. Springer; 2016. p. 26–34.

[29]. Wang L, Li G, Adeli E, et al. Anatomy-guided joint tissue segmentation and topological correction for 6-month infant brain MRI with risk of autism. Hum Brain Mapp 2018;39(6):2609–23. doi:10.1002/hbm.24027. [PubMed: 29516625]

[30]. de Brebisson A, Montana G. Deep neural networks for anatomical brain segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops; 2015. p. 20–8.

[31]. Henschel L, Conjeti S, Estrada S, et al. FastSurfer-a fast and accurate deep learning based neuroimaging pipeline. Neuroimage 2020;219:117012. doi:10.1016/j.neuroimage.2020.117012. [PubMed: 32526386]

[32]. Li W, Wang G, Fidon L, et al. On the compactness, efficiency, and representation of 3D convolutional networks: brain parcellation as a pretext task. In: International conference on information processing in medical imaging. Springer; 2017. p. 348–60.

[33]. Wang H, Yushkevich PA. Multi-atlas segmentation with joint label fusion and corrective learning-an open source implementation. Front Neuroinform 2013;7:27. doi:10.3389/fninf.2013.00027. [PubMed: 24319427]

[34]. Fischl B. Freesurfer. Neuroimage 2012;62(2):774–81. doi:10.1016/j.neuroimage.2012.01.021. [PubMed: 22248573]

[35]. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale. 2020. arXiv:201011929.

[36]. Wang W, Xie E, Li X, et al. Pyramid vision transformer: a versatile backbone for dense prediction without convolutions. Proceedings of the IEEE/CVF international conference on computer vision; 2021. p. 568–578. doi:10.48550/arXiv.2102.12122.

[37]. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Adv Neural Inf Process Syst 2017;30.

[38]. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging 2010;29(6):1310–20. doi:10.1109/TMI.2010.2046908. [PubMed: 20378467]

[39]. Hazlett HC, Gu H, McKinstry RC, et al. Brain volume findings in 6-month-old infants at high familial risk for autism. Am J Psychiatry 2012;169(6):601–8. doi:10.1176/appi.ajp.2012.11091425. [PubMed: 22684595]

[40]. Bogovic JA, Jedynak B, Rigg R, et al. Approaching expert results using a hierarchical cerebellum parcellation protocol for multiple inexpert human raters. Neuroimage 2013;64:616–29. doi:10.1016/j.neuroimage.2012.08.075. [PubMed: 22975160]

[41]. Howell BR, Styner MA, Gao W, et al. The UNC/UMN baby connectome project (BCP): an overview of the study design and protocol development. Neuroimage 2019;185:891–905. doi:10.1016/j.neuroimage.2018.03.049. [PubMed: 29578031]

[42]. Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation. Springer; 2016. p. 424–32. doi:10.48550/arXiv.1606.06650.

[43]. Hatamizadeh A, Tang Y, Nath V, et al. UNETR: transformers for 3D medical image segmentation. Proceedings of the IEEE/CVF winter conference on applications of computer vision; 2022. p. 574–584. doi:10.48550/arXiv.2103.10504.

[44]. Dobko M, Kolinko DI, Viniavskyi O, et al. Combining CNNs with transformer for multimodal 3D MRI brain tumor segmentation with self-supervised pretraining. 2021. arXiv:211007919.

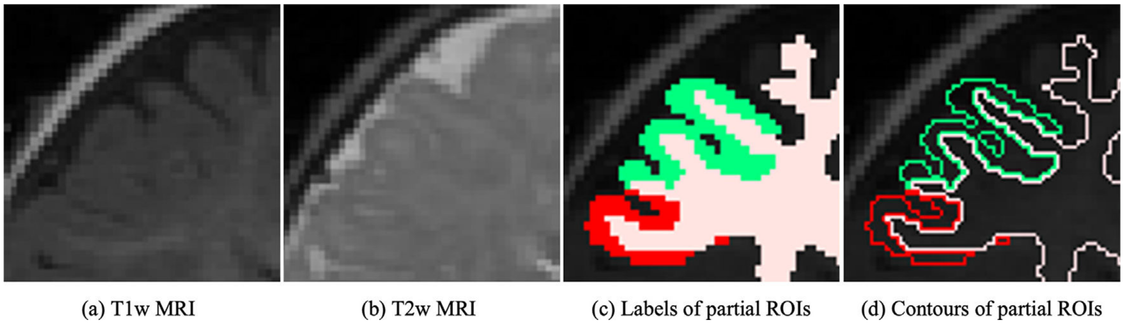
[45]. Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision; 2021. p. 10012–22.

