Flexibly Distilled 3D Rectified Flow with Anatomical Constraints for Developmental Infant Brain MRI Prediction

Haifeng Wang[1,3], Zehua Ren[2,3], Heng Chang[1,3], Xinmei Qiu[1,3], Fan Wang[2,3], Chunfeng Lian[1,3,4,(✉)], and Jianhua Ma[2,3,4,(✉)]

[1] School of Mathematics and Statistics, Xian Jiaotong University, Xi'an, China
[2] Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xian Jiaotong University, Xi'an, China
[3] Research Center for Intelligent Medical Equipment and Devices (IMED), Xi'an
[4] Pazhou Lab (Huangpu), Guangzhou, China
{chunfeng.lian, jhma}@xjtu.edu.cn

Abstract. Longitudinal prediction of infant brain MRIs is crucial for individualized neurodevelopment tracking and disorder forecasting. However, existing methods, such as diffusion-based generative models, often struggle to capture the complex spatiotemporal dynamics of developing brains, leading to unreliable predictions that lack subject-specific, anatomically consistent growth patterns. To address this, we propose a **Flexibly Distilled 3D Rectified Flow (FDRF)** framework, which integrates anatomical constraints for dual-stream predictions of volumetric images and tissue maps along developmental trajectories. Our framework features an age-conditioned feature fusion module for controllable prediction with targeted age appearances and employs anatomical constraints derived from segmentation labels and high-frequency image details to ensure subject-level spatiotemporal consistency. Additionally, we introduce a flexible distillation of rectified flow, enabling a unified one-step generative model for high-fidelity cross-time predictions while preserving individualized anatomical details. Given 6-month MRIs and tissue maps as the input, our model reliably predicts their spatiotemporal growths at 12 and 24 months, outperforming existing diffusion-based baselines by relatively large margins. Our codes can be found at https://github.com/ladderlab-xjtu/FDRF.

Keywords: Rectified Flow · Controllable Distillation · Longitudinal Prediction · Infant Brain MRI.

1 Introduction

The early development of the infant brain is crucial for lifelong neurological health and cognitive function [3,7]. It is a period of rapid structural and functional changes that are fundamental for the establishment of normal neural networks and the acquisition of various skills. Understanding these normal development patterns is of great clinical significance, particularly for the early detection

and diagnosis of neurodevelopmental disorders such as autism spectrum disorder (ASD) [11,2].

Thus, longitudinal prediction of infant brain MRIs along the development trajectory is of great value for individualized neurodevelopment and associated disorder forecasting. In this context, the emergence of diffusion models offers a promising solution. In recent years, diffusion models have shown great potential in medical image generation due to their ability to generate high-quality images with realistic details. These models progressively denoise a signal from pure noise, capturing complex data distributions and generating high-fidelity samples. Key developments in diffusion models include Denoising Diffusion Probabilistic Models (DDPM) [5], Denoising Diffusion Implicit Models (DDIM) [15], Latent Diffusion Models (LDM) [12], and Flow Matching [8], each enhancing sampling efficiency, computational complexity, and generalization performance. These advancements in diffusion models have led to significant progress in medical image generation, addressing challenges such as data scarcity and privacy concerns. For instance, the SequenceMorph framework [17] provides an unsupervised approach for motion tracking on cardiac image sequences, with potential for extension to other medical imaging domains. The Treatment-aware Diffusion Probabilistic Model [9] integrates treatment information to predict longitudinal MRI and diffuse glioma growth, supporting personalized treatment planning. The Sequence-Aware Diffusion Model (SADM) [18] leverages temporal dependencies in longitudinal data for high-fidelity medical image generation, advancing the field of longitudinal medical image analysis.

Despite these advances, existing diffusion-based methods for infant brain MRI generation or prediction still face several key limitations. Specifically, they struggle with age control in longitudinal generation, anatomical structure quality, and sampling efficiency. In terms of age control during longitudinal generation, existing methods lack flexibility in precisely controlling the generated images to present age-specific spatiotemporal appearances.

