SVRMamba: Slice-to-Volume Reconstruction from Multiple MRI Stacks with Slice Sequence Guided Mamba

Jiangjie Wu[1], **Hongjiang Wei**[2], **Yuyao Zhang**[1*]

[1]School of Information Science and Technology, ShanghaiTech University, Shanghai, China
[2]School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
wujj@shanghaitech.edu.cn, hongjiang.wei@sjtu.edu.cn, zhangyy8@shanghaitech.edu.cn

Abstract

In fetal magnetic resonance imaging (MRI), slice-to-volume reconstruction (SVR) involves the computational creation of a 3D volume from multiple stacks of 2D slices. This process is challenging due to slice misalignment and image noise. Current state-of-the-art (SOTA) SVR methods typically employ coarse-to-fine techniques that iteratively refine slice-to-volume motion correction and 3D volume reconstruction. However, both processes are inherently inefficient, making these methods time-consuming and prone to errors. This often results in less robust and accurate outcomes, primarily due to insufficient modeling of the spatial relationships between slices. Typically, 2D fetal MRI slices are acquired using the interleave sequence, which first acquires the odd slices and then the even slices in one stack. To this end, we propose a novel Mamba-based framework called SVRMamba, which integrates slice-to-volume reconstruction with slice sequence-guided state space modeling. Specifically, our approach reformulates Mamba's unidirectional scanning into a slice sequence-guided odd-even directional scanning method and marks the slice positions using sequence embedding tokens. This enables the network to learn the slice relationships and spatial sequences, enhancing fetal MRI SVR motion correction performance. We further integrate a convolutional neural network (CNN)-based interpolation network that generates a noise-suppressed 3D reconstruction by leveraging the predicted motion for each slice. This framework notably enhances 3D fetal brain SVR, delivering substantial improvements in both reconstruction speed and overall performance. Extensive experiments conducted on various benchmark and clinical datasets demonstrate that SVRMamba significantly outperforms existing SOTA methods, delivering comparable results with a remarkable sixtyfold increase in reconstruction speed.

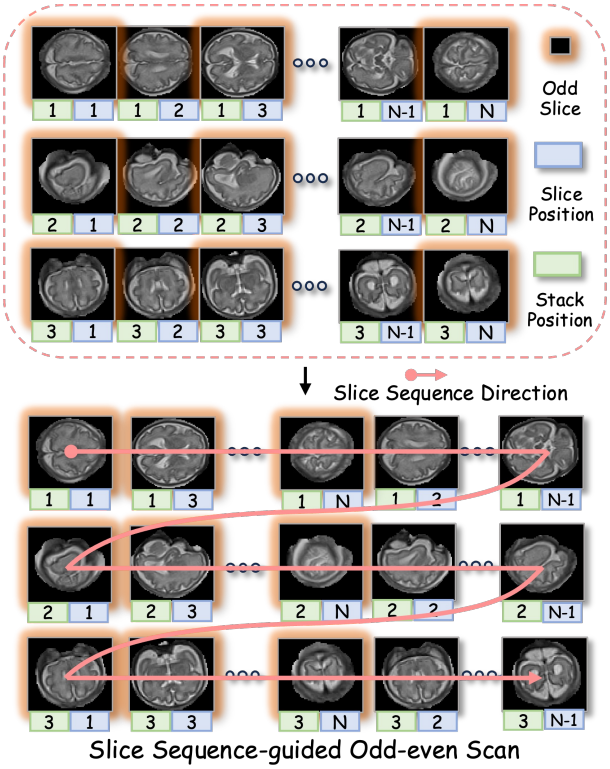

Figure 1: Slice sequence-guided odd-even scan. The pink dotted box highlights the original input slices and its corresponding slice-stack position.

Introduction

Non-invasive imaging of the brain, particularly in cases where the subject may exhibit severe uncontrollable motion, such as in fetal populations, often relies on two-dimensional (2D) magnetic resonance imaging (MRI) (Coakley et al. 2004; Davidson et al. 2021; Aviles Verdera et al. 2023). This technique involves acquiring multiple stacks of 2D brain slices and assembling them into a 3D volume through a process known as slice-to-volume reconstruction (SVR). Unlike 3D MRI acquisitions, where a single motion event can corrupt the entire k-space (Fourier) data, 2D techniques confine this corruption to individual slices, preserving the integrity of the k-space data for the remaining slices. SVR can then be applied to the non-corrupted slices, enabling artifact-free brain imaging and supporting various downstream tasks, such as brain morphometry (Coakley et al. 2004; Payette et al. 2023; Namburete et al. 2023), brain cortex development study (Zöllei et al. 2020; Wu et al. 2021a; Chen et al. 2023), and atlas creation(Evans et al. 2012; Gholipour et al.

*Corresponding author.

