

# An Automated Localization, Segmentation and Reconstruction Framework for Fetal Brain MRI

Michael Ebner<sup>1,\*</sup>, Guotai Wang<sup>1,\*</sup>, Wenqi Li<sup>1</sup>, Michael Aertsen<sup>2</sup>,  
Premal A. Patel<sup>1</sup>, Rosalind Aughwane<sup>1,3</sup>, Andrew Melbourne<sup>1</sup>, Tom Doel<sup>1</sup>,  
Anna L. David<sup>3,4</sup>, Jan Deprest<sup>1,3,4</sup>, Sébastien Ourselin<sup>1,5</sup>, and  
Tom Vercauteren<sup>1,4,5</sup>

<sup>1</sup>Translational Imaging Group, WEISS, University College London, UK

<sup>2</sup>Dept. of Radiology, University Hospitals KU Leuven, Belgium

<sup>3</sup>Inst. for Women’s Health, University College London, UK

<sup>4</sup>Dept. of Obstetrics and Gynaecology, University Hospitals KU Leuven, Belgium

<sup>5</sup>School of Biomedical Engineering & Imaging Sciences, King’s College London, UK

{[michael.ebner.14](mailto:michael.ebner.14@ucl.ac.uk), [guotai.wang.14](mailto:guotai.wang.14@ucl.ac.uk)}@ucl.ac.uk

\*Authors contributed equally

**Abstract.** Reconstructing a high-resolution (HR) volume from motion-corrupted and sparsely acquired stacks plays an increasing role in fetal brain Magnetic Resonance Imaging (MRI) studies. Existing reconstruction methods are time-consuming and often require user interaction to localize and extract the brain from several stacks of 2D slices. In this paper, we propose a fully automatic framework for fetal brain reconstruction that consists of three stages: 1) brain localization based on a coarse segmentation of a down-sampled input image by a Convolutional Neural Network (CNN), 2) fine segmentation by a second CNN trained with a multi-scale loss function, and 3) novel, single-parameter outlier-robust super-resolution reconstruction (SRR) for HR visualization in the standard anatomical space. We validate our framework with images from fetuses with variable degrees of ventriculomegaly associated with spina bifida. Experiments show that each step of our proposed pipeline outperforms state-of-the-art methods in both segmentation and reconstruction comparisons. Overall, we report automatic SRR reconstructions that compare favorably with those obtained by manual, labor-intensive brain segmentations. This potentially unlocks the use of automatic fetal brain reconstruction studies in clinical practice.

## 1 Introduction

Fetal Magnetic Resonance Imaging (MRI) has become increasingly important for prenatal diagnosis as a complementary tool to prenatal sonography. To mitigate the effect of fetal (and maternal) motion, fast imaging methods such as Single-Shot Fast Spin Echo (SSFSE) are used to acquire thick, low-resolution (LR) 2D slices that can largely freeze in-plane motion. However, in order to assess and quantify fetal pathology, it is highly desirable to reconstruct a single isotropic, high-resolution (HR) volume. Existing reconstruction toolkits generally rely on

an iterative motion-correction/reconstruction approach [3, 6, 9]. Since the position and orientation of the fetal brain vary significantly between patients in relation to maternal structures, localizing the fetal brain and obtaining a segmented mask to exclude the surrounding tissues is crucial for achieving accurate motion-correction. At present, this usually requires manual localization of the fetal brain and uses manual or semi-automatic methods to obtain fetal brain masks, which is laborious and time consuming. To avoid this, [1] reconstructs the entire field-of-view by breaking each slice into patches. However, in addition to increased computational requirements, this leads to non-rigid motion-correction and therefore suboptimal outcomes for rigidly moving regions such as the fetal brain. We believe that a fully automatic reconstruction pipeline based on *automatic* brain localization, extraction (segmentation) and *robust* reconstruction steps is favored to achieve accurate fetal brain reconstructions and potential clinical translation.