Regarding the fidelity of generated images, the anatomical details are still not satisfactory, especially at the subject level. The high-frequency details of individualized anatomical structures are often limited, which hinders the reliable quantification and prediction of brain development.

Moreover, many existing generation methods suffer from inefficient sampling processes (e.g., DDPM [5], DDIM [15], and flow matching [8]). The long sampling time hampers their real-world applications, for which efficiency and speed are critical factors.

To address these challenges, we propose a Flexibly Distilled 3D Rectified Flow (FDRF) framework, integrating anatomical constraints, for dual-stream predictions of volumetric images and tissue maps along early neurodevelopmental trajectories (from 6 to 24 months of age). Our FDRF makes four primary contributions. (1) We introduce a month-adaptive feature fusion module that integrates age conditions into image structural features, enabling precise control over the age of generated images and addressing the difficulty of predicting developmental images for specific months. (2) We develop a dual-stream generation

architecture that concurrently predicts brain images and tissue maps, ensuring better consistency and accuracy at the subject level in anatomical structures. (3) We employ a rectified flow [10] training strategy followed by distillation, allowing one-step sampling in inference and significantly improving the efficiency compared to iterative diffusion-based methods such as DDPM, DDIM, and flow matching. (4) We incorporate edge-based regularization and tissue segmentation loss as anatomical constraints into the distillation procedure, leading to high-fidelity predictions that preserve individualized spatiotemporal details.

2 Methods

2.1 Preliminaries: Rectified Flow

Rectified Flow [10] is a method for solving the transport mapping problem, which is an improvement over flow matching [8]. Specifically, for two distributions $\mathbf{z}_0 \sim \pi_0$ and $\mathbf{z}_1 \sim \pi_1$ in $\mathbb{R}^d$, the goal is to find a mapping $T : \mathbb{R}^d \to \mathbb{R}^d$ such that when $\mathbf{z}_0 \sim \pi_0$, $T(\mathbf{z}_0) \sim \pi_1$. Diffusion models [5,15] address this type of transport mapping problem by converting it into a continuous-time process governed by a class of stochastic differential equations (SDE). However, this approach suffers from long sampling times and instability issues.

To address these issues, rectified flow introduces an ordinary differential equation (ODE) model that transfers π_0 to π_1 via a straight line:

$$\mathrm{d}\mathbf{z}_t = v(\mathbf{z}_t, t)\mathrm{d}t, \quad \text{with} \quad \mathbf{z}_t = \mathbf{z}_0 + t(\mathbf{z}_1 - \mathbf{z}_0),\ t \in [0, 1], \tag{1}$$

where v denotes the velocity vector field. In practice, rectified flow employs a neural network v_θ to estimate the velocity vector field. To obtain the parameterized velocity vector field v_θ, we need to solve the following straightforward least squares regression problem:

$$\min_\theta \int_0^1 \mathbb{E} \left\| (\mathbf{z}_1 - \mathbf{z}_0) - v_\theta(\mathbf{z}_t, t) \right\|^2 \mathrm{d}t. \tag{2}$$

Thus, the loss function for training the neural network is given by:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{z}_0 \sim \pi_0, \mathbf{z}_1 \sim \pi_1} \left\| (\mathbf{z}_1 - \mathbf{z}_0) - v_\theta(\mathbf{z}_t, t) \right\|^2, \tag{3}$$

where $t \sim \mathcal{U}[0, 1]$. By leveraging the optimized v_θ as a path predictor, rectified flow successfully links the two distributions through almost linear paths. The forward Euler method enables rectified flow to achieve high-quality results from the training data with minimal computational steps.

2.2 FDRF: Flexibly Distilled 3D Rectified Flow

In this section, we introduce the Flexibly Distilled 3D Rectified Flow (FDRF), a novel pipeline designed to generate infant brain MRI and corresponding tissue segmentation labels from 6 months to 12 and 24 months. Our approach leverages a combination of structural and age features within a Rectified Flow framework, enhanced by edge and segmentation losses during the distillation phase, to achieve high-quality one-step sampling.