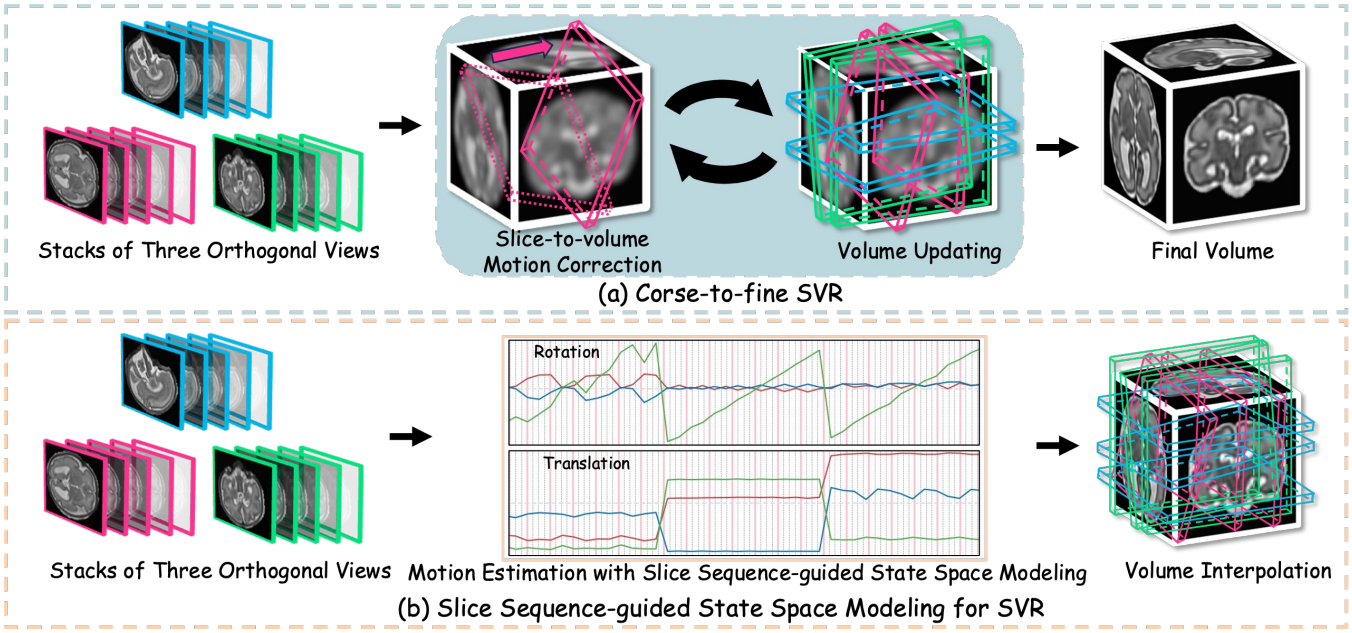

Figure 2: **Motivation**. Visual comparisons of different SVR. (a) Corse-to-fine SVR combines slice-to-volume motion correction and volume updating to solve motion parameters leading to time-consuming and limited motion correction due to inadequate modeling of spatial relationships between slices. (b) The proposed SVR using slice sequence-guided state space modeling achieves effective spatial relationship modeling and fast volume construction.

2017; Wu et al. 2021b).

Coarse-to-fine SVR methods (Kuklisova-Murgasova et al. 2012; Ebner et al. 2020; Shi et al. 2022; Wu et al. 2023; Xu et al. 2023; Uus et al. 2024) have achieved state-of-the-art (SOTA) performance by combining slice-to-volume motion correction and volume updating to solve motion parameters, as shown in Figure 2 (a). For instance, classical optimization-based SVR approaches (Alansary et al. 2017; Ebner et al. 2020; Gholipour, Estroff, and Warfield 2010; Kuklisova-Murgasova et al. 2012; Uus et al. 2024) iteratively perform slice-to-volume motion correction and volume updating to address motion parameters and reconstruct the underlying volume. Some methods (Hou et al. 2018; Shi et al. 2022; Xu et al. 2022) incorporate pre-trained convolutional neural networks to provide initial motion estimates, further enhancing reconstruction performance. Recently, learning-based SVR methods, such as NeSVoR and ASSURED (Xu et al. 2023; Wu et al. 2023), have utilized joint updates of slice motion parameters and volumes to achieve superior results. Despite these advancements, existing models are time-consuming and often fail to reconstruct robust 3D volumes in challenging clinical scenarios with severe motion corruption and noise. This limitation arises partly from inadequate modeling of spatial relationships between slices. Furthermore, the coarse-to-fine approach separately corrects the motion of each slice, neglecting the context of overall slice relationships. This method relies on a 3D volume to share information across slices, which can introduce more errors if the 3D volume is corrupted.

To address these challenges, we propose SVRMamba, a novel model that employs slice sequence information and state space modeling for robust 3D volume reconstruction. Mamba (Gu and Dao 2023), a new state space modeling method with global receptive field capability, is typically designed for long-range data (Gu and Dao 2023; Zhang et al. 2024; Wang et al. 2024; Behrouz and Hashemi 2024; Li et al. 2024b; Lieber et al. 2024). Few studies have adapted Mamba for 3D vision tasks (Zhu et al. 2024; Liang et al. 2024; Gong et al. 2024; Liu et al. 2024a). In this work, we explore Mamba's potential for the 3D SVR task. While Mamba-based models excel in modeling long-range sequences (Gu and Dao 2023; Zhang et al. 2024; Wang et al. 2024; Behrouz and Hashemi 2024; Li et al. 2024b; Lieber et al. 2024), they struggle to capture local-relation information, such as the semantics between stacks and slices. We propose a Slice Sequence-guided odd-even directional sCan (SSC) to effectively capture the semantic and spatial relationships between slices, as shown in Figure 1. Slice positional embedding can effectively capture stack-slice relations, so we integrate Slice Sequence Embedding (SSE) into state space modeling, significantly enhancing the representation by considering intricate slice relations. In particular, to effectively reconstruct the volume and suppress noise, we train a fully convolutional neural network (CNN) to interpolate the volume constructed from the slices using the predicted motion parameters. Our contributions are as follows:

1. Instead of the coarse-to-fine SVR, we are the first to integrate state space modeling and an interpolation network to reconstruct robust 3D MRI volumes. Our key innovation is converting Mamba's unidirectional scanning into slice sequence-guided odd-even directional scanning for

motion correction, followed by an interpolation network to generate the 3D volume.

2. We propose a simple yet effective slice sequence-guided state space network to capture spatial relations between slices and stacks using a slice sequence embedding token that marks both slice and stack positions.
3. We introduce a fully convolutional interpolation network for 3D volume reconstruction from the multiple stacks that effectively suppress noise using CNN architectures.
4. Extensive experiments on multiple challenging benchmarks and clinical scenarios demonstrate SVRMamba's superiority over current SOTAs, achieving a sixty-fold acceleration in reconstruction time while maintaining comparable performance.

Related Works

Slice-to-volume Reconstruction

Various approaches have been developed for SVR (Rousseau et al. 2006; Gholipour, Estroff, and Warfield 2010; Kuklisova-Murgasova et al. 2012; Kainz et al. 2015; Alansary et al. 2017; Hou et al. 2018; Tourbier et al. 2015; Ebner et al. 2020; Shi et al. 2022; Wu et al. 2023; Xu et al. 2023; Uus et al. 2024). Many of these approaches use a coarse-to-fine strategy to estimate motion parameters. Early work by Rousseau et al. (Rousseau et al. 2006) involved predicting the 3D MRI volume by iteratively updating slice motion parameters and the volume itself from orthogonal slice stacks, aiming to refine the final volume. To extend the range of motion correction, supervised learning-based methods (Hou et al. 2018; Shi et al. 2022; Xu et al. 2022) predict slice positions to provide better initialization for optimization. For instance, Shi et al. (Shi et al. 2022) employed a CNN to predict motion parameters for NiftyMIC (Ebner et al. 2020). Recent advancements include unsupervised learning methods that update slice motion parameters and volume jointly, enhancing 3D MRI reconstruction through implicit representation approaches. Methods like NeSVoR and ASSURED (Xu et al. 2023; Wu et al. 2023) use multi-layer perceptrons (MLPs) and CNNs to improve reconstruction performance. Despite these advancements, coarse-to-fine SVR methods can be time-consuming and prone to local minima due to insufficient modeling of slice spatial relationships. Bundling slices into a single 3D volume may amplify errors, as corrupted volumes affect the accuracy of motion parameter predictions. In contrast, our approach redefines the SVR problem as a 3D image registration and interpolation challenge (Balakrishnan et al. 2018, 2019; Mok and Chung 2020; Young et al. 2024). We introduce a slice sequence-guided state space model to capture slice relationships and spatial sequences, and use a fully CNN-based interpolation network to recover the volume directly from the stacked slices, improving both reconstruction speed and performance.

