D^{3}M: Deformation-Driven Diffusion Model for Synthesis of Contrast-Enhanced MRI with Brain Tumors

Haowen Pang[1], Peng Zhang[1], Xiaoming Hong[1], Shannan Chen[2,3], and Chuyang Ye[1(✉)]

[1] School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing, China
chuyang.ye@bit.edu.cn

[2] College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China

[3] Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Northeastern University, Shenyang, China

Abstract. *Contrast-enhanced magnetic resonance images* (CEMRIs) provide valuable information for brain tumor diagnosis and treatment planning. However, CEMRI acquisition requires contrast agent injection, which poses problems such as health risks, high costs, and environmental concerns. To address these drawbacks, researchers have synthesized CEMRIs from *non-contrast magnetic resonance images* (NCMRIs) to remove the need for contrast agents. However, CEMRI synthesis from NCMRIs is highly ill-posed, where false positive and false negative enhancement can be produced, especially for brain tumors. In this study, we propose a *deformation-driven diffusion model* (D^{3}M) for CEMRI synthesis with brain tumors from NCMRIs. Instead of modeling enhancement errors as intensity errors, we formulate them as incorrect interpretation of tumor subcomponents, where enhanced tumors are misinterpreted as non-enhanced tumors and vice versa. In this way, the enhancement can be geometrically corrected with spatial deformation. This reduces the difficulty of CEMRI synthesis, as the intensity error is usually large to correct whereas the geometry correction is relatively small. Specifically, we first introduce a *multi-step spatial deformation module* (MSSDM) in D^{3}M. MSSDM performs image deformation to adjust the enhancement, displacing enhanced regions to remove false positive and false negative enhancement. Moreover, as the denoising process of diffusion models is stepwise, MSSDM is applied at these multiple diffusion steps. Second, to further guide the spatial deformation, we incorporate an auxiliary task of segmenting the enhanced tumor, which aids the model understanding of contrast enhancement. Accordingly, we introduce a *dual-stream image-mask decoder* (DSIMD) that jointly produces intermediate enhanced images and masks of enhanced tumors. Results on two public datasets demonstrate that D^{3}M outperforms existing methods in CEMRI synthesis.

Chuyang Ye is the corresponding author. Email: chuyang.ye@bit.edu.cn.

Keywords: Image synthesis · Diffusion model · Contrast-enhanced MRI.

1 Introduction

Contrast-enhanced magnetic resonance images (CEMRIs) provide valuable information for brain tumor diagnosis and treatment planning [5,22]. However, CEMRI acquisition requires injection of contrast agents, which brings concerns about patient health risks [15,25], high economic costs [27], and environmental harm [2,8]. Given these concerns, removal of the need for contrast agents while preserving image contrast for disease assessment is highly desirable.

To address the problem above, researchers have synthesized CEMRIs from *non-contrast magnetic resonance images* (NCMRIs), such as T1-weighted, T2-weighted, and FLAIR images [6,17,23,26,27]. For example, Preetha et al. [17] propose a 3D *convolutional neural network* (CNN) based on U-Net and a conditional *generative adversarial network* (GAN) method inspired by Pix2Pix [10] to synthesis CEMRIs from NCMRIs. Gui et al. [6] propose a conditional autoregressive vision model for the synthesis problem. In recent years, diffusion models have achieved great success in medical image synthesis with realistic synthesis results [16,24,26,29]. State-of-the-art diffusion models [14,18] can be directly applied to CEMRI synthesis, and they can also be adapted specifically for CEMRI synthesis. For example, Xu et al. [26] propose a diffusion model based on common-unique decomposition for liver CEMRI synthesis, and the method outperforms other non-diffusion-based synthesis models.

Compared with other image synthesis tasks, CEMRI synthesis from NCMRIs remains highly ill-posed as NCMRIs only provide ambiguous evidence about the enhanced regions [6]. Consequently, existing methods, including those based on diffusion models, still produce noticeable false positive and false negative enhancement results, where high intensities are produced for non-enhanced regions and low intensities are produced for regions that should be enhanced, respectively. The erroneous enhancement is particularly severe for tumor regions, where these models fail to capture the intricate morphology of tumor subcomponents.