Several works investigate automatic fetal brain localization and segmentation. In [7], these are performed by an automatic reconstruction pipeline using a Support Vector Machine and Random Forests, but these are limited by hand-crafted features and testing inefficiency. The template-based automatic method proposed in [11] takes hours for localization. Convolutional Neural Networks (CNN) are used in [10] for fast slice-based fetal brain segmentation but can easily obtain false positives and show poor performance for challenging cases.

Robust super-resolution reconstruction (SRR) methods for fetal MRI have been proposed previously to prevent slice misregistrations from affecting the SRR outcome [3, 6]. However, in [3] no complete outlier rejection is achieved and the method in [6] relies on multiple hyperparameters to be tuned in order to achieve optimal reconstructions while both require time-consuming optimization methods due to their resulting non-convex problem formulation.

We propose a novel framework for automatic fetal brain reconstruction from fetal MRI as shown in Fig. 1. First, we propose a coarse segmentation-based localization using a CNN. Second, we use a multi-scale loss function for training a second CNN to obtain a fine segmentation of the fetal brain. Third, we introduce an effective complete outlier-rejection approach for robust SRR that relies only on a single hyperparameter and retains a linear least-squares formulation. The proposed framework is validated by producing HR 3D volumes from MR images of fetuses with spina bifida (SB) aperta who were assessed prior to and after in-utero open surgical repair. This procedure is performed in selected fetuses and improves neurological outcome compared to a postnatal repair [8].

## 2 Methods

**Localization based on Coarse Segmentation.** Differently from traditional top-down object localization methods using sliding window classification or bounding box regression, we use a bottom-up strategy for fetal brain localization based on a coarse segmentation by a CNN with pixel-level prediction. The framework is theoretically applicable to different CNN models, and we select P-Net [12] for its compactness and efficiency. We refer to this localization task network as

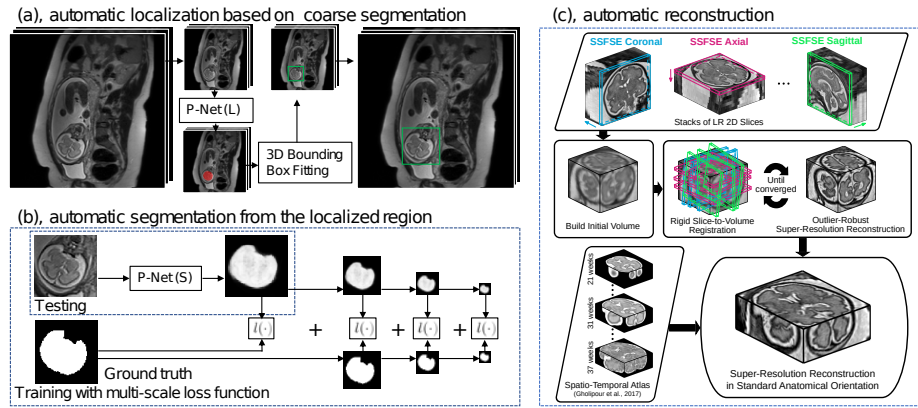


Fig. 1. The proposed fully automatic framework for fetal brain reconstruction.

**P-Net (L).** Let  $I$  denote a stack of slices and  $I_i$  represent the  $i$ -th 2D slice of  $I$ . To reduce the inference time in this stage, we down-sample  $I_i$  to  $I'_i$  with a given size  $96 \times 96$ . We use  $I'$  to denote the whole down-sampled stack. As shown in Fig. 1(a), to get a 3D bounding box of the fetal brain, we first use the P-Net (L) to obtain a segmentation of  $I'$  by stacking the 2D segmentations, i.e., a coarse segmentation, which is smoothed by morphological closing and opening. Then we fit a 3D bounding box to the largest connected component of the smoothed segmentation. The 3D bounding box is rescaled to the original space of  $I$  and expanded by a margin of 5 mm, and used as the localization result.

