Controllable Flow Matching for 3D Contrast-Enhanced Brain MRI Synthesis from Non-Contrast Scans

Heng Chang[1,2], Yu Shang[3], Haifeng Wang[1,2], Yuxia Liang[4], Haoyu Wang[2,5], Fan Wang[6], Chen Niu[7] (✉), and Chunfeng Lian[1,2(✉)]

[1] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China
chunfeng.lian@xjtu.edu.cn
[2] Research Center for Intelligent Medical Equipment and Devices (IMED), Xi'an Jiaotong University, Xi'an 710049, China
[3] School of Future Technology, Xi'an Jiaotong University, Xi'an, China
[4] Department of Health Medicine, First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi 710061 P.R. China
[5] School of Software Engineering, Xi'an Jiaotong University, Xi'an, China
[6] Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China
[7] Positron Emission Tomography/Computed Tomography Center, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi 710061 P.R. China
niuchen.xjtu@mail.xjtu.edu.cn

Abstract. Magnetic resonance imaging (MRI) enhanced by the gadolinium-based contrast agents (GBCAs) is crucial in the assessment and management of cancer. However, the use of GBCAs introduces additional costs and raises potential safety concerns, including the risk of gadolinium accumulation in brain. Several generative learning methods based on GANs and diffusion models have been proposed to generate contrast-enhanced MRI from non-contrast-enhanced MRI. However, GANs face challenges such as gradient vanishing and mode collapse. Diffusion models also face several challenges, such as generation instability and long sampling times. In this paper, we propose a controllable flow matching (CFM) model for efficient synthesis of 3D contrast-enhanced brain MRI with fine-grained details of targets of interests. CFM adopts a straight-line generation path, enabling the generation of images in a single step. We design a multi-stage training strategy integrating controllable constraints to ensure that such a single-step sampling generating contrast-enhanced MRI meet specific controllable conditions. Our CFM model has been evaluated on both the BraTS2023 and an in-house datasets. Experimental results demonstrate that CFM led to state-of-the-art image generation and tumor delineation performance with promising generalizability. Our codes can be found at https://github.com/ladderlab-xjtu/CFM.

Keywords: Contrast-Enhanced Brain MRI Synthesis · Flow Matching · One-Step Generation · Controllable Generation.

1 Introduction

Magnetic resonance imaging (MRI) enhanced by the gadolinium-based contrast agents (GBCAs) is crucial for cancer evaluation and management [14]. For instance, contrast-enhanced T1-weighted (T1ce) MRI provides sharp tumor boundary of glioma and clearly differentiates active tumor regions from necrotic tissue. However, GBCAs injections, especially when with multiple repetitions, increase scanning costs and raise safety concerns, such as gadolinium depositions in the brain [3]. To maintain clinical benefits while minimizing risks, developing alternative techniques for achieving similar imaging enhancement with reduced or no GBCAs is of significant practical value.

To this end, various generative learning methods have been proposed for synthesizing contrast-enhanced MRI from non-contrast images. While generative adversarial networks (GANs) [6,20] face challenges like gradient vanishing and mode collapse, recent research has shifted toward diffusion models due to their effectiveness [5,16]. For instance, cross-conditioned diffusion models [18] enable single-step generation, and adaptive latent diffusion models (LDMs) [9] facilitate image transformation in latent space. Controllable frameworks like ControlNet [19] and ControlNet++ [10] have also been explored for targeted generation. However, diffusion models struggle with long sampling and inference times as well as associated error accumulation in 3D scan synthesis with fine-grained details. As an efficient and flexible alternative, flow matching (FM) models [11,12], which leverage ordinary differential equations (ODEs) for deterministic cross-distribution mapping, offer stable training, fewer sampling steps, and one-step generation. Despite their success in natural image generation, FM models still face limitations in 3D brain MRI synthesis, particularly in image detail and training efficiency.

In this paper, we propose a **controllable flow matching (CFM)** model for high-fidelity synthesis of 3D contrast-enhanced brain MRI, with a focus on detailed clinical targets such as glioma lesions. Our work makes three key contributions:

- Our CFM features a straight-line path model, enabling efficient one-step sampling for 3D medical image translation.
- Our CFM can flexibly integrate customized constraints, such as precise 3D tumor delineation, into training to ensure controllable translation that focuses on clinical targets.
- A multi-stage training strategy enhances the efficiency of CFM, achieving lesion-aware, controllable image generation.