Network Architecture FDRF synthesizes brain images along with the corresponding tissue segmentation maps for 12 or 24 months from their 6-month counterparts. Specifically, for $(\mathbf{x}_0, \mathbf{y}_0) \in \pi_0$ representing the MRI image and segmentation label of a 6-month-old infant, and $(\mathbf{x}_1, \mathbf{y}_1) \in \pi_1$ representing those of infants aged 12 or 24 months, FDRF employs a velocity vector field U-Net [13] equipped with a feature fusion module (FFM) to model deformation and segmentation evolution from π_0 to π_1, as shown in Fig. 1. This design enables accurate and efficient mapping of distributions without relying on classifier-free guidance (CFG) [6] strategies, effectively capturing the essential characteristics of infant brain MRI and segmentation labels across different ages.

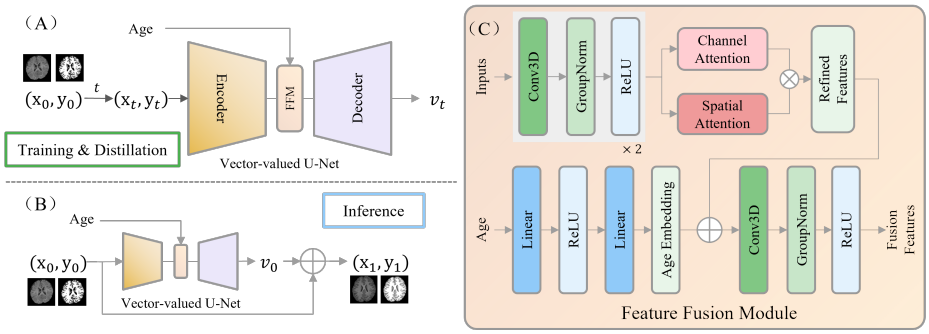

Fig. 1. The overall framework of FDRF. (A) During training, the vector-valued U-Net is optimized with $t \sim \mathcal{U}[0, 1]$, while in Flexible Distillation, $t = 0$ is used to enable one-step sampling; (B) Given the age, 6-month-old brain MRI, and tissue labels, the famework samples 12-month or 24-month-old brain MRI and tissue labels; (C) The feature fusion module (FFM) integrates age features and structural features.

Loss Functions To optimize the velocity vector field network, we start by using the standard Mean Squared Error (MSE) loss during the training phase. Given that $(x_0, y_0) \sim \pi_0, (x_1, y_1) \sim \pi_1$, the MSE loss is defined as:

$$\mathcal{L}_{\text{MSE}}(\theta) = \mathbb{E} \left\| (\mathbf{z}_1 - \mathbf{z}_0) - v_\theta(\mathbf{z}_t, t) \right\|^2, \quad \text{with } t \sim \mathcal{U}[0, 1], \tag{4}$$

where $\mathbf{z}_0 = (\mathbf{x}_0, \mathbf{y}_0), \mathbf{z}_1 = (\mathbf{x}_1, \mathbf{y}_1)$ and $\mathbf{z}_t = t\mathbf{z}_1 + (1-t)\mathbf{z}_0$. In addition to the MSE loss, we introduce two specialized loss functions: Edge loss and Segmentation loss. These loss functions are designed to enhance the network's performance by focusing on specific aspects of the velocity vector field. Before delving into the details of these specialized loss functions, we define the predicted image $\hat{\mathbf{x}}$ and the predicted segmentation mask $\hat{\mathbf{y}}$ as follows:

$$(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = (\mathbf{x}_0, \mathbf{y}_0) + v_\theta(\mathbf{x}_0, \mathbf{y}_0, 0). \tag{5}$$

Edge Loss This loss function focuses on preserving the structural details of MRI images by penalizing differences in the gradient of the predicted and target images.

$$\mathcal{L}_{\text{edge}}(\theta) = \mathbb{E}\left\|\Delta\mathbf{x}_1 - \Delta\hat{\mathbf{x}}\right\|^2, \tag{6}$$

where Δ denotes the Laplacian filtering operation.

