



# Multi-scale Self-supervised Learning for Multi-site Pediatric Brain MR Image Segmentation with Motion/Gibbs Artifacts

Yue Sun<sup>1,2</sup>, Kun Gao<sup>2</sup>, Weili Lin<sup>2</sup>, Gang Li<sup>2</sup>, Sijie Niu<sup>1</sup>(✉), and Li Wang<sup>2</sup>(✉)

<sup>1</sup> Department of Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan 250022, China

<sup>2</sup> Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, USA

l.i.wang@med.unc.edu

**Abstract.** Accurate tissue segmentation of large-scale pediatric brain MR images from multiple sites is essential to characterize early brain development. Due to imaging motion/Gibbs artifacts and multi-site issue (or domain shift issue), it remains a challenge to accurately segment brain tissues from multi-site pediatric MR images. In this paper, we present a multi-scale self-supervised learning (M-SSL) framework to accurately segment tissues for multi-site pediatric brain MR images with artifacts. Specifically, we first work on the downsampled images to estimate coarse tissue probabilities and build a global anatomic guidance. We then train another segmentation model based on the original images to estimate fine tissue probabilities, which are further integrated with the global anatomic guidance to refine the segmentation results. In the testing stage, to alleviate the multi-site issue, we propose an iterative self-supervised learning strategy to train a *site-specific* segmentation model based on a set of reliable training samples automatically generated for a to-be-segmented site. The experimental results on pediatric brain MR images with real artifacts and multi-site subjects from the iSeg-2019 challenge demonstrate that our M-SSL method achieves better performance compared with several state-of-the-art methods.

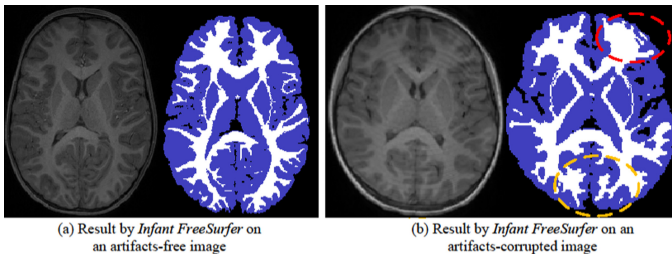
**Keywords:** Pediatric brain segmentation · Motion/Gibbs artifacts · Deep learning · Multi-site issue

## 1 Introduction

Accurate segmentation of the pediatric brain MR images into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) is one of the most pivotal steps to characterize early brain development [1, 2]. However, compared with adult brain MR images, pediatric brain MR images exhibit low tissue contrast caused by the ongoing myelination and severe imaging artifacts caused by head motion, creating challenging tasks for tissue segmentation [3]. Therefore, existing tools developed for adult brains, e.g., BrainSuite

[4], FSL [5], FreeSurfer [6], and HCP pipeline [7], often perform poorly on the pediatric brain MR images.

Recently, many efforts based on convolutional neural networks have been devoted to the pediatric brain segmentation and achieved encouraging results. For instance, Wang et al. [8] designed an anatomy-guided densely-connected U-Net architecture to perform the infant brain segmentation task, with the anatomical prior as guidance to improve the segmentation accuracy. Nie et al. [9] proposed to train fully convolutional networks for each modality image, and then fuse their high-layer features together to obtain final segmentation. Zöllei et al. [10] proposed an automated segmentation and surface extraction pipeline (named as *Infant FreeSurfer*) for T1-weighted (T1w) neuroimaging data of infants aged 0–2 years. However, these previous works may not handle the motion/Gibbs artifacts. For example, as shown in Fig. 1, *Infant FreeSurfer* could achieve an accurate tissue segmentation result on an artifacts-free image (Fig. 1(a)), but fails on an image in presence of motion/Gibbs artifacts, as shown in Fig. 1(b).



**Fig. 1.** Impact of artifacts for tissue segmentation. Results by *Infant FreeSurfer* [10] on pediatric brain T1w MR images without/with motion/Gibbs artifacts at 24 months of age. The red and orange dashed ellipses indicate some unreliable results, which are mainly caused by motion/Gibbs artifacts as shown in (b).