2 Preliminaries

2.1 Flow Matching

Given two probability distributions $x_0 \sim p_0$ and $x_1 \sim p_1$,with their corresponding samples Z_0 and Z_1, FM constructs a mapping between two probability

distributions through an ODE-based continuous motion system:

$$\frac{\mathrm{d}}{\mathrm{d}t}Z_t = u(Z_t, t), \quad \forall t \in [0, 1], \tag{1}$$

where Z_t is a point on the path from Z_0 to Z_1, corresponding to time t, and u is the velocity vector field at the point Z_t. As t progresses from 0 to 1, Z_0 moves along the velocity vector field u calculated based on Eq.(1), reaching Z_1. The optimization objective in FM is defined as:

$$\min \int_0^1 E_{Z_0 \sim p_0, Z_1 \sim p_1} \left[\|u(Z_t) - v(Z_t, t)\|^2 \right] \mathrm{d}t, \tag{2}$$

where $v(\cdot, \cdot)$ is a velocity prediction neural network. Based on Eq.(2), FM optimizes the v using paired sampled data from p_0 and p_1. Detailed formulas and proofs can be found in relevant flow matching papers[11,12].

2.2 Straight-Line Path

In FM, the straight-line path is considered as the optimal transport path, based on which we formulate the mapping from the source domain to the target domain. Suppose we have samples $X_0 \sim \pi_0$ and $X_1 \sim \pi_1$ from two distributions(e.g., T1W and T1ce MRI, respectively), then the ODE expression for the straight-line path is given by:

$$\frac{\mathrm{d}}{\mathrm{d}t}X_t = X_1 - X_0, \forall t \in [0, 1], \text{ where } X_t = tX_1 + (1 - t)X_0. \tag{3}$$

The FM with a straight-line path is illustrated in Fig. 1.(A). According to Eq.(2), the FM optimization objective can be specified as:

$$\min \int_0^1 E_{X_0 \sim \pi_0, X_1 \sim \pi_1} \left[\|(X_1 - X_0) - v(X_t, t)\|^2 \right] \mathrm{d}t. \tag{4}$$

In the actual training process of this paper, the loss function for predicting velocity can be set as:

$$\mathcal{L}_v = \mathcal{L}_{\text{mse}}(X_1 - X_0, v_\theta(tX_1 + (1 - t)X_0, t)), \text{ with } t \sim \text{Uniform}([0, 1]), \tag{5}$$

where $v_\theta(\cdot, \cdot)$ is a velocity prediction network, parameterized by θ. $X_1 - X_0$ represents the theoretically constant velocity v, which serves as the optimization target for $v_\theta(\cdot, \cdot)$. During training, X_0 and X_1 are paired images from two different modalities, and t is sampled uniformly between $[0, 1]$ for each training step.

Since the path is a straight line, the velocity does not change with t. During the sampling phase for inference, we can directly predict X_1 from X_0. Such an efficient one-step generation is as:

$$X_1 = X_0 + v_\theta(X_0, 0). \tag{6}$$

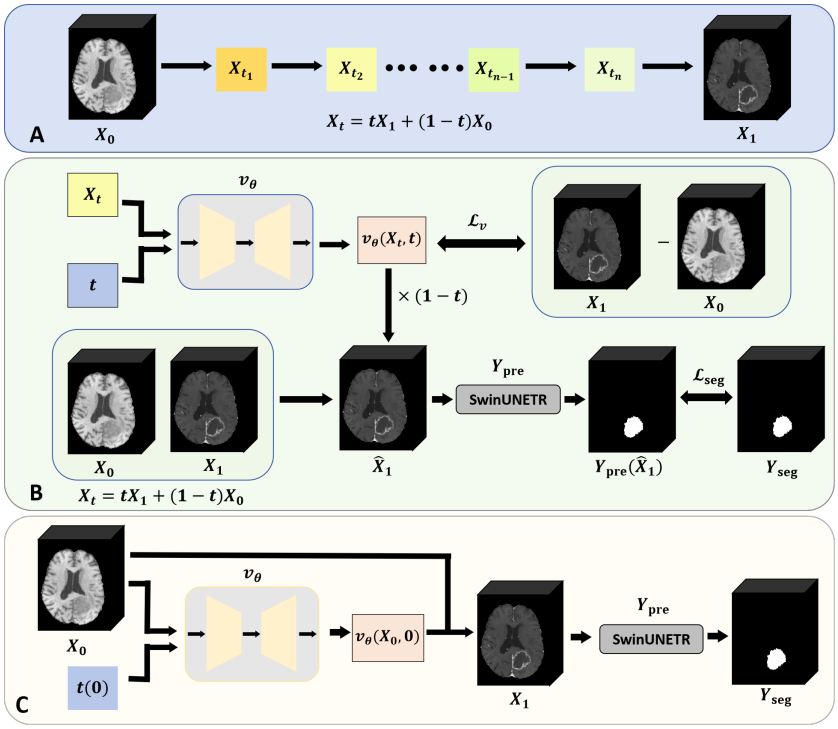

Fig. 1. (A) illustrates the generation process of flow matching along a straight-line path. (B) illustrates how the velocity loss and segmentation loss are calculated during the training phase. These two losses are used to optimize the CFM model. (C) illustrates the sampling phase, where images are generated in one step. Finally, segmentation labels are obtained from the generated images, achieving controllable generation.

3 Controllable Flow Matching

3.1 Controllable Generation

In this paper, our CFM extends the original FM for 3D T1w-to-T1ce brain MRI translation, by leveraging the accurate segmentation of tumor lesions as the controllable constraints to enhance the fine-grained glioma details in the generated T1ce images.

Specifically, in the training phase as shown in Fig. 1.(B), we generate a synthetic image $\hat{X}_1$ via a one-step sampling in terms of the image X_t and the predicted velocity $v_\theta(X_t, t)$, since the velocity does not change with t:

$$X_1 \approx \hat{X}_1 = X_t + (1-t)v_\theta(X_t, t). \tag{7}$$

For controllable generation of T1ce images with fine-grained details of glioma, we integrate a supplementary tumor segmentation task into the training procedure. It provides key constraints to enhance the synthetic quality, considering that

the high-fidelity imaging of glioma details is the precondition for its delineation. Specifically, a fundamental Dice segmentation loss is employed, such as:

$$\mathcal{L}_{\text{seg}} = \mathcal{L}_{\text{dice}}(Y_{\text{pre}}(\hat{X}_1), Y_{\text{seg}}), \text{ with } t \sim \text{Uniform}([0,1]), \tag{8}$$

where $Y_{\text{pre}(\cdot)}$ is a the pre-trained segmentation model, and Y_{seg} represents the ground truth segmentation labels.

After training, we adopt a one-step generation strategy along the straight-line path to get the synthetic T1ce image by our CFM, as shown in Eq.(6). Finally, the generated medical images can be passed through the pre-trained segmentation model to obtain the annotations of the clinical target. The entire sampling process is illustrated in Fig. 1.(C), where the segmentation masks are obtained from the generated T1ce images.

3.2 Multi-Stage Training

The details of the multi-stage training algorithm for our proposed CFM are shown in Algorithm 1, where T_{seg} and T_{distill} are measured in epochs. *In the first stage*, only the velocity loss is added to ensure the generation of approximate target-domain images. *In the second stage*, the segmentation loss is included and combined with the velocity loss, jointly optimizing v_θ to ensure controllable generation. *In the third stage*, we set the constant to 0, which is referred to as the distillation phase [12]. By focusing on training the model at critical moments, this stage is critical in enhancing the synthetic details by one-step sampling generation in inference.