Segmentation Loss This loss function ensures the accuracy of tissue segmentation labels by comparing the generated and ground-truth segmentation masks using a combination of Dice loss ($\mathcal{L}_{\text{dice}}$) and cross-entropy loss ($\mathcal{L}_{\text{ce}}$). These losses are widely used in the literature for evaluating segmentation performance [19,4,1]. The segmentation loss is formulated as follows:

$$\mathcal{L}_{\text{seg}}(\theta) = \mathcal{L}_{\text{dice}}(\mathbf{y}_1, \hat{\mathbf{y}}) + \mathcal{L}_{\text{ce}}(\mathbf{y}_1, \hat{\mathbf{y}}). \tag{7}$$

Flexible Distillation for One-Step Sampling During the training phase, we exclusively utilize MSE loss function, as shown in Eq. 4. To further enhance the efficiency of the velocity field and enable one-step sampling, we design a flexible distillation process. This process leverages the inherent property of the rectified flow framework, which promotes straight trajectories between distributions due to its linear interpolation formulation ($\mathbf{z}_t = \mathbf{z}_0 + t(\mathbf{z}_1 - \mathbf{z}_0)$). This property allows the velocity field v_θ to directly approximate the displacement $\mathbf{z}_1 - \mathbf{z}_0$ at $t = 0$, enabling one-step sampling without iterative ODE solving.

To achieve this, the flexible distillation process is designed with a specific loss function that optimizes the velocity field at $t = 0$. The loss function is formulated as follows:

$$\mathcal{L}(\theta) = \underbrace{\mathbb{E}\left\|(\mathbf{z}_1 - \mathbf{z}_0) - v_\theta(\mathbf{z}_0, 0)\right\|^2}_{\text{MSE at } t=0} + \lambda_1\mathcal{L}_{\text{edge}}(\theta) + \lambda_2\mathcal{L}_{\text{seg}}(\theta), \tag{8}$$

where $\mathbf{z}_0 = (\mathbf{x}_0, \mathbf{y}_0)$ and λ_1, λ_2 is hyperparameters that balance the two losses. Fixing $t = 0$, the distillation phase explicitly forces the network to internalize the full deformation and segmentation dynamics into a single-step prediction. This leverages the linearity of rectified flow paths to collapse the continuous-time ODE into a direct mapping:

$$(\mathbf{x}_1, \mathbf{y}_1) = (\mathbf{x}_0, \mathbf{y}_0) + v_\theta(\mathbf{x}_0, \mathbf{y}_0, 0), \tag{9}$$

where $v_\theta(\mathbf{x}_0, \mathbf{y}_0, 0)$ predicts the displacement from π_0 to π_1. The edge loss ($\mathcal{L}_{\text{edge}}$) and segmentation loss ($\mathcal{L}_{\text{seg}}$) ensure that high-frequency anatomical boundaries and tissue labels remain sharp and consistent in the one-step output, compensating for potential oversmoothing caused by distillation.

3 Experiments and Results

3.1 Datasets and Implementation

Dataset We employed longitudinal T1-weighted magnetic resonance imaging (MRI) data and corresponding tissue segmentation labels from the Infant Brain

Imaging Study (IBIS) dataset [14], a multi-center neurodevelopmental cohort focusing on infants at high likelihood for autism spectrum disorder. The segmentation labels were initially generated by iBEAT [16] and subsequently refined by three radiologists. The dataset comprises paired scans from the same subjects at 6 months and either 12 or 24 months of age, forming two distinct longitudinal trajectories (6→12 months and 6→24 months). Each subject contributes at least one paired time point, with some subjects contributing pairs for both trajectories. This results in a total of 452 MRI-segmentation pairs (242 for 6→12 months, 210 for 6→24 months) from 284 individual subjects. The dataset was split into training, validation, and test sets at a 75%, 5%, and 20% ratio, respectively.

