2024 IEEE International Symposium on Biomedical Imaging (ISBI) | 979-8-3503-1333-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISBI56570.2024.10635389

AUTOMATED FETAL BRAIN VOLUME RECONSTRUCTION FROM MOTION-CORRUPTED STACKS WITH DEEP LEARNING

*Laifa Ma[1], Weili Lin[1], He Zhang[2], Gang Li[1]**

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, USA
[2] Department of Radiology, Obstetrics and Gynecology Hospital, Fudan University, Shanghai, China

ABSTRACT

The fetal brain MRI 3D volume is critical for malformation diagnosis and development assessment. However, the inevitable fetal motion during MRI acquisition can result in loss of structural continuity and corrupted volumetric information, making it challenging to reconstruct a high-quality 3D volume of the fetal brain from multiple motion-corrupted stacks. Conventional 3D volume reconstruction methods based on slice-to-volume registration are limited to capturing relatively small motion, time-consuming, and sensitive to volume initialization. To address these challenges, we propose a novel end-to-end deep learning method for automated fetal brain MRI 3D volume reconstruction. Firstly, a multi-scale feature fusion model is proposed to solve arbitrary motion correction of 2D slices. It can capture both fine-grained details and global context information at different scales. By fusing features of different scales, the model can accurately estimate fetal motion. Secondly, an initial 3D volume is estimated by point spread function, and an outlier removal strategy is used to reject failed slices. Next, we design a residual-based encoder-decoder model, which can learn an enriched set of features, to improve the quality of the initial 3D volume and reconstruct high-resolution fetal brain MRI 3D volume. We specifically design a perceptual loss, which measures differences based on perceptual features to generate texture and structure information, and introduce an adversarial loss for encouraging anatomically more realistic details. We performed experiments on both simulated and real fetal MR images. The results demonstrate the proposed method is able to solve arbitrary motion correction of 2D slices and reconstruct high-resolution 3D volumes with high efficiency and accuracy, compared to state-of-the-art methods.

Index Terms— Fetal brain MRI, deep learning, motion correction, 3D reconstruction

1. INTRODUCTION

Fetal brain magnetic resonance imaging (MRI) can provide detailed and accurate 3D information of brain morphology, making it particularly critical for fetal brain malformation diagnosis and development assessment [1]. However, the inevitable fetal motion and maternal breathing during MRI acquisition can result in loss of structural continuity and corrupted volumetric information [2]. Although fast imaging techniques can overcome the in-plane motion of individual 2D slices [3], the motion between slices still exists. Therefore, there is an urgent need to reconstruct high-quality fetal brain 3D volumes from multiple motion-corrupted 2D slices. This would enable accurate diagnosis and comprehensive assessment of fetal brain development, and potentially guide interventions to optimize neurodevelopmental outcomes.

The earliest method proposed by Roussea et al. [4] employed the gradient ascent method to maximize the normalized mutual information for slice motion correction, and scattered data interpolation is used to reconstruct fetal MRI 3D volumes. Gholipour et al. [5] combined the slice acquisition model and the robust M-estimation method to reconstruct the high-resolution 3D volume. This method was further improved by Murgasova et al. [6] by integrating intensity matching and robust statistics of 2D slices. Kim et al. [7] proposed a slice intersection-based motion correction method. They formulated the slice-to-volume registration (SVR) problem as a traditional optimization problem by measuring the matching intersection intensity between slices. Michael et al. [8] proposed the two-step iterative SVR and outlier-robust super-resolution reconstruction (SRR) method, which adopts an outlier removal strategy to remove misregistered slices. Bernhard et al. [9] proposed a fast multi-GPU accelerated method for SVR, achieving a high speedup ratio using GPU parallel computing compared to the method using a single CPU. Although these conventional 3D volume reconstruction methods have proven effectiveness, they have three major drawbacks: 1) limited to capturing relatively small motion; 2) sensitive to volume initialization because they require a target volume with small motion between slices as the initial volume; and 3) time consuming [10].

