Joint Motion Estimation with Geometric Deformation Correction for Fetal Echo Planar Images Via Deep Learning

Author name(s) withheld EMAIL(S) WITHHELD
Address withheld

Editors: Under Review for MIDL 2024

Abstract

In this paper, we introduce a novel end-to-end predictive model for efficient fetal motion correction using deep neural networks. Diverging from conventional methods that estimate fetal brain motions and geometric distortions separately, our approach introduces a newly developed joint learning framework that not only reliably estimates various degrees of rigid movements, but also effectively corrects local geometric distortions of fetal brain images. Specifically, we first develop a method to learn rigid motion through a closed-form update integrated into network training. Subsequently, we incorporate a diffeomorphic deformation estimation model to guide the motion correction network, particularly in regions where local distortions and deformations occur. To the best of our knowledge, our study is the first to simultaneously track fetal motion and address geometric deformations in fetal echo-planar images. We validated our model using real fetal functional magnetic resonance imaging data with simulated and real motions. Our method demonstrates significant practical value to measure, track, and correct fetal motion in fetal MRI. The code will be released online.

Keywords: Deep Learning, Fetal MRI, Rigid Motion Correction, Geometric Deformation.

1. Introduction

Motion estimation is a critical procedure, meticulously designed to correct image artifacts induced by object motion, especially in the realm of medical image analysis for fetal imaging. Its critical role spans across various fields, including image segmentation (Ebner et al., 2020; Faghihpirayesh et al., 2022), reconstruction (Gholipour et al., 2010; Kuklisova-Murgasova et al., 2012; Xu et al., 2023), and pose estimation (Salehi et al., 2018; Golland and Adalsteinsson, 2020), among others. Effective motion compensation and correction techniques significantly contribute to the overall accuracy and efficiency of fetal imaging studies (Malamateniou et al., 2013). Estimating fetal motion presents inherent challenges, notably the unpredictable nature of fetal head movements and local geometric distortions due to accumulated imaging errors. To overcome these challenges, existing motion correction methods mostly fall into two distinct categories: i) iterative optimization methods based on mathematical models, and ii) predictive models learned by deep neural networks.

One of the initial paradigms of 3D fetal brain reconstruction using motion correction was a three-step model (Rousseau et al., 2006) involving multi-resolution slice alignment for motion correction, intensity correction for inhomogeneity, and super-resolution reconstruction through scattered data interpolation. To address motion errors from 2D slice misalignments, a slice-to-volume registration (SVR) model was introduced (Jiang et al., 2007), incorporating a scattered data interpolation method with a novel multi-level B-spline kernel. A motion correction technique was further presented to align image stacks based

© 2024 CC-BY 4.0, Anonymous.

on 2D slice intersections to facilitate the reconstruction of a high isotropic resolution 3D volume (Kim et al., 2009). A generative (forward) model of slice acquisition and its inverse problem solution were first proposed in (Gholipour et al., 2010). This approach featured a robust M-estimation solution to reduce the effects of corrupted (outlier) slices in a super-resolution reconstruction framework. Building upon these developments, an SVR approach was presented to encompass complete outlier removal through robust statistics based on the expectation-maximization (EM) algorithm (Kuklisova-Murgasova et al., 2012). Tourbier et al. (Tourbier et al., 2015) later introduced total variation regularization, efficiently solved using the primal-dual hybrid gradient method. While these advancements have enhanced the accuracy and efficiency of motion correction in image reconstruction, all gradient-based optimization methods for slice realignments have a very limited capture range, often failing in cases of large and rapid fetal movements.

Motivated by the progress in deep learning, a spectrum of models has emerged, with a particular focus on advancing predictive motion estimation (Mahendran et al., 2017). Hou et al. developed a method for predicting 3D rigid transformations of arbitrarily oriented 2D slices (Hou et al., 2018). This pursuit expanded with the introduction of a real-time fetal motion tracking system, leveraging the strengths of diverse neural networks, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to predict motion directly from input images (Salehi et al., 2018; Singh et al., 2020; Evan et al., 2022).

To address non-rigid deformations in fetal MRI, two distinct approaches were presented. These include the patch-to-volume reconstruction technique (Alansary et al., 2017) and the deformable SVR technique (Uus et al., 2020), both designed to correct non-rigid motion to reconstruct deformable anatomy such as fetal body parts. In the domain of automated fetal MRI reconstruction, Ebner et al. developed a toolkit that featured slice-level outlier rejection, fetal brain localization, and volume reconstruction (Ebner et al., 2020).