Implementation details The FDRF model processed input 3D images and labels with a shape of $(160, 192, 160)$. It underwent 400 epochs of training, followed by 200 epochs of distillation. The training was performed on two NVIDIA RTX 3090 GPUs, utilizing a batch size of 1. The Adam optimizer was employed with a learning rate of 1×10^{-5}, and the loss function weights were set to $\lambda_1 = 0.7$ and $\lambda_2 = 0.2$. Notably, a unified model was trained for longitudinal predictions of growth brains at 12 and 24 months of age, in terms of 6-month inputs.

3.2 Main Results

Comparison Methods In this study, we compared two versions of our methods, FDRF1 and FDRF2, with several state-of-the-art techniques. Specifically, we included DDIM [15], which employs DDIM with a CFG [6] strategy using 50 sampling steps; RF (CFG), combining rectified flow [10] with CFG; RF (FFM), incorporating rectified flow with the feature fusion module (FFM) instead of CFG. FDRF1 builds on RF (FFM) by adding an edge loss function $\mathcal{L}_{\text{edge}}$ (Eq. 6) during the distillation phase, while FDRF2 further enhances FDRF1 with an additional segmentation loss function $\mathcal{L}_{\text{seg}}$ (Eq. 7) during the same phase. It is important to highlight that the only method marked with an asterisk (*) focuses exclusively on generating MRI images from MRI inputs, specifically RF (FFM)*, without incorporating segmentation masks. Conversely, the other methods integrate the joint generation of images and segmentation masks. The visual results of these methods are presented in Fig. 2, where the superior performance of FDRF2 can be clearly observed in terms of image quality and segmentation accuracy.

Image Prediction Results The performance of the proposed methods, RF (FFM), FDRF1, and FDRF2, was evaluated in terms of image quality at 12 and 24 months. As shown in Table 1, the results show that our methods consistently outperformed the other methods in terms of the **mean square error (MSE)** and the highest **peak signal-to-noise ratio (PSNR)** and the **structural similarity index measure (SSIM)**. For example, at the 12-month time point, FDRF2 achieved an MSE of 0.0023, a PSNR of 26.5034, and an SSIM of 0.9384,

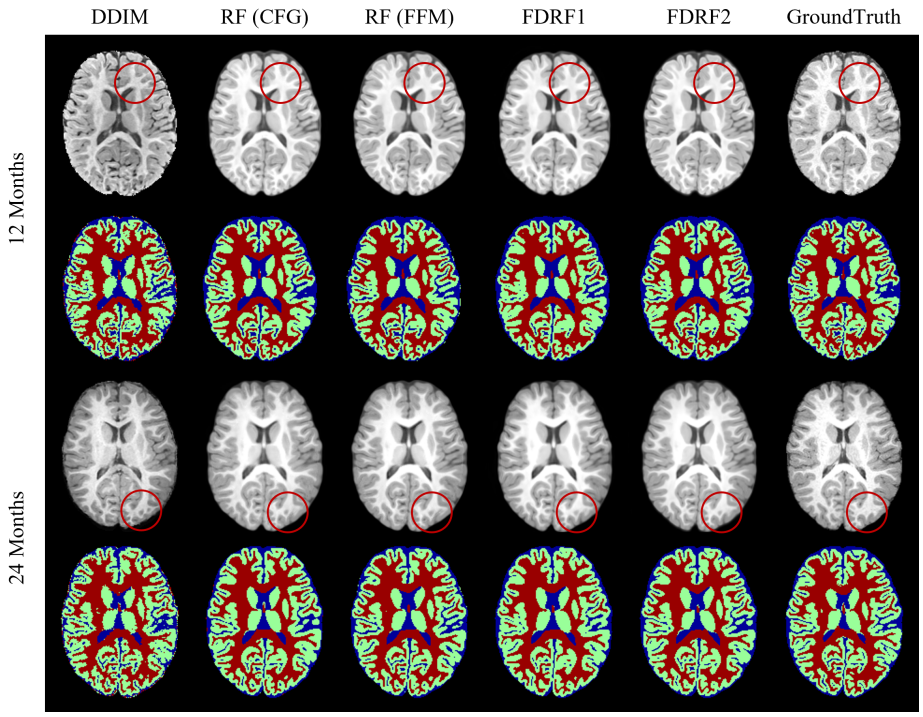

Fig. 2. The upper part shows the brain MRI image and tissue segmentation of IBIS1048 generated at 12 months, while the lower part corresponds to those at 24 months.