In recent years, deep learning methods have been used to improve the capture range of slice motion correction and speed up the fetal brain MRI 3D reconstruction. SVRnet [10] used the convolutional neural network (CNN) to predict the

∗ is the corresponding author (gang_li@med.unc.edu).

initial transformation information of fetal brain MRI 2D slices, and then used it to initialize the traditional reconstruction method to obtain high-quality brain 3D volumes. Of note, they developed the novel anchor point-based coordinate system to represent the position of fetal 2D motion-corrupted slices in the 3D standard atlas space. Pei et al. [11] proposed a U-shaped network architecture to predict fetal MRI 2D slices motion, which is assisted by a tissue segmentation task to improve the fetal motion estimation accuracy. Xu et al. [12] proposed an Implicit Neural Representation method for slice-to-volume reconstruction. Shi et al. [13] proposed an affinity fusion-based iterative motion correction framework for fetal brain MRI 2D slices. It learns sequential slice motion from multiple slices and uses affinity fusion to integrate features between 2D slices for 3D volume reconstruction. Although these deep learning-based can partially alleviate the problems of conventional iterative SVR method, they only focused on the fetal brain MRI 2D slice motion correction and the time consuming SVR process remains a challenge.

To address above challenges, we propose, to the best of our knowledge, the first end-to-end learning-based fetal brain MRI 3D volume reconstruction method. The main contributions are as follows: 1) a multi-scale feature fusion model is proposed to detect and correct arbitrary motion of 2D slices. By fusing features of different scales, the model can accurately estimate the random fetal motion. 2) An initial 3D volume is estimated by point spread function (PSF), and an outlier removal strategy is used to reject failed slices. 3) A residual-based encoder-decoder model with a specifically designed perceptual loss improves the quality of the initial 3D volume and reconstructs the high-resolution fetal brain MRI 3D volume with anatomically more realistic details ensured by using an adversarial loss.

2. METHOD

Give n motion-corrupted fetal brain MRI 2D slices from multi-stacks, $\boldsymbol{x} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$, the goal is to estimate corresponding transformation parameters $\boldsymbol{T} = [\boldsymbol{T}_1, \ldots, \boldsymbol{T}_n]$ in the standard atlas space, and reconstruct a high-quality fetal brain MRI 3D volume $\boldsymbol{v}$ according to the transformation parameters of these slices. To achieve this goal, we propose an end-to-end highly efficient and effective fetal brain MRI 3D volume reconstruction method based on deep learning, as shown in Fig. 1, which consists of three steps: 1) motion estimation, 2) initial volume reconstruction, and 3) high-quality high-resolution volume reconstruction.

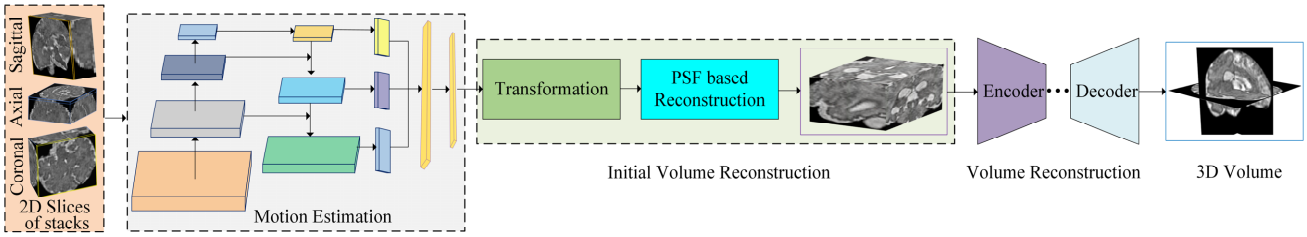

Fig. 1. The framework of the proposed automated fetal brain MRI 3D volume reconstruction.

2.1 Motion Correction Network

Inspired by the Feature Pyramid Network and ResNet [14], we propose a multi-scale feature fusion model to predict the motion of fetal brain MRI 2D slices. It includes a bottom-up path, a top-down path, and lateral connections, as shown in Fig. 2. The bottom-up path is the feature extraction backbone of the model, which is composed of basic residual blocks to generate feature maps at four different scales, with channels of 256, 512, 1024, and 2048, respectively. Each residual block consists of successive convolution of 1×1, 3×3, and 1×1. In Fig. 2, each convolution is followed by group normalization and ReLU activation function.

The top-down path generates higher resolution features through up-sampling. Then, the feature maps of the bottom-up and top-down paths are fused by lateral connections. Low-resolution feature maps are up-sampled by a factor of 2. The channel of feature maps from bottom-up pathway is reduced to 256 using the 1×1 convolution layer, and then added with up-sampled feature maps. Finally, the fusion layer, which consists of the 3×3 convolution, group normalization (GN), ReLU activation function, and adaptive maximum pooling, is used to generate feature maps with channels 256, namely, F3, F4, and F5. Then they are flattened and concatenated to form feature vectors. Two fully connected layers FC1 and FC2 are used to estimate transformation parameters of 2D slices in standard atlas space.