Moreover, the collaborative use of multi-domain images (acquired from different imaging sites) is more prevalent recently, which makes the segmentation task more difficult. Generally, a model trained on a specific-site dataset often performs well on testing subjects from the same site, but poorly on subjects from other sites with different protocols/scanners. This is called the “multi-site issue” or “domain-shift issue” problem in medical image analysis. For instance, a MICCAI grand challenge on 6-month infant brain MRI segmentation from multiple sites (i.e., iSeg-2019, <https://iseg2019.web.unc.edu/>) reported and discussed this critical issue [1].

To accurately segment multi-site pediatric images with artifacts, we present a multi-scale self-supervised learning (M-SSL) framework in this paper. In the training stage, inspired by [8], we first train a segmentation model based on the downsampled images to estimate coarse tissue probabilities and build a global anatomic guidance. We then train another segmentation model based on the original images to estimate fine tissue probabilities. The global anatomic guidance and the fine tissue probabilities are integrated as inputs to train a final segmentation model. In the testing stage, to alleviate the multi-site issue, we propose an iterative self-supervised learning (SSL) strategy to train a *site-specific* segmentation model based on a set of reliable training samples, which are

automatically generated and iteratively updated, for a to-be-segmented site. The main contributions of this paper are summarized as follows:

1. We propose a framework to accurately segment multi-site pediatric brain MR images with motion/Gibbs artifacts.
2. We leverage downsampled tissue segmentations to build a global anatomic guidance, which alleviates the motion/Gibbs artifacts.
3. We propose an iterative SSL strategy to train *site-specific* segmentation models to minimize the multi-site issue.

## 2 Dataset and Proposed M-SSL Method

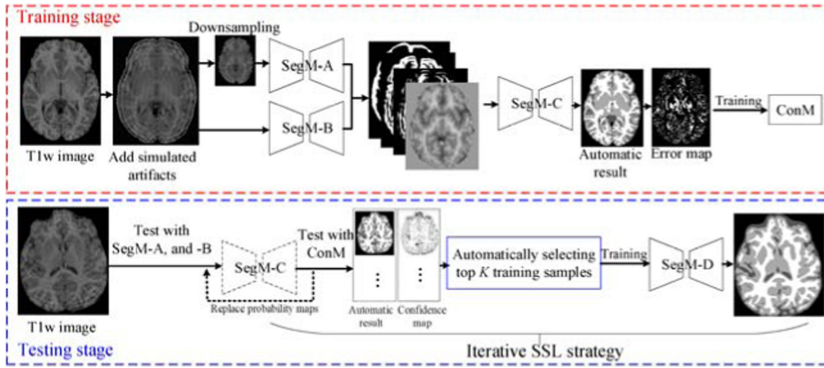
**Dataset and Preprocessing.** T1w pediatric brain MR images used in this study for training were from the UNC/UMN Baby Connectome Project (BCP) [11]. They were acquired at around 24 months of age on Siemens Prisma scanners with 160 sagittal slices using parameters: TR/TE = 2400/2.2 ms and voxel resolution =  $0.8 \times 0.8 \times 0.8 \text{ mm}^3$ . We randomly selected 5 subjects with manual labels as a training dataset. For validation, T1w MR images with real artifacts were from University of Houston, which were acquired with 160 sagittal slices using parameters: TR/TE = 1900/2.98 ms and voxel resolution =  $1.0 \times 1.0 \times 1.0 \text{ mm}^3$ . For image preprocessing, the resolution of all images was resampled into  $0.8 \times 0.8 \times 0.8 \text{ mm}^3$ , then in-house tools were used to perform skull stripping, intensity inhomogeneity correction, and cerebellum removal.

### 2.1 The Proposed Method

We propose a multi-scale self-supervised training (M-SST) framework to accurately segment tissues for multi-site pediatric MR images with motion/Gibbs artifacts, consisting of training and testing stages as shown in Fig. 2. We first elaborate on the training stage, consisting of training three segmentation models and a confidence model to detect reliability of automated segmentation results. Then, we design the testing stage to train a *site-specific* segmentation model based on a set of reliable training samples for the to-be-segmented site. Finally, we introduce the implementation details of the proposed method.