[46]. Zhou HY, Guo J, Zhang Y, et al. nnFormer: interleaved transformer for volumetric segmentation. 2021. arXiv:210903201.

[47]. Chen B, Liu Y, Zhang Z, et al. TransAttUnet: multi-level attention-guided U-Net with transformer for medical image segmentation. 2021. arXiv:210705274.

[48]. Dubost F, Yilmaz P, Adams H, et al. Enlarged perivascular spaces in brain MRI: automated quantification in four regions. Neuroimage 2019;185:534–44. doi:10.1016/j.neuroimage.2018. [PubMed: 30326293]

[49]. Gotmare A, Keskar NS, Xiong C, et al. A closer look at deep learning heuristics: learning rate restarts, warmup and distillation. 2018. arXiv:181013243.



(a) T1w MRI

(b) T2w MRI

(c) Labels of partial ROIs

(d) Contours of partial ROIs

Figure 1.

Isointense infant images (5–8 months old) exhibit extremely-low tissue contrast due to inherent myelination and maturation, with (a) T1w MRI, (b) T2w MRI, (c) regions of interest (ROIs), and (d) contours of selected ROIs. The color-coding represents different ROIs, with green for the left middle frontal cortex, red for the left triangular portion of the inferior frontal gyrus, and flesh pink for the left cerebral white matter.

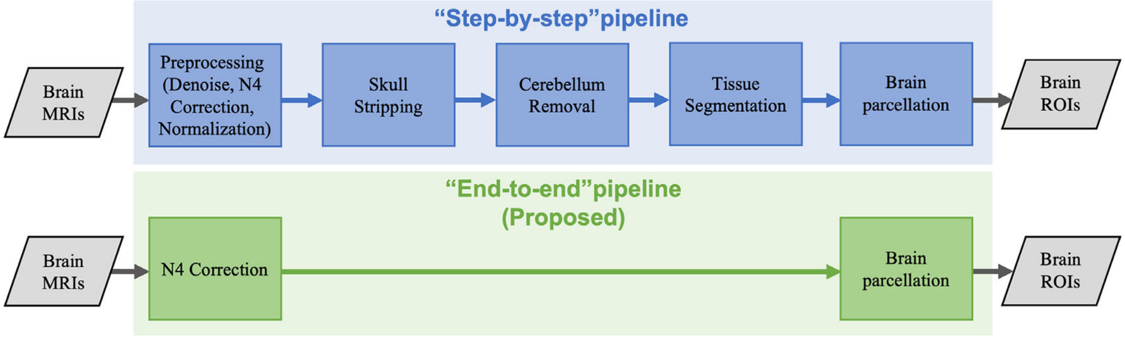


Figure 2.

Conventional “step-by-step” pipeline *v.s.* the proposed “end-to-end” pipeline.

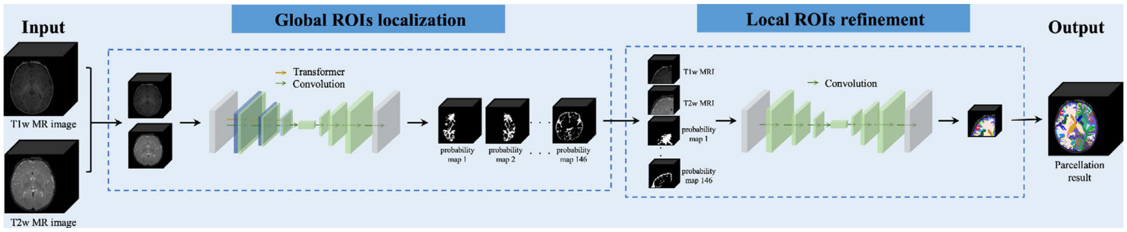


Figure 3.