State Space Models

State space sequence models (SSMs) (Gu, Goel, and Ré 2021; Gu et al. 2021), originally inspired by classical state-space models (Kalman 1960), have emerged as a promising architecture for sequence modeling. Mamba (Gu and Dao 2023) introduces a selective SSM architecture, integrating time-varying parameters into the SSM framework and proposing a hardware-aware algorithm to facilitate efficient training and inference processes. Recent works have adapted Mamba for visual learning, leveraging its global receptive field and dynamic weights. For example, some approaches have been applied Mamba to classification, segmentation, and object detection tasks by scanning input image patches (Liu et al. 2024b; Zhu et al. 2024; Yue and Li 2024). VMamba (Liu et al. 2024b) enhanced Mamba by adding vertical scanning directions, creating a cross-scan. Zhang et al. (Zhang et al. 2024) developed a Mamba model for motion generation, scanning unidirectionally along the temporal sequence and bidirectionally along channel dimensions in a hierarchical manner. Vision Mamba (Zhu et al. 2024) marks image sequences with position embeddings and compresses the visual representation using bidirectional state space models. Inspired by Vision Mamba, we designed a slice sequence-guided odd-even scan and marked slices with sequence embedding tokens for both slice and stack positions. This approach aims to effectively address the multi-stack SVR problem by accurately capturing spatial relationships between slices.

Method

We propose a novel Mamba-based method that incorporates slice sequence-guided state space modeling to learn the relationships between slices within stacks, coupled with a CNN-based interpolation network that produces a noise-suppressed 3D reconstruction as a byproduct (Figure 3). First, we introduce the concept of state space models (SSMs). Next, we detail the principles behind the proposed slice sequence-guided state space network and the CNN-based interpolation network.

Preliminaries

S6 Models Selective Scan Structured State Space Sequence (S6) models (Gu and Dao 2023) are a category of sequence models known for their superior ability to handle sequences. These models extend the previously proposed S4 models (Gu, Goel, and Ré 2021), which map one-dimensional input functions or sequences, denoted as $x(t) \in \mathcal{R}$, through hidden states $h(t) \in \mathcal{R}^N$ to an output $y(t) \in \mathcal{R}$. This process is represented by a linear Ordinary Differential Equation (ODE):

$$\begin{aligned} h'(t) &= \boldsymbol{A}h(t) + \boldsymbol{B}x(t) \\ y(t) &= \boldsymbol{C}h(t) + Dx(t) \\ h_t &= \overline{\boldsymbol{A}}h_{t-1} + \overline{\boldsymbol{B}}x_t, \\ y_t &= \boldsymbol{C}h(t) \end{aligned} \tag{1}$$

where $\mathbf{A} \in \mathcal{R}^{N\times N}$, $\mathbf{B} \in \mathcal{R}^{N\times 1}$, $\mathbf{C} \in \mathcal{R}^{N\times 1}$ and $D \in \mathcal{R}^1$ are the weighting parameters. For practical computation, these continuous dynamical systems are discretized. This is done by transforming $\mathbf{A}$ and $\mathbf{B}$ into discrete parameters $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ using the zero-order hold (ZOH) discretization rule:

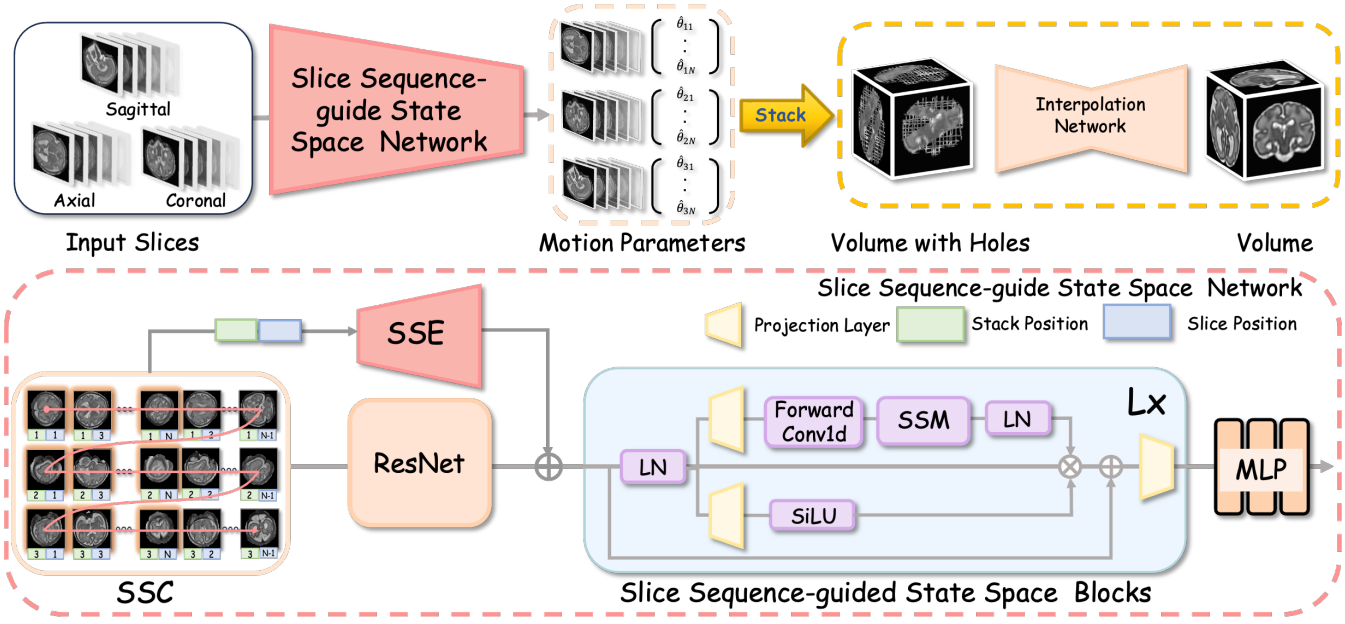

Figure 3: **Overview of the SVRMamba Model Architecture.** The model processes slices from three views to predict motion parameters using the Slice Sequence-guided State Space Network, as shown in the pink box at the bottom. First, the slices' positions are marked and aligned according to the SSC. Next, image feature tokens are extracted from the slices using ResNet, while slice position tokens are generated through SSE. These tokens are combined and fed into the Slice Sequence-guided State Space Blocks (SSB), which use state space modeling to capture spatial relationships between slices. An MLP is then employed within the SSB to predict motion parameters. Using these motion parameters, all slices are stacked into a volume, after which an interpolation network fills any holes and suppresses noise, resulting in a high-quality 3D reconstruction.

$$\begin{aligned} \overline{\mathbf{A}} &= \exp(\Delta \mathbf{A}) \\ \overline{\mathbf{B}} &= (\Delta \mathbf{A})^{-1}(\exp(\Delta \mathbf{A}) - \mathbf{I}) \cdot \Delta \mathbf{B} \end{aligned} \tag{2}$$

Where Δ represents the discrete step size. Since the weighting parameters and discretization rules remain fixed over time, S4 models can be viewed as linear time-invariant systems. Mamba expands on S4 models by extending the projection matrices to scan the entire input sequence through a selective scan.