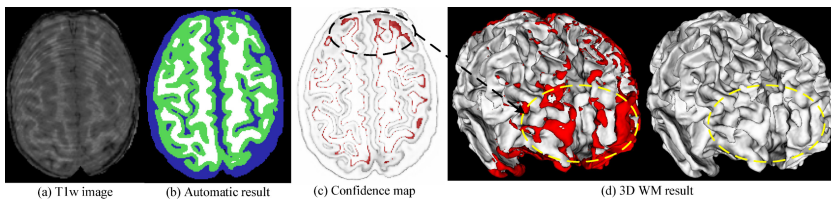
### 2.2 Training Stage

The architecture of a segmentation model can be chosen from U-Net [12], V-Net [13], U-Net++ [14], ADU-Net [8], and nnU-Net [15] et al. In this paper, we adopt ADU-Net as the segmentation architecture, which demonstrates outstanding performance on pediatric brain segmentation. As shown in Fig. 2, in the training stage, benefiting from downsampling and simulated motion/Gibbs artifacts, we first train two segmentation models (named as SegM-A and SegM-B) to generate global anatomic guidance and the fine tissue probabilities, which are integrated as inputs to train the third segmentation model (named as SegM-C) later. Next, an error map, defined as the differences between



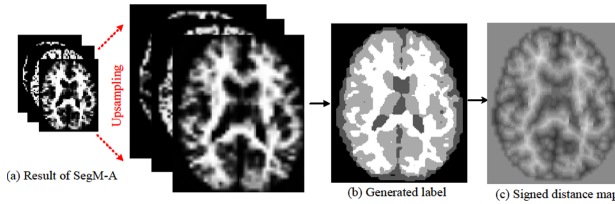
**Fig. 2.** Illustration of our M-SSL method for brain segmentation of pediatric MR images affected by motion/Gibbs artifacts. **Training stage:** Downsampled and original images with simulated artifacts, are input to train SegM-A to build a global anatomical guidance in the downsampled image space and SegM-B for 3 tissue probability maps in the original image space, respectively. Then a four-channel input (one signed distance map from SegM-A and three probability maps from SegM-B) are automatically generated for the training of SegM-C. Finally, a ConM is trained to evaluate the reliability of those automated segmentations at the voxel level. **Testing stage:** After inputting testing subjects into the trained SegM-A, -B and -C, we can obtain automated segmentation results. Then an iterative SSL strategy is proposed to train a *site-specific* segmentation model SegM-D for the to-be-segmented site.

ground truth and automated segmentations from SegM-C, is regarded as targets to train a confidence model (named as ConM) that is able to automatically detect reliability of automated segmentation results at the voxel level. Figure 3 presents the effectiveness of the confidence map evaluated on the automated segmentation of a testing subject, where the testing subject is not included in the training dataset of ConM. It can be seen that some gyral shapes (circled by a yellow dotted ellipse) in the right figure of Fig. 3(d) are not reasonable. The ConM can effectively detect these unreasonable regions, as indicated by the red color in Fig. 3(c), and also in the left figure of Fig. 3(d).



**Fig. 3.** The effectiveness of confidence map on a testing subject acquired at 24 months of age. (a) T1w testing image, (b) segmentation result by SegM-C, (c) the corresponding confidence map (generated by ConM, where some unreliable (darker) regions of WM are marked with red color), and (d) the 3D WM rendering result (the red region corresponds with the red region of (c)). Note that the testing subject is excluded in the training dataset of ConM.

**Global Anatomic Guidance.** Prior knowledge, e.g., the cortical thickness is within a certain range, could be employed as an anatomical guidance for the tissue segmentation [1, 8]. Considering the artifacts as high-frequency noise, we can alleviate the artifacts by simply downsampling the original images. Moreover, the downsampled images allow for a large receptive field during the network training. Therefore, instead of estimating the anatomical guidance from the original images with artifacts, we work on the downsampled images to train the SegM-A. Based on the trained model SegM-A, we then upsample the segmentation result into the original image space and construct a signed distance map with respect to the boundary of WM/GM to incorporate the cortical thickness as an anatomical guidance. In detail, after upsampling the result from SegM-A (see Fig. 4(a)), we can derive a label image (see Fig. 4(b)). Based on the label image, it is straightforward to construct a signed distance map with respect to the boundary of WM/GM, as shown in Fig. 4(c). Basically, the function value at each voxel is the shortest distance to its nearest point on the boundary of WM/GM, taking positive value for voxels inside of WM, and negative value for voxels outside of WM.



**Fig. 4.** (a) Shows the probability maps estimated by the SegM-A, then the maps are upsampled to the original size. (b) Is the generated label according to the upsampled probability maps. (c) Illustrates the signed distance map with respect to the WM/GM boundary.