Algorithm 1 Controllable Flow Matching

Input: Training Data $(X_0, X_1, Y_{\text{seg}})$, T_{seg}, T_{distill}.
Stage 1: $T < T_{\text{seg}} < T_{\text{distill}}$.
  Training: $\mathcal{L}_v = \mathcal{L}_{\text{mse}}(X_1 - X_0, v_\theta(tX_1 + (1-t)X_0, t))$, with $t \sim \text{Uniform}([0,1])$.
Stage 2: $T_{\text{seg}} < T < T_{\text{distill}}$.
  Training: $\mathcal{L} = \mathcal{L}_v + \mathcal{L}_{\text{seg}}$
  $\mathcal{L}_v = \mathcal{L}_{\text{mse}}(X_1 - X_0, v_\theta(tX_1 + (1-t)X_0, t))$, with $t \sim \text{Uniform}([0,1])$.
  $\mathcal{L}_{\text{seg}} = \mathcal{L}_{\text{dice}}(Y_{\text{pre}}(X_t + (1-t)v_\theta(X_t, t)), Y_{\text{seg}})$, where $X_t = tX_1 + (1-t)X_0$.
Stage 3: $T_{\text{distill}} < T$.
  Training: $\mathcal{L} = \mathcal{L}_v + \mathcal{L}_{\text{seg}}$
  $\mathcal{L}_v = \mathcal{L}_{\text{mse}}(X_1 - X_0, v_\theta(X_0, 0))$, with $t = 0$.
  $\mathcal{L}_{\text{seg}} = \mathcal{L}_{\text{dice}}(Y_{\text{pre}}(X_0 + v_\theta(X_0, 0)), Y_{\text{seg}})$.

4 Experiments

4.1 Datasets and Implementation

BraTS2023 dataset This publicly accessible dataset [1,2,7,8,13] contains 1251 3D brain MRI images acquired from glioma patients. Each subject includes

four types of MRI sequences: T1w, T1ce, T2w, and FLAIR, with ground-truth annotations of tumor lesions. In our experiments, we selected the T1w and T1ce MRIs, using T1w as the source domain and T1ce as the target domain. We used the segmentation images of enhancing tumor (ET) and whole tumor (TC: enhancing tumor plus necrotic tumor) as the controllable constraints. The data was splitted into 3:1:1 for training/validation/testing.

In-house dataset The in-house dataset in this study consists of 73 paired samples, each containing T1w and T1ce images. This dataset was not used during the training phase but was exclusively used for testing the model. It differs in equipment and scanning methods from BraTS2023, making it valuable for assessing the model's generalization ability.

Implementation details In the CFM model, the velocity prediction network is a U-Net [15] with an attention module [17] added to the deepest layer, and the pre-trained segmentation model used is SwinUNETR [4]. The parameters of the SwinUNETR model are frozen during the training phase of the generation model and are utilized during the testing phase to evaluate the segmentation performance on the generated images. In the training process, the parameters were set as $T_{\text{seg}} = 20$ epochs and $T_{\text{distill}} = 150$ epochs. For 3D medical image data, the input size is $160 \times 192 \times 96$, ensuring that the input image includes the majority of the brain and tumor regions. The model was trained on one NVIDIA A6000 GPU with a unit batch size of 1, using the Adam optimizer with a learning rate of 2e-5.

4.2 Results

Our CFM was comparted with two GAN-based generative models, i.e., Pix2Pix [6] and CycleGAN [20], a transformer-based model SwinUNETR [4], a diffusion-based models DDIM [16] and a controllable diffusion model ControlNet++ [10] and the original FM model [11,12], which follows a straight-line path without controllable constraints. The FM model also adopts the distillation strategy outlined in Algorithm 1. These competing methods were reimplemented for 3D image synthesis following the official source codes. All experimental environments were kept the same for fair comparisons.

Generation Results The quantitative metrics for evaluating the generation results are Structural Similarity Index (SSIM%), Peak Signal-to-Noise Ratio (PSNR) and Mean Absolute Error (MAE). Average inference time per instance(second) is employed as a metric for assessing the model's inference speed. As shown in Table 1, our CFM model achieves optimal generative metrics on both the BraTS2023 and in-house datasets. In the inference phase, to achieve better generation results, we set the sampling steps of DDIM and ControlNet++ to 10. As a one-step generation model, the inference time of CFM is significantly lower

Table 1. Quantitative comparison on the BraTS2023 and In-house datasets

Datasets	Methods	SSIM↑	PSNR↑	MAE↓	Inference Time↓
BraTS2023	Pix2Pix[6]	85.04	28.5777	0.0298	0.1997
	CycleGAN[20]	85.78	28.6067	0.0290	0.2024
	SwinUNETR[4]	84.86	28.0441	0.0322	0.2034
	DDIM[16]	86.49	29.0958	0.0322	3.7511
	Controlnet++[10]	85.41	28.4801	0.0303	3.7427
	FM[11,12]	87.60	29.5386	0.0272	0.1421
	Ours	**87.65**	**29.5674**	**0.0271**	**0.1421**
In-house	Pix2Pix[6]	81.23	27.7121	0.0358	0.2351
	CycleGAN[20]	82.54	27.8499	0.0352	0.2471
	SwinUNETR[4]	83.46	27.9946	0.0346	0.2217
	DDIM[16]	81.14	27.5210	0.0358	4.3477
	Controlnet++[10]	81.37	27.0894	0.0392	4.5413
	FM[11,12]	83.82	27.9231	0.0343	0.1742
	Ours	**84.17**	**28.1270**	**0.0333**	**0.1742**

