Self-Supervised Transfer Learning for Infant Cerebellum Segmentation with Multi-Domain MRIs

Yue Sun, Kun Gao, Shihui Ying, Senior Member, IEEE, Weili Lin, Kathryn L. Humphreys, Gang Li, Senior Member, IEEE, Sijie Niu*, Mingxia Liu*, Senior Member, IEEE, Li Wang*, Senior Member, IEEE

Abstract—Cerebellum is a rapidly developing and critical brain structure during the early postnatal stages. Accurate segmentation of the cerebellum into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) is essential to characterize early cerebellum development. Compared with adult cerebellum, infant cerebellum MRI exhibits extremely low tissue contrast and severe partial volume effects in magnetic resonance imaging (MRI), posing a huge challenge for manual and automatic segmentation of the cerebellum. The latest trend of collaborative use of multi-domain infant images (acquired from different imaging sites or time-points) has made the segmentation task more difficult. In this paper, we propose a novel self-supervised transfer learning (SSTL) framework for infant cerebellum segmentation. The experimental results on 180 subjects with T1- and T2-weighted MRIs from two cohorts demonstrate that our SSTL achieves better segmentation results compared with several state-of-the-art methods. We further apply our method to an infant cerebrum segmentation challenge (i.e., iSeg-2019) with multi-site data, and achieve significantly better Dice ratios compared with top-ranked methods.

Index Terms—Cerebellum segmentation, infant, self-supervised learning, deep learning, multi-site data

I. INTRODUCTION

The human cerebellum is critical for motor control [1]. Recent findings also suggest the cerebellum plays an essential role in the complex neural underpinnings of brain disorders, with behavioral implications beyond the motor domain [2]. During the first two years of life, the cerebellum undergoes its most dramatic development, with its volume increases by 240% from 2 weeks to 1 year of age and 15% from 1 to 2 years of age [3]. Volpe et al. [5] reported that cerebellar abnormality is a relatively less recognized but likely important cause of neurodevelopmental disability in small premature infants. Therefore, exploring the early development of the cerebellum, especially during the first two years after birth, is fundamental to assist in the diagnosis and treatment of neurodevelopmental disorders. As a key component of diagnosis and treatment pipeline, the accurate segmentation of the infant cerebellum into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) helps characterize the early cerebellum development.

Most existing studies on brain tissue segmentation focus on the cerebrum [6]–[10], while there are very few works proposed for cerebellum tissue segmentation [11]–[16], especially for pediatric cerebellum where three practical challenges prevent segmentation. First, due to the low tissue contrast, the manual annotation is extremely challenging, especially for infant subjects younger than 3 months of age. Consequently, the number of training subjects is often limited or unavailable.

Fig. 1 shows exemplary T1-weighted (T1w) and T2-weighted (T2w) MR images of infant cerebellum at four time-points. The segmentation results (the third row) are obtained by ADU-Net [4], with model trained on 24-month-old subjects. Directly applying the segmentation model to younger (e.g., ≤3 months of age) infants cannot generate satisfactory results. The corresponding tissue intensity distributions for the T1-weighted images are shown in the last row, where the intensity distributions are highly overlapped for the images of younger infants.
month-old subjects exhibit relatively high tissue contrast, while the images of newborn subjects show relatively low tissue contrast. This makes manual and automatic segmentation of the infant cerebellum very challenging. Second, the collaborative use of multi-domain infant images (acquired from different imaging sites or time-points) makes the segmentation task more difficult. As reported in a MICCAI grand challenge on 6-month infant brain MRI segmentation from multiple sites (i.e., iSeg-2019) [8], a model trained on a specific-site dataset performs well on test subjects from the same site, but poorly on subjects from other sites with different imaging protocols/scanners. This is called the “domain shift” problem in medical image analysis, which is caused by the difference in distribution between different domains such as different sites. Data from different time-points also have the domain shift problem, because infant subjects have different tissue contrasts and data distributions at different time-points. In this work, a domain denotes a specific site or time-point. As shown in Fig. 1, when we directly apply a model trained on data of 24-month-old subjects to testing images of younger infants, the segmentation results are not satisfactory with many anatomical errors. Due to extremely low tissue contrast and severe partial volume effect, there are often anatomical errors in the segmentation results, e.g., “hole” or “handle” (see red dashed oval circles and rectangle boxes) in Fig. 2.

To address these challenges, we propose a self-supervised transfer learning (SSTL) framework for infant cerebellum segmentation with multi-domain MR images. Note that a short conference version of this work appeared in International Workshop on Machine Learning in Medical Imaging (2020) [17], which is a semi-supervised method for cerebellum segmentation but only validated on a few subjects and focused on a single imaging site. In this journal paper, we propose to transfer manual labels of 24-month-old subjects to younger infants for reliable segmentation in a self-supervised learning manner, and perform extensive experiments on multiple imaging sites. Our motivation is based on the fact that cerebellum at early time-points (e.g., ≤3 months of age) exhibits extremely low tissue contrast, while 24-month-old cerebellum shows high tissue contrast in MRIs. A previous study has also proven that 24-month-old cerebellum can be automatically or manually segmented due to its high tissue contrast [13]. To alleviate the domain shift issue, we propose to automatically generate a set of reliable training samples for different time-points and imaging sites. To be specific, we first borrow the manual labels from the late time-point with high contrast (or a specific site) to train a segmentation model. Then the segmentation model is directly applied to test imaging data from remaining time-points/sites. We further utilize a confidence map to automatically identify anatomical errors of automatic segmentations and generate a set of reliable training samples for each time-point/site. Lastly, we train the segmentation model guided by generated training data and a spatially-weighted cross-entropy loss function. Experiments on two cohorts and the iSeg-2019 challenge suggest that our method outperforms several state-of-the-art methods in infant brain segmentation.

The rest of this paper is organized as follows. We first briefly review related studies on supervised segmentation methods in Section II. Then, we introduce studied subjects and our proposed method in Section III and Section IV, respectively. Experiments and analysis are given in Section V. We further discuss several key components of the proposed method in Section VI and conclude this paper in Section VII.