Noteworthy progress has been achieved in the domain of sensorless imaging, particularly in 3D volume reconstruction from 2D freehand ultrasound images, leveraging deep implicit representations (Yeung et al., 2021). Xu et al. incorporated transformers (Vaswani et al., 2017) trained on synthetically transformed data, streamlining automatic relevance detection between slices (Xu et al., 2022). Advancing this approach, Xu et al. further introduced a reconstruction technique that incorporates implicit neural representations, enhancing image reconstruction performance (Xu et al., 2023). Evan et al. introduced KeyMorph, an unsupervised deep learning framework designed for robust and interpretable multi-modal medical image registration (Evan et al., 2022). Utilizing anatomically-consistent keypoints, KeyMorph aims to enhance the registration process and increase alignment accuracy. Moyer et al. introduced a real-time method for tracking rigid motion by employing equivariant neural networks (Moyer et al., 2021). The effectiveness of this approach in capturing significant rigid motions is derived from the intrinsic rotation-equivariant nature of equivariant filters (Cohen and Welling, 2016).

Despite the notable performance achieved by the aforementioned methods, there is currently a lack of an end-to-end predictive motion correction framework that can adequately address both uncontrollable fetal motion and geometric distortions. This gap in the existing literature motivated our endeavor to develop a comprehensive framework proficient in efficiently managing fetal motions and geometric distortions through deep learning. The core contributions of our proposed motion correction method are summarized in three folds:

- **Pioneering contribution**: Our study introduces the first comprehensive end-to-end predictive motion tracking framework adept at handling both significant rigid motion and geometric distortions in fetal echo planar imaging (EPI).
- **Significance**: Our proposed framework not only achieves motion correction accuracy comparable to current state-of-the-art methods but also ensures faster and more stable convergence. This attribute is of significant value, especially in real-time, automated fetal head steering systems, where frequent and substantial motions occur.
- **Theoretical advancement for broad applicability**: The theoretical tool developed through our joint learning algorithm demonstrates wide applicability, offering advantages in various fetal imaging applications. This includes real-time fetal brain segmentation, facilitating population-based studies of fetal brain development using mean template estimation, and improving EPI image reconstruction with advanced distortion correction techniques.

2. Methodology

In this section, we first outline the theoretical basis needed to address rigid motions and deformations in fetal imaging. We start by reviewing the principles of rigid motion estimation. Next, we focus on correcting geometric distortions, specifically using Large Deformation Diffeomorphic Metric Mapping (LDDMM), chosen for its effectiveness in creating smooth, invertible mappings that maximally maintain the correctness of topological information (Beg et al., 2005). Finally, we describe the design of our proposed network architecture and the approach to its training.

2.1. Rigid Motion Estimation

Rigid motion estimation aims to identify the best translation $\mathcal{T}$ and rotation $\mathcal{R}$ parameters that define a rigid transformation $Q(\mathcal{T}, \mathcal{R})$ between a pairwise images. This process involves minimizing the Euclidean distance d between the source image S and the target image T,

$$Q^*(\mathcal{T}, \mathcal{R}) = \arg\min_{Q} \; d[S \circ Q(\mathcal{T}, \mathcal{R}), T], \tag{1}$$

where $\circ$ is the composition operator that resamples S using the rigid transformation $Q(\mathcal{T}, \mathcal{R})$. When this operator is applied to any vector $\mathbf{v}$, it yields a transformed vector $Q(\mathbf{v})$ of the form $Q(\mathbf{v}) = \mathcal{R}\mathbf{v} + \mathcal{T}$. Here, $\mathcal{R}^T = \mathcal{R}^{-1}$ indicating that $\mathcal{R}$ is an orthogonal matrix.

The transformation function Q can be calculated using low-dimensional representations, such as point clouds or key points of images (Moyer et al., 2021; Evan et al., 2022). This process is quantified by the energy function $\mathrm{E}(Q)$, as shown in the equation,

$$\mathrm{E}(Q) = \|S \circ Q(\mathcal{T}, \mathcal{R}, \bar{S}, \bar{T}) - T\|_F, \tag{2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, and $\bar{S}$, $\bar{T}$ represent the low-dimensional representations of the source and target respectively. We derive the closed-form solution for both translation and rotation parameters,

$$\mathcal{T} = \bar{T} - \mathcal{R}\bar{S}, \quad \mathcal{R} = V \cdot U^T, \quad \text{s.t. } \det(\mathcal{R}) = 1, \tag{3}$$

where $U\Sigma V^* = \bar{S} \cdot \bar{T}^T$, U and V^* are real orthogonal matrices, Σ is a diagonal matrix with non-negative real numbers on the diagonal. Setting the determinant of $\mathcal{R}$ to 1 guarantees that it accurately reflects a rigid transformation.