Table 2. Quantitative Comparison of Segmentation Tasks in terms of Dice Score on BraTS2023

Methods	TC↑	ET↑	Avg↑
Pix2Pix[6]	0.4360	0.2317	0.3339
CycleGAN[20]	0.3796	0.1813	0.2805
SwinUNETR[4]	0.2759	0.1028	0.1893
DDIM[16]	0.6171	0.4330	0.5251
Controlnet++[10]	0.6636	0.4696	0.5667
FM[11,12]	0.6166	0.4409	0.5288
Ours	**0.7115**	**0.5415**	**0.6265**
T1_GT	0.5309	0.3934	0.4621
T1c_GT	0.9056	0.8788	0.8922

than that of DDIM and ControlNet++, which require multiple sampling steps. Our CFM model shows significantly better generative metrics on the in-house dataset compared to other models, indicating that CFM, through controllable generation, focuses on important generative regions (tumor delineation) in the external validation dataset, thereby demonstrating strong generalization ability.

Segmentation Results The segmentation results are evaluated using the Dice score. Table 2 presents the segmentation performance of generated T1c images using a pre-trained segmentation model. Our CFM model achieves optimal performance in both TC and ET segmentation, demonstrating the superiority of controllable generation. Although ControlNet++ also integrates segmentation loss during training as a controllable generation model, the larger error in its

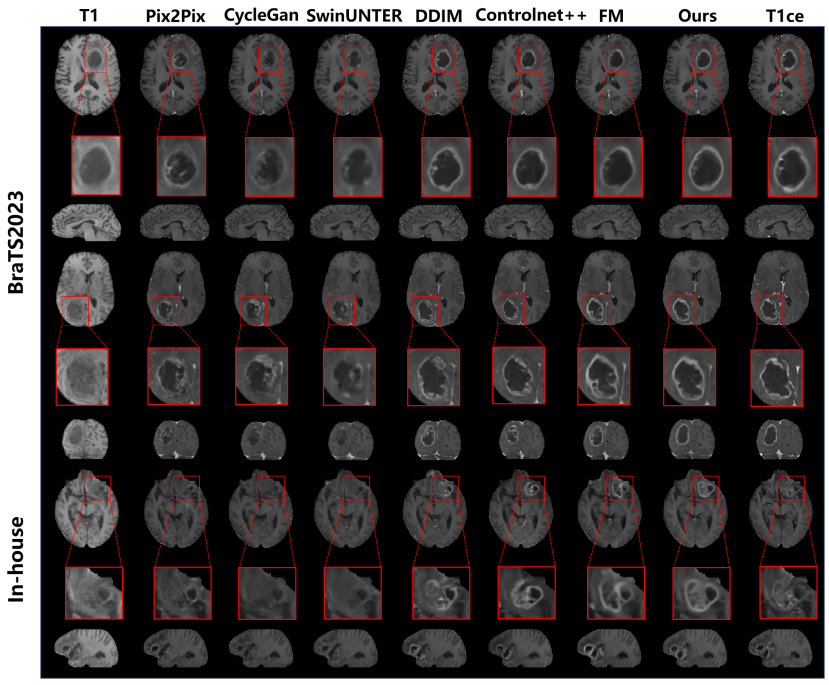

Fig. 2. Visualization of generated 3D T1ce images: The first six rows represent results from the BraTS2023 dataset, and the last three rows represent results from the in-house dataset

single-step backpropagation leads to suboptimal segmentation results compared to CFM.

Qualitative results As shown in Fig. 2, the T1ce images generated by CFM with controllable constraints focus more on clinical targets, specifically the clear distinction between active tumor regions and necrotic tissue. As shown in the visualization results, diffusion models produce noisier outputs due to error accumulation from multiple sampling steps. Fig. 3 demonstrates that the T1ce images generated by CFM from T1 can produce more detailed segmentation labels when processed through a pre-trained segmentation model, achieving the clinical objective, without the GBCAs.