II. RELATED WORK

Supervised learning methods have been widely used for medical image segmentation. They are generally categorized into fully-supervised, semi-supervised, and weakly-supervised methods according to the availability of manual annotations. Fully-supervised methods usually show outstanding performance [18], benefiting from a large number of voxel-level annotations. However, the voxel-wise annotation is highly time-consuming and labor-expensive, and often requires strong domain knowledge in medical image segmentation [19], [20].

To alleviate the effort of manual labeling, many semi-supervised and weakly-supervised approaches have been proposed for segmentation [19], [21]–[27]. For example, Bai et al. [24] introduced a semi-supervised learning approach trained on both labeled and unlabeled data, where a conditional random field was used to refine the segmentation for unlabeled data. Nie et al. [23] proposed an attention based semi-supervised deep learning framework, which utilized the confidence map to train a model with unlabeled data. Different from the above methods leveraging pseudo labels (just picking up the class which has the maximum predicted probability every weights update for unlabeled data [28]), Zhang et al. [26] employed an adversarial constrained convolutional neural network (CNN) for weakly supervised segmentation, by imposing inequality constraints of prior knowledge. Wang et al. [27] proposed a double-uncertainty weighted method for semi-supervised segmentation with a teacher-student model, where teacher model can provide guidance for student model by penalizing their inconsistent prediction on both labeled and unlabeled data. However, most existing semi-supervised methods seldom consider a selection strategy from pseudo
labels and weakly-supervised methods still have a performance gap compared with methods supervised by voxel-level labels.

Self-supervised learning has recently attracted much attention, which refers to learning methods in which ConvNets are explicitly trained with automatically generated labels [29]. Shimoda et al. [30] proposed a self-supervised difference detection model, which estimated noise from the results of the mapping functions by predicting the difference between the segmentation masks before and after the mapping. Xue et al. [31] developed a target domain self-supervision for domain adaptation by constructing an edge generation auxiliary task to assist primary segmentation task. The self-supervised methods in medical images often refine the automatic results of testing images, which are applicable to solve a transfer learning problem, but most of the existing methods lack of focusing on the reliability of automatically generated labels.

III. DATA AND MRI PREPROCESSING

In this study, the T1w and T2w infant brain MRIs were randomly chosen from two cohorts. The first one contains MRIs from 175 subjects (18 labeled images are used for training, 25 labeled images and 132 unlabeled images are used for validation) from the UNC/UMN Baby Connectome Project (BCP) [32], where imaging data were acquired from subjects at ≤ 3, 6, 9, 12, 18, and 24 months of age using a Siemens Prisma scanner. Another cohort contains MRIs of five 6-month-old infant subjects collected using a Philips scanner, which is only used to test the generalization ability of the proposed method and all competing methods. Detailed information of these two cohorts is listed in Table I. Note that only imaging data of eighteen 24-month-old subjects from BCP with manual labels are used as training data in this work. During scanning, infants were naturally sleeping, fitted with ear protection, and their heads were secured in a vacuum-fixation device. For image preprocessing, the resolution of all images was resampled into 0.8 × 0.8 × 0.8 mm$^3$, and T2w images were linearly aligned with their corresponding T1w images. We performed skull stripping, intensity inhomogeneity correction, and removal of the cerebrum leveraging an infant-dedicated pipeline (i.e., iBEAT V2.0 Cloud$^3$). Note that all studied subjects in this paper are cross-sectional.

Accurate manual segmentation, providing labels for training and testing, is of great importance for learning-based segmentation methods. For MRIs scanned by the Siemens Prisma scanner, we manually annotated images of 18 subjects at 24 months of age to train a segmentation model. Due to the low tissue contrast and extremely folded tiny structures, annotation editing for images from 18 months to ≤ 3 months becomes increasingly difficult and time-consuming. Therefore, only a limited number (i.e., 5 subjects per time-point) of younger subjects are manually annotated at each of five time-points (i.e., 18, 12, 9, 6, and ≤ 3 months of age) as testing data for quantitative comparison, while the remaining data have no manual label and are used for the segmentation quality assessment by two raters.

IV. METHODOLOGY

To alleviate the effort of manual annotation and generate accurate segmentation for multi-domain infant cerebellum, we develop a self-supervised transfer learning (SSTL) framework for infant cerebellum MRI segmentation, as illustrated in Fig. 3. The proposed SSTL consists of two steps: 1) supervised segmentation based on imaging data with manual labels, and 2) self-supervised transfer learning for cross-site (or cross-time-point) segmentation. In the first step, we train a segmentation model based on images of those 24-month-old subjects with manual labels, considering the high tissue contrast of MRIs at this time point. An error map is calculated based on the difference between automatic segmentation and manual labels. This error map is further used to train a confidence network to detect anatomical errors of the automatic segmentation. In the second step, we utilize the confidence map (predicted by the trained confidence network) to generate a set of reliable training samples for the remaining sites (or time-points). Finally, we train the segmentation model guided by our proposed spatially-weighted cross-entropy loss to refine the segmentation. This step is performed in a self-supervised transfer learning manner, aiming to alleviate the domain shift between different sites/time-points and improve the generalization ability of our trained model. More details can be found in the following.

A. Supervised Learning with Limited Manual Labels at 24 Months of Age

Various network architectures can be used for supervised segmentation, such as U-Net [33], V-Net [34], U-Net++ [35], ADU-Net [4], and nnU-Net [36]. In this work, we employ ADU-Net [4] as the backbone segmentation model because it performs well in infant brain segmentation, with improvements of Dice ratio on CSF, GM and WM: 3.1%, 3.3% and 1.4% comparing with U-Net [33] on testing 18 subjects. Specifically, the ADU-Net has a similar architecture to U-Net [33], with seven dense blocks [37] are introduced into the down-sampling path and the up-sampling path. Each dense block consists of three BN-ReLU-Conv-Dropout operations, in which each convolution (Conv) layer includes 16 kernels and the dropout rate is 0.1. The final layer in the ADU-Net is a Conv layer, followed by a softmax non-linearity to provide the per-class probability for each voxel in MRI. As shown in Fig. 3, a cross-entropy loss $L_{seg}$ is used in the ADU-Net, defined as

$$L_{seg} = -\sum_{i=1}^{C} y_i \ln x_i$$ (1)
where $C$ is the number of categories ($C = 4$ in this work, e.g., background, CSF, GM, and WM), $x_i$ denotes the predicted probability map, $y_i$ is the target of segmentation.

