Segmentation of the Cortical Plate in Fetal Brain MRI with a Topological Loss

Priscille de Dumast[1,2(✉)], Hamza Kebiri[1,2], Chirine Atat[1], Vincent Dunet[1], Mériam Koob[1], and Meritxell Bach Cuadra[1,2]

[1] Department of Radiology, Lausanne University Hospital (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland
priscille.guerrierdedumast@unil.ch
[2] CIBM Center for Biomedical Imaging, Lausanne, Switzerland

Abstract. The fetal cortical plate undergoes drastic morphological changes throughout early *in utero* development that can be observed using magnetic resonance (MR) imaging. An accurate MR image segmentation, and more importantly a topologically correct delineation of the cortical gray matter, is a key baseline to perform further quantitative analysis of brain development. In this paper, we propose for the first time the integration of a topological constraint, as an additional loss function, to enhance the morphological consistency of a deep learning-based segmentation of the fetal cortical plate. We quantitatively evaluate our method on 18 fetal brain atlases ranging from 21 to 38 weeks of gestation, showing the significant benefits of our method through all gestational ages as compared to a baseline method. Furthermore, qualitative evaluation by three different experts on 26 clinical MRIs evidences the out-performance of our method independently of the MR reconstruction quality. Finally, as a proof of concept, 3 fetal brains with abnormal cortical development were assessed. The proposed topologically-constrained framework outperforms the baseline, thus, suggesting its additional value to also depict pathology.

Keywords: Fetal brain · Cortical plate · Deep learning · Topology · Magnetic resonance imaging

1 Introduction

The early *in utero* brain development involves complex intertwined processes, reflected in both physiological and structural changes [23]. The developing cortical plate specifically undergoes drastic morphological transformations throughout gestation. Nearly all gyri are in place at birth, even though the complexification of their patterns carries on after birth [18]. T2-weighted (T2w) magnetic resonance imaging (MRI) offers a good contrast between brain tissues, hence allowing to assess the brain growth and detect abnormalities *in utero*. In the

C. H. Sudre et al. (Eds.): UNSURE 2021/PIPPI 2021, LNCS 12959, pp. 200–209, 2021.
https://doi.org/10.1007/978-3-030-87735-4_19

clinical context, fetal MRI is performed with fast, 2D orthogonal series in order to minimize the effect of unpredictable fetal motion but results in low out-of-plane spatial resolution and significant partial volume effect. In order to combine these multiple series, advanced imaging techniques based on super-resolution (SR) algorithms [10,24] allow the reconstruction of 3D high-resolution motion-free isotropic volumes. Together with improved visualization, these SR volumes open up to more accurate quantitative analysis of the growing brain anatomy. Consequently, based on 3D reconstructed volumes, multiple studies explored semi-automated fetal brain tissue segmentation [19] and cortical folding patterns *in-utero* [6,25]. Cortical plate is crucial in early brain development as pathological conditions, e.g. ventriculomegaly, are proved to manifest along with altered foldings [2]. However, cortical plate segmentation remains challenging as it undergoes significant changes due to the brain growth and maturation, respectively modifying the morphology and the image contrast [19]. Furthermore, being a thin layer easily altered by partial volume effect in MRI, anatomical topology is prone to be incorrectly represented by automatic segmentation methods.

In this respect, we present a fully automated and topologically correct age-invariant segmentation method of the cortical plate. In [4,5], the first topological-based segmentation of the fetal cortex was introduced, based on geometrical constraints that integrated anatomical and topological priors. Regrettably, their topological correctness was not further evaluated and qualitative results on only 6 fetuses were presented. More recently, deep learning (DL) methods have also focused on fetal brain MRI cortical gray matter segmentation. Using a neonatal segmentation framework as initialization, [12] proposes a multi-scale approach for the segmentation of the developing cortex, while [9] implements a two-stage segmentation framework with an attention refinement module. Nevertheless, while the segmentation accuracy of these recent DL methods is promising, none of these works assess the topological correctness of their results. In fact, these works report high overlap metrics but illustrated results show lack of topological consistency with notably discontinuous/broken cortical ribbons.

To our knowledge, only two works explore topological fidelity of the segmentation in different applications. In [15], they proposed a topological loss for neuronal membrane segmentation. More recently, topological constraints for MR cardiac image segmentation have been presented [3], although prior topological knowledge is required. In this paper we integrate for the first time a topological constraint, from [15], in a deep image segmentation framework to overcome the limitation of disjoint cortical plate segmentation in fetal MRI and further improve DL architectures (Fig. 1).