In this study, we propose a *deformation-driven diffusion model* (D^3M) for improved CEMRI synthesis with brain tumors from NCMRIs. Instead of modeling enhancement errors as intensity errors, we formulate them as incorrect interpretation of tumor subcomponents, where enhanced regions are misinterpreted as non-enhanced ones and vice versa. Such reformulation allows the enhancement to be geometrically corrected via spatial deformation. This geometric perspective reduces the difficulty of CEMRI synthesis, as intensity errors for enhancement are typically large and challenging to correct, whereas geometric correction is relatively small and more manageable. Specifically, in D^3M we first introduce a *multi-step spatial deformation module* (MSSDM). MSSDM deforms the synthesized image to adjust the enhancement, displacing enhanced regions to correct false positive and false negative enhancement. Unlike traditional post-processing deformation methods, MSSDM is directly embedded within the denoising process and alternates with image generation at different diffusion steps. This tight

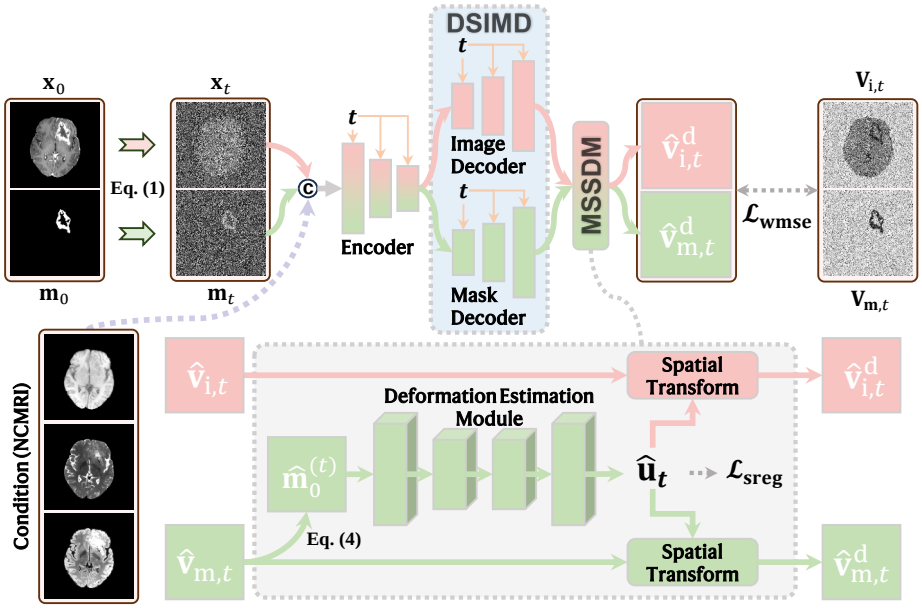

Fig. 1. An overview of the network architecture of D^{3}M.

integration not only prevents severe error accumulation that is difficult to correct, but also promotes the joint optimization of image generation and geometric correction. Second, to further guide the spatial deformation, we incorporate an auxiliary task of segmenting the enhanced part of the tumor, which improves the model understanding of contrast enhancement. To obtain the segmentation, we introduce a *dual-stream image-mask decoder* (DSIMD) in D^{3}M, which jointly produces intermediate enhanced images and masks of enhanced tumors. These masks are used by MSSDM for spatial deformation. D^{3}M is trained by minimizing a standard weighted mean squared error loss with deformation smoothness regularization. Experiments were performed on two public datasets for validation, and the results demonstrate that our method outperforms existing approaches, especially for brain tumor regions. The codes of our method are available at https://github.com/PangHaowen-hub/D3M.

2 Methods

2.1 Problem Formulation and Method Overview

We aim to train a synthesis model to synthesize CEMRIs (contrast-enhanced T1-weighted images) with brain tumors from NCMRIs, including commonly used T1-weighted, T2-weighted, and FLAIR images. Moreover, for training data we use auxiliary information of annotations of enhanced tumors to guide the model

training. Once training is completed, only the NCMRIs are needed for inference, without requiring enhanced tumor masks.