We input imaging data (paired T1w and T2w MRIs) and corresponding manual labels from eighteen 24-month-old subjects into the ADU-Net, thus generating four probability maps of tissues with the same size of inputs. The final segmentation result of each voxel is determined by the softmax strategy. However, if directly applying the trained model to other time-points/sites, we cannot obtain the satisfactory results due to the domain shift issue as shown in Fig. 1. To address this issue, we develop a self-supervised transfer learning strategy to automatically generate a set of reliable training samples based on automatic tissue segmentations of other time-points/sites.

Inspired by [23], we design a confidence network to evaluate the reliability of automatically generated segmentation for each voxel, and employ the U-Net structure [33] (with the contracting and expansive paths) to perform this task. The error map, defined as the differences between manual labels and automatic segmentations, is regarded as targets to train this confidence network. We design a loss $L_{cp}$ to learn whether the segmentation results are reasonable,

$$L_{cp} = -a \cdot y \ln x + b \cdot (1 - y) \ln(1 - x)$$

where $x$ is the predicted error map, and $y$ is the target of error map. $a$ and $b$ are two constant parameters, set as $a = 0.1, b = 1$. In this work, we apply the confidence model trained on 24-month labels to all age groups based on a fact that the anatomical segmentation errors (“hole” and “handle”) are often present in the label space [38]–[41], independent of sites/time-points. Fig. 4 depicts the reliability of segmentation results evaluated by the confidence map (the first column), where the corresponding segmentation results and manual labels are shown in the second and third columns, respectively. Note that the darker the color in the confidence map, the less reliable the segmentations. From Fig. 4, we can see some unreasonable results (i.e., anatomical errors) circled by red dashed lines, which are reflected as dark color in the confidence map. Therefore, we can utilize the confidence map to evaluate the reliability of tissue segmentations generated by an automated segmentation model such as ADU-Net.

B. Self-supervised Transfer Learning for Cross-Domain Segmentation

In practical applications, testing MRIs of subjects are often acquired from multiple sites (with different imaging protocols/scanners) or time-points, resulting in the domain shift issue where the MRI intensity distribution of training and test
subjects are different. As shown in Fig. 1, directly applying the model trained on one domain (e.g., 24 months of age) to test images from different domains may introduce anatomical errors in the segmentation results such as “hole” and “handle”. Based on the fact that the arbor vitae is a complete and folded tree-like appearance, it should be free of anatomical errors. To this end, we propose a self-supervised transfer learning strategy to alleviate the anatomical errors of segmentation results, by generating a set of reliable training samples from unlabeled data at each time-point/site. Specifically, we first apply ADU-Net (trained on 24-month-old infants) to imaging data from other time-points/sites to derive their automatic segmentations, and then apply the confidence network to evaluate the reliability of the automatic segmentations. With the guidance of the confidence map, we select the top $K$ subjects (MRIs with reliable segmentations) as the training samples for each time-point/site, even though these subjects have no manual labels.

With self-supervision of those generated training data, we can train a segmentation model for each specific time-point/site. One naive way to generate a set of training samples is extracting 3D image patches and corresponding segmentation patches from each of top $K$ subjects. However, this assumes that the segmentation result for each voxel is accurate, which is not always true as shown in the first row of Fig. 4. Taking the reliability of segmentation result for each voxel into account, we incorporate the confidence map into a spatially-weighted cross-entropy loss function for cross-domain segmentation, which is defined as

$$L_{seg-weights} = -w \sum_{i=1}^{C} y_i \ln x_i$$

(3)

where $w$ denotes the weights from the confidence map. With the proposed spatially-weighted cross-entropy loss, one can explicitly make use of confidence map for network training. This encourages the segmentation model to pay more attention to more reliable labels, thus helping avoid anatomical errors.

**C. Implementation Details**

In our implementation, we randomly extracted 1,000 MR image patches (size: $32 \times 32 \times 32$) from each training image, where the T1w and T2w images are treated as two channels in the proposed network. The kernels were initialized by Xavier. We used SGD strategy for network optimization. The learning rate was set as 0.005 and multiplied by 0.1 after each epoch. Please note that we separately optimize segmentation and confidence models in this paper. In the infant cerebellum segmentation task, we treat imaging data of 24-month-old subjects with manual labels as the source domain and testing data of 18-month-old subjects as the to-be-analyzed target domain. We can then apply our self-supervised transfer learning framework on the data of 18-month-old subjects for segmentation. Then, we regard data of 18-month-old subjects as a source domain and that of 12-month-old subjects as a target domain. That is, we gradually propagate labels from 24-month-old subjects to younger subjects (i.e., $24 \rightarrow 18 \rightarrow 12 \rightarrow 9 \rightarrow 6 \rightarrow 0 \sim 3$), by treating two adjacent time-points as the source domain and the target domain. Finally, we obtain a specific segmentation model for each time-point/site to help reduce the distribution gap between different time-points.

**V. EXPERIMENTS**

**A. Experimental Setup**

To demonstrate the advantage of our SSTL method in handling multi-site data, we validate it on infant images from multiple time-points (e.g., $\leq 3$ months to 18 months) and the confidence ranking of automatic cerebellum segmentations on images of 12-month-old subjects. We can observe that the confidence ranking is consistent with the accuracy of segmentations. This implies that our confidence network is effective in evaluating the reliability of automatic segmentations generated by the ADU-Net. In this way, we can generate a set of training samples for each time-point/site, even though these subjects have no manual labels.
imaging sites. Our method is a general framework, which can be extended to other applications such as infant cerebrum segmentation. Accordingly, we validate our method on multi-site infant cerebrum data from the iSeg-2019 challenge [8].