2.2. Geometric Deformation Correction via LDDMM

In this section, we provide an overview of the LDDMM algorithm for image registration (Beg et al., 2005). This algorithm is employed to address deformable geometric distortions between the rigid motion-corrected image $S \circ Q(\mathcal{T}, \mathcal{R}, \bar{S}, \bar{T})$ and the target image T. For simplicity, we denote $\hat{\mathbf{S}}$ as $S \circ Q(\mathcal{T}, \mathcal{R}, \bar{S}, \bar{T})$.

Given both $\hat{\mathbf{S}}$ and T, defined on a d-dimensional torus domain $\Gamma = \mathbb{R}^d/\mathbb{Z}^d$ ($\hat{\mathbf{S}}(x), T(x) : \Gamma \to \mathbb{R}$), the objective of diffeomorphic image registration is to find the shortest path to generate time-varying diffeomorphisms $\{\psi_t(x)\} : t \in [0, 1]$ such that $\hat{\mathbf{S}} \circ \psi_1$ is similar to T. This is typically solved by minimizing an explicit energy function over the initial transformation fields v_0 (Vialard et al., 2012), which can be expressed as follows,

$$\mathrm{E}(v_0) = \frac{1}{\sigma^2} \|\hat{\mathbf{S}} \circ \psi_1(v_0) - T\|_2^2 + (Lv_0, v_0), \tag{4}$$

where σ^2 is noise variance in images and $L : V \to V^*$ is a symmetric, positive-definite differential operator that maps a tangent vector $v_t \in V$ into its dual space as a momentum vector $m_t \in V^*$. This is typically denoted as $m_t = Lv_t$ or $v_t = Km_t$, with K being an inverse operator of L. The notation $(\cdot, \cdot)$ denotes the pairing of a momentum vector with a tangent vector, which is similar to an inner product.

The geodesic shooting process states that the geodesic path $\{\psi_t\}$ can be uniquely determined by integrating a given initial velocity v_0 forward in time by using the Euler-Poincaré differential (EPDiff) equation (Arnol'd, 1966; Miller et al., 2006) as:

$$\frac{\partial v_t}{\partial t} = -K\left[(Dv_t)^T \cdot m_t + Dm_t \cdot v_t + m_t \cdot \operatorname{div} v_t\right], \quad \frac{d\psi_t}{dt} = -D\psi_t \cdot v_t, \tag{5}$$

where the operator D denotes a Jacobian matrix and $\cdot$ represents element-wise matrix multiplication. Here, div is the divergence.

2.3. Network Architecture and Training

We develop a deep learning model that explicitly estimates both rigid motion and geometric distortions by integrating a combination of Eq. (2) and Eq. (4) into our objective function. We show that such joint estimation results in improved accuracy and robustness of the model. Our framework consists of two modules: an enhanced rigid motion correction neural network that corrects rigid motions with closed-form update, and an unsupervised learning of geometric distortion correction through an image registration network. An illustration of our joint learning framework is shown in Fig. 1. Below, we provide a detailed description of our network architecture and objective function.

Enhanced motion correction network. Let $\Theta = (\mathcal{T}, \mathcal{R}, \bar{S}, \bar{T})$ represent the encoder parameters that learn rigid parameters and key features from image spaces, with $Q(\Theta)$ denoting the transformation function yielded by the learned rigid parameters from the key