2 Methodology

2.1 Topological Loss

Our approach, is based on the topological loss function proposed in [15]. The topology-preserving loss compares the predicted likelihood to the ground truth segmentation using the concept of persistent homology [11]. In a nutshell, homology structures are obtained by filtration to all possible threshold values of the

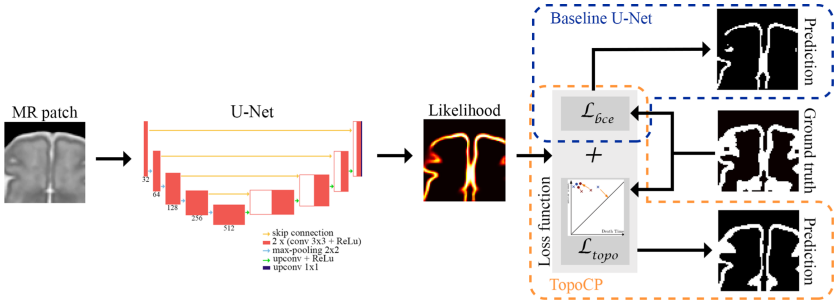

Fig. 1. Figure adapted from [15]. TopoCP, integrates a topological loss based on persistent homology to a 2D U-Net segmentation of cortical plate fetal MRI.

predicted likelihood and reported in a persistence diagram (Fig. 1). Both 0-dimensional and 1-dimensional Betti numbers [13], corresponding respectively to the number of connected components and the number of holes, are tracked. The persistence diagrams of the likelihood and the ground truth are matched, finding the best one-to-one structure correspondence, and the topological loss is computed as the distance between the matched pairs. We refer the reader to the original paper for advanced technical details [15].

2.2 Network Architecture

The topological loss introduced above is indeed compatible with any deep neural network providing a pixel-wise prediction. We chose as baseline the well-established U-Net [22] image segmentation method, as it recently proved its ability to deal with 2D fetal brain MRI tissue segmentation [17]. The baseline 2D U-Net uses a binary cross-entropy loss function $\mathcal{L}_{bce}$. The proposed framework *TopoCP* is based on a 2D U-Net trained using

$$\mathcal{L} = \mathcal{L}_{bce} + \lambda_{topo}\mathcal{L}_{topo}, \tag{1}$$

where $\mathcal{L}_{topo}$ is the topological term in [15] and λ_{topo} the weight of the contribution of $\mathcal{L}_{topo}$ in the final loss.

The 2D U-Net architecture is composed of encoding and decoding paths. The encoding path in our study is composed of 5 repetitions of the followings: two 3×3 convolutional layers, followed by a rectified linear unit (ReLu) activation function and a 2×2 max-pooling downsampling layer. Feature maps are hence doubled from 32 to 512. In the expanding path, 2×2 upsampled encoded features concatenated with the corresponding encoding path are 3×3 convolved and passed through ReLu. The network prediction is computed with final 1×1 convolution. Both Baseline and TopoCP are implemented in Tensorflow. In TopoCP, the topological loss is implemented in C++ and built as a Python library.

2.3 Training Strategy

The publicly available dataset Fetal Tissue Annotation and Segmentation Dataset (FeTA) is used in the training phase [20,21]. Discarding pathological and non-annotated brains, our training dataset results in 15 healthy fetal brains (see details summarized in Table 1). Both networks are fed with 64×64 patches of axial orientation (see Fig. 1), containing cortical gray matter. Intensities of all image patches are standardized and data augmentation is performed by randomly flipping and rotating patches (by $n \times 90°$, $n \in [\![0;3]\!]$). As in [15], to overcome the high computational cost of persistent homology, we adopted the following optimization strategy: 1) our baseline model was trained over 23 epochs with a learning rate decay scheduled at epochs 11, 16, 17, 22 and a decay factor of 0.5, initialized at 0.0001; 2) from the pretrained model in the first step, both networks were fine-tuned over 35 epochs, with a learning rate decay scheduled at epochs 14, 23 for Baseline U-Net and none for TopoCP. TopoCP was trained with $\lambda_{topo} = 1$. A 7-fold cross-validation approach was used to determine the epochs for learning rate decay.