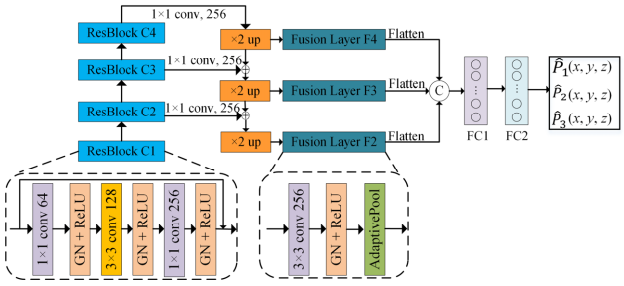

Fig. 2. The framework of the multi-scale feature fusion model used to estimate the arbitrary motion of 2D slices.

The multi-scale feature fusion model can capture both fine-grained details and global context information at different scales, thus enabling the model to handle different

2D slices with different appearance and improving its robustness and ability to predict transformation parameters. Similar to SVRnet [10], the Cartesian coordinate anchor points are used to define transformation parameters, where three non-collinear points are used to determine the position of a slice in the standard space, i.e., the bottom-left corner P_1, center point P_2, and bottom-right corner P_3. To train the model, the loss function is defined as:

$$\mathcal{L}_T = \|\hat{P}_1 - P_1\|_1 + \|\hat{P}_2 - P_2\|_1 + \|\hat{P}_3 - P_3\|_1, \quad (1)$$

Where $\hat{P}_1$, $\hat{P}_2$, and $\hat{P}_3$ are the predicted values of three anchor points, and P_1 , P_2 , and P_3 are the ground truth in the standard anatomical space.

2.2 Robust Initial Volume Reconstruction

According to the n motion-corrupted fetal brain MRI 2D slices $\boldsymbol{x}$, and their corresponding estimated transformation parameters $\boldsymbol{T}$, we apply a slice acquisition model:

$$\boldsymbol{x}_k = \boldsymbol{B}_k \boldsymbol{T}_k \boldsymbol{v}\ ;\ k = 1, \ldots, n, \quad (2)$$

where $\boldsymbol{x}_k$ is the kth slice from stacks, $\boldsymbol{T}_k$ is the corresponding transformation matrix, which is recovered by predicted three anchor points, $\boldsymbol{B}_k$ is the point spread function (PSF). PSF is approximated as a 3D Gaussian function, the variance covariance matrix is $\sum = \mathrm{diag}((\frac{1.2s_1}{2.355})^2, (\frac{1.2s_2}{2.355})^2, (\frac{s_3}{2.355})^2)$, where s_1 and s_2 are the in-plane pixel spacing, and s_3 is the slice thickness [9]. Then we can achieve the initial fetal brain MRI 3D volume $\boldsymbol{v}$ by solving the inverse problem of Eq. (2).

In order to reject slices that failed to predict the transformation parameters, an outlier removal strategy is defined as follow:

$$\boldsymbol{I}_\lambda := \{1 \le k \le n : \mathrm{Sim}(\boldsymbol{x}_k, \boldsymbol{B}_k \boldsymbol{T}_k \boldsymbol{v}) \ge \lambda\}, \quad (3)$$

where $\boldsymbol{I}_\lambda$ is a set of accepted slices whose similarity between them and the slices produced by the slice acquisition model is greater than or equal to the threshold λ, Sim stands for calculating the similarity between two 2D slices (we used we used Normalized Cross-Correlation), λ is set to 0.9, and n is total 2D slices. The initial 3D volume is then updated according to Eq. (2) and $\boldsymbol{I}_\lambda$.

2.3 High-quality 3D Volume Reconstruction

The initial fetal MRI 3D volume is typically blurred and noisy. To improve its quality and further reconstruct high-quality 3D volume, we design a residual-based encoder-decoder model, which can learn an enriched set of textural and structural features, as shown in Fig. 3(a).

The model consists of an encoder and a decoder. The encoder has three 3×3×3 convolutions with stride 2 and residual blocks between them. The decoder consists of three up-sampling steps of feature maps, where the first up-sampling is followed by three residual blocks, the second and third are followed by one residual block. To avoid the checkerboard effect caused by up-sampling, we use the trilinear interpolation and a 3×3×3 convolution to perform the up-sampling. At the last layer, feature maps are mapped onto the dimensions of fetal brain MRI 3D volume using 1×1×1 convolution. In addition, we also introduce skip connections between encoder and decoder, enabling richer semantic features representation.