As the intensity errors of false positive and false negative enhancement can be large, we rethink them as misinterpretation of enhanced tumors, so that the enhancement can be corrected geometrically with spatial deformation. The geometric correction allows more effective suppression of enhancement errors, as the required displacement is usually small and easier to estimate. Based on this formulation, we develop D^{3}M, a diffusion model driven by spatial deformation, for CEMRI synthesis with brain tumors. An overview of D^{3}M is shown in Fig. 1, where we propose two major components, MSSDM and DSIMD. MSSDM is responsible for performing the deformation that corrects enhancement at each diffusion step, whereas DSIMD produces enhanced images and masks of enhanced tumors that guide MSSDM. The detailed designs are described below.

2.2 Training and Inference Procedures of D^{3}M

Like typical diffusion models [29], D^{3}M processes 2D slices and concatenates them into 3D volumes of CEMRIs. The geometric correction is integrated in D^{3}M for both training and inference procedures. The training stage learns the denoising process and geometric correction, and the inference stage uses multiple steps to gradually synthesize the CEMRI with the help of geometric correction.

Training. The training procedure consists of both forward and reverse processes. In the forward process, the noisy CEMRI $\mathbf{x}_t$ and noisy enhanced tumor mask $\mathbf{m}_t$ at the t-th step are given by:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \cdot \mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t} \cdot \boldsymbol{\epsilon}_{\mathrm{i},t} \quad \text{and} \quad \mathbf{m}_t = \sqrt{\bar{\alpha}_t} \cdot \mathbf{m}_0 + \sqrt{1-\bar{\alpha}_t} \cdot \boldsymbol{\epsilon}_{\mathrm{m},t}, \tag{1}$$

where $\mathbf{x}_0$ and $\mathbf{m}_0$ are the original CEMRI and mask, respectively, $\boldsymbol{\epsilon}_{\mathrm{i},t}$ and $\boldsymbol{\epsilon}_{\mathrm{m},t}$ are independent images of zero-mean, unit-variance Gaussian noise, and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$ with α_s representing the image noise schedule parameters following [3].

In the reverse process, at the t-th step, instead of directly estimating the noise or denoised image, D^{3}M follows [19] and learns to estimate the image velocity term $\mathbf{v}_{\mathrm{i},t}$ and mask velocity term $\mathbf{v}_{\mathrm{m},t}$, which are defined as

$$\mathbf{v}_{\mathrm{i},t} = \sqrt{\bar{\alpha}_t} \cdot \boldsymbol{\epsilon}_{\mathrm{i},t} - \sqrt{1-\bar{\alpha}_t} \cdot \mathbf{x}_0 \quad \text{and} \quad \mathbf{v}_{\mathrm{m},t} = \sqrt{\bar{\alpha}_t} \cdot \boldsymbol{\epsilon}_{\mathrm{m},t} - \sqrt{1-\bar{\alpha}_t} \cdot \mathbf{m}_0. \tag{2}$$

The estimation is achieved with an encoder $E(\cdot)$ and DSIMD $D(\cdot)$:

$$\hat{\mathbf{v}}_{\mathrm{i},t}, \hat{\mathbf{v}}_{\mathrm{m},t} = D(E(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}, t), t), \tag{3}$$

where $\hat{\mathbf{v}}_{\mathrm{i},t}$ and $\hat{\mathbf{v}}_{\mathrm{m},t}$ are the estimated image and mask velocity, respectively, and $\mathbf{c}$ is the conditional images (NCMRIs). Compared to direct estimation of the noise or denoised image, estimating the velocity terms enables the model to learn a more efficient and stable reverse diffusion process, thereby improving both the performance and stability of the model [19]. Note that $E(\cdot)$ and $D(\cdot)$ are

shared among diffusion steps, and their detailed design will be presented later in Section 2.3; also, mask velocity is estimated in addition to image velocity, as it is incorporated to guide model training.