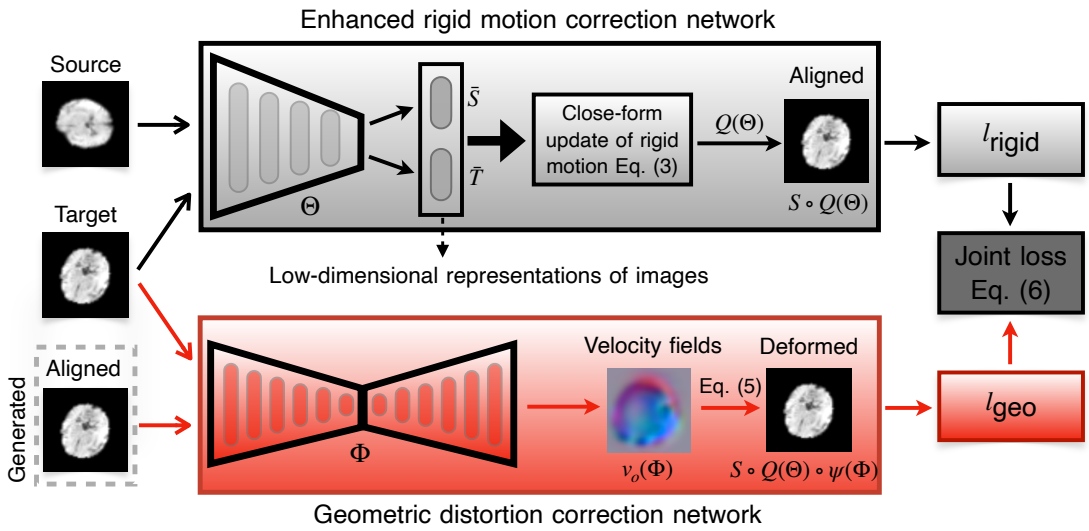

Figure 1: An illustration of our proposed **joint correction network (JCN)** architecture. The top module performs rigid motion correction on pairs of images using closed-form updates derived from their key point representations. The bottom module corrects geometric distortions in the aligned and target images, employing a geometric loss function with a regularization term in diffeomorphic transformation space. Both models are interrelated—the geometrically corrected data is fed back as augmented data to enhance the accuracy of rigid motion correction. This, in turn, aligns better with the geometric distortion network, facilitating the correction of local distortions in corresponding positions of the fetal brain.

points of images. The rigid correction loss is computed between the aligned and target images. Θ is characterized using one of two backbone structures, (i) a 7-layer equivariant neural network to effectively capture rotation-equivaraint representations using 3D steerable CNNs (Weiler et al., 2018; Moyer et al., 2021), and (ii) a conventional 9-layer CNN as employed in KeyMorph (Evan et al., 2022).

Geometric distortion correction based on image registration network. Let Φ denote the parameters of an encoder-decoder in our geometric learning network, where $\psi(\Phi)$ represents the deformation fields and $v_0(\Phi)$ represents the velocity fields learned by the network. Except for estimating deformations by LDDMM (Hinkle, 2018), we also provide another deep learning model that learns stationary velocity fields (Balakrishnan et al., 2019) meanwhile maintains a comparable model accuracy. Advanced predictive image registration models, including TransMorph (Chen et al., 2022) and DiffuseMorph (Kim et al., 2022), can easily be integrated into our framework.

Joint correction network and objective function. Our joint correction network (JCN) comes in two versions: JCN: EF, which employs equivariant filters, and JCN: KM, based on the CNN architecture that is similar to KeyMorph. Our total objective function,

which serves as the network's training loss, combines both the motion correction component (Eq. (2)) and the geometric deformation estimation (Eq. (4)), formulated as follows,

$$l(\Theta,\Phi) = \underbrace{\lambda\|S\circ Q(\Theta)-T\|_F}_{l_{\text{rigid}}} + \underbrace{\frac{1}{\sigma^2}\|S\circ Q(\Theta)\circ\psi(\Phi)-T\|_2^2 + (Lv_0(\Phi),v_0(\Phi))}_{l_{\text{geo}}}$$
$$+\,\text{reg}(\Theta,\Phi), \qquad \text{s.t. Eq. (3), \& Eq. (5).} \tag{6}$$

Here, reg$(\cdot)$ is a regularization term constrained on network parameters, and λ is a weighting factor balancing the contributions of both networks. Other image dissimilarity metrics, such as normalized cross correlation (Avants et al., 2008) and mutual information (Wells III et al., 1996; Wang et al., 2023) can be adopted in our developed objective function.