The architecture of the proposed global-to-local framework for infant brain parcellation is composed of two key stages: the global ROIs localization stage and the local ROIs refinement stage. The former utilizes a combination of transformer and convolution to identify the preliminary position of the ROIs in the whole brain space. The latter then refines the boundaries of the ROIs using convolution.

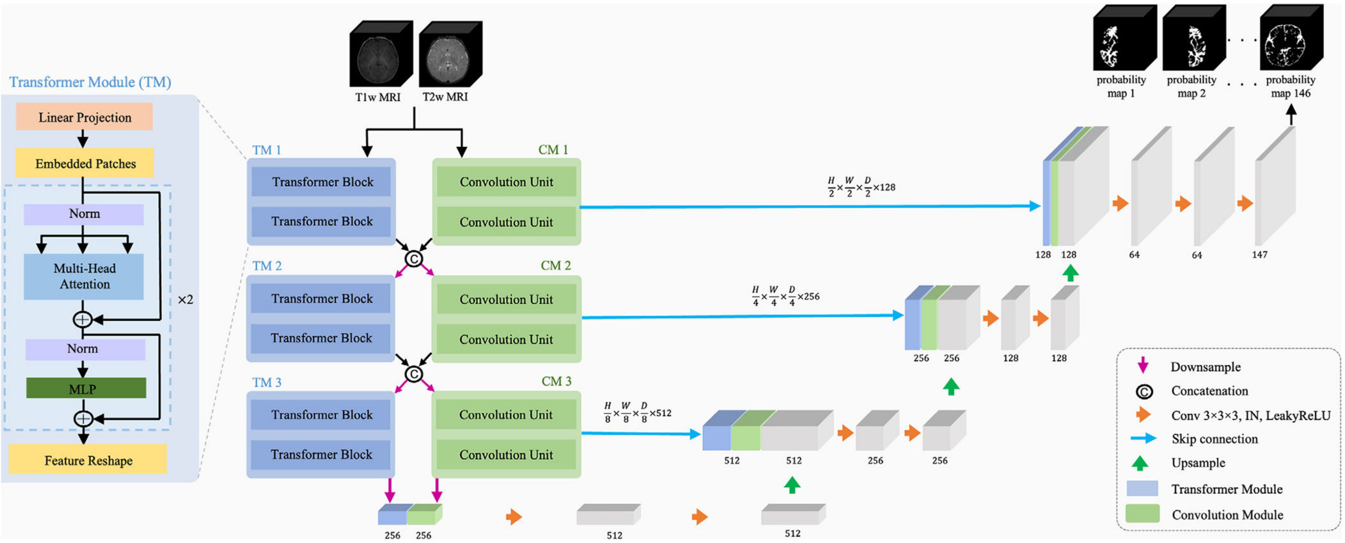


Figure 4.

The architecture of the global ROIs localization network, with the integration of transformer and convolution.