In the experiments, we compare our method with four state-of-the-art methods for medical image segmentation, including 1) volBrain [42], 2) Infant FreeSurfer [43], 3) Multi-atlas-based method [44], 4) ASD-Net [23], and 5) ADU-Net [4]. The volBrain is an automated MRI Brain Volumetry System3, Infant FreeSurfer is an automated segmentation and surface extraction pipeline4, ASD-Net is an attention based semi-supervised deep learning framework, and the ADU-Net architecture is the backbone of our segmentation model. In the volBrain system, we choose the CERES pipeline [14] to automatically analyze the cerebellum, which wins a MICCAI cerebellum segmentation challenge. Specifically, the CERES [14] applies a patch-based multi-atlas method to perform cerebellum segmentation, leveraging cerebellar atlases from adults. According to the guidance in the website, we submitted the T1w images without any preprocessing (e.g., skull striping, registration, denoising) to automatically obtain the GM and WM results. Infant FreeSurfer designs a multi-atlas label fusion segmentation framework for T1-weighted (T1w) neuroimaging data of infants aged 0–2 years. Multi-atlas-based method is a label fusion segmentation framework where manual label information from training images could be used to segment cerebellum of new infant images, where demons algorithm [45] is used to perform the registration procedure in this paper. The ASD-Net [23] uses a generative adversarial network (GAN) to predict labels and confidence map, then trains the segmentation model with automatically generated labels. The ADU-Net [4] designs a U-Net like architecture with seven dense blocks are introduced into the down-sampling path and the up-sampling path. Codes/Packages of the competing methods are kindly provided by the authors. Specifically, we first train models with manual labels at the 24-month time-point for the ASD-Net and ADU-Net methods, respectively. For each remaining site (or time-point), we directly applied the trained model at 24-month-old subjects of the ADU-Net method to the younger testing subjects, while trained a GAN model of the ASD-Net with all automatically generated labels from the quantitative validation subjects by considering there is not a training sample selection strategy.

For a fair comparison, these competing methods and our
proposed SSTL share the same training data and testing data. When performing SSTL, we empirically set $K = 10$ in the experiments and also study its influence in Section VI. The Dice ratio is used as the metric to evaluate the segmentation results achieved by our method and those competing methods.

### B. Cross-time-point Cerebellum Segmentation Results on Infant Subjects from Siemens Prisma scanner (BCP)

We first compare our method with five competing methods in cross-time-point cerebellum segmentation on the BCP cohort with the Siemens Prisma scanner. The 24-month-old subjects are used as training data, while those at early time-points (i.e., 18-, 12-, 9-, 6-, and $\leq 3$–month-old) are used as the testing data. Figure 6 presents the visual comparison of segmentation results, where the T1w and T2w images, segmentation results generated by six methods, and the corresponding manual labels are shown from top to bottom.

It can be observed from Fig. 6 that, compared with results of five competing methods, the cerebellum segmentations achieved by the proposed SSTL method are much more consistent with the manual labels. As shown in the second row to the fourth row of Fig. 6, the segmentation results of volBrain, Infant FreeSurfer and multi-atlas-based method are especially coarse and inaccurate. The possible reason are: volBrain employs adult atlases for infant cerebellum segmentation, while infant cerebellum shows low tissue contrast compared with the adult cerebellum; the atlas of cerebellum used for Infant FreeSurfer does not have enough details of tissues compared with manual labels of last row in Fig. 6; the multi-atlas-based method is also not able to generate satisfactory results by only applying limited manual labels of 24-month-old data as atlas, since all studied subjects in this paper are cross-sectional. Similar to the above competing methods, the ASD-Net cannot generate reasonable segmentation results (as shown in the fifth row of Fig. 6). It may be because ASD-Net (with only a simple discriminator network) cannot accurately detect unreliable regions in the cerebellum with very complex tissue structures. The ADU-Net method produces better results than volBrain, Infant FreeSurfer, multi-atlas-based method and ASD-Net, but it is still unsatisfactory (see the sixth row of Fig. 6), especially for subjects less than 3 months of age. The underlying reason is that ADU-Net directly applies models trained on 24-month-old subjects to younger subjects, ignoring the distribution gap between different time-points.

1) Quantitative Analysis: For quantitative comparison, we report the Dice ratio and 95th-percentile Hausdorff Distance (HD95) yielded by six methods for segmenting three tissues (i.e., CSF, GM, and WM) in the cerebellum MRIs of 25 infant subjects in Table II. In addition, the Wilcoxon signed-rank test is also performed to evaluate the statistical difference between our SSTL method and each of five competing methods. Since the volBrain/Infant FreeSurfer method cannot well segment CSF, we do not report the corresponding results in Table II. From this table, we can clearly see that the proposed method consistently outperforms four competing methods at all five time-points when segmenting three types of tissues, while the ADU-Net usually performs better than volBrain, Infant FreeSurfer, multi-atlas-based method and ASD-Net. It’s worth noting that the Dice ratio gradually decreases when segmenting images of younger subjects. For instance, the Dice ratio of WM results achieved by ADU-Net at 18 and $\leq 3$ months of age are 93.52% and 85.96%, respectively, while those of SSTL are 94.39% and 91.40%, respectively. For the HD95 metric, our SSTL always achieves lower values than the competing methods, especially for the most challenging task of segmenting infant cerebellum younger than 3 months of age.

2) Segmentation Quality Assessment: Besides the above testing data from 25 subjects with manual labels to quantitatively evaluate the segmentation results, we test images of 132 subjects at the same five time-points to demonstrate the robustness of our method by visually accessing the fitting accuracy of results. The 132 automatic results are categorized into three groups, i.e., “good”, “fair” and “poor”, by two raters. Figure 7 presents (a) the number distribution of 132 subjects at different time-points, (b) the characteristic rendering WM results of different groups, and (c) the proportion of automatic results in each group by two raters. We can see that the results obtained by our proposed method are rated as “good” and “fair” in 81.42% (averaged by two raters), which reflects our method can obtain encouraging results when performed on a large number of testing data.

We also make an inter-rater comparison to determine how many segmentations are consistency categorized the same group between the two raters. Results indicated that 124 subjects of total 132 subjects (93.94%) were rated same, which suggests the segmentation quality assessment was relatively consistent across raters.