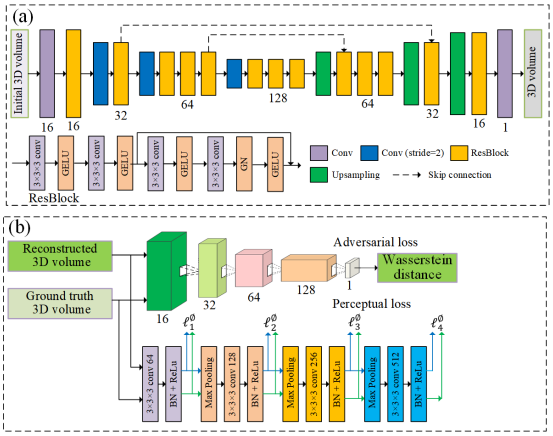

Fig. 3. The framework of the fetal brain MRI 3D volume reconstruction model. (a) The model architecture. (b) Module for calculating the adversarial and perceptual loss.

To optimize the model for achieving better performance, we employ the perceptual loss [15], which measures the difference between two 3D volumes in terms of the feature map similarity, rather than pixel-wise differences to preserve more textural and structural details. Herein, we design a simple model to extract features, as shown in Fig. 3(b), which consists of convolutions, batch normalization, ReLU activation and maximum pooling. We input the ground truth fetal brain MRI 3D volume and the reconstructed 3D volume into this network, and then compute the perceptual distance between their activated feature maps, which is defined as follow:

$$\mathcal{L}_p^{\theta,i}(\boldsymbol{\mu}, \boldsymbol{v}) = \frac{1}{C_i D_i H_i W_i} \|\theta(\boldsymbol{\mu}) - \theta(\boldsymbol{v})\| \quad (4)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{v}$ are the ground truth 3D volume and the reconstructed 3D volume, respectively, i is the index of the activated feature maps, C, D, H, and W represent the shape of feature maps, and θ is the network generating the feature maps.

To produce anatomically more realistic details on the reconstructed 3D volume, the adversarial loss is introduced as in WGAN, which has shown to be effective [16]. We designed a simple discriminator network to calculate the Wasserstein distance between the ground truth 3D volume and the reconstructed 3D volume, as shown in Fig. 3(b). It consists of 3×3×3 convolutions, batch normalization, and Leaky ReLU activation function. The adversarial loss is defined as follow:

$$\mathcal{L}_{adv} = \mathrm{E}_{\boldsymbol{\mu} \sim \mathrm{P}_{\boldsymbol{\mu}}}[D(\boldsymbol{\mu})] - \mathrm{E}_{v \sim \mathrm{P}_v}[D(G(\boldsymbol{v}))], \quad (5)$$

where $\boldsymbol{v}$ is the reconstructed initial 3D volume, G is the proposed reconstruction model, D is the discriminator model.

The total loss function is a combination of the motion transformation prediction loss, perceptual loss, and adversarial loss:

$$\mathcal{L}=\mathcal{L}_T + \beta\mathcal{L}_p + \gamma\mathcal{L}_{adv} \quad (6)$$

where β and γ are used to balance the contributions of different losses terms.

3. EXPERIMENT AND RESULT

3.1 Datasets and Implementation Details

The proposed method is evaluated on two independent fetal brain MRI datasets, including 215 T2-weighted fetal brain MRI volumes between 21 and 38 weeks of gestational age. This retrospective study was approved by the institutional review board (Obstetrics and Gynecology Hospital, Fudan University, China). The fetal volumes were reconstructed using NiftyMIC [8], and then aligned onto a fetal brain atlas with 0.8 mm isotropic resolution [17]. The datasets are divided into training (160), validation (30), and test (25) sets. Similar to [18], based on the number of control points and the cubic spline interpolation, we simulated the motion trajectory of the fetus in three orthogonal views. We set the rotation angle between stacks to be -90° to 90° and the rotation angle between slices from -10° to 10°. The translation parameters are -8 mm to 8 mm between stacks and -4 mm to 4 mm between slices. The slice thickness is 2 to 4 mm, slices size is 140 × 140, the gap between slices is 3 to 5. Based on the slice acquisition model, in training, the stacks are randomly generated, while for testing, 6 stacks are randomly generated for each test subject. The intensity of stacks was normalized between -1 to 1.