3. Experimental Setup

Data. Our study includes $1,881$ pairs of 3D EPIs from fMRI time series of 15 subjects undergone fetal MRI scans using a Siemens 3T scanner (Skyra or Prisma) between August 2015 to September 2021. The study was approved by the institutional review board, and written informed consent was obtained from all research MRI participants. The dataset covers a gestational age range from 22.57 to 38.14 weeks (mean 32.39 weeks). Imaging parameters included a slice thickness of 2 to $3mm$, repetition time (TR) of 2 to 5.6 seconds (mean 3.1 seconds), echo time (TE) of 0.03 to 0.08 seconds (mean 0.04 seconds), and a flip angle (FA) of 90 degrees. The fetal brains were extracted from the scans prior to resampling using a real-time deep learning segmentation method (Faghihpirayesh et al., 2023). The extracted brain scans were resampled to an image size of 96^3 with a voxel resolution of $3mm^3$ and underwent intensity normalization.

Baselines & Evaluation Metrics. We compared various motion correction methods including conventional iterative rigid registration (in ITK) to estimate rigid transformation as 3D versors (Ibanez et al., 2003) and rigid motion estimation by deep learning such as DeepPose (Salehi et al., 2018) and KeyMorph (Evan et al., 2022), which only estimate rigid motions. We compare their performance with our joint models JCN: KM and JCN:EF. To show the benefit of our joint learning, we report the best Dice score of disjoint approaches, where the rigid motion estimation is treated as a preprocessing step before geometric distortion estimation. We evaluated performance through visual comparison and by computing translational and angular errors on simulated motions. We also manually added translations (in mms) and rotations (in degrees) to real fMRI scans containing natural fetal motions by changing their maximum motion values in three directions. To better demonstrate our model's stability, we show the Dice coefficient comparison by computing the overlapping regions between aligned image and the target across three levels of motions.

Hyperparameters. We set the dimension for low-dimensional key points to 128 when computing the close-form update of rigid transformation in Eq. (3) and set $\alpha = 3$ for the operator L with using 5 as time steps of Euler integration for geodesic shooting (Eq. (5)). The noise variance was fixed at $\sigma = 0.02$. For network training, we set batch size as 4, weight decay as 0.0001 for a L_2 regularization, and a learning rate of $\eta = 1e-5$ for 300 epochs with

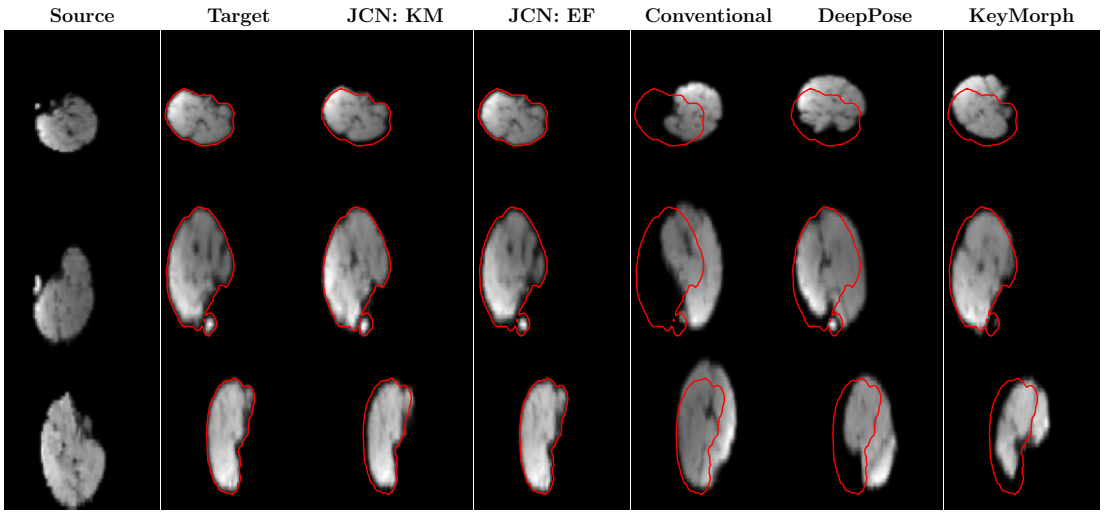

Figure 2: Fetal motion correction visualizations for all methods. Left to right: source, target (the contour is highlighted as red), motion-corrected images by our methods (JCN: KM and JCN: EF), the iterative conventional method, DeepPose, and KeyMorph. **Dice scores**, **0.95 for JCN: KM**, **0.97 for JCN: EF**, 0.32 for conventional, 0.63 for DeepPose, and 0.73 for KeyMorph. Brief Dice score comparison between the best model of disjoint ones and ours, 0.90/ **0.97**. Motion correction efficiency with average time consumption: **0.491s** per pair.