### C. Cross-site Cerebellum Segmentation Results on Infant Subjects from Philips Scanner

We also evaluate the performance of different methods in a cross-site cerebellum segmentation task. Specifically, we directly apply our self-trained segmentation model on images of 6-month-old subjects in Section V-B to five 6-month-old infant images acquired from a Philips scanner, then train a segmentation model according to our SSTL strategy. As for the ASD-Net, we utilize the 6-month-old segmentation model in Section V-B to test Philips data, then train a segmentation model with automatically generated labels, and ADU-Net method test Philips data with the 24-month-old segmentation model in Section V-B. The visual segmentation results are reported in Fig. 8 (a), where 3 competing methods (i.e., volBrain, ASD-Net and ADU-Net) cannot generate reasonable results. In contrast, our method can obtain much better results in this cross-site segmentation task. This could due to our self-supervised transfer learning strategy that can leverage knowledge from BCP to another cohort with a different scanner for reducing inter-site difference in their data distribution.

The quantitative results in terms of Dice ratio are shown in Fig. 8 (b). For comparison, we also show the results
**TABLE II**

Dice ratio (DICE) and 95th-percentile Hausdorff distance (HD95) of cerebellum segmentation results on testing images of 25 subjects at 18, 12, 9, 6, and ≤3 months of age from BCP. The symbol "+" indicates that our proposed SSTL method is significantly better than volBrain, Infant FreeSurfer, multi-atlas-based method, ASD-Net and ADU-Net (with p-value < 0.05).

<table>
<thead>
<tr>
<th>Age (m)</th>
<th>Method</th>
<th>CSF (%)</th>
<th>DICE (%)</th>
<th>HD95 (mm)</th>
<th>WM (%)</th>
<th>DICE (%)</th>
<th>HD95 (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>volBrain</td>
<td>N/A</td>
<td>78.23±1.10⁺</td>
<td>16.34±3.87⁺</td>
<td>54.43±3.87⁺</td>
<td>25.23±2.14⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Infant FreeSurfer</td>
<td>N/A</td>
<td>76.5±0.29⁺</td>
<td>13.32±0.27⁺</td>
<td>61.88±7.77⁺</td>
<td>14.97±10.16⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multi-atlas-based method</td>
<td>N/A</td>
<td>74.05±2.40⁺</td>
<td>6.70±3.47⁺</td>
<td>84.04±0.34⁺</td>
<td>13.24±13.60⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ASD-Net</td>
<td>N/A</td>
<td>84.87±2.49⁺</td>
<td>9.86±1.14⁺</td>
<td>89.15±0.83⁺</td>
<td>11.31±9.48⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ADU-Net</td>
<td>N/A</td>
<td>87.57±1.70⁺</td>
<td>8.07±0.54⁺</td>
<td>91.57±0.43⁺</td>
<td>9.55±2.39⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>91.36±1.11⁺</td>
<td>7.42±0.83⁺</td>
<td>93.40±0.75⁺</td>
<td>9.25±12.00⁺</td>
<td>94.39±0.69⁺</td>
<td>5.67±0.97⁺</td>
</tr>
<tr>
<td>12</td>
<td>volBrain</td>
<td>N/A</td>
<td>70.80±13.54⁺</td>
<td>18.55±21.04⁺</td>
<td>49.86±4.70⁺</td>
<td>22.81±0.10⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Infant FreeSurfer</td>
<td>N/A</td>
<td>74.70±1.78⁺</td>
<td>12.61±6.87⁺</td>
<td>65.10±6.12⁺</td>
<td>10.43±0.33⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multi-atlas-based method</td>
<td>N/A</td>
<td>72.41±6.33⁺</td>
<td>7.79±16.07⁺</td>
<td>83.63±1.52⁺</td>
<td>14.11±11.01⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ASD-Net</td>
<td>N/A</td>
<td>84.95±2.41⁺</td>
<td>10.01±6.03⁺</td>
<td>89.09±1.07⁺</td>
<td>14.85±5.10⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ADU-Net</td>
<td>N/A</td>
<td>84.06±2.40⁺</td>
<td>9.33±2.26⁺</td>
<td>89.83±0.88⁺</td>
<td>12.80±24.95⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>90.5±1.25⁺</td>
<td>8.14±6.59⁺</td>
<td>93.08±0.47⁺</td>
<td>9.54±23.50⁺</td>
<td>94.14±0.48⁺</td>
<td>5.07±0.80⁺</td>
</tr>
<tr>
<td>9</td>
<td>volBrain</td>
<td>N/A</td>
<td>79.56±1.72⁺</td>
<td>15.73±0.40⁺</td>
<td>52.65±0.23⁺</td>
<td>22.15±0.19⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Infant FreeSurfer</td>
<td>N/A</td>
<td>76.03±1.79⁺</td>
<td>15.03±0.01⁻</td>
<td>63.59±1.07⁺</td>
<td>10.44±0.08⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multi-atlas-based method</td>
<td>N/A</td>
<td>70.40±4.27⁺</td>
<td>7.11±1.26⁺</td>
<td>81.53±0.29⁺</td>
<td>16.08±2.92⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ASD-Net</td>
<td>N/A</td>
<td>80.20±2.67⁺</td>
<td>15.10±1.33⁺</td>
<td>87.05±0.75⁺</td>
<td>7.52±2.35⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ADU-Net</td>
<td>N/A</td>
<td>87.82±2.16⁺</td>
<td>9.49±4.85⁺</td>
<td>91.00±6.88⁺</td>
<td>5.88±1.19⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>88.43±6.84⁺</td>
<td>6.81±0.49⁺</td>
<td>93.57±5.24⁺</td>
<td>10.55±2.72⁺</td>
<td>95.83±1.01⁺</td>
<td>6.16±0.35⁺</td>
</tr>
<tr>
<td>6</td>
<td>volBrain</td>
<td>N/A</td>
<td>77.76±1.24⁺</td>
<td>11.53±65.08⁺</td>
<td>50.96±3.91⁺</td>
<td>22.38±0.13⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Infant FreeSurfer</td>
<td>N/A</td>
<td>47.63±0.49⁺</td>
<td>15.85±7.99⁺</td>
<td>33.95±1.43⁺</td>
<td>22.74±10.40⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multi-atlas-based method</td>
<td>N/A</td>
<td>42.22±5.40⁺</td>
<td>10.81±0.15⁺</td>
<td>55.97±1.59⁺</td>
<td>13.62±18.53⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ASD-Net</td>
<td>N/A</td>
<td>75.66±0.69⁺</td>
<td>16.75±5.54⁺</td>
<td>85.29±1.91⁺</td>
<td>15.47±8.98⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ADU-Net</td>
<td>N/A</td>
<td>76.68±0.76⁺</td>
<td>10.47±7.40⁺</td>
<td>86.30±0.79⁺</td>
<td>15.41±4.00⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>86.34±0.52⁺</td>
<td>10.68±0.62⁺</td>
<td>90.91±1.05⁺</td>
<td>13.54±6.12⁺</td>
<td>92.64±1.59⁺</td>
<td>5.19±0.55⁺</td>
</tr>
<tr>
<td>≤3</td>
<td>volBrain</td>
<td>N/A</td>
<td>76.79±0.39⁺</td>
<td>20.79±48.98⁺</td>
<td>52.17±0.43⁺</td>
<td>25.92±68.32⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Infant FreeSurfer</td>
<td>N/A</td>
<td>46.13±5.83⁺</td>
<td>15.37±6.12⁺</td>
<td>31.87±5.06⁺</td>
<td>32.48±0.15⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multi-atlas-based method</td>
<td>N/A</td>
<td>41.00±17.68⁺</td>
<td>10.26±0.86⁺</td>
<td>52.56±17.22⁺</td>
<td>17.41±19.66⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ASD-Net</td>
<td>N/A</td>
<td>76.52±2.70⁺</td>
<td>14.65±5.17⁺</td>
<td>79.46±2.36⁺</td>
<td>12.21±16.50⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ADU-Net</td>
<td>N/A</td>
<td>77.43±3.62⁺</td>
<td>9.56±2.28⁺</td>
<td>82.13±0.97⁺</td>
<td>11.46±15.32⁺</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>86.82±1.51⁺</td>
<td>9.71±2.99⁺</td>
<td>88.22±0.93⁺</td>
<td>8.88±6.59⁺</td>
<td>91.40±1.62⁺</td>
<td>5.60±1.09⁺</td>
</tr>
</tbody>
</table>