The models were implemented with PyTorch and trained on NVIDIA TITAN Xp GPU using Adam optimizer (learning rate=0.0001). We first conducted pre-training for the motion correction model, and the number of iterations was 25,000. Then, we conducted joint training for the motion correction model and the reconstruction model, and the number of iterations was 1,000. Data augmentation methods are also performed, including the gamma transformation, contrast jitter, and Gaussian noise.

We adopted NiftyMIC [8] and SVRTK [6] as baselines for comparison. The standard metrics: Structural Similarity (SSIM) and Peak Signal-to-Noise Ratio (PSNR), are used to evaluate the performance of these methods. The run time is used to evaluate the efficiency of these methods

3.2. Results

Table 1 shows the mean and standard deviations of quantitative results obtained by different methods on the simulated test fetal brain MRI 2D slices in multi-stacks, the best results are marked in bold. The experiments demonstrate that the proposed method is superior to comparison methods. Specifically, in terms of PSNR, our method improves by 7.14 and 2.96 compared to SVRTK and NiftyMIC, respectively. According to SSIM, our method improves by 0.194 and 0.237, respectively. On average, the proposed end-to-end learning-based fetal MRI 3D volume method takes only 5.37 seconds to reconstructs a high-quality 3D volume. While NiftyMIC and SVRTK require 901.8 seconds and 117.2 seconds, respectively. Compared with these methods, the proposed method obtained the speedup of 167.93 and 21.82 times, respectively.

Method	PSNR	SSIM	Run time(s)
Ours	**22.767** (1.99)	**0.866** (0.056)	**5.37** (0.001)
NiftyMIC	19.806 (1.83)	0.672 (0.114)	901.8 (124.3)
SVRTK	15.624 (2.23)	0.629 (0.124)	117.2 (34.2)

Table 1. The results achieved by different method.

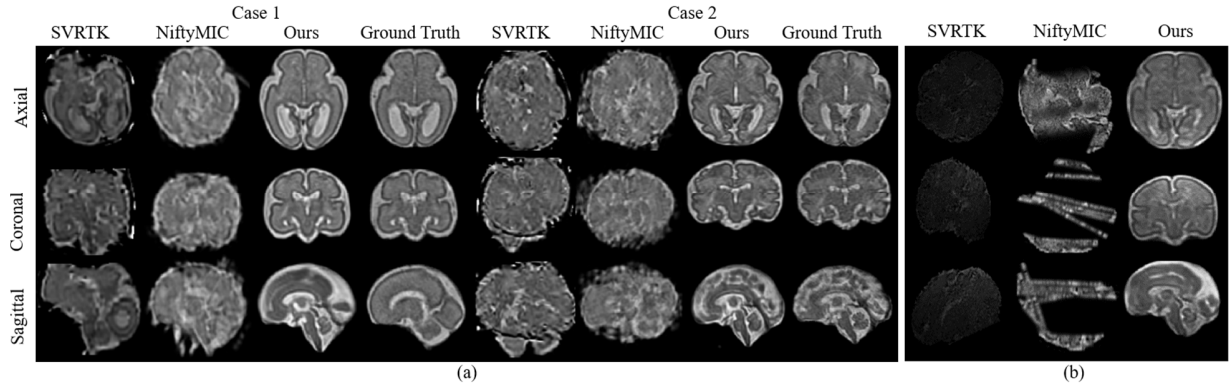

Fig. 4. (a) Two reconstructed 3D volumes and ground truth examples using different methods. (b) Reconstructed volumes obtained by different method on three real stacks.

Fig. 4(a) shows the reconstruction results obtained by different methods. In Case 1 and Case 2, the gestational weeks are 26 and 32, respectively. The results show that our method can achieve better results compared with state-of-the-art methods. We also used real fetal data to demonstrate that the proposed model. The slice thickness is 2 mm and the in-

plane resolution 1×1 mm^2. For preprocessing, the intensity bias field was corrected and the fetal brain was segmented using NiftyMIC [8]. Fig. 4(b) shows that our proposed method obtains better results than comparison methods.