SVRMamba

Problem Formulation Given multiple stacks of 2D MRI slice denoted as $I = \{I_{ij} \mid i = 1, 2, 3; j = 1, 2, \ldots, N\}$, our goal is to estimate the motion parameters $\hat{\boldsymbol{\theta}} = \left\{\hat{\theta}_{ij} \mid i = 1, 2, 3; j = 1, 2, \ldots, N\right\}$ that relate the slices within a canonical 3D space, and use these predicted motion parameters to stack the slices into a 3D volume $\widehat{\boldsymbol{V}}$. We aim to learn a mapping function $f\left(I_{ij}\right) = \left\{\hat{\theta}_{ij}\right\}$ that regresses rigid motion parameters from the image I_{ij}, where $\hat{\theta}_{ij} = (d_x, d_y, d_z, r_x, r_y, r_z)_{ij}$. Here, $\hat{\boldsymbol{d}}_{ij} = (d_x, d_y, d_z)_{ij}$ represents translations, and $\hat{\boldsymbol{r}}_{ij} = (r_x, r_y, r_z)_{ij}$ represents Euler rotation angles along the three axes. Using the predicted motion parameters, we then stack all 2D MRI slices into an intermediate volume $\widetilde{V}$. Typically, a 3D volume reconstructed from three orthogonal stacks may contain holes if regions of the underlying volume are missed during slice acquisition due to subject motion. To address this, we feed $\widetilde{\boldsymbol{V}}$ into a CNN-based IN to generate the complete 3D volume $\widehat{\boldsymbol{V}}$.

Model Architecture. An overview of the proposed SVRMamba architecture is illustrated in Figure 3. The process begins by feeding the slices from three views into the slice sequence-guided state space network to predict the motion parameters. This network leverages slice spatial relationships by modeling the state space using the proposed slice sequence-guided state space blocks, incorporating SSC and SSE. After estimating the motion parameters, the slices are stacked into a volume. Finally, an interpolation network is employed to fill holes and produce a noise-suppress volume.

Slice Sequence-guided State Space Network In this network, the input slices from the three views are first aligned using the SSC. Next, the aligned slice images are fed into a ResNet to extract feature tokens $\mathbf{t}$. These tokens are then flattened into 1-D feature tokens, and a learnable slice sequence spatial position embedding is added to preserve the spatial positional information of the slice sequence. The tokens are combined with the sequence embedding as follows:

$$\mathbf{T}_0 = [\mathbf{t}_{11}; \mathbf{t}_{12}; \cdots; \mathbf{t}_{3N}] + \mathbf{E}_{pos} \tag{3}$$

This token $\mathbf{T}_0$ is then sent to the slice sequence-guided state space blocks for further processing.

Slice Sequence Scan and Slice Sequence Embedding

To achieve robust reconstruction and effectively utilize feature tokens, we reformulate Mamba's unidirectional scanning as a slice sequence-guided odd-even scan, specifically adapting it for 3D SVR tasks. Unlike video-based Mamba models (Chen et al. 2024; Li et al. 2024a), which primarily learn temporal features from one frame sequences, our SVRMamba focuses on the spatial sequence within different stacks. We found that incorporating slice sequence embedding into the Mamba framework significantly enhances motion parameters prediction performance by better modeling slice-stack relationships. The position embedding is initialized using a sinusoidal function:

$$\begin{aligned} \mathbf{E}_{pos}(pos, 2c) &= \sin(pos \times w) \\ \mathbf{E}_{pos}(pos, 2c+1) &= \cos(pos \times w) \end{aligned} \tag{4}$$

$$w = \frac{1}{10000^{2c/C_{out}}} \tag{5}$$

where pos represents the stack and slice indices along the slice series, c is the index along the output channels, and C_{out} is the total number of output channels. The term w is a multiplier that is learnable during network training.

Slice Sequence-guided State Space Block (SSB) The SSB is designed to follow a specific slice sequence pattern, taking into account both slice-to-slice and slice-to-stack relationships. The token sequence $\mathbf{T}_{l-1}$ is passed to the l-th layer of the SSB, producing the output $\mathbf{T}_l$. The final output motion token $\mathbf{T}_l$ is then normalized and fed into an MLP to obtain the final predicted motion parameters $\hat{\boldsymbol{\theta}}$:

$$\begin{aligned} \mathbf{T}_l &= \mathrm{SSB}\left(\mathbf{T}_{1-1}\right) + \mathbf{T}_{1-1}, \\ \mathbf{z} &= \mathrm{Norm}\left(\mathbf{T}_l\right), \\ \hat{\boldsymbol{\theta}} &= \mathrm{MLP}(\mathbf{z}), \end{aligned} \tag{6}$$

where L is the number of layers, and Norm denotes the normalization layer.

Noise-suppressed Interpolation Network Unlike natural images, MR images usually have limited variability and follow a predictable distribution for 3D volumes. This characteristic makes interpolation methods particularly effective for filling gaps in reconstructions and adhering to regularity constraints, such as those required for pathology. To reduce image noise, we use CNN-based architectures with the spectral bias (Liu et al. 2023) as the core of our interpolation network. The reconstructed 3D volume, $\widehat{V}$, is obtained through:

$$\widehat{\boldsymbol{V}} = f_{IN}(\widetilde{\boldsymbol{V}}) \tag{7}$$

Here, f_{IN} represents the interpolation network, which we implement using a standard 3D U-Net model. It is important to note that searching for the best interpolation network model is not the focus of this study.