Fig. 7. Segmentation quality assessment for data of 132 subjects at different time-points by two raters. (a) Number distribution of 132 subjects at different time-points. (b) Rendering WM results of three groups, i.e., “good”, “fair” and “poor”. (c) Proportion of automatic results in each group categorized by two raters, with the average result.
of different methods when segmenting 6-month-age subjects from BCP with a Siemens scanner in Fig. 8 (b). It can be seen from this figure that compared with three competing methods, the Dice ratio of our method is more consistent across two imaging sites. This demonstrates that our SSTL method has a strong generalization ability.

D. Generality: Cross-site Cerebrum Segmentation Results on Infant Subjects from iSeg-2019 Challenge

Our framework is general and can be applied to other tasks, especially for those with multi-site datasets. In this group of experiment, we further validate the proposed SSTL method in the cerebrum segmentation task in the iSeg-2019 challenge. In iSeg-2019, the training imaging data are from Multi-visit Advanced Pediatric Brain Imaging Study [46], while the testing imaging data are from three different sites (i.e., BCP, Stanford University, and Emory University). That is, the training and testing data were scanned with different imaging protocols/scanners, with more details given in [8].

In this experiment, we compare our method with top 3 methods from three participating teams (i.e., QL111111, Tao_SMU, and FightAutism) in this challenge. The segmentation results achieved by those top 3 methods and our proposed method are shown in Fig. S1 of the Supplementary Materials. We can observe that the result generated by our method is more consistent with manual label, compared with three competing methods. This conclusion can also be drawn from the quantitative results (Dice ratio and HD95) in the top of Fig. S1, which suggests that the proposed method achieves the highest Dice ratio in segmenting WM when testing images are from three imaging sites. In particular, our method has the least difference of Dice ratio results among three sites, which proves its effectiveness in dealing with the domain shift problem. Besides, there is a significant difference between our method and QL111111, Tao_SMU, and FightAutism for segmenting GM and WM in terms of Dice ratio ($p$-value < 0.01).

VI. DISCUSSION

A. Influence of Gradual Label Propagation

In this paper, we propose a self-supervised transfer learning (SSTL) method for infant cerebellum segmentation to deal with the domain shift issue caused by different scanners/protocols in different time-points/sites. The experimental results in Figs. 6-8 and Table II demonstrate that our proposed SSTL framework can effectively improve the accuracy of infant cerebellum tissue segmentation and outperforms several state-of-the-art methods, especially for the younger infants. One outstanding challenge in segmenting the cerebellum of younger infants (e.g., 0-month-old) is that the tissue contrast in MRI is particularly low, which brings many difficulties to both manual editing and automatic segmentation methods. This challenge still remains to be further investigated to improve the accuracy. To this end, we design a gradual label propagation strategy in SSTL to propagate labels from 24- to 0-month-old subjects (i.e., 24 → 18 → 12 → 9 → 6 → 3 → 0), aiming to generate reliable segmentations for early time-point cerebellum. In the following, we investigate whether our gradual propagation strategy is better than the direct propagation strategy (e.g., 24 → 0).