**Fine Segmentation.** After localization, we use a second CNN to obtain a fine segmentation of the fetal brain. We consistently use the P-Net structure [12] for the fine segmentation and refer to it as **P-Net (S)**. As shown in Fig. 1(b), P-Net (S) works on the extracted region of  $I$  based on the localization result, rather than the entire image of  $I$ , to reduce false positives. The commonly adopted logistic loss and Dice loss function for image segmentation use a sum of pixel-wise losses, without considering the relationship between neighboring pixels. This can result in noise and spatial inconsistency in the prediction. To alleviate this problem, we propose a training loss function across multiple scales  $L_S(Y, \hat{Y}) := \frac{1}{S} \sum_{s=1}^S l(Y_s, \hat{Y}_s)$  where  $Y$  is the prediction of an image given by the segmentation CNN,  $\hat{Y}$  is the corresponding ground truth and  $S$  is the number of scales.  $Y_s$  and  $\hat{Y}_s$  are the rescaled versions of  $Y$  and  $\hat{Y}$  at scale  $s$ , respectively. Here, the rescale operation at scale  $s$  is implemented by average pooling with a kernel size  $2^{s-1} \times 2^{s-1}$ .  $l(\cdot)$  is a basic loss function, e.g, logistic or Dice loss. When  $S > 1$ ,  $l(\cdot)$  encourages the prediction to be close to the ground truth at a higher level.

**Outlier-Robust Super-Resolution Reconstruction.** Taking advantage of these brain segmentations, we deploy a mask-guided rigid motion correction for

all masked slices  $\mathbf{y}_k$  followed by an outlier-robust SRR step to recover the most likely HR volume  $\mathbf{x}^i$  which satisfies the slice acquisition model  $\mathbf{y}_k^i \approx \mathbf{A}_k^i \mathbf{x}^i$  [3] with a linear slice-acquisition operator  $\mathbf{A}_k^i$  for reasonably well motion-corrected slices  $\mathbf{y}_k^i$  at iteration  $i$ . The SRR is solved using a maximum a-posteriori formulation

$$\mathbf{x}^i := \arg \min_{\mathbf{x}} \left( \sum_{k \in \mathcal{K}_\sigma^i} \frac{1}{2} \|\mathbf{y}_k^i - \mathbf{A}_k^i \mathbf{x}\|_{\ell^2}^2 + \frac{\alpha}{2} \|\nabla \mathbf{x}\|_{\ell^2}^2 \right) \quad \text{s. t.} \quad \mathbf{x} \geq 0 \quad (1)$$

for a slice-index set  $\mathcal{K}_\sigma^i := \{1 \leq k \leq K : \text{Sim}(\mathbf{y}_k^i, \mathbf{A}_k^i \mathbf{x}^{i-1}) \geq \sigma\}$  containing only slices in high agreement with their simulated counterparts projected from the previous HR iterate according to a similarity measure  $\text{Sim}$  and parameter  $\sigma > 0$ .  $\alpha \geq 0$  denotes a regularization parameter and  $\nabla$  the differential operator. We thus have a convex SRR problem with complete outlier removal in a linear least-squares formulation that is efficiently solvable using matrix-free operations [2].

### 3 Experiments and Results

**Data and Implementation.** The automatic reconstruction framework was applied to a cohort of 16 fetuses with SB, scanned at University Hospitals KU Leuven before (B1) and after (B2) surgical treatment at the gestational age (GA) of  $23.47 \pm 0.92$  weeks and  $25.73 \pm 1.28$  weeks, respectively. For each study, 3 to 10 SSFSE stacks in different planes were collected, with pixel size 0.63 mm to 1.58 mm and slice thickness 3 mm to 6 mm.

For detection and segmentation, separate cohorts of 30 healthy and 16 fetuses with SB (before treatment) with GA  $29.51 \pm 4.46$  weeks and GA  $23.60 \pm 3.11$ , respectively, were used for training (126 stacks) and validation (12 stacks). The groups B1 (119 stacks) and B2 (105 stacks) were used for testing. Manual segmentations of the fetal brains were used as the ground truth for the segmentation task. The bounding box of the manual segmentation was extended by 5 mm and used as the ground truth for fetal brain localization. Stack intensities were normalized by its mean and standard deviation. Our CNNs were implemented in TensorFlow using NiftyNet [5]. We used  $S = 4$  scales for  $L_S$  and used Dice loss as the basic loss function. We used Adaptive Moment Estimation (Adam) for training, with initial learning rate  $10^{-3}$ , weight decay  $10^{-7}$  and  $10k$  iterations.