Loss Function

The loss function for the proposed network consists of two components: 1) motion loss for the transformation parameters $\mathcal{L}_1$, and 2) volume interpolation loss for the motion-corrected volume $\mathcal{L}_2$. For the motion loss, given that 3D rigid motion is represented by a special Euclidean group denoted as $SE(3)$, we use the geodesic distance (Salehi et al. 2018) as the loss metric. This measures the distance on the 6-dimensional non-Euclidean manifold, rather than using Euclidean norms. The motion loss $\mathcal{L}_1$ for the transformation parameters $\hat{\boldsymbol{\theta}}$ is defined as:

$$\mathcal{L}_1(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^{3} \sum_{j=1}^{N} \left(\left\| \hat{\boldsymbol{R}}_{ij}^T \boldsymbol{R}_{ij} \right\|_2 + \left\| \hat{\boldsymbol{d}}_{ij} - \boldsymbol{d}_{ij} \right\|_2 \right) \tag{8}$$

where $\hat{\boldsymbol{R}}_{ij}$ and $\hat{\boldsymbol{d}}_{ij}$ are the predicted rotation matrix and displacement parameters of slice I_{ij} derived from the predicted motion parameters $\hat{\theta}_{ij}$. The terms $\boldsymbol{R}_{ij}$ and $\boldsymbol{d}_{ij}$ denote the true rotation matrix and displacement parameters, respectively. The loss for the reconstructed volume $\mathcal{L}_2$ is defined as:

$$\mathcal{L}_2(\boldsymbol{V}, \widehat{\boldsymbol{V}}) = \|\boldsymbol{V} - \widehat{\boldsymbol{V}}\|_2 \tag{9}$$

where $\widehat{\boldsymbol{V}}$ is the interpolated volume, and $\boldsymbol{V}$ is the ground truth volume. The final loss function for SVRMamba is a weighted combination of $\mathcal{L}_1$ and $\mathcal{L}_2$:

$$\mathcal{L} = \mathcal{L}_1(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) + \mathcal{L}_2(\boldsymbol{V}, \widehat{\boldsymbol{V}}) \tag{10}$$

Experiments

Dataset

We evaluated SVRMamba using three publicly available and widely-used datasets, which also serve as the ground truth (GT) for comparison: 1) **FeTA dataset** (Payette et al. 2021): This dataset includes 80 fetal brain MRI volumes. The dataset is divided into 68 volumes for training and 12 volumes for testing. 2) **FBA dataset** (Wu et al. 2021a): This dataset contains 14 fetal brain MRI volumes with a 0.8 mm isotropic resolution, split into 6 volumes for training and 8 volumes for testing. 3) **Spina Bifida Aperta (SBA) dataset** (Fidon et al. 2021): This dataset comprises 14 pathological fetal brain MRI volumes with a 0.8 mm isotropic resolution, divided into 6 volumes for training and 8 volumes for testing. We created a single training dataset consisting of 80 volumes (68 from FeTA, 6 from SBA, and 6 from FBA) and three separate test datasets (12 volumes from FeTA, 8 from SBA, and 8 from FBA). For the training dataset, we prepared paired data (simulated three-view orthogonal stacks, motion parameters, and GTs) by applying random horizontal flips, rotations, and translations to the registered GT volumes to simulate the various motions. For each subject, three orthogonal stacks were simulated with an in-plane resolution of 0.8 mm and a slice thickness of 3-4 mm. Detailed information about data simulation and training procedures is provided in the supplementary materials.

Methods	Time (s)	SBA				FETA				FBA			
		PSNR↑	SSIM↑	MT↓	MR↓	PSNR↑	SSIM↑	MT↓	MR↓	PSNR↑	SSIM↑	MT↓	MR↓
NiftyMIC	1642.8	19.01	0.60	1.29	1.16	18.70	0.65	2.02	2.89	13.06	0.37	11.73	8.29
SVRTK	241.6	21.58	0.74	0.51	0.62	18.98	0.67	2.23	2.96	16.69	0.50	6.08	6.75
NeSVoR	183.9	22.69	0.76	0.47	0.48	21.61	0.79	1.68	1.62	15.02	0.45	6.38	5.19
SVoRT+NeSVoR	187.4	22.60	0.75	0.39	0.53	21.23	0.78	1.24	1.34	22.76	0.78	2.82	2.39
SVRMamba	**3.0**	**23.84**	**0.83**	**0.24**	**0.36**	**23.46**	**0.83**	**0.31**	**0.41**	**23.61**	**0.84**	**0.58**	**1.51**

Table 1: Comparison with SOTAs on three test datasets. The best results are bold.

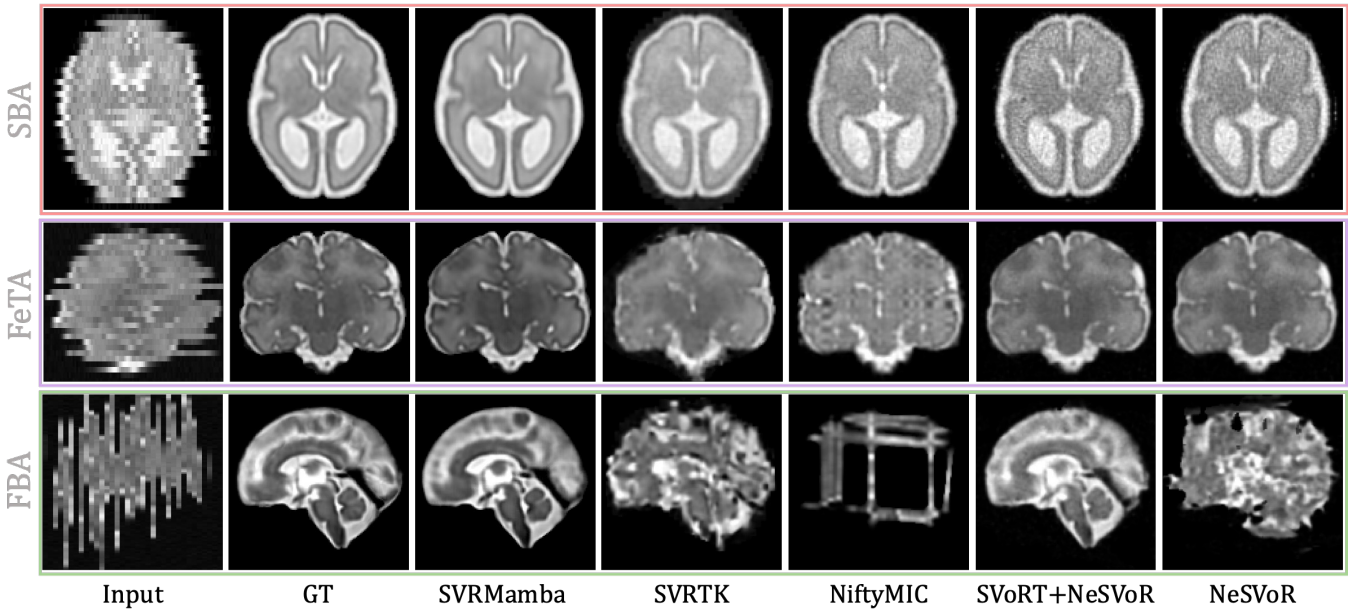

Figure 4: **Qualitative comparison** of the proposed SVRMamba with SOTAs on three test datasets.