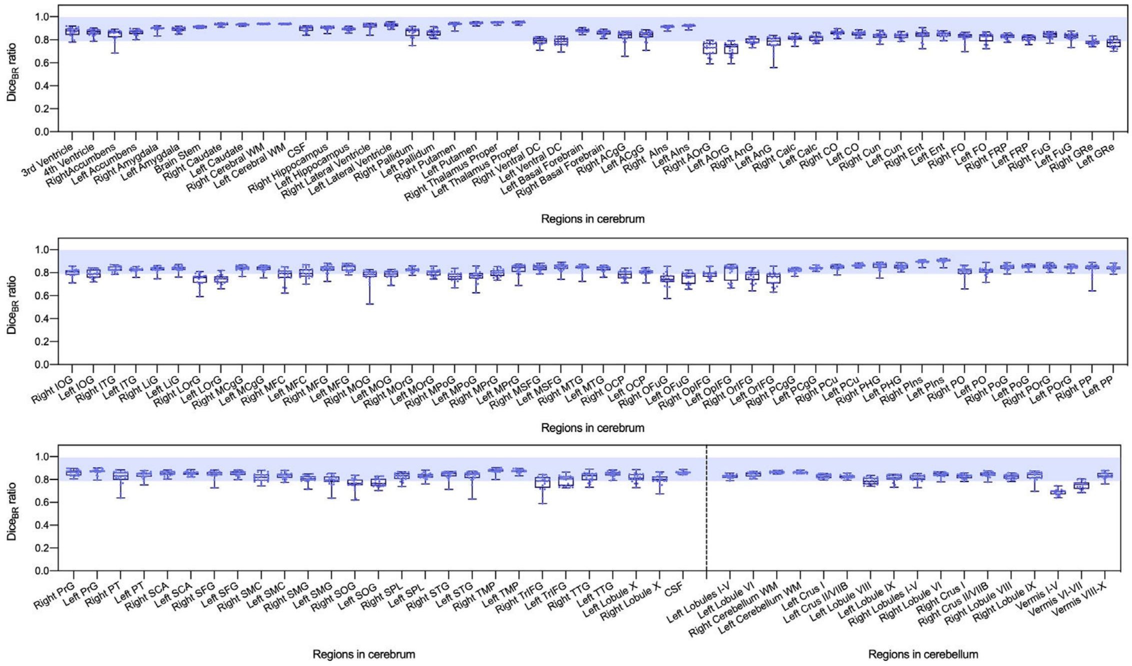


Figure 5.

The average *Dice$_{BR}$* ratio for 170 subjects from the NDAR dataset for the cerebrum and cerebellum using the proposed pipeline.