The SRR algorithm was applied to the B1 and B2 cases using the obtained automatic segmentations for the rigid slice-to-volume registrations. The image data was preprocessed using ITK bias field and linear intensity correction. Three two-step rigid slice-to-volume registration and outlier-robust SRR iterations were performed whereby  $\text{Sim}$  was set to be normalized cross-correlation (NCC) and  $\sigma$  was empirically set to be 0.6, 0.65 and 0.7 per iteration to account for increasing accuracy in (1), respectively. The regularization parameter  $\alpha = 0.02$  was determined using L-curve studies. Final SRRs were reconstructed in the standard anatomical space by using a brain-volume matched template from a spatiotemporal atlas [4]. The outlier-robust SRR part was implemented in Python using ITK and the LSMR solver for (1). It is made open-source on GitHub<sup>1</sup>.

<sup>1</sup> <https://github.com/gift-surg/NiftyMIC>

**Table 1.** Quantitative evaluation of different methods for fetal brain localization with \* denoting significant differences based on a paired  $t$ -test ( $p < 0.05$ ).

	IoU (%)		Centroid distance (mm)		Stack-level runtime(s)
	Group B1	Group B2	Group B1	Group B2	
Keraudren et al. [7]	72.87±10.37	69.18±11.73	7.31±4.58	7.68±2.83	15.03±3.54
<b>P-Net (L)</b>	<b>84.74±5.55*</b>	<b>83.67±5.04*</b>	<b>3.70±2.10*</b>	<b>4.51±2.45*</b>	<b>2.35±1.02*</b>

**Table 2.** Quantitative evaluation of fetal brain segmentation with \* denoting significant better performance compared with P-Net (S) based on a paired  $t$ -test ( $p < 0.05$ ).

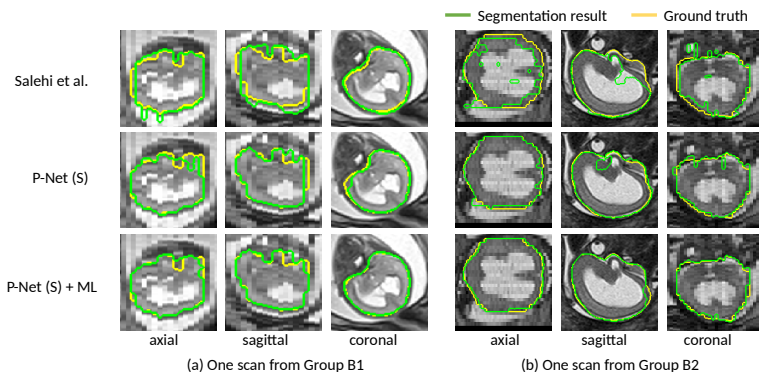
	Dice (%)		Hausdorff (mm)		Stack-level runtime(s)
	Group B1	Group B2	Group B1	Group B2	
Salehi et al. [10]	90.12±4.19	88.57±4.24	14.75±7.51	12.08±6.63	<b>1.98±0.76</b>
P-Net (S)	91.56±3.33	90.93±4.95	10.93±5.66	9.83±5.29	3.65±1.34
<b>P-Net (S) + ML</b>	<b>93.87±2.79*</b>	<b>92.94±4.14*</b>	<b>6.94±4.29*</b>	<b>7.84±3.26*</b>	3.66±1.33