3 Evaluation

3.1 Quantitative Evaluation

Data. In the training dataset (FeTA), label maps were sparse (annotations were performed on every 2^{nd} to 3^{rd} slice) and their interpolation resulted in *noisy* labels with topological inconsistencies. Therefore, we rather evaluate our method on an independent pure testing dataset, presenting a topologically accurate segmentation. The normative spatiotemporal MRI atlas of the fetal brain [14] provides 3D high-quality isotropic smooth volumes along with tissue label maps, including more than fifty anatomical regions, for all gestational age between 21 and 38 weeks (see Table 1 for details). Atlas labels were merged to match the tissue classes represented in our training dataset.

Analysis. Though inferred segmentation rely on 2D patches, performance of the methods is evaluated on the whole 3D segmentation. Three complementary types of evaluation metrics are used: 1) the overlap between the ground truth and the predicted segmentation is quantified with the Dice similarity coefficient (DSC) [8]; 2) a boundary-distance-based metric is measured to evaluate the contours: the 95^{th} percentile of the Hausdorff distance (HD95) [16]; 3) finally, the topological correctness is quantified with the error of a topological invariant: the Euler characteristic (EC), defined as a function of the k-dimensional (k-dim) Betti numbers (B_k), topologically invariant themselves. The 3D Euler characteristic is defined as:

$$EC = B_0 - B_1 + B_2, \tag{2}$$

where B_0 counts the number of connected components, B_1 the number of holes (tunnels) and B_2 counts number of void/cavities encapsulated in the binary

objects. Topology errors are defined as the absolute difference of the ground truth and the prediction measures. For completeness, k-dim Betti errors (BE) are also reported. To assess the significance of the observed differences between the two methods, we perform a Wilcoxon rank sum test for each metrics. p-values were adjusted for multiple comparisons using Bonferroni correction and statistical significance level was set to 0.05.

3.2 Qualitative Evaluation

Data. In order to better represent the diversity of the cortical variability and to prove the generalization of our approach to SR reconstructions of clinical acquisitions, we introduce a second pure testing set of T2w SR images of 26 healthy fetuses. Two subsets were created, from a consensus of three experts evaluation, based on the quality of the reconstructed 3D volumes: 1) excellent (N = 16) and 2) acceptable (N = 10) - with remaining motion artifacts or partial volume effects. Additionally, as a proof of concept, three subjects with cortical plate pathologies were segmented (schizencephaly (1); polymicrogyria (1); corpus callosum agenesis (CCA) and schizencephaly (1)). MR image patches were preprocessed for intensity standardization with no further intensity-based domain adaptation performed. Nevertheless, prior to the segmentation inference, clinical images were resampled to match the resolution of the training data using ANTs [1] in order to present a similar field of view (see Table 1).

Analysis. Three experienced raters (two radiologists and one engineer) performed independently a qualitative analysis of the baseline and TopoCP segmentations. For healthy subjects, randomly-ordered segmentation of axial slices from healthy subjects were presented. The experts were asked to indicate if they preferred either the segmentation A or B or if they were of equivalent quality. The inter-rater reliability was assessed with their percentage agreement before considering a consensus evaluation resulting from the majority voting of the experts' evaluations. For the pathological cases, three radiologists, blindly assessed the whole 3D volume to ensure that the pathological area was included.

Table 1. Summary of the data used for training and quantitative and qualitative evaluation.

Dataset	Field strength	Vendor	Num. of subjects	Gestational age (weeks)	Reconstruction method	Resolution (mm^3)
Training	1.5T; 3T	General electric	15	[22.6–33.4] (28.7 ± 3.5)	mialSRTK [7,24]	$0.5 \times 0.5 \times 0.5$
Evaluation quantitative	1.5T; 3T	Siemens; Philips	18	21–38	Gholipour et al., 2017 [14]	$0.8 \times 0.8 \times 0.8$
Evaluation qualitative	1.5T	Siemens	29	[18–25] (27.8 ± 4.1)	mialSRTK [7,24]	$1.12 \times 1.12 \times 1.12$

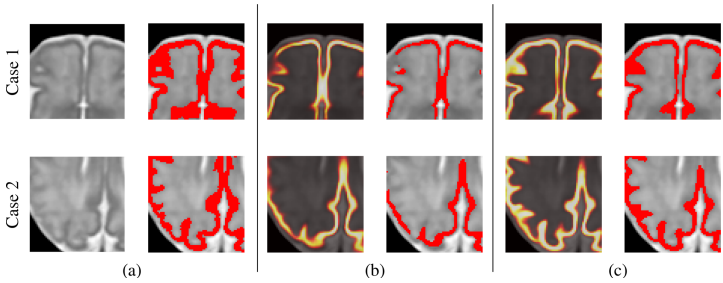

Fig. 2. Segmentation results on 35 weeks of gestation atlas. (a) T2w (left) and ground truth segmentation overlaid (right). (b) Baseline U-Net and (c) TopoCP: predicted likelihood (left) and estimated segmentation (right). Likelihood probabilities: 0 1. Case 1 illustrates a net improvement in the segmentation of the midsagittal area and frontal cortical foldings. Case 2 shows a more accurate detection of the deep sulci with TopoCP.