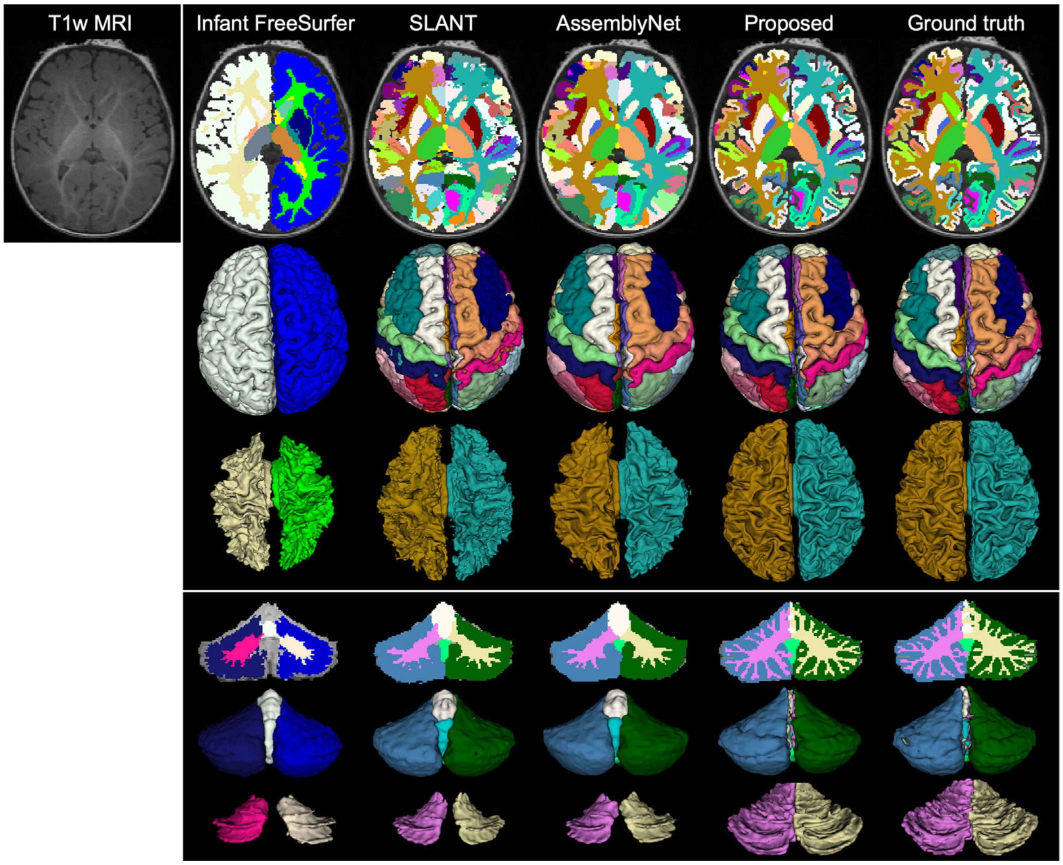


Figure 6.

Examples of an infant brain MR image from the NDAR dataset and the parcellation results for the cerebrum and cerebellum obtained using Infant FreeSurfer, SLANT, AssemblyNet, the proposed pipeline, and the corresponding ground truth.

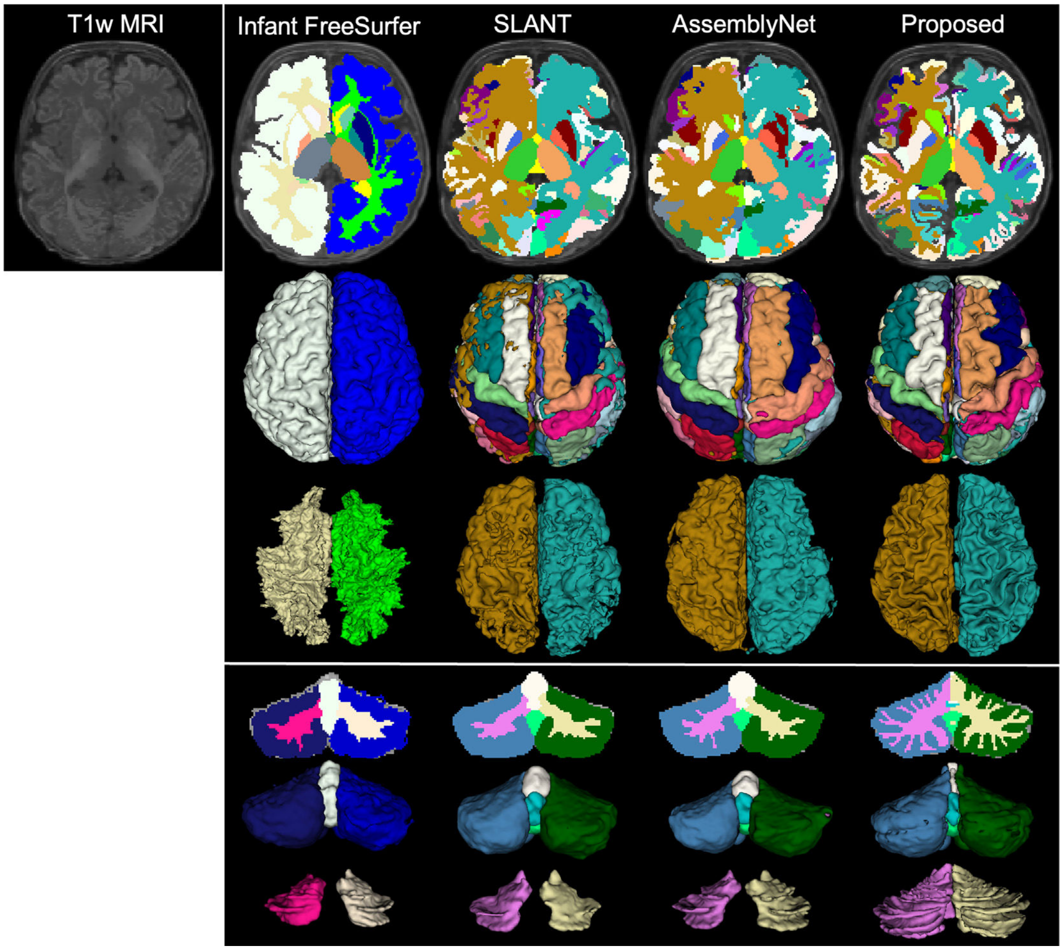


Figure 7.