**Localization and Segmentation Results.** Table 1 summarises the localization results of P-Net (L) compared against the method of Keraudren et al. [7] using the Intersection over Union (IoU) score and the centroid distance between the localized and the ground truth bounding box. A visual comparison is provided in the supplementary material. P-Net (L) was trained using the manual segmentations as pixel-wise annotations. Fig. 2 presents a visual comparison of three methods for fetal brain segmentation applied to Group B1 and Group B2 respectively: 1) Salehi et al.<sup>2</sup> [10], applying the U-Net to the whole input image for segmentation without a localization stage, 2) P-Net (S) trained with the basic Dice loss function (at a single scale), and 3) P-Net (S) + ML where P-Net (S) was trained with our proposed multi-scale loss function. Both P-Net (S) and P-Net (S) + ML were applied to the output of P-Net (L). The method of Salehi et al. [10] has a lower performance than our coarse-to-fine segmentation methods. P-Net (S) + ML achieves a better spatial consistency with reduced noises in the segmentation than P-Net (S). Quantitative evaluation of these segmentation methods is presented in Table 2, which shows that the proposed multi-scale loss for training leads to higher segmentation accuracy than P-Net (S).

**Outlier-Robust SRR Results.** Out of the 32 performed reconstructions, six were discarded<sup>3</sup> from the final evaluation leaving 13 cases for each B1/B2 group. For each reconstruction 3/5.7/6/8 (min/mean/median/max) stacks were used. Three different configurations were tested to analyze the input-mask-sensitivity of our proposed SRR algorithm: Reconstructions were performed using the output rectangular masks of P-Net (L), the output masks of P-Net (S) + ML and

<sup>2</sup> We followed the implementation at <https://bitbucket.org/bchradiology/u-net/src> and re-trained the model with our own training images.

<sup>3</sup> Four cases with successful SRRs failed at final template-space alignment step; two failed at SRR due to heavy motion that could not be corrected for by any method.



**Fig. 2.** Visual comparison of different methods for fetal brain segmentation.

**Table 3.** Slice similarities  $\text{Sim}(\mathbf{y}_k^i, \mathbf{A}_k^i \mathbf{x}^i)$  for the respective  $\mathcal{K}_{0.7}^i$ -slices,  $i = 3$ .

	NCC		RMSE		SSIM	
	Group B1	Group B2	Group B1	Group B2	Group B1	Group B2
SRR (M)	<b><u>0.94±0.06</u></b>	<b><u>0.94±0.05</u></b>	<b><u>23.93±8.18</u></b>	<b><u>30.25±26.03</u></b>	<b><u>0.71±0.10</u></b>	<b><u>0.71±0.11</u></b>
<b>SRR (S)</b>	<b><u>0.94±0.05</u></b>	<b><u>0.94±0.06</u></b>	24.05±8.54	30.80±26.28	<b><u>0.71±0.10</u></b>	0.70±0.12
SRR (L)	0.87±0.15	0.87±0.14	29.50±14.07	38.56±35.23	0.60±0.16	0.60±0.17

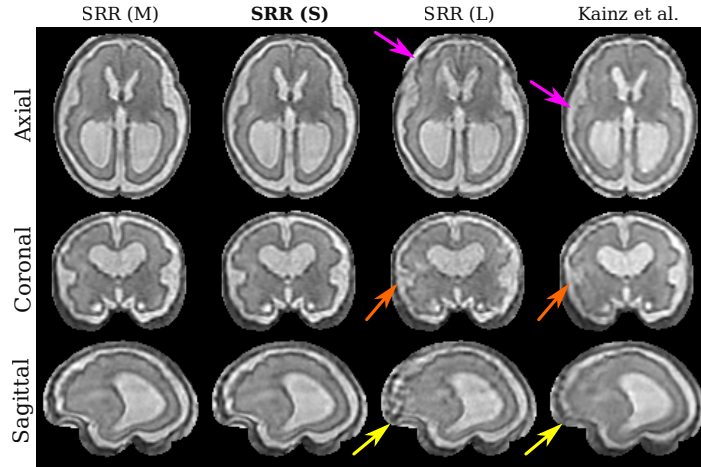
the manual segmentations which we refer to as SRR (L), SRR (S) and SRR (M), respectively. Table 3 shows the slice similarities  $\text{Sim}(\mathbf{y}_k^i, \mathbf{A}_k^i \mathbf{x}^i)$  for the final iteration  $i = 3$  using NCC, root mean squared error (RMSE) and structural similarity (SSIM). The proposed SRR (S) algorithm achieves reconstructions of almost identical self-consistency as SRR (M). In absence of a ground-truth, an additional subjective quality assessment in a clinical context was made where two blinded pediatric radiologists assessed all reconstructions side-by-side. We compared this to reconstructions obtained by the state-of-the-art toolkit developed by Kainz et al. [6] using the manual segmentations as input masks. Table 4 underlines that SRR (S) and SRR (M) achieve reconstructions of high quality that are subjectively almost indistinguishable. The comparison against Kainz et al. [6] confirms the effectiveness of our proposed outlier-robust SRR framework which is also illustrated in Fig. 3.