Adam optimizer. We divided the data into training (1283 pairs from 11 subjects), validation (299 pairs from 2 subjects), and testing (299 pairs from 2 subjects). The best-performing networks were saved based on validation performance across all models. All experiments were conducted with an NVIDIA RTX A6000 GPU.

4. Results

Fig. 2 presents a motion correction study for pairwise fMRI scans characterized by significant motions. It shows that the conventional and DeepPose methods fail to correctly align images, while our approach excels, closely matching the target image. Our method effectively handles both rigid motion and local geometric deformations, leveraging estimated velocity fields for more accurate correction.

Fig. 3 illustrates the comparative analysis of motion correction errors in translation and rotation across different methods. It highlights the enhanced performance of our proposed technique in correcting fetal motions, surpassing the state-of-the-art methods.

Fig. 4 statistically demonstrates the dice coefficient comparisons for fMRI time sequences under various motion conditions. It demonstrates that our method consistently achieves higher dice scores, regardless of the motion degree. This highlights the robustness and stability of our model, particularly in challenging clinical scenarios with significant motion occurrences.

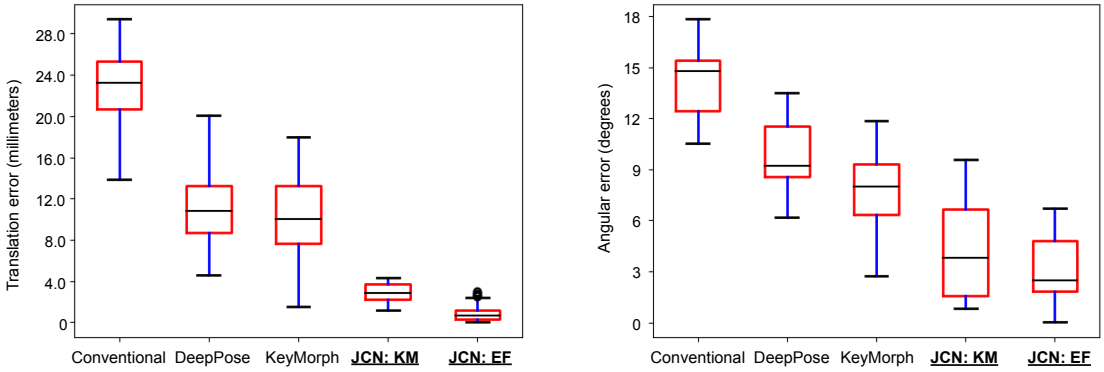

Figure 3: Motion correction comparison of translation and rotation errors (averaged for all directions) of 299 real fMRI scans with simulated motions.

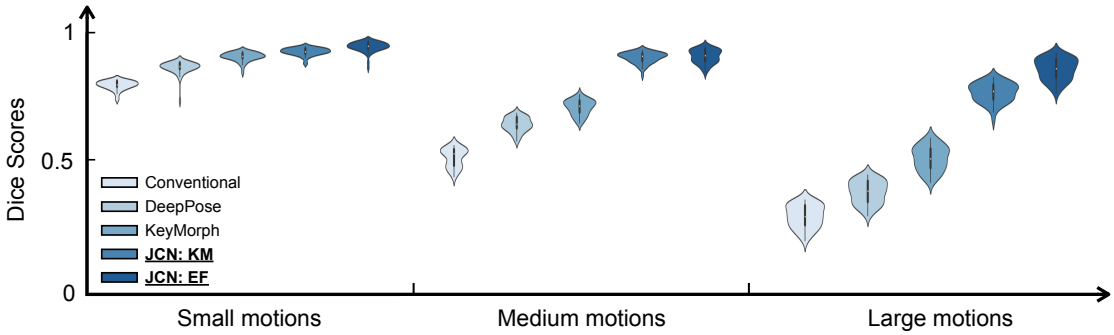

Figure 4: Motion correction performance by dice evaluation on real fMRI with different degrees of motions, small ($\mathcal{T}_{max}$ = 10mm, $\mathcal{R}_{max} = 5°$), medium ($\mathcal{T}_{max}$ = 20mm, $\mathcal{R}_{max} = 10°$) and large motions ($\mathcal{T}_{max}$ = 30mm, $\mathcal{R}_{max} = 20°$). The dice score of our best model, for motion levels from left to right are, 0.97, 0.93, 0.92.