4 Results

Figure 2 shows the ground truth of two representative patches with their predicted likelihood and segmentation overlaid on the T2w SR image. These results illustrate the benefits of TopoCP on the estimated probability maps, detecting more subtle variation of the cortex. The improved likelihood echoes with a better segmentation. A summary of the 3D performance (Sect. 3.1) metrics on the fetal brain atlas is presented in Table 2. TopoCP outperformed the Baseline U-Net in both similarity- and distance-based evaluation metrics. Corrected p-values between both methods (shown in italics) indicate that our method significantly improves the baseline segmentation. Regarding the topological correctness, the holistic EC error shows significant improvement with TopoCP. The 1-dim BE is the most improved Betti Error and with the highest impact on the global topological assessment. We recall that it represents the error of bored cortical ribbon compared to the ground truth, which is the initial problem addressed. Besides, it should be noted that the 0-dim BE is deteriorated with TopoCP. Visual inspection shows the presence of small isolated false positives in the deep gray matter area. Although, these false positives do not echo with impaired similarity and distance-based metrics. We hypothesise that this behaviour would be due to the fact that training was done on positive (cortex-aware) patches only. We believe these false positive can be reduced with the integration of negative patches in the training phase. Nonetheless, the 3D topology of the cortical plate with TopoCP is much closer to the reality than with Baseline U-Net (see Fig. 3a). Moreover, we observe large standard deviations in the topology-based metrics, although they are slightly reduced with TopoCP (Table 2). Figure 3b shows that the performance metrics varies over the gestational age. For both methods, we observe better performances in the middle of the gestational age range, which we explain as this corresponds to the age range present in the training set (see Table 1). Furthermore, third trimester fetuses benefits more from TopoCP than

Table 2. Performances (mean ± standard deviation), best score for each metric in bold. p-values (in italics) of paired Wilcoxon rank sum test adjusted with Bonferroni multiple comparisons correction, between both methods for each metric.

	DSC↑		HD95 (mm)↓		0-d BE↓		1-d BE↓		2-d BE↓		EC Error↓	
	Baseline U-Net	TopoCP	Baseline U-Net	TopoCP	Baseline U-Net	TopoCP	Baseline U-Net	TopoCP	Baseline U-Net	TopoCP	Baseline U-Net	TopoCP
3D	0.57 ± 0.07	**0.72 ± 0.05**	3.5 ± 0.87	**2.58 ± 0.96**	**10.1 ± 10.8**	13.3 ± 9.6	61 ± 30.3	**35.4 ± 23.7**	8.5 ± 13.4	**8.0 ± 12.2**	60.1 ± 33.4	**30.0 ± 25.3**
	2e-07		*0.0053*		*0.076*		*0.00099*		*0.96*		*0.00075*	

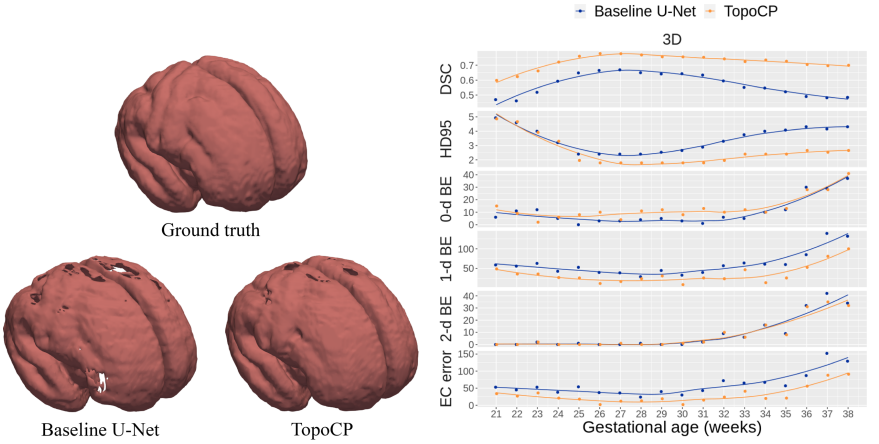

Fig. 3. (a) 3D rendering of 28 weeks-old atlas cortical plate segmentation from both automatic methods compared to the ground truth. (b) Performance metrics at the subject-level computed on the whole 3D volume for all atlas images.

others. TopoCP is more valuable to older fetuses, as they undergo the more complex cortical gyrification patterns. While the overlap metric constantly improves throughout gestation, distance error is mainly enhanced from the third trimester. The topological loss has a stronger positive effect on the topological errors for old subjects, although the whole range of gestational age presented benefits from it.