Experiment Setup

Implementation Detail We implemented SVRMamba using PyTorch and ran it on an NVIDIA RTX Titan GPU. The model was trained using the Adam optimizer, with a cosine annealing learning rate scheduler that had a minimum learning rate of 0.0001 over 2×10^5 iterations. For the SBA test dataset, the motion parameters were sampled with 3D rotation parameters uniformly and independently from $U(-2, 2)$ degrees and 3D translation from $U(-2, 2)\text{mm}$. For the FeTA test dataset, the motion parameters were sampled with 3D rotation parameters from $U(-6, 6)$ degrees and 3D translation from $U(-4, 4)\text{mm}$. For the FBA test dataset, the motion parameters were sampled with 3D rotation parameters from $U(-15, 15)$ degrees and 3D translation from $U(-6, 6)\text{mm}$.

Evaluation Metric To assess the accuracy of the predicted motion parameters across different models, we used the Mean Absolute Error (MAE) for both rotation angles (MR) and translations (MT). Additionally, we evaluated the volume reconstruction by comparing the reconstructed volumes with the GT using various quantitative metrics, including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

Comparison with State-of-the-Art Method

We compared SVRMamba against four SOTA SVR methods, each of which provided open-source implementations: 1) SVRTK: a classic SVR toolkit (Uus et al. 2024), 2) NiftyMIC: an automatic SVR framework (Ebner et al. 2020), 3) NeSVoR: a SVR method based on implicit neural representation (Xu et al. 2023), and 4) SVoRT+NeSVoR: an approach utilizing SVoRT (Xu et al. 2022) to provide initialized transformations for NeSVoR. To ensure fair comparisons, we conducted hyperparameter tuning for each method. We randomly selected one subject from the simulated fetal dataset and tuned the hyperparameters to minimize the mean squared error between the reconstructed volume and the GT. Once optimized, these hyperparameters were fixed and applied to the test datasets.

Table 1 displays the quantitative results of our comparison, demonstrating that SVRMamba outperforms all other methods across all evaluation metrics. Figure 4 shows the qualitative results of 3D volume reconstruction using various methods on three test datasets. For the SBA dataset, which features moderate motion and severe noise, all methods manage to reconstruct the volume, but only SVRMamba effectively reduces noise due to its interpolation network.

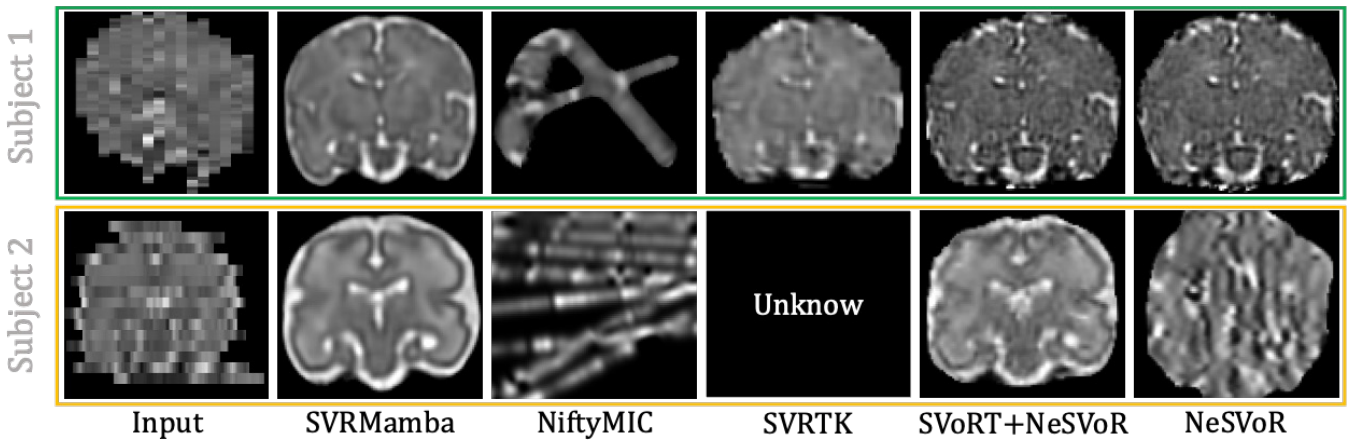

Figure 5: Comparison of SVRMamba with SOTAs on clinical data.

On the FeTA dataset, characterized by moderate motion and noise, SVRMamba delivers superior brain volume reconstruction with less noise compared to other methods. In the FBA dataset, despite severe motion, SVRMamba still achieves better brain volume reconstruction with minimal noise. This enhanced performance is attributed to its effective learning of spatial slice relationships through the slice sequence-guided state space model and improved noise suppression with the CNN-based IN.

To assess SVRMamba's generalization to real clinical scenarios, even when trained on synthetic datasets, we tested it on two real T2-weighted clinical datasets with three orthogonal view stacks of slices. These datasets, retrospectively sourced from archived MRI data, present challenges due to real-world noise and uncontrollable fetal motion. Detailed parameters for these datasets are provided in the supplementary materials. Figure 5 shows a qualitative comparison of SVRMamba's 3D volume reconstruction against SOTA models. In challenging clinical scenarios, SVRMamba consistently delivers superior reconstruction quality, preserving more of the brain structure and reducing noise more effectively than other models.

Ablation Study

We conducted the following ablation experiments:

- **w/o SSC**: We replaced SSC with unidirectional scanning.
- **w/o SSE**: We removed SSE from the slice sequence-guided state space network.
- **w/o IN**: We removed IN and instead used the widely adopted Point Spread Function (PSF)-based interpolation method to directly reconstruct the volume.

The qualitative results are provided in the supplementary materials. Table 2 presents the quantitative results of ablation studies. The slice sequence-guided state space network, which models the slice-stack relationship and learns the spatial sequences of slices through state space modeling, degraded significantly when SSC or SSE was excluded. This demonstrates that SSC and SSE are crucial for capturing local context and effectively guiding the state space network.

Dataset	Methods	PSNR↑	SSIM↑	MT↓	MR↓
SBA	SVRMamba	**23.84**	**0.83**	**0.24**	**0.36**
	w/o SSC	23.41	0.81	0.26	0.44
	w/o SSE	22.68	0.78	0.32	0.59
	w/o IN	22.46	0.77	0.24	0.36
FeTA	SVRMamba	**23.46**	**0.83**	**0.31**	**0.41**
	w/o SSC	22.59	0.80	0.35	0.61
	w/o SSE	20.64	0.72	0.48	0.92
	w/o IN	21.73	0.76	0.31	0.41
FBA	SVRMamba	**23.61**	**0.84**	**0.58**	**1.51**
	w/o SSC	22.95	0.80	0.75	1.80
	w/o SSE	21.72	0.77	1.32	3.51
	w/o IN	22.46	0.77	0.58	1.51

Table 2: Ablation experiments on three datasets.

An even more substantial drop in performance was observed when the proposed IN was replaced with the conventional interpolation method, underscoring the importance of IN in achieving high-quality 3D reconstruction.