5. Conclusion

This paper presents a pioneering predictive model for fetal motion correction using deep neural networks. In contrast to conventional approaches that independently estimate fetal brain motions and geometric distortions, our proposed method adopts an efficient and robust joint learning framework. This framework excels in achieving optimal and consistent performance across various degrees of fetal motions. The model demonstrates notable effectiveness on fMRI data with both simulated and real motions, showcasing significant potential in real-time tracking and steering systems for fetal head motion. An intriguing avenue for future exploration involves applying our developed model to image reconstruction for EPI data, which poses more challenging fetal motion artifacts.

Acknowledgments

Acknowledgments withheld.

References

Amir Alansary, Martin Rajchl, Steven G McDonagh, Maria Murgasova, Mellisa Damodaram, David FA Lloyd, Alice Davidson, Mary Rutherford, Joseph V Hajnal, Daniel Rueckert, et al. Pvr: patch-to-volume reconstruction for large area motion correction of fetal mri. *IEEE transactions on medical imaging*, 36(10):2031–2044, 2017.

V. I. Arnol'd. Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. *Ann. Inst. Fourier*, 16:319–361, 1966.

Brian B Avants, Charles L Epstein, Murray Grossman, and James C Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, 2008.

Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019.

MIRZA Faisal Beg, Michael I Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(2):139–157, 2005.

Junyu Chen, Eric C Frey, Yufan He, William P Segars, Ye Li, and Yong Du. Transmorph: Transformer for unsupervised medical image registration. *Medical image analysis*, 82: 102615, 2022.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016.

Michael Ebner, Guotai Wang, Wenqi Li, Michael Aertsen, Premal A Patel, Rosalind Aughwane, Andrew Melbourne, Tom Doel, Steven Dymarkowski, Paolo De Coppi, et al. An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain mri. *NeuroImage*, 206:116324, 2020.

M Yu Evan, Alan Q Wang, Adrian V Dalca, and Mert R Sabuncu. Keymorph: Robust multi-modal affine registration via unsupervised keypoint detection. In *International Conference on Medical Imaging with Deep Learning*, pages 1482–1503. PMLR, 2022.

Razieh Faghihpirayesh, Davood Karimi, Deniz Erdoğmuş, and Ali Gholipour. Deep learning framework for real-time fetal brain segmentation in mri. In *International Workshop on Preterm, Perinatal and Paediatric Image Analysis*, pages 60–70. Springer, 2022.

Razieh Faghihpirayesh, Davood Karimi, Deniz Erdoğmuş, and Ali Gholipour. Fetal-bet: Brain extraction tool for fetal mri. *arXiv preprint arXiv:2310.01523*, 2023.

Ali Gholipour, Judy A Estroff, and Simon K Warfield. Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain mri. *IEEE transactions on medical imaging*, 29(10):1739–1758, 2010.

Polina Golland and Elfar Adalsteinsson. 3d fetal pose estimation with adaptive variance and conditional generative adversarial network. In *Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis: First International Workshop, ASMUS 2020, and 5th International Workshop, PIPPI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4-8, 2020, Proceedings*, volume 12437, page 201. Springer Nature, 2020.

Jacob D Hinkle. Lagomorph. Technical report, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), 2018.

Benjamin Hou, Bishesh Khanal, Amir Alansary, Steven McDonagh, Alice Davidson, Mary Rutherford, Jo V Hajnal, Daniel Rueckert, Ben Glocker, and Bernhard Kainz. 3-d reconstruction in canonical co-ordinate space from arbitrarily oriented 2-d images. *IEEE transactions on medical imaging*, 37(8):1737–1750, 2018.

Luis Ibanez, Will Schroeder, Lydia Ng, Josh Cates, et al. The itk software guide, 2003.

Shuzhou Jiang, Hui Xue, Alan Glover, Mary Rutherford, Daniel Rueckert, and Joseph V Hajnal. Mri of moving subjects using multislice snapshot images with volume reconstruction (svr): application to fetal, neonatal, and adult brain studies. *IEEE transactions on medical imaging*, 26(7):967–980, 2007.

Boah Kim, Inhwa Han, and Jong Chul Ye. Diffusemorph: Unsupervised deformable image registration using diffusion model. In *European conference on computer vision*, pages 347–364. Springer, 2022.

Kio Kim, Piotr A Habas, Francois Rousseau, Orit A Glenn, Anthony J Barkovich, and Colin Studholme. Intersection based motion correction of multislice mri for 3-d in utero fetal brain image formation. *IEEE transactions on medical imaging*, 29(1):146–158, 2009.