4. CONCLUSION

In this work, we propose an automated end-to-end fetal brain MRI 3D volume reconstruction method using deep learning. Firstly, a multi-scale feature fusion model is proposed to solve arbitrary motion correction of 2D slices. Secondly, an initial 3D volume is estimated by point spread function, and an outlier removal strategy is used to reject failed slices. Next, we design a residual-based encoder-decoder model, which can learn an enriched set of features, to improve the quality of the initial 3D volume and reconstruct a high-resolution fetal brain MRI 3D volume. We performed experiments on both simulated and real fetal MR images. The results demonstrate the proposed method is able to reconstruct high-resolution 3D volumes with high efficiency and accuracy compared to state-of-the-art methods.

5. ACKNOWLEDGMENTS

This work was supported in part by NIH grants (MH116225, MH117943 and MH123202 to G.L.).

6. REFERENCES

[1] S. B. Mulkey, D. I. Bulas, G. Vezina, et al., “Sequential neuroimaging of the fetus and newborn with in utero Zika virus exposure,” *JAMA pediatrics*, vol. 173, no. 1, pp. 52–59, 2019.

[2] A. Gholipour, J. A. Estroff, and S. K. Warfield, “Robust super-resolution volume reconstruction from slice acquisitions: Application to fetal brain MRI,” *IEEE Trans Med Imaging*, vol. 29, no. 10, pp. 1739–1758, 2010.

[3] S. N. Saleem, “Fetal MRI: an approach to practice: a review,” *Journal of advanced research*, vol. 5, no. 5, pp. 507–523, 2014.

[4] F. Rousseau, O. A. Glenn, B. Iordanova, C. Rodriguez-Carranza, et al., “Registration-based approach for reconstruction of high-resolution in utero fetal MR brain images,” *Academic radiology*, vol. 13, no. 9, pp. 1072–1081, 2006.

[5] Ali, Gholipour, A. Judy, K. Simon, et al., “Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain MRI,” *IEEE Trans Med Imaging*, vol. 29, no. 10, pp. 1739-1758, 2012.

[6] M. Kuklisova-Murgasova, G. Quaghebeur, M.A. Rutherford, et al., “Reconstruction of fetal brain MRI with intensity matching and complete outlier removal,” *Med Image Anal*, vol. 16, no. 8, pp. 1550–1564, 2012.

[7] K. Kim, P. A. Habas, F. Rousseau, et al., “Intersection based motion correction of multislice MRI for 3-D in utero fetal brain image formation,” *IEEE Trans Med Imaging*, vol. 29, no. 1, pp. 146–158, 2009.

[8] M. Ebner, G. Wang, W. Li, et al., “An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI,” *Neuroimage*, vol. 206, pp. 116324, 2020.

[9] B. Kainz, M. Steinberger, W. Wein, et al., “Fast volume reconstruction from motion corrupted stacks of 2D slices,” *IEEE Trans Med Imaging*, vol. 34, no. 9, pp. 1901-1913, 2015.

[10] B. Hou, B. Khanal, A. Alansary, S. McDonagh, et al., “3D reconstruction in canonical coordinate space from arbitrarily oriented 2D images,” *IEEE Trans Med Imaging*, vol. 37, no. 8, pp. 1737–1750, 2018.

[11] Y. Pei, F. Zhao, Z. Tao, et al., "PETS-Nets: Joint Pose Estimation and Tissue Segmentation of Fetal Brains Using Anatomy-guided Networks," *IEEE Trans Med Imaging*, 2023. In Press.

[12] J. Xu, D. Moyer, B. Gagoski, J. E. Iglesias, et al., “NeSVoR: Implicit Neural Representation for slice-to-volume reconstruction in MRI,” *IEEE Trans Med Imaging*, vol. 42, no. 6, pp. 1707-1719, 2023.

[13] W. Shi, H. Xu, C. Sun, et al., “AFFIRM: affinity fusion-based framework for iteratively random motion correction of multi-slice fetal brain MRI,” *IEEE Trans Med Imaging*, vol. 42, no. 1, pp. 209-219, 2022.

[14] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” In CVPR, pp. 770-778, 2016.

[15] J. Johnson, A. Alahi, L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” In *ECCV*, pp. 694-711, 2016.

[16] A. Arjovsky, S. Chintala, “Wasserstein generative adversarial networks,” In *ICML*, pp. 214-223, 2017.

[17] A. Gholipour, C. K. Rollins, C. Velasco-Annis, et al., “A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth,” *Sci. Rep.*, vol. 7, no. 1, pp. 1–13, 2017.

[18] L. Ma, L. Chen, F. Zhao, et al., "Geometric constrained deep learning for motion correction of fetal brain MR images," In *ISBI*, pp. 1-5, 2023.