Conclusion

We introduce SVRMamba, a novel framework designed for reconstructing 3D volumes from multiple 2D slice stacks. Our approach combines slice sequence-guided state space modeling with a CNN-based interpolation network to deliver high-quality 3D reconstructions. By reformulating Mamba's unidirectional scanning into a slice sequence-guided odd-even directional scanning and incorporating sequence embedding tokens, our framework effectively learns slice relationships and spatial sequences, improving motion correction. Additionally, the CNN-based interpolation network enhances reconstruction quality by reducing noise. Comprehensive experiments on benchmark and clinical datasets show that SVRMamba significantly outperforms existing SOTA methods, achieving a sixtyfold increase in reconstruction speed while maintaining comparable accuracy.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62071299, the Guizhou Senior Innovative Talent Project (QKHPTRC-GCC[2022]041–1), and the Guizhou Provincial Science and Technology project (QKHZC[2019]2810).

References

Alansary, A.; Rajchl, M.; McDonagh, S. G.; Murgasova, M.; Damodaram, M.; Lloyd, D. F.; Davidson, A.; Rutherford, M.; Hajnal, J. V.; Rueckert, D.; et al. 2017. PVR: patch-to-volume reconstruction for large area motion correction of fetal MRI. *IEEE transactions on medical imaging*, 36(10): 2031–2044.

Aviles Verdera, J.; Story, L.; Hall, M.; Finck, T.; Egloff, A.; Seed, P. T.; Malik, S. J.; Rutherford, M. A.; Hajnal, J. V.; Tomi-Tricot, R.; et al. 2023. Reliability and feasibility of low-field-strength fetal MRI at 0.55 T during pregnancy. *Radiology*, 309(1): e223050.

Balakrishnan, G.; Zhao, A.; Sabuncu, M. R.; Guttag, J.; and Dalca, A. V. 2018. An unsupervised learning model for deformable medical image registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 9252–9260.

Balakrishnan, G.; Zhao, A.; Sabuncu, M. R.; Guttag, J.; and Dalca, A. V. 2019. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8): 1788–1800.

Behrouz, A.; and Hashemi, F. 2024. Graph mamba: Towards learning on graphs with state space models. *arXiv preprint arXiv:2402.08678*.

Chen, G.; Huang, Y.; Xu, J.; Pei, B.; Chen, Z.; Li, Z.; Wang, J.; Li, K.; Lu, T.; and Wang, L. 2024. Video mamba suite: State space model as a versatile alternative for video understanding. *arXiv preprint arXiv:2403.09626*.

Chen, L.; Wang, Y.; Wu, Z.; Shan, Y.; Li, T.; Hung, S.-C.; Xing, L.; Zhu, H.; Wang, L.; Lin, W.; et al. 2023. Four-dimensional mapping of dynamic longitudinal brain subcortical development and early learning functions in infants. *Nature Communications*, 14(1): 3727.

Coakley, F. V.; Glenn, O. A.; Qayyum, A.; Barkovich, A. J.; Goldstein, R.; and Filly, R. A. 2004. Fetal MRI: a developing technique for the developing patient. *American Journal of Roentgenology*, 182(1): 243–252.

Davidson, J. R.; Uus, A.; Matthew, J.; Egloff, A. M.; Deprez, M.; Yardley, I.; De Coppi, P.; David, A.; Carmichael, J.; and Rutherford, M. A. 2021. Fetal body MRI and its application to fetal and neonatal treatment: an illustrative review. *The Lancet Child & Adolescent Health*, 5(6): 447–458.

Ebner, M.; Wang, G.; Li, W.; Aertsen, M.; Patel, P. A.; Aughwane, R.; Melbourne, A.; Doel, T.; Dymarkowski, S.; De Coppi, P.; et al. 2020. An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI. *NeuroImage*, 206: 116324.

Evans, A. C.; Janke, A. L.; Collins, D. L.; and Baillet, S. 2012. Brain templates and atlases. *Neuroimage*, 62(2): 911–922.

Fidon, L.; Viola, E.; Mufti, N.; David, A. L.; Melbourne, A.; Demaerel, P.; Ourselin, S.; Vercauteren, T.; Deprest, J.; and Aertsen, M. 2021. A spatio-temporal atlas of the developing fetal brain with spina bifida aperta. *Open Research Europe*, 1.

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

Gholipour, A.; Rollins, C. K.; Velasco-Annis, C.; Ouaalam, A.; Akhondi-Asl, A.; Afacan, O.; Ortinau, C. M.; Clancy, S.; Limperopoulos, C.; Yang, E.; et al. 2017. A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth. *Scientific reports*, 7(1): 476.

Gong, H.; Kang, L.; Wang, Y.; Wan, X.; and Li, H. 2024. nnmamba: 3d biomedical image segmentation, classification and landmark detection with state space model. *arXiv preprint arXiv:2402.03526*.

Gu, A.; and Dao, T. 2023. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.

Gu, A.; Goel, K.; and Ré, C. 2021. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*.

Gu, A.; Johnson, I.; Goel, K.; Saab, K.; Dao, T.; Rudra, A.; and Ré, C. 2021. Combining recurrent, convolutional, and continuous-time models with linear state space layers. *Advances in neural information processing systems*, 34: 572–585.

Hou, B.; Khanal, B.; Alansary, A.; McDonagh, S.; Davidson, A.; Rutherford, M.; Hajnal, J. V.; Rueckert, D.; Glocker, B.; and Kainz, B. 2018. 3-D reconstruction in canonical coordinate space from arbitrarily oriented 2-D images. *IEEE transactions on medical imaging*, 37(8): 1737–1750.

Kainz, B.; Steinberger, M.; Wein, W.; Kuklisova-Murgasova, M.; Malamateniou, C.; Keraudren, K.; Torsney-Weir, T.; Rutherford, M.; Aljabar, P.; Hajnal, J. V.; et al. 2015. Fast volume reconstruction from motion corrupted stacks of 2D slices. *IEEE transactions on medical imaging*, 34(9): 1901–1913.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems.

Kuklisova-Murgasova, M.; Quaghebeur, G.; Rutherford, M. A.; Hajnal, J. V.; and Schnabel, J. A. 2012. Reconstruction of fetal brain MRI with intensity matching and complete outlier removal. *Medical image analysis*, 16(8): 1550–1564.

Li, K.; Li, X.; Wang, Y.; He, Y.; Wang, Y.; Wang, L.; and Qiao, Y. 2024a. Videomamba: State space model for efficient video understanding. *arXiv preprint arXiv:2403.06977*.

Li, L.; Wang, H.; Zhang, W.; and Coster, A. 2024b. Stg-mamba: Spatial-temporal graph learning via selective state space model. *arXiv preprint arXiv:2403.12418*.