Qualitative assessment of healthy fetuses indicates a good inter-rater agreement of 74%. Figure 4a shows the consensus of the experts' blind evaluation of the cortical plate segmentation on SR volumes based on T2w clinical acquisitions. For both excellent and acceptable sets, TopoCP was selected as giving the best segmentation (overall on 81% of the slices), showing the robustness of our method to the SR quality. Figure 4b illustrates a representative slice segmented with both methods. Similarly, all raters preferred TopoCP segmentation in the three pathological cases (CCA and schizencephaly shown in Fig. 5).

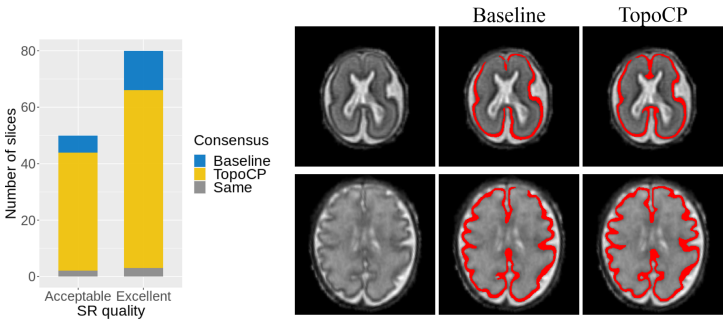

Fig. 4. (a) Experts' qualitative evaluation results in the comparison of Baseline U-Net and TopoCP automatic segmentations. (b) Segmentation results on 23 (top) and 32 (bottom) gestational weeks fetuses.

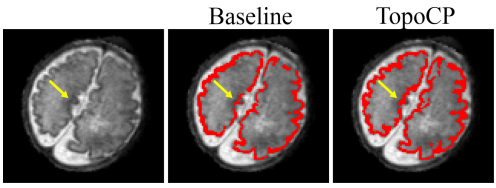

Fig. 5. Segmentation results of a 33 weeks old subject with corpus callosum agenesis and schizencephaly. Yellow arrows indicate the pathological area, where TopoCP is better performing. (Color figure online)

5 Discussion and Conclusion

This work assesses for the first time the integration of a topological constraint in DL-based segmentation of the fetal cortical plate on MRI. Our results on a wide range of gestational ages (21 to 38 weeks) (measured with 3D topology error) and qualitative assessment on 29 clinical subjects (including 3 with cortical pathologies) demonstrate the resulting improved topological correctness of the fetal cortex, despite noisy training labels and 2D inference. Our approach can possibly be extended to 3D, although, one should note that an increase in the input dimension will echo to an increase of the computational cost. In this study, we arbitrarily set to 1 the weight of the topological loss, as done in [15]. We acknowledge the loss contribution has its influence in the training phase and should be fine tuned for improved performance. By testing our method on different acquisitions than those of the training phase, we observe that the segmentation quality of our method seems robust to different scanners and reconstruction methods. Nevertheless, the main drawback of our work is its sensitivity to the resolution of the input image. Resampling of both the input image and result segmentation introduces interpolation that might embed the final results. We hypothesize that training on images of various resolutions would make our method more robust

to this parameter. We briefly presented preliminary results showing the benefits of TopoCP in the segmentation of pathological cortical plates. While all training images were of neurotypical fetal brains, we assume pathological brains could be added to training set to better represent the variability of fetal cortical plates. Finally, we emphasize the genericity of this loss, which can be applied to any segmentation network providing a pixel-wise prediction. We believe that pairing up the topological loss with state-of-the-art methods would considerably improve the resulting segmentation, even in a multi-class task.

Acknowledgments. This work was supported by the Swiss National Science Foundation (project 205321-182602). We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), Ecole polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE) and Geneva University Hospitals (HUG).