In addition, we show the segmentation results for images of 0-month-old subjects achieved by our SSTL with direct propagation and gradual propagation in Fig. 9. We can observe from Fig. 9 that SSTL with gradual propagation yields much better segmentation results than that of direct propagation. This can be due to the large distribution differences of images between 24- and 0-month-old subjects. Therefore, compared with the direct propagation strategy, the proposed gradual propagation strategy is more reasonable to transfer labels of 24-month-old data to 0-month-old data for cerebellum segmentation.
B. Cross-Validation on MRIs of 24-month-old Infants

In previous experiments, we use all images of eighteen 24-month-old subjects from BCP with manual labels as training data. Now we study the performance of the segmentation network (i.e., ADU-Net) in the proposed SSTL framework in segmenting the cerebellum of 24-moth-old infants, by performing 2-fold cross-validation. That is, all data of eighteen 24-month-old subjects from BCP are randomly partitioned into two folds. Each fold is alternatively used as training data, while the other fold is treated as testing data. The Dice ratio results (%) achieved by the ADU-Net are 93.39 ± 8.58, 92.60 ± 4.57 and 94.02 ± 3.96 for CSF, GM and WM, respectively. Recalling the results in Table II, we can see that the results for segmenting the cerebellum of younger infants (e.g., 6-month-old) are generally worse than those of 24-month-old infants. This further validates the necessity to design robust cerebellum segmentation models for young infants, and our SSTL provides a general transfer learning solution to address this issue.

We also explore the performance of our method in improving the segmentation accuracy for cerebellum images of 24-month-old subjects. To this end, based on the above cross-validation results for 24-month-old subjects, we further employ the SSTL strategy to select a set of reliable training samples from the testing results for each of the two folds. As a result, top $K$ subjects (corresponding MRIs with segmentations) are automatically generated to as the training inputs of a new segmentation model. The eighteen 24-month-old images with manual labels are used as the testing data in this experiment, and the new trained model achieves the performance improvement of Dice ratio (%) (i.e. CSF: 94.14 ± 8.28, GM: 93.39 ± 4.45 and WM: 94.59 ± 3.66).

C. Influence of Number of Unlabeled Data

Based on the confidence map, we select top $K$ unlabeled training data for self-supervised learning. We now evaluate the performance of our SSTL leveraging different values of $K$ within the range of $[1, 2, \cdots, 11]$. Herein, we take 18-month-old time-point for an example and train segmentation models leveraging the top $K$ unlabeled subjects (with their automatic segmentations as output), as well as leave-one-out-cross-validation to optimize parameters of models. The corresponding average Dice ratios and standard deviations are reported in Fig. 10.

Fig. 10 suggests that as $K$ increases, the Dice ratio results for three tissues gradually increase, and the best results are reached when $K = 10$. This implies that leveraging reliable unlabeled data (selected by our method) helps improve the segmentation performance. But when $K > 10$, the performance becomes worse. The possible reason is that introducing many less reliable subjects and their labels will degrade the performance. We would like to mention that the optimal value of $K$ can be determined case by case based on different data. If the initial segmentations are basically reasonable, we can select more unlabeled training images (i.e., a larger $K$) for self-supervised transfer training; and vice versa.

D. Influence of Automatically Generated Labels

In this work, we manually labeled 5 imaging data at each early time-point as testing samples for validation, while there are many unlabeled data at these time-points. To augment the quality of training data, we develop a self-supervised transfer learning strategy to select a set of $K$ reliable unlabeled imaging data as training samples for each early time-point. Now we investigate whether these selected unlabeled data help boost the segmentation performance. To this end, we compare the segmentation results of our segmentation network (i.e., ADU-Net) in both fully supervised and semi-supervised learning settings. For fully supervised learning at the 6 months of age, we train the ADU-Net leveraging $N$ ($N = [1, \cdots, 4]$) imaging data with manual labels, and test the model on the remaining $5 - N$ images with manual segmentation. For semi-supervised learning, we use both the $N$ labeled samples and our selected top $K = 10$ reliable segmentations as labels (with the corresponding MRIs) as training data, and the remaining $5 - N$ labeled images as the testing data. Note that all test images will not participate in model training.

Figure 11 reports the segmentation results for WM at the 6-month-old time-point. As we can see from Fig. 11, compared with supervised learning, the semi-supervised learning strategy greatly improves the accuracy of WM segmentation in terms of Dice ratio, especially when the number of training subjects with manual labels is highly limited. In other words, our selected top $K$ unlabeled subjects are beneficial to improve the segmentation performance.
VII. Conclusion

In this paper, we proposed a self-supervised transfer learning (SSTL) framework for automatic segmentation of infant cerebellum with multi-domain MRIs. We first construct a set of imaging data from 24-month-old infants as training samples (with manual labels), and then propagate their manual labels to younger infants for cerebellum segmentation. Experimental results on 180 subjects from two cohorts and the infant cerebrum segmentation challenge (i.e., iSeg-2019) demonstrate that SSTL outperforms several state-of-the-art methods.

E. Influence of Spatially-Weighted Cross-Entropy Loss

To study the influence of our spatially-weighted cross-entropy loss in Eq. (3), we compare SSTL with its counterpart leveraging only the conventional cross-entropy loss (i.e., without leveraging the spatially-weighted constraint). In Fig. 12, we report the results of WM segmentation achieved by our method leveraging the conventional cross-entropy loss and the proposed spatially-weighted cross-entropy loss in zoomed views, as well as corresponding Dice ratio and HD95 evaluations. We can see from Fig. 12 (a) that leveraging the conventional cross-entropy loss, there are many anatomical errors (as indicated by red lines in the first column), whereas these errors are largely avoided by leveraging our designed loss (as shown in the second column). The same conclusion can also be summarized from Fig. 12 (b), according to higher Dice ratio and lower HD95 of results with our proposed loss.

In Supplementary Materials, we discuss the effectiveness of confidence map for cross-site and cross-time-point segmentation, the influence of self-supervised transfer learning and the evaluation and influence of confidence maps.

F. Limitations and Future Work

Several limitations need to be considered to further improve the current framework. First, the size of the training samples is still relatively small. It is desired to collect more neuroimaging data from multi-site MRI studies and use generative models to augment the training samples. Second, since the arbor vitae is a complete and folded tree-like appearance and seems similar with each other, the automatically generated labels are feasible to construct an atlas for assisting to improve the segmentation accuracy, like 0-month-old subjects. Third, although cerebellum and cerebrum data are used to verify the generality of our SSTL method in this paper, we will apply the method to more segmentation tasks of other anatomical structures in our future work. Besides, a fundamental solution, such as modeling spatial and appearance trajectory of infant imaging, will be our future work for the cerebellum segmentation of infant MRIs. To this end, we may need to identify the correspondence between longitudinal scans for each individual subject. Therefore, the tissue segmentation generated by our method will help find the correspondence accurately, since it is well known that the segmentation-based registration is more accurate than intensity-based registration [6].