Liang, D.; Zhou, X.; Wang, X.; Zhu, X.; Xu, W.; Zou, Z.; Ye, X.; and Bai, X. 2024. Pointmamba: A simple

state space model for point cloud analysis. *arXiv preprint arXiv:2402.10739*.

Lieber, O.; Lenz, B.; Bata, H.; Cohen, G.; Osin, J.; Dalmedigos, I.; Safahi, E.; Meirom, S.; Belinkov, Y.; Shalev-Shwartz, S.; et al. 2024. Jamba: A hybrid transformer-mamba language model. *arXiv preprint arXiv:2403.19887*.

Liu, J.; Yu, R.; Wang, Y.; Zheng, Y.; Deng, T.; Ye, W.; and Wang, H. 2024a. Point mamba: A novel point cloud backbone based on state space model with octree-based ordering strategy. *arXiv preprint arXiv:2403.06467*.

Liu, Y.; Li, J.; Pang, Y.; Nie, D.; and Yap, P.-T. 2023. The devil is in the upsampling: Architectural decisions made simpler for denoising with deep image prior. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12408–12417.

Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; and Liu, Y. 2024b. VMamba: Visual State Space Model. *arXiv preprint arXiv:2401.10166*.

Mok, T. C.; and Chung, A. 2020. Fast symmetric diffeomorphic image registration with convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4644–4653.

Namburete, A. I.; Papież, B. W.; Fernandes, M.; Wyburd, M. K.; Hesse, L. S.; Moser, F. A.; Ismail, L. C.; Gunier, R. B.; Squier, W.; Ohuma, E. O.; et al. 2023. Normative spatiotemporal fetal brain maturation with satisfactory development at 2 years. *Nature*, 623(7985): 106–114.

Payette, K.; de Dumast, P.; Kebiri, H.; Ezhov, I.; Paetzold, J. C.; Shit, S.; Iqbal, A.; Khan, R.; Kottke, R.; Grehten, P.; et al. 2021. An automatic multi-tissue human fetal brain segmentation benchmark using the fetal tissue annotation dataset. *Scientific data*, 8(1): 167.

Payette, K.; Li, H. B.; de Dumast, P.; Licandro, R.; Ji, H.; Siddiquee, M. M. R.; Xu, D.; Myronenko, A.; Liu, H.; Pei, Y.; et al. 2023. Fetal brain tissue annotation and segmentation challenge results. *Medical Image Analysis*, 88: 102833.

Rousseau, F.; Glenn, O. A.; Iordanova, B.; Rodriguez-Carranza, C.; Vigneron, D. B.; Barkovich, J. A.; and Studholme, C. 2006. Registration-based approach for reconstruction of high-resolution in utero fetal MR brain images. *Academic radiology*, 13(9): 1072–1081.

Salehi, S. S. M.; Hashemi, S. R.; Velasco-Annis, C.; Ouaalam, A.; Estroff, J. A.; Erdogmus, D.; Warfield, S. K.; and Gholipour, A. 2018. Real-time automatic fetal brain extraction in fetal MRI by deep learning. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, 720–724. IEEE.

Shi, W.; Xu, H.; Sun, C.; Sun, J.; Li, Y.; Xu, X.; Zheng, T.; Zhang, Y.; Wang, G.; and Wu, D. 2022. AFFIRM: Affinity Fusion-based Framework for Iteratively Random Motion correction of multi-slice fetal brain MRI. *IEEE Transactions on Medical Imaging*.

Tourbier, S.; Bresson, X.; Hagmann, P.; Thiran, J.-P.; Meuli, R.; and Cuadra, M. B. 2015. An efficient total variation algorithm for super-resolution in fetal brain MRI with adaptive regularization. *NeuroImage*, 118: 584–597.

Uus, A. U.; Neves Silva, S.; Aviles Verdera, J.; Payette, K.; Hall, M.; Colford, K.; Luis, A.; Sousa, H. S.; Ning, Z.; Roberts, T.; et al. 2024. Scanner-based real-time 3D brain+ body slice-to-volume reconstruction for T2-weighted 0.55 T low field fetal MRI. *medRxiv*, 2024–04.

Wang, C.; Tsepa, O.; Ma, J.; and Wang, B. 2024. Graph-mamba: Towards long-range graph sequence modeling with selective state spaces. *arXiv preprint arXiv:2402.00789*.

Wu, J.; Chen, L.; Li, Z.; Wang, L.; Wang, R.; Wei, H.; and Zhang, Y. 2023. ASSURED: A Self-Supervised Deep Decoder Network for Fetus Brain MRI Reconstruction. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, 1–5. IEEE.

Wu, J.; Sun, T.; Yu, B.; Li, Z.; Wu, Q.; Wang, Y.; Qian, Z.; Zhang, Y.; Jiang, L.; and Wei, H. 2021a. Age-specific structural fetal brain atlases construction and cortical development quantification for Chinese population. *Neuroimage*, 241: 118412.

Wu, J.; Yu, B.; Wang, L.; Yang, Q.; and Zhang, Y. 2021b. Longitudinal Chinese population structural fetal brain atlases construction: toward precise fetal brain segmentation. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2745–2749. IEEE.

Xu, J.; Moyer, D.; Gagoski, B.; Iglesias, J. E.; Grant, P. E.; Golland, P.; and Adalsteinsson, E. 2023. NeSVoR: Implicit Neural Representation for Slice-to-Volume Reconstruction in MRI. *IEEE Transactions on Medical Imaging*.

Xu, J.; Moyer, D.; Grant, P. E.; Golland, P.; Iglesias, J. E.; and Adalsteinsson, E. 2022. SVoRT: Iterative transformer for slice-to-volume registration in fetal brain MRI. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, 3–13. Springer.

Young, S. I.; Balbastre, Y.; Fischl, B.; Golland, P.; and Iglesias, J. E. 2024. Fully Convolutional Slice-to-Volume Reconstruction for Single-Stack MRI. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11535–11545.

Yue, Y.; and Li, Z. 2024. Medmamba: Vision mamba for medical image classification. *arXiv preprint arXiv:2403.03849*.

Zhang, Z.; Liu, A.; Reid, I.; Hartley, R.; Zhuang, B.; and Tang, H. 2024. Motion mamba: Efficient and long sequence motion generation with hierarchical and bidirectional selective ssm. *arXiv preprint arXiv:2403.07487*.

Zhu, L.; Liao, B.; Zhang, Q.; Wang, X.; Liu, W.; and Wang, X. 2024. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*.

Zöllei, L.; Iglesias, J. E.; Ou, Y.; Grant, P. E.; and Fischl, B. 2020. Infant FreeSurfer: An automated segmentation and surface extraction pipeline for T1-weighted neuroimaging data of infants 0–2 years. *Neuroimage*, 218: 116946.