References

1. Avants, B., et al.: A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage **54**(3), 2033–2044 (2011). https://doi.org/10.1016/j.neuroimage.2010.09.025
2. Benkarim, O.M., et al.: Cortical folding alterations in fetuses with isolated non-severe ventriculomegaly. NeuroImage Clin. **18**, 103–114 (2018). https://doi.org/10.1016/j.nicl.2018.01.006
3. Byrne, N., Clough, J.R., Montana, G., King, A.P.: A persistent homology-based topological loss function for multi-class CNN segmentation of cardiac MRI. In: Puyol Anton, E., et al. (eds.) STACOM 2020. LNCS, vol. 12592, pp. 3–13. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-68107-4_1
4. Caldairou, B., et al.: Segmentation of the cortex in fetal MRI using a topological model. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, USA, pp. 2045–2048. IEEE, March 2011. https://doi.org/10.1109/ISBI.2011.5872814
5. Caldairou, B., et al.: Data-driven cortex segmentation in reconstructed fetal MRI by using structural constraints. In: Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W. (eds.) CAIP 2011. LNCS, vol. 6854, pp. 503–511. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23672-3_61
6. Clouchoux, C., et al.: Quantitative in vivo MRI measurement of cortical development in the fetus. Brain Struct. Funct. **217**(1), 127–139 (2012). https://doi.org/10.1007/s00429-011-0325-x
7. Deman, P., et al.: meribach/mevislabFetalMRI: MEVISLAB MIAL super-resolution reconstruction of fetal brain MRI v1.0 (2020). https://doi.org/10.5281/zenodo.3878564
8. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)
9. Dou, H., et al.: A deep attentive convolutional neural network for automatic cortical plate segmentation in fetal MRI (2020). arXiv: 2004.12847
10. Ebner, M., et al.: An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI. NeuroImage **206** (2020). https://doi.org/10.1016/j.neuroimage.2019.116324

11. Edelsbrunner, H., et al.: Topological persistence and simplification. Discret. Comput. Geom. **28**(4), 511–533 (2002). https://doi.org/10.1007/s00454-002-2885-2
12. Fetit, A.E., et al.: A deep learning approach to segmentation of the developing cortex in fetal brain MRI with minimal manual labeling. In: Medical Imaging with Deep Learning (MIDL) (2020). https://openreview.net/forum?id=SgZo6XA-l
13. Gardner, M.: The Sixth Book of Mathematical Games from Scientific American. WH Freeman, New York (1984)
14. Gholipour, A., et al.: A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth. Sci. Rep. **7**(1), 476 (2017). https://doi.org/10.1038/s41598-017-00525-w
15. Hu, X., et al.: Topology-preserving deep image segmentation. In: Advances in Neural Information Processing Systems, vol. 32, pp. 5657–5668. Curran Associates, Inc. (2019)
16. Huttenlocher, D.P., et al.: Comparing images using the Hausdorff distance. IEEE Trans. Pattern Anal. Mach. Intell. **15**(9), 850–863 (1993). https://doi.org/10.1109/34.232073
17. Khalili, N., et al.: Automatic brain tissue segmentation in fetal MRI using convolutional neural networks. Magn. Reson. Imaging **64**, 77–89 (2019). https://doi.org/10.1016/j.mri.2019.05.020
18. Lenroot, R.K., Giedd, J.N.: Brain development in children and adolescents: insights from anatomical magnetic resonance imaging. Neurosci. Biobehav. Rev. **30**(6), 718–729 (2006). https://doi.org/10.1016/j.neubiorev.2006.06.001
19. Makropoulos, A., et al.: A review on automatic fetal and neonatal brain MRI segmentation. NeuroImage **170**, 231–248 (2018). https://doi.org/10.1016/j.neuroimage.2017.06.074
20. Payette, K., Jakab, A.: Fetal tissue annotation dataset feta, February 2021. https://doi.org/10.5281/zenodo.4541606. https://doi.org/10.5281/zenodo.4541606
21. Payette, K., et al.: A comparison of automatic multi-tissue segmentation methods of the human fetal brain using the feta dataset (2020). arXiv: 2010.15526
22. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
23. Tierney, A.L., Nelson, C.A.: Brain development and the role of experience in the early years. Zero Three **30**(2), 9–13 (2009)
24. Tourbier, S., et al.: An efficient total variation algorithm for super-resolution in fetal brain MRI with adaptive regularization. NeuroImage **118**, 584–597 (2015). https://doi.org/10.1016/j.neuroimage.2015.06.018
25. Wright, R., et al.: Automatic quantification of normal cortical folding patterns from fetal brain MRI. NeuroImage **91**, 21–32 (2014). https://doi.org/10.1016/j.neuroimage.2014.01.034