## 4 Discussion and Conclusion

In this work, we present a fully automated pipeline for fetal brain MRI reconstruction benefiting from deep-learning based automatic fetal brain localization and segmentation. We propose CNN-based coarse segmentation for robust localization and training with a multi-scale loss function for a fine segmentation of the fetal brain. Compared with Keraudren et al. [7], our localization method does not need prior information such as gestational age and achieved superior performance in less time. Unlike [10] which takes a whole image input to a CNN,

**Table 4.** Summary of clinical evaluation. Anatomical clarity was assessed on the cerebellar structure, the aqueduct and the interhemispheric fissure. SRR quality rates introduced artifactual structures and edge uncertainty. The respective average scoring range is shown in square brackets (the higher, the better). Radiologists’ preference ranks subjectively from least (0) to most preferred (3) reconstruction. A \* denotes a significant difference compared to SRR (M) based on a Wilcoxon signed-rank test ( $p < 0.05$ ).

	Anat. Clarity [0 . . . 4]		SRR Quality [0 . . . 2]		Radiologists’ Pref.	
	Group B1	Group B2	Group B1	Group B2	Group B1	Group B2
SRR (M)	2.32±0.55	<b>3.01±0.68</b>	<b>0.88±0.30</b>	1.10±0.35	<b>1.96±0.75</b>	<b>2.04±0.69</b>
<b>SRR (S)</b>	<b>2.36±0.52</b>	<b>3.01±0.75</b>	0.81±0.51	1.10±0.42	1.54±0.85	1.58±0.81
SRR (L)	1.92±0.69*	2.69±0.84*	0.54±0.43*	0.85±0.43*	0.58±0.64*	0.77±0.67*
Kainz et al. [6]	2.01±0.75*	2.92±0.77	0.83±0.33	<b>1.15±0.36</b>	1.92±1.02	1.62±0.79



**Fig. 3.** Visual comparison for one single case showing the SRR (M)/(S) similarity and the effectiveness of the proposed outlier-robust SRR algorithm to prevent artifactual structures such as shown with arrows.

our segmentation method follows a coarse-to-fine way, and leads to higher segmentation accuracy. Moreover, we propose an alternative robust-outlier rejection method during the SRR step which, in contrast to [3, 6], leads to a simple, yet effective, linear least-squares formulation with a single hyperparameter. Overall, our experiments show automatic fetal brain MRI reconstructions that are comparable to manual segmentation-based reconstructions, effectively eliminating the need of any manual intervention. In the future, we aim to apply this framework to quantify the impact of spina bifida repair surgeries by measuring the resolution of the Chiari type II malformation and the degree of ventriculomegaly.

**Acknowledgements.** This work is supported by Wellcome Trust [WT101957; 203145Z/16/Z], EPSRC [EP/L016478/1; NS/A000027/1; NS/A000050/1], and the NIHR UCLH BRC.