### 2.3 Testing Stage

Due to the multi-site issue, the trained model in the source site cannot be directly applied to the to-be-segmented site. To alleviate the multi-site issue, we propose an iterative self-supervised learning (SSL) strategy to train a *site-specific* segmentation model for the to-be-segmented site. Based on the idea that the better the probability maps input to SegM-C, the better the outputs of SegM-C, we therefore replace the three tissue probability maps (part of the input of SegM-C) with the output of SegM-C. By iteratively updating the probability maps, the results of SegM-C are gradually refined (i.e., at Round  $N$ ). Then, we apply the SSL method [2] to further refine the results at Round  $N$ , which can effectively refine the segmentation results of testing subjects from multiple sites. In detail, we utilize the SSL method to automatically generate a set of reliable training samples from the testing subjects, which are used to train a *site-specific* segmentation model (i.e., SegM-D). Finally, the testing subjects are directly input to the trained SegM-D to derive final results.

## 2.4 Implementation Details

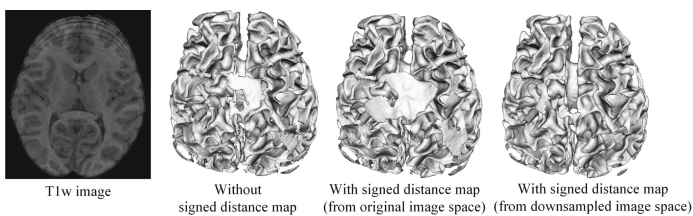
In our experiment, we set 1.0 and 0.35 as the weight parameters for simulated motion and Gibbs artifacts [16] respectively, to preprocess MR images in the training stage of Fig. 2. Then, we randomly extract 1,000 patches (size:  $32 \times 32 \times 32$ ) from each training subject. The loss for segmentation models (i.e., SegM-A, SegM-B, and SegM-C) is cross-entropy, and a spatially-weighted cross-entropy loss [2] is used for training SegM-D. The loss of ConM is multi-task cross-entropy. The kernels are initialized by Xavier [38], and we use SGD optimization strategy. The learning rate is 0.005 and multiplies by 0.1 after each epoch.

## 3 Experimental Results

To demonstrate the performance of our proposed M-SSL method, we first make ablation studies to verify the importance of each component, like downsampling, signed distance map, and the proposed iterative SSL strategy. Then, we validate our method on pediatric MR images with real motion/Gibbs artifacts. Finally, the method is applied on multi-site pediatric brain subjects from the iSeg-2019 challenge [1] to report quantitative analysis.

### 3.1 Ablation Study

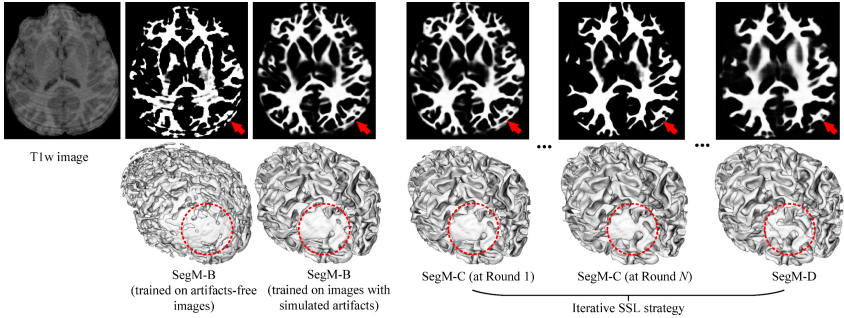
**Influence of Downsampling and the Signed Distance Map.** To validate the effectiveness of downsampling and signed distance map, we make an ablation study to compare the results obtained by SegM-C trained without/with signed distance map (generated from original image space or downsampled image space) as shown in Fig. 5. Obviously, compared with the results in the second and third columns of Fig. 5, the gyrus of WM tissue (the fourth column in Fig. 5) are clearer and more reasonable with the guidance of the signed distance map (generated from downsampled image space).



**Fig. 5.** The importance of downsampling and the signed distance map. From left to right: T1w image and the WM results obtained by SegM-C trained without/with the signed distance map (generated from original image space or downsampled image space).