Examples of an infant brain MR image in the BCP dataset and corresponding parcellation results for cerebrum and cerebellum produced by Infant FreeSurfer, SLANT, AssemblyNet and the proposed pipeline.

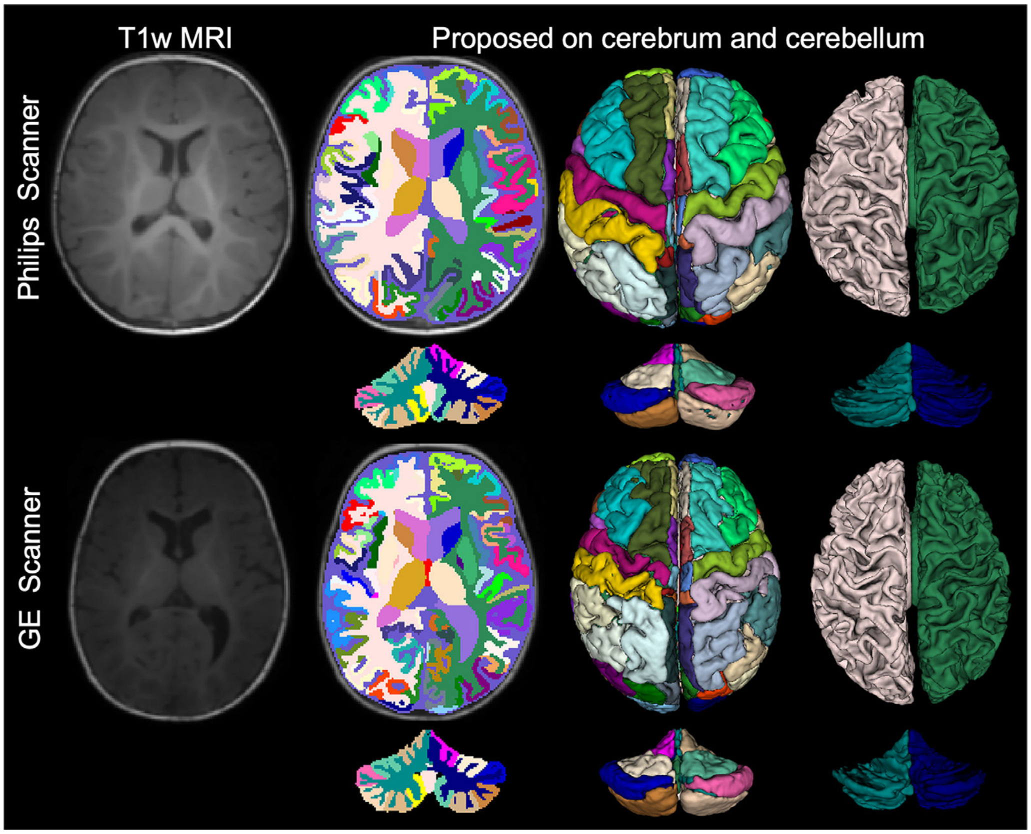


Figure 8.

Examples of two infant brain MR images scanned by a GE scanner and a Philips scanner, and their parcellation results on cerebrum and cerebellum produced by the proposed pipeline.

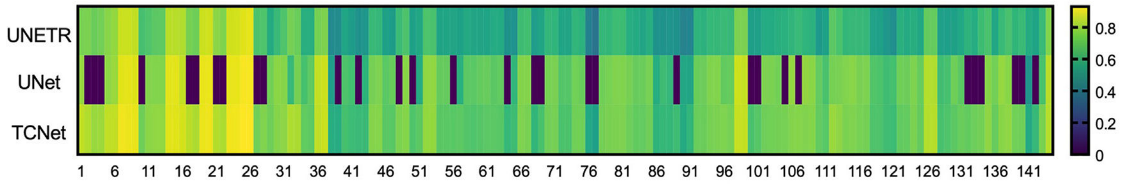


Figure 9.