which are the lowest MSE and the highest values of PSNR and SSIM among all methods. Similarly, at the 24-month time point, FDRF2 achieved an MSE of 0.0027, a PSNR of 25.9495, and an SSIM of 0.9084, indicating superior performance in generating higher-quality MRI images with fewer artifacts and better structural preservation compared to the other methods. It should be noted that the performance of RF (FFM) exceeds that of RF (FFM)* in all key metrics. This observation underscores the importance of employing a joint image and segmentation generation architecture, which simultaneously produces MRI images and tissue segmentation labels. Such an architecture ensures better consistency and accuracy in anatomical details, thereby enhancing the overall quality of the generated images.

Beyond quality, our single-step sampling strategy offers **substantial computational efficiency**. Our models achieved an average inference time of just **0.08 seconds** per image in the 92-sample test set, a significant improvement over DDIM(CFG)'s 50-step sampling, which required **30.88 seconds**.

Tissue Prediction Results The tissue segmentation performance of the different methods at 12 and 24 months was evaluated using the Dice coefficient. As

Table 1. Comparison of Model Performance at 12 and 24 Months.

Methods	12 Months			24 Months			Sample Time↓
	MSE↓	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑	
DDIM [15]	0.0030	25.2921	0.9149	0.0034	24.8439	0.8871	30.88 s
RF (CFG) [10]	0.0024	26.3310	0.9290	0.0030	25.5375	0.8991	0.48 s
RF (FFM)*	0.0042	24.1587	0.9214	0.0051	23.6301	0.8890	0.08 s
RF (FFM) (Ours)	0.0024	26.3415	0.9306	0.0029	25.6403	0.9020	0.08 s
FDRF1 (Ours)	0.0023	26.4785	0.9332	0.0029	25.6181	0.9051	0.08 s
FDRF2 (Ours)	**0.0023**	**26.5034**	**0.9348**	**0.0027**	**25.9495**	**0.9084**	0.08 s

Table 2. Dice coefficients for tissue segmentation generated by various methods at 12 and 24 months.

Methods	12 Months			24 Months		
	GM	WM	Average	GM	WM	Average
DDIM [15]	0.7782	0.8077	0.7566	0.7447	0.7591	0.6966
RF (CFG) [10]	0.8128	0.8518	0.7938	0.7770	0.7969	0.7340
RF (FFM) (Ours)	0.8097	0.8599	0.7989	0.7725	0.8060	0.7370
FDRF1 (Ours)	0.8224	**0.8608**	0.8003	0.7836	0.8081	0.7405
FDRF2 (Ours)	**0.8223**	0.8596	**0.8003**	**0.7846**	**0.8080**	**0.7418**

shown in Table 2, our methods, RF (FFM), FDRF1 and FDRF2, demonstrated superior performance in segmenting grey matter (GM) and white matter (WM) tissues. This superior performance can be attributed to our joint training framework, which effectively generates both high-quality images and accurate labels, thereby providing robust anatomical supervision for image generation.

Ablation Analysis The experimental results highlight the effectiveness of our methodological innovations. The edge loss function $\mathcal{L}_{\text{edge}}$ (Eq. 6) and segmentation loss function $\mathcal{L}_{\text{seg}}$ (Eq. 7) during knowledge distillation enhance structural fidelity. Simultaneous MRI image and tissue segmentation label generation ensures anatomical consistency. FDRF2, integrating these innovations, demonstrates superior performance in preserving critical anatomical structures across developmental timepoints. This joint image-segmentation generation approach provides robust anatomical supervision, crucial for longitudinal infant brain development studies.