**Importance of Simulated Artifacts and Iterative SSL Strategy.** In the training stage of Fig. 2, we use simulated motion/Gibbs artifacts to preprocess the intensity images to train SegM-B. Then, we propose an iterative SSL strategy to refine the segmentation results during the testing stage as discussed in Sect. 2.3. We are wondering whether these

manners are helpful to improve the accuracy of testing results or not. Figure 6 shows the results of a testing subject with real motion/Gibbs artifacts generated by SegM-B and the iterative SSL strategy. First, we can see the SegM-B trained on images with simulated artifacts is more robust to deal with testing images with real artifacts (the second and third columns in Fig. 6). Second, by leveraging the proposed iterative SSL method, some ring-like tissue results caused by artifacts are gradually alleviated indicated by red arrows (see the fourth to sixth columns in Fig. 6). Therefore, the simulated artifacts added for training subjects and the proposed iterative SSL strategy are helpful to improve the accuracy of testing results.



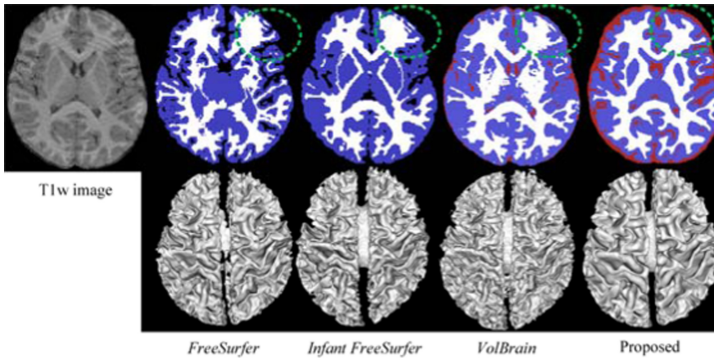
**Fig. 6.** Comparison of segmentation results for testing subjects with real artifacts. From left to right: T1w image, the results generated by SegM-B (trained on images without/with simulated artifacts), SegM-C and SegM-D (by iterative SSL strategy). Some regions are indicated by red arrows to show the difference, and the corresponding 3D rendering WM results are circled by red dotted circles.

### 3.2 Comparison Results on Pediatric Brain Images with Real Artifacts

We first verify the performance of our method on 9 brain T1w images with real motion and Gibbs artifacts at 24 months of age. In this experiment, we compare with three state-of-the-art pipelines/tools, including 1) *FreeSurfer* [6], 2) *Infant FreeSurfer* [10], and 3) *volBrain* [17]. Thanks to the freely releasing these pipelines/tools from pioneers, we can directly apply these pipelines/tools to derive tissue segmentation results according to their manuals. Figure 7 presents exemplary tissue segmentation results of one testing image with severe artifacts, obtained by three competing methods and our method. We can observe that most results generated by competing methods show ring-like tissue shapes circled by green dotted ellipses, which are mainly caused by artifacts. However, our results have smoother and more reasonable tissue segmentations as shown in the last column of Fig. 7. The qualitative comparison clearly demonstrates the advantage of the proposed method in terms of accuracy.

### 3.3 Comparisons on Multi-site Infant Subjects in the ISeg-2019 Challenge

To quantitatively validate of our proposed method, we test the multi-site brain images in the iSeg-2019 challenge. According to the review article [1], we choose one testing site



**Fig. 7.** Tissue segmentation comparison between *FreeSurfer* [6], *Infant FreeSurfer* [10], *volBrain* [17] and our proposed method on a pediatric brain MR image with severe artifacts. The first column shows pediatric T1w images with artifacts. The second to fifth columns are the corresponding segmentation results generated by three competing methods and our method, which are shown with 2D slices and 3D rendering WM tissues, respectively.

(i.e., Stanford University, exhibits different distribution in comparison of other sites) to test our method, which consists of five testing subjects, one of them are significantly affected by motion/Gibbs artifacts. We compare our method with top 3 methods (i.e., QL111111, Tao SMU, and FightAutism) in the challenge as reported in Table 1, in terms of CSF, GM and WM results. From Table 1, our method achieves the highest Dice ratio in terms of GM and WM results, with significant difference compared with others ( $p$ -value  $< 0.05$ ).