Maria Kuklisova-Murgasova, Gerardine Quaghebeur, Mary A Rutherford, Joseph V Hajnal, and Julia A Schnabel. Reconstruction of fetal brain mri with intensity matching and complete outlier removal. *Medical image analysis*, 16(8):1550–1564, 2012.

Siddharth Mahendran, Haider Ali, and René Vidal. 3d pose regression using convolutional neural networks. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 2174–2182, 2017.

C Malamateniou, SJ Malik, SJ Counsell, JM Allsop, AK McGuinness, T Hayat, Kathryn Broadhouse, RG Nunes, AM Ederies, JV Hajnal, et al. Motion-compensation techniques in neonatal and fetal mr imaging. *American Journal of Neuroradiology*, 34(6):1124–1136, 2013.

Michael I Miller, Alain Trouvé, and Laurent Younes. Geodesic shooting for computational anatomy. *Journal of mathematical imaging and vision*, 24(2):209–228, 2006.

Daniel Moyer, Esra Abaci Turk, P Ellen Grant, William M Wells, and Polina Golland. Equivariant filters for efficient tracking in 3d imaging. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24*, pages 193–202. Springer, 2021.

Francois Rousseau, Orit A Glenn, Bistra Iordanova, Claudia Rodriguez-Carranza, Daniel B Vigneron, James A Barkovich, and Colin Studholme. Registration-based approach for reconstruction of high-resolution in utero fetal mr brain images. *Academic radiology*, 13 (9):1072–1081, 2006.

Seyed Sadegh Mohseni Salehi, Shadab Khan, Deniz Erdogmus, and Ali Gholipour. Real-time deep pose estimation with geodesic loss for image-to-template rigid registration. *IEEE transactions on medical imaging*, 38(2):470–481, 2018.

Ayush Singh, Seyed Sadegh Mohseni Salehi, and Ali Gholipour. Deep predictive motion tracking in magnetic resonance imaging: application to fetal imaging. *IEEE transactions on medical imaging*, 39(11):3523–3534, 2020.

Sébastien Tourbier, Xavier Bresson, Patric Hagmann, Jean-Philippe Thiran, Reto Meuli, and Meritxell Bach Cuadra. An efficient total variation algorithm for super-resolution in fetal brain mri with adaptive regularization. *NeuroImage*, 118:584–597, 2015.

Alena Uus, Tong Zhang, Laurence H Jackson, Thomas A Roberts, Mary A Rutherford, Joseph V Hajnal, and Maria Deprez. Deformable slice-to-volume registration for motion correction of fetal body and placenta mri. *IEEE transactions on medical imaging*, 39(9): 2750–2759, 2020.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

François-Xavier Vialard, Laurent Risser, Daniel Rueckert, and Colin J Cotter. Diffeomorphic 3d image registration via geodesic shooting using an efficient adjoint calculation. *International Journal of Computer Vision*, 97(2):229–241, 2012.

Jian Wang, Jiarui Xing, Jason Druzgal, William M Wells III, and Miaomiao Zhang. Metamorph: Learning metamorphic image transformation with appearance changes. In *International Conference on Information Processing in Medical Imaging*, pages 576–587. Springer, 2023.

Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco S Cohen. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. *Advances in Neural Information Processing Systems*, 31, 2018.

William M Wells III, Paul Viola, Hideki Atsumi, Shin Nakajima, and Ron Kikinis. Multimodal volume registration by maximization of mutual information. *Medical image analysis*, 1(1):35–51, 1996.

Junshen Xu, Daniel Moyer, P Ellen Grant, Polina Golland, Juan Eugenio Iglesias, and Elfar Adalsteinsson. Svort: iterative transformer for slice-to-volume registration in fetal brain mri. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 3–13. Springer, 2022.

Junshen Xu, Daniel Moyer, Borjan Gagoski, Juan Eugenio Iglesias, P Ellen Grant, Polina Golland, and Elfar Adalsteinsson. Nesvor: Implicit neural representation for slice-to-volume reconstruction in mri. *IEEE Transactions on Medical Imaging*, 2023.

Pak-Hei Yeung, Linde Hesse, Moska Aliasi, Monique Haak, Weidi Xie, Ana IL Namburete, et al. Implicitvol: sensorless 3d ultrasound reconstruction with deep implicit representation. *arXiv preprint arXiv:2109.12108*, 2021.