4 Conclusion

Our proposed Flexibly Distilled 3D Rectified Flow (FDRF) method effectively addresses the challenge of longitudinal MRI image synthesis for infant brains, achieving high-quality image generation while preserving anatomical structures.

The feature fusion module ensures robust performance across different developmental timepoints. The method's accuracy and robustness, demonstrated through the publicly available IBIS dataset, highlight its potential to advance neurodevelopmental image synthesis and clinical applications. Future work will focus on optimizing the FDRF pipeline and exploring its broader applications in medical imaging. Our code will be released publicly in the near future.

Acknowledgments. This work was supported in part by Science and Technology Innovation 2030-Major Projects (No. 2022ZD0209000), NSFC Grant (Nos. 12326616), Natural Science Basic Research Program of Shaanxi (No. 2024JC-TBZC09), and Shaanxi Provincial Key Industrial Innovation Chain Project (No. 2024SF-ZDCYL-02-10).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
2. Ecker, C., Suckling, J., Deoni, S.C., Lombardo, M.V., Bullmore, E.T., Baron-Cohen, S., Catani, M., Jezzard, P., Barnes, A., Bailey, A.J., et al.: Brain anatomy and its relationship to behavior in adults with autism spectrum disorder: a multicenter magnetic resonance imaging study. Archives of general psychiatry **69**(2), 195–209 (2012)
3. Gilmore, J.H., Knickmeyer, R.C., Gao, W.: Imaging structural and functional brain development in early childhood. Nature Reviews Neuroscience **19**(3), 123–137 (2018)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
5. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
6. Ho, J., Salimans, T.: Classifier-free diffusion guidance. arXiv preprint arXiv:2207.12598 (2022)
7. Johnson, M.H.: Functional brain development in humans. Nature Reviews Neuroscience **2**(7), 475–483 (2001)
8. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. arXiv preprint arXiv:2210.02747 (2022)
9. Liu, Q., Fuster-Garcia, E., Hovden, I.T., MacIntosh, B.J., Grødem, E.O., Brandal, P., Lopez-Mateu, C., Sederevičius, D., Skogen, K., Schellhorn, T., et al.: Treatment-aware diffusion probabilistic model for longitudinal mri generation and diffuse glioma growth prediction. IEEE Transactions on Medical Imaging (2025)
10. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. arXiv preprint arXiv:2209.03003 (2022)
11. Lord, C., Elsabbagh, M., Baird, G., Veenstra-Vanderweele, J.: Autism spectrum disorder. The lancet **392**(10146), 508–520 (2018)
12. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)

13. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2015)
14. Shi, F., Yap, P.T., Wu, G., Jia, H., Gilmore, J.H., Lin, W., Shen, D.: Infant brain atlases from neonates to 1-and 2-year-olds. PloS one **6**(4), e18746 (2011)
15. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
16. Wang, L., Wu, Z., Chen, L., Sun, Y., Lin, W., Li, G.: ibeat v2. 0: a multisite-applicable, deep learning-based pipeline for infant cerebral cortical surface reconstruction. Nature protocols **18**(5), 1488–1509 (2023)
17. Ye, M., Yang, D., Huang, Q., Kanski, M., Axel, L., Metaxas, D.N.: Sequencemorph: A unified unsupervised learning framework for motion tracking on cardiac image sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(8), 10409–10426 (2023)
18. Yoon, J.S., Zhang, C., Suk, H.I., Guo, J., Li, X.: Sadm: Sequence-aware diffusion model for longitudinal medical image generation. In: International Conference on Information Processing in Medical Imaging. pp. 388–400. Springer (2023)
19. Zhao, R., Qian, B., Zhang, X., Li, Y., Wei, R., Liu, Y., Pan, Y.: Rethinking dice loss for medical image segmentation. In: 2020 IEEE International Conference on Data Mining (ICDM). pp. 851–860. IEEE (2020)