## References

1. Alansary, A., Rajchl, M., McDonagh, S.G., Murgasova, M., Damodaram, M., Lloyd, D.F.A., Davidson, A., Rutherford, M., Hajnal, J.V., Rueckert, D., Kainz, B.: PVR: Patch-to-Volume Reconstruction for Large Area Motion Correction of Fetal MRI. *IEEE Trans. Med. Imaging* **36**(10), 2031–2044 (2017)
2. Ebner, M., Chouhan, M., Patel, P.A., Atkinson, D., Amin, Z., Read, S., Punwani, S., Taylor, S., Vercauteren, T., Ourselin, S.: Point-Spread-Function-Aware Slice-to-Volume Registration: Application to Upper Abdominal MRI Super-Resolution. In: *Reconstr. Segmentation, Anal. Med. Images. RAMBO 2016*, vol. 10129, pp. 3–13. Springer International Publishing (2017)
3. Gholipour, A., Estroff, J.A., Warfield, S.K.: Robust Super-Resolution Volume Reconstruction From Slice Acquisitions: Application to Fetal Brain MRI. *IEEE Trans. Med. Imaging* **29**(10), 1739–1758 (2010)
4. Gholipour, A., Rollins, C.K., Velasco-Annis, C., Ouaalam, A., Akhondi-Asl, A., Afacan, O., Ortinau, C.M., Clancy, S., Limperopoulos, C., Yang, E., Estroff, J.A., Warfield, S.K.: A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth. *Sci. Rep.* **7**(1), 476 (2017)
5. Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D.I., Wang, G., Eaton-Rosen, Z., Gray, R., Doel, T., Hu, Y., Whyntie, T., Nachev, P., Modat, M., Barratt, D.C., Ourselin, S., Cardoso, M.J., Vercauteren, T.: NiftyNet: a deep-learning platform for medical imaging. *Comput. Methods Programs Biomed.* **158**, 113–122 (2018)
6. Kainz, B., Steinberger, M., Wein, W., Kuklisova-Murgasova, M., Malamateniou, C., Keraudren, K., Torsney-Weir, T., Rutherford, M., Aljabar, P., Hajnal, J.V., Rueckert, D.: Fast Volume Reconstruction From Motion Corrupted Stacks of 2D Slices. *IEEE Trans. Med. Imaging* **34**(9), 1901–1913 (2015)
7. Keraudren, K., Kuklisova-Murgasova, M., Kyriakopoulou, V., Malamateniou, C., Rutherford, M., Kainz, B., Hajnal, J., Rueckert, D.: Automated fetal brain segmentation from 2D MRI slices for motion correction. *Neuroimage* **101**, 633–643 (2014)
8. Ovaere, C., Eggink, A., Richter, J., Cohen-Overbeek, T.E., Van Calenbergh, F., Jansen, K., Oepkes, D., Devlieger, R., De Catte, L., Deprest, J.A.: Prenatal Diagnosis and Patient Preferences in Patients with Neural Tube Defects around the Advent of Fetal Surgery in Belgium and Holland. *Fetal Diagn. Ther.* **37**(3), 226–234 (2015)
9. Rousseau, F., Glenn, O.A., Iordanova, B., Rodriguez-Carranza, C., Vigneron, D.B., Barkovich, J.A., Studholme, C.: Registration-Based Approach for Reconstruction of High-Resolution In Utero Fetal MR Brain Images. *Acad. Radiol.* **13**(9), 1072–1081 (2006)
10. Salehi, S.S.M., Hashemi, S.R., Velasco-Annis, C., Ouaalam, A., Estroff, J.A., Erdogmus, D., Warfield, S.K., Gholipour, A.: Real-time automatic fetal brain extraction in fetal MRI by deep learning. *arXiv Prepr. arXiv1710.09338* (2017)
11. Tourbier, S., Velasco-Annis, C., Taimouri, V., Hagmann, P., Meuli, R., Warfield, S.K., Bach Cuadra, M., Gholipour, A.: Automated template-based brain localization and extraction for fetal brain MRI reconstruction. *Neuroimage* **155**, 460–472 (2017)
12. Wang, G., Li, W., Zuluaga, M.A., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., Vercauteren, T.: Interactive Medical Image Segmentation using Deep Learning with Image-specific Fine-tuning. *IEEE Trans. Med. Imaging* pp. 1–1 (2018)