Like standard diffusion models, $\hat{\mathbf{v}}_{\mathrm{i},t}$ and $\hat{\mathbf{v}}_{\mathrm{m},t}$ contain false positive and false negative enhancement. Thus, we propose to further incorporate geometric correction applied to $\hat{\mathbf{v}}_{\mathrm{i},t}$ and $\hat{\mathbf{v}}_{\mathrm{m},t}$ in the reverse process. To this end, we design MSSDM to estimate the deformation field that displaces the enhanced regions for correction. As the enhanced tumor mask is less sensitive to noise than the velocity terms and intensity images, MSSDM takes the current intermediate segmentation mask $\hat{\mathbf{m}}_0^{(t)}$ of the enhanced tumor to estimate the deformation field $\hat{\mathbf{u}}_t = U(\hat{\mathbf{m}}_0^{(t)})$, where $U(\cdot)$ represents the deformation estimation module in MSSDM, and $\hat{\mathbf{m}}_0$ is computed based on Eqs. (1) and (2) as

$$\hat{\mathbf{m}}_0^{(t)} = \sqrt{\bar{\alpha}_t} \cdot \mathbf{m}_t - \sqrt{1-\bar{\alpha}_t} \cdot \hat{\mathbf{v}}_{\mathrm{m},t}. \tag{4}$$

MSSDM then applies $\hat{\mathbf{u}}_t$ to $\hat{\mathbf{v}}_{\mathrm{i},t}$ and $\hat{\mathbf{v}}_{\mathrm{m},t}$ to correct erroneous enhancement and produces deformed image and mask velocity terms $\hat{\mathbf{v}}_{\mathrm{i},t}^{\mathrm{d}}$ and $\hat{\mathbf{v}}_{\mathrm{m},t}^{\mathrm{d}}$, respectively. Mathematically, for each voxel p in $\hat{\mathbf{v}}_{\mathrm{i},t}^{\mathrm{d}}$, we find its corresponding location $p' = p + \hat{\mathbf{u}}_t(p)$ in $\hat{\mathbf{v}}_{\mathrm{i},t}$ and have $\hat{\mathbf{v}}_{\mathrm{i},t}^{\mathrm{d}}(p) = \hat{\mathbf{v}}_{\mathrm{i},t}(p')$. Because image values are defined at integer locations, we perform linear interpolation to obtain $\hat{\mathbf{v}}_{\mathrm{i},t}^{\mathrm{d}}(p)$:

$$\hat{\mathbf{v}}_{\mathrm{i},t}^{\mathrm{d}}(p) = \phi_{\hat{\mathbf{u}}_t}(\hat{\mathbf{v}}_{\mathrm{i},t}(p)) = \hat{\mathbf{v}}_{\mathrm{i},t}(p+\hat{\mathbf{u}}_t(p)) = \sum_{q \in \mathcal{N}(p')} \hat{\mathbf{v}}_{\mathrm{i},t}(q) \prod_{d \in \{x,y\}} (1-|p'_d - q_d|), \tag{5}$$

where $\phi_{\hat{\mathbf{u}}_t}(\cdot)$ denotes the deformation based on $\hat{\mathbf{u}}_t$, $\mathcal{N}(p')$ represents the neighbor voxels of p', and d iterates over the spatial dimensions x and y. $\hat{\mathbf{v}}_{\mathrm{m},t}^{\mathrm{d}}$ is obtained in the same way. The detailed design of MSSDM is presented in Section 2.3.

We train D^{3}M by minimizing the sum of a standard weighted mean squared error loss $\mathcal{L}_{\mathrm{wmse}}$ [19] of the synthesis result at each step and deformation smoothness regularization $\mathcal{L}_{\mathrm{sreg}}$ [1] applied to the deformation field at each step.

Inference. The trained D^{3}M is used for CEMRI synthesis via the reverse process. The synthesis begins with two independent zero-mean, unit-variance Gaussian noise images. At the t-th step of the reverse process, the estimates $\hat{\mathbf{x}}_t$ and $\hat{\mathbf{m}}_t$ of $\mathbf{x}_t$ and $\mathbf{m}_t$ respectively from step $t+1$ are fed into the network along with the NCMRIs and the time step to produce $\hat{\mathbf{v}}_{\mathrm{i},t}$ and $\hat{\mathbf{v}}_{\mathrm{m},t}$. Next, the intermediate estimate $\hat{\mathbf{m}}_0^{(t)}$ of the enhanced tumor mask is computed as $\hat{\mathbf{m}}_0^{(t)} = \sqrt{\bar{\alpha}_t} \cdot \hat{\mathbf{m}}_t - \sqrt{1-\bar{\alpha}_t} \cdot \hat{\mathbf{v}}_{\mathrm{m},t}$ and passed through MSSDM to obtain the deformation field $\hat{\mathbf{u}}_t$. Similar to $\hat{\mathbf{m}}_0^{(t)}$, an intermediate estimate $\hat{\mathbf{x}}_0^{(t)}$ of the CEMRI is computed as $\hat{\mathbf{x}}_0^{(t)} = \sqrt{\bar{\alpha}_t} \cdot \hat{\mathbf{x}}_t - \sqrt{1-\bar{\alpha}_t} \cdot \hat{\mathbf{v}}_{\mathrm{x},t}$. Finally, the estimate $\hat{\mathbf{x}}_{t-1}$ of $\mathbf{x}_{t-1}$ with geometric correction is computed as