The average $Dice_{BR}$ ratios of parcellation results of all subjects corresponding to 146 brain regions in the global ROIs localization stage for UNETR, UNet, and TCNet on the NDAR dataset.

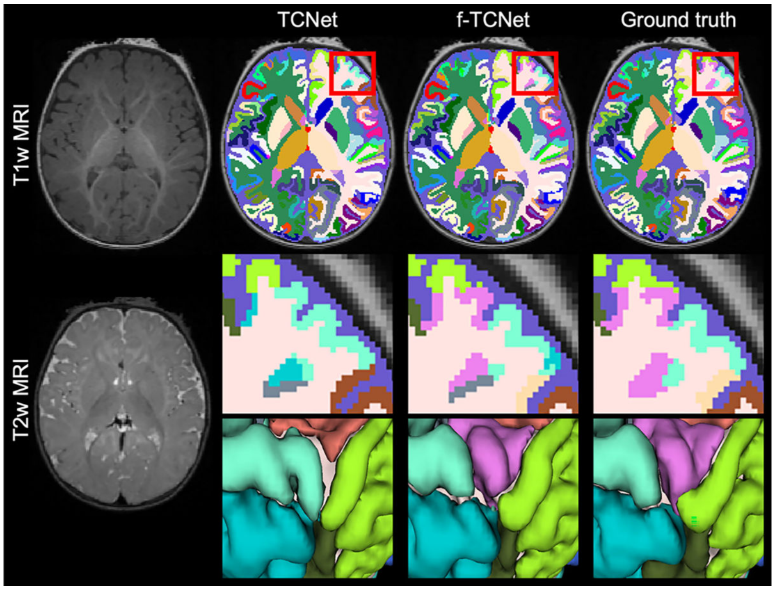


Figure 10.

Examples of an infant brain MR image from the NDAR dataset and resulting parcellation produced by TCNet, f-TCNet and the corresponding ground truth.

Table 1

Existing parcellation works for brain MRIs

Methods	Key techniques	6-month-old subjects	End-to-end	Number of ROIs (*n*)
SegNet [30]	Deep learning	×	✓	134
FastSurfer [31]	Deep learning	×	✓	95
HighRes3DNet [32]	Deep learning	×	✓	155
PICSL [33]	Registration	×	✓	134
FreeSurfer [34]	Registration	×	×	45
SLANT [22]	Deep learning	×	✓	134
AssemblyNet [23]	Deep learning	×	✓	134
iBEAT [25]	Registration	✓	×	90
Infant FreeSurfer [24]	Registration	✓	×	45
Proposed pipeline	Deep learning	✓	✓	146

Table 2

Training and testing datasets

Groups	Scanner (3T)	Modality	TR/TE (ms)	Resolution (mm^3)	Number of subjects (n)	Age (months)
Training	Siemens (NDAR)	T1w	2400/3.16	1.0 × 1.0 × 1.0	160	6
		T2w	3200/499	1.0 × 1.0 × 1.0		
	Siemens (NDAR)	T1w	2400/3.16	1.0 × 1.0 × 1.0	170	6
		T2w	3200/499	1.0 × 1.0 × 1.0		
Testing	Siemens (BCP)	T1w	2400/2.24	0.8 × 0.8 × 0.8	83	6
		T2w	3200/564	0.8 × 0.8 × 0.8		
	GE	T1w	6.512/2.64	1.0 × 1.0 × 1.0	5	6
		T2w	3.65/102.48	1.0 × 1.0 × 1.0		
	Philips	T1w	10/4.6	0.9982 × 0.9982 × 1.0	5	6
		T2w	2500/310	0.7639 × 0.7639 × 0.8		

Table 3

Configurations of the Transformer Modules and Convolution Modules in the global ROIs localization network