5 Conclusions

In this paper, we propose a controllable flow matching (CFM) method for synthesizing 3D contrast-enhanced brain MRI. By setting the FM generation path as a straight line, CFM enables one-step image generation. The model incorporates controllable constraints for targeted clinical generation, specifically

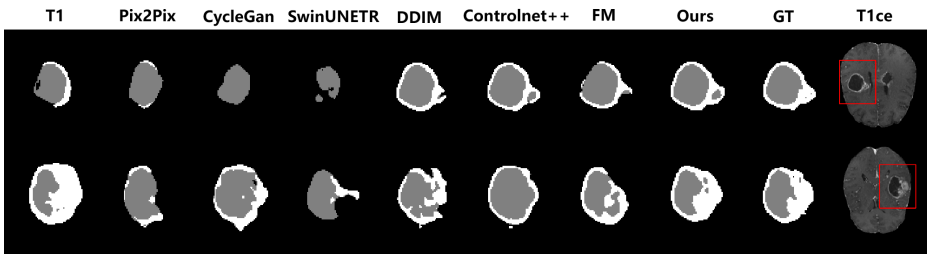

Fig. 3. Visualization of generated images for segmentation

generating T1ce images from T1w, making tumor regions easier to segment without additional GBCAs. Future extensions could include more controllable constraints or cyclic consistency to improve performance, showing CFM's strong scalability for various controllable generation tasks.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (No. 12326616), the Natural Science Basic Research Program of Shaanxi, China (No. 2024JC-TBZC-09), Shaanxi Provincial Key Industrial Innovation Chain Project (No. 2024SF-ZDCYL-02-10), and Tianyuan Fund for Math.ematics of the National Natural Science Foundation of China (No. 12426105).

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

References

1. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F.C., Pati, S., et al.: The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314 (2021)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Scientific data **4**(1), 1–13 (2017)
3. Gulani, V., Calamante, F., Shellock, F.G., Kanal, E., Reeder, S.B.: Gadolinium deposition in the brain: summary of evidence and recommendations. The Lancet Neurology **16**(7), 564–570 (2017)
4. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI brainlesion workshop. pp. 272–284. Springer (2021)
5. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
6. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
7. Karargyris, A., Umeton, R., Sheller, M.J., Aristizabal, A., George, J., Wuest, A., Pati, S., Kassem, H., Zenk, M., Baid, U., et al.: Federated benchmarking of medical artificial intelligence with medperf. Nature machine intelligence **5**(7), 799–810 (2023)

8. Kazerooni, A.F., Khalili, N., Liu, X., Haldar, D., Jiang, Z., Anwar, S.M., Albrecht, J., Adewole, M., Anazodo, U., Anderson, H., et al.: The brain tumor segmentation (brats) challenge 2023: focus on pediatrics (cbtn-connect-dipgr-asnr-miccai brats-peds). ArXiv pp. arXiv–2305 (2024)
9. Kim, J., Park, H.: Adaptive latent diffusion model for 3d medical image to image translation: Multi-modal magnetic resonance imaging study. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7604–7613 (2024)
10. Li, M., Yang, T., Kuang, H., Wu, J., Wang, Z., Xiao, X., Chen, C.: Controlnet++: Improving conditional controls with efficient consistency feedback: Project page: liming-ai. github. io/controlnet_plus_plus. In: European Conference on Computer Vision. pp. 129–147. Springer (2024)
11. Lipman, Y., Chen, R.T.Q., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023. OpenReview.net (2023)
12. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. In: The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023. OpenReview.net (2023)
13. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
14. Preetha, C.J., Meredig, H., Brugnara, G., Mahmutoglu, M.A., Foltyn, M., Isensee, F., Kessler, T., Pflüger, I., Schell, M., Neuberger, U., et al.: Deep-learning-based synthesis of post-contrast t1-weighted mri for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study. The Lancet Digital Health **3**(12), e784–e794 (2021)
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
16. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021. OpenReview.net (2021)
17. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
18. Xing, Z., Yang, S., Chen, S., Ye, T., Yang, Y., Qin, J., Zhu, L.: Cross-conditioned diffusion model for medical image to image translation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 201–211. Springer (2024)
19. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3836–3847 (2023)
20. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)