$$\hat{\mathbf{x}}_{t-1} = \phi_{\hat{\mathbf{u}}_t}\left(\sqrt{\bar{\alpha}_{t-1}} \cdot \hat{\mathbf{x}}_0^{(t)} + \sqrt{1-\bar{\alpha}_{t-1}} \cdot \hat{\boldsymbol{\epsilon}}_{\mathrm{i},t}\right). \tag{6}$$

Here, the terms in the parenthesis are determined following [21], and $\hat{\boldsymbol{\epsilon}}_{\mathrm{i},t}$ is the predicted image noise computed based on Eq. (1) as $\hat{\boldsymbol{\epsilon}}_{\mathrm{i},t} = \frac{\hat{\mathbf{x}}_t - \sqrt{\bar{\alpha}_t}\cdot\hat{\mathbf{x}}_0^{(t)}}{\sqrt{1-\bar{\alpha}_t}}$. Similarly, $\hat{\mathbf{m}}_{t-1}$, i.e., the estimate of $\mathbf{m}_{t-1}$, is obtained in the same form as Eq. (6). The process is repeated until $t = 0$.

2.3 Implementation Details

Following [7], the encoder of D^{3}M is based on PixelCNN++ [20] with the Wide ResNet backbone [28]. The deformation estimation module in MSSDM uses a CNN architecture similar to U-Net, comprising an encoder and decoder with skip connections. Both the encoder and decoder employ 3×3 convolutions, followed by LeakyReLU ($\alpha = 0.2$). The encoder progressively reduces the spatial dimension by half at each layer (four layers in total), whereas the decoder combines upsampling, convolutions, and skip connections to restore the spatial resolution. The image decoder and mask decoder in DSIMD share the same network structure, which is the PixelCNN++ decoder [20].

All images are normalized by clipping the intensity values between the 0.5th and 99.5th percentiles, followed by rescaling to the range [0,1]. The input image size is 256×256. The number of training iterations is 200,000. The Adam optimizer [12] is used with a batch size of 16 and a learning rate of 8×10^{-5}.

3 Results

3.1 Data Description

We evaluated D^{3}M on two public datasets: BraSyn [13] and BraTS-PEDs [11]. BraSyn and BraTS-PEDs consist of brain magnetic resonance images of 1,470 and 307 patients with brain tumors, respectively, including aligned T1-weighted, T2-weighted, FLAIR, and contrast-enhanced T1-weighted images. For BraSyn and BraTS-PEDs, tumor masks (including enhanced tumor masks) are available for 1,251 and 216 patients, respectively. These masks were manually annotated and reviewed by clinical experts to ensure high-quality labels, and the corresponding patients were used for model training. Specifically, for BraSyn, the training/validation/test sets comprised images of 1,001/250/219 patients, respectively; for BraTS-PEDs, the training/validation/test sets comprised images of 173/43/91 patients, respectively.

3.2 Comparison with Existing Image Synthesis Methods

We compared our method with representative existing image synthesis methods, including Pix2Pix [10], ResViT [4], Palette [18], and I^{2}SB [14]. Pix2Pix is a GAN-based image synthesis method with a CNN architecture. ResViT is a multimodal medical image synthesis model combining vision transformers with convolutional operators and adversarial learning. Palette is a basic diffusion model for image synthesis. I^{2}SB is a Schrödinger bridge diffusion model, which improves upon