Transformer modules		Convolution modules	
TM1	Tran (16, 12, 768), Tran (16, 12, 768)	CM1	Conv (2, 64, 3, 1), Conv (64, 64, 3, 1)
	In (2, 96 ×96 ×96)		In (2, 96 ×96 ×96)
	Out (64, 96 ×96 ×96)		Out (64, 96 ×96 ×96)
Max pooling1	In (128, 96 ×96 ×96)		In (128, 96 ×96 ×96)
	Out (128, 48 ×48 ×48)		Out (128, 48 ×48 ×48)
TM2	Tran (8, 8, 128), Tran (8, 8, 128)	CM2	Conv (128, 128, 3, 1), Conv (128, 128, 3, 1)
	In (128, 48 ×48 ×48)		In (128, 48 ×48 ×48)
	Out (128, 48 ×48 ×48)		Out (128, 48 ×48 ×48)
Max pooling2	In (256, 48 ×48 ×48)		In (256, 48 ×48 ×48)
	Out (256, 24 ×24 ×24)		Out (256, 24 ×24 ×24)
TM3	Tran (4, 4, 16), Tran (4, 4, 16)	CM3	Conv (256, 256, 3, 1), Conv (256, 256, 3, 1)
	In (256, 24 ×24 ×24)		In (256, 24 ×24 ×24)
	Out (256, 24 ×24 ×24)		Out (256, 24 ×24 ×24)
Max pooling3	In (512, 24 ×24 ×24)		In (512, 24 ×24 ×24)
	Out (512, 12 ×12 ×12)		Out (512, 12 ×12 ×12)

Tran (patch_size, head_num, hidden_size), Conv (inchannel, outchannel, kernel_size, stride), In (channel, H×W×D), Out (channel, H×W×D) TM: transformer modules, CM: convolution modules.

Table 4

The average $Dice_{BR}$ and $Dice_{WB}$ ratios for the cerebrum and cerebellum by different methods on the NDAR dataset (%)

Methods	Cerebrum		Cerebellum	
	$Dice_{BR}$	$Dice_{WB}$	$Dice_{BR}$	$Dice_{BR}$
Infant FreeSurfer	32.90±1.50	47.25 ±0.79	45.27±1.13	50.65±1.27
Proposed	87.40±6.76	89.59 ±6.67	85.12±5.43	86.70±5.39
SLANT	57.73±2.89	62.91 ±2.92	61.91±1.21	70.80±0.95
AssemblyNet	62.35±3.08	67.86 ±2.83	63.16±1.75	72.25±1.05
Proposed	81.47±1.51	87.00 ±1.22	79.91±2.94	86.77±0.84

Table 5

The average $Dice_{BR}$ and $Dice_{WB}$ ratios of all brain regions of all subjects for the NDAR dataset (%)

Methods	Modality	Processing unit	Feature extraction	$Dice_{BR}$	$Dice_{WB}$
UNETR	T1w, T2w	Whole-brain (2 × 2 × 2 mm^3)	Convolution	61.02±4.87	71.82±3.06
UNet	T1w, T2w	Whole-brain (2 × 2 × 2 mm^3)	Transformer	59.94±1.04	79.06±1.07
TCNet	T1w, T2w	Whole-brain (2 × 2 × 2 mm^3)	Convolution + Transformer	72.27±1.62	80.65±1.17

Table 6

The average $Dice_{BR}$ and $Dice_{WB}$ ratios on the NDAR dataset (%)

Methods	Modality	Global ROIs localization stage			Local ROIs refinement stage			$Dice_{BR}$	$Dice_{WB}$
		Processing unit	Network selection	Feature extraction	Processing unit	Network selection	Feature extraction		
TCNet	T1w,T2w	-	-	-	patch (1 × 1 × 1 mm^3)	TCNet	Convolution + Transformer	75.97±5.84	87.23±6.73
f-TCNet	T1w,T2w	whole-brain image (2 × 2 × 2 mm^3)	TCNet	Convolution + Transformer	patch (1 × 1 × 1 mm^3)	UNet	Convolution	81.88±0.95	87.54 ±0.70

Table 7

The average $Dice_{BR}$ ratios using our pipeline on the NDAR dataset (%)

Modality	Method	$Dice_{BR}$ in cerebrum	$Dice_{BR}$ in cerebellum
T1w	Proposed	81.47±1.51	79.91±2.94
T1w, T2w	Proposed	81.96±0.99	81.29±1.43