REFERENCES

Self-Supervised Transfer Learning for Infant Cerebellum Segmentation with Multi-Domain MRIs – Supplementary Materials

Yue Sun, Kun Gao, Shihui Ying, Senior Member, IEEE, Weili Lin, Kathryn L. Humphreys, Gang Li, Senior Member, IEEE, Sijie Niu*, Mingxia Liu*, Senior Member, IEEE, Li Wang*, Senior Member, IEEE

In what follows, we show additional experimental results, verify the effectiveness of confidence map on cross-site and cross-time-point segmentations, study the influence of self-supervised transfer learning strategy and show the evaluation and influence of confidence maps.

I. ADDITIONAL RESULTS

As mentioned in the main text, we report the results of Dice ratio and HD95 in Fig. S1.

II. EFFECTIVENESS OF CONFIDENCE MAP FOR CROSS-SITE AND CROSS-TIME-POINT SEGMENTATION

For cross-site and cross-time-point applications, i.e., five 6-month-old testing images scanned with a Philips scanner (see Section V-C), we directly applied our self-trained segmentation model on these 6-month-old images from BCP to obtain automated segmentations. Then, following our self-supervised transfer learning strategy, we applied the confidence model (trained on 6-month-old automated segmentations from BCP) to automatically evaluate the reliability of segmentations for images from the Philips scanner. For clarity, we have now shown top 2 and bottom 2 ranked confidence maps generated by our method in the following Fig. S2. From Fig. S2, we can observe that the confidence ranking reflects the reliability of segmentation results. Therefore, our designed confidence model is effective in evaluating the reliability of automated segmentations, even though for those cross-site and cross-time-point images.

III. IMPROVEMENT OF SELF-SUPERVISED TRANSFER LEARNING STRATEGY

In this work, we randomly selected five BCP MRIs for five time-points as testing data (see Section V-B, Table II), which are not included in the top $K$ samples. Therefore, we can calculate the Dice ratio values for these images to verify the performance of models before and after re-training on the remaining samples. We take 12-month-old time-point for an example and report the Dice ratios of segmentations in the following Fig. S3 (a). We can observe from Fig. S3 (a) that, the Dice ratio of automated segmentation by model after training with Top $K$ samples (red color) are always higher than that of model before training with top $K$ samples (blue color). Therefore, the model after training with selected samples shows a better performance in terms of Dice ratio, with an improvement of $8\%$ for CSF, $7\%$ for GM and $5\%$ for WM. As for change of the corresponding confidence scores, we have reported the average confidence values for 12-month-old cerebellum segmentations (top $K$ and the remaining data) in the following Fig. S3 (b), which are generated by models before (blue color) and after (red color) training with selected samples. It can be seen from Fig. S3 (b) that the average confidence values indicated by red bar are always higher than that of blue bar, which demonstrates that the confidence models are effective to evaluate the reliability of automated segmentations.

IV. EVALUATION AND INFLUENCE OF CONFIDENCE MAPS

The performance of the confidence model is critical to automatically select reliable training samples for training a new segmentation model at each time-point/site. Therefore, we perform a 2-fold cross-validation experiment to quantitatively analyze the performance of confidence model, i.e., calculating the Dice ratio between error maps (i.e., ground truth) and predicted confidence maps for eighteen 24-month-old subjects. That is, all data from eighteen 24-month-old subjects of BCP are randomly partitioned into two folds. Each fold is alternatively used as training data, while the other fold is treated as testing data. Fig. S4 shows the Dice ratios between the error map and the predicted confidence map for eighteen 24-month-old subjects. We can see that the most Dice ratio values are above $90\%$, which indicates the predicted confidence map can effectively detect the unreliable regions.

*Corresponding authors: S. Niu (sjniu@hotmail.com), M. Liu (mxliu@med.unc.edu), and L. Wang (li_wang@med.unc.edu).

Y. Sun, K. Gao, and L. Wang were supported by NIH grants (Nos. MH109773 and MH117943). G. Li was supported in part by NIH grant (No. MH117943). M. Liu was supported in part by NIH grant (No. AG041721).
Furthermore, to demonstrate the improvement contributed by the confidence maps, we made an ablation experiment to compare the difference of site-specific segmentation models trained with/without the confidence maps. With the confidence maps, we automatically selected top $K$ reliable training samples, then trained a site-specific segmentation model for each site/time-point, with the proposed spatially-weighted cross-entropy loss. Without the confidence maps, we considered all testing data as training samples, then leveraged the cross-entropy loss to train a site-specific segmentation model for each site/time-point. We validated on 30 images with manual labels (i.e., 25 BCP subjects and 5 Philips subjects), and presented the corresponding Dice ratios in Fig. S5. We can see that by applying the confidence maps, the generated results (red bars) are more accurate than the results obtained by models without the confidence maps (blue bars). The significant differences are listed in Table S1, in which we can see the results with the confidence maps are significantly better than the results without the confidence maps, for all age groups. Although the statistical power is limited, the quantitative results are able to demonstrate the improved performance with the confidence maps.
Fig. S3. Comparison of (a) Dice ratios and (b) average confidence values of segmentation results for testing subjects at 12-month-old time-point, with the segmentation models before and after re-training on the top $K$ samples.

Fig. S4. Dice ratio between error map (i.e., ground truth) and predicted confidence map for eighteen 24-month-old subjects by performing a 2-fold cross-validation.

Fig. S5. Dice ratios of tissue results for 6 site-specific segmentation models trained without/with confidence map, at 30 testing subjects.

<table>
<thead>
<tr>
<th>Age (in month)</th>
<th>BCP</th>
<th>Philips</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>GM</td>
<td>0.0030</td>
<td>0.0100</td>
</tr>
<tr>
<td>WM</td>
<td>0.0203</td>
<td>0.0117</td>
</tr>
</tbody>
</table>