**Table 1.** Dice ratio (%) of cross-site brain segmentation results on five testing subjects from the iSeg-2019 challenge. “+” indicates that our proposed method is significantly better than the top three methods with  $p$ -value  $< 0.05$ .

| Method | QL111111         | Tao_SMU          | FightAutism      | Proposed                             |
|--------|------------------|------------------|------------------|--------------------------------------|
| CSF    | $82.77 \pm 1.15$ | $82.89 \pm 1.18$ | $82.28 \pm 1.17$ | <b><math>82.48 \pm 1.02</math></b>   |
| GM     | $79.95 \pm 2.01$ | $79.08 \pm 2.48$ | $79.00 \pm 1.78$ | <b><math>81.93 \pm 1.27^+</math></b> |
| WM     | $85.90 \pm 2.75$ | $84.13 \pm 3.14$ | $84.31 \pm 2.45$ | <b><math>87.67 \pm 1.69^+</math></b> |

## 4 Conclusion

To conclude, we propose a multi-scale self-supervised learning (M-SSL) framework to accurately segment tissues for multi-site pediatric brain MR images with motion/Gibbs artifacts. According to the above experiments, the M-SSL method can achieve encouraging results compared with several state-of-the-art methods. In future work, we will further improve our method and test on more multi-site subjects with artifacts.



**Acknowledgements.** This work was supported in part by National Institutes of Health grants MH109773 and MH117943, National Natural Science Foundation of China under Grant No. 61701192, 61872419, No.61873324, the Natural Science Foundation of Shandong Province, China, under Grant No. ZR2017QF004, No. ZR2019MF040, ZR2019MH106. This work utilizes approaches developed by an NIH grant (1U01MH110274) and the efforts of the UNC/UMN Baby Connectome Project Consortium.

## References

1. Sun, Y., et al.: Multi-site infant brain segmentation algorithms: the iSeg-2019 challenge. *IEEE Trans. Med. Imaging* **40**(5), 1363–1376 (2021)
2. Sun, Y., Gao, K., Niu, S., Lin, W., Li, G., Wang, L.: Semi-supervised transfer learning for infant cerebellum tissue segmentation. In: Liu, M., Yan, P., Lian, C., Cao, X. (eds.) *MLMI 2020*. LNCS, vol. 12436, pp. 663–673. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-59861-7\\_67](https://doi.org/10.1007/978-3-030-59861-7_67)
3. Wang, L., et al.: Benchmark on automatic 6-month-old infant brain segmentation algorithms: the iSeg-2017 challenge. *IEEE Trans. Med. Imaging* **38**(9), 2219–2230 (2019)
4. Shattuck, D.W., et al.: BrainSuite: an automated cortical surface identification tool. *Med. Image Anal.* **6**(2), 129–142 (2002)
5. Jenkinson, M., et al.: FSL. *Neuroimage* **62**(2), 782–790 (2012)
6. Fischl, B.: FreeSurfer. *Neuroimage* **62**(2), 774–781 (2012)
7. Glasser, M.F., et al.: The minimal preprocessing pipelines for the human connectome project. *Neuroimage* **80**, 105–124 (2013)
8. Wang, L., et al.: Volume-based analysis of 6-month-old infant brain MRI for autism biomarker identification and early diagnosis. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) *MICCAI 2018*. LNCS, vol. 11072, pp. 411–419. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-00931-1\\_47](https://doi.org/10.1007/978-3-030-00931-1_47)
9. Nie, D., et al.: Fully convolutional networks for multi-modality isointense infant brain image segmentation. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1342–1345 (2016)
10. Zöllei, L., et al.: Infant FreeSurfer: an automated segmentation and surface extraction pipeline for T1-weighted neuroimaging data of infants 0–2 years. *NeuroImage* **218**, 116946 (2020)
11. Howell, B.R., et al.: The UNC/UMN Baby Connectome Project (BCP): an overview of the study design and protocol development. *Neuroimage* **185**, 891–905 (2019)
12. Milletari, F., et al.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571 (2016)
13. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: a nested U-net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) *DLMIA/ML-CDS -2018*. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-00889-5\\_1](https://doi.org/10.1007/978-3-030-00889-5_1)
14. Isensee, F., et al.: nnU-Net: Breaking the Spell on Successful Medical Image Segmentation (2019). <https://arxiv.org/abs/1904.08128>
15. Czervionke, L.F., et al.: Characteristic features of MR truncation artifacts. *Am. J. Neuroradiol.* **151**(6), 1219–1228 (1988)
16. Manjón, J.V., et al.: volBrain: an online MRI brain volumetry system. (in English). *Front. Neuroinf.* **10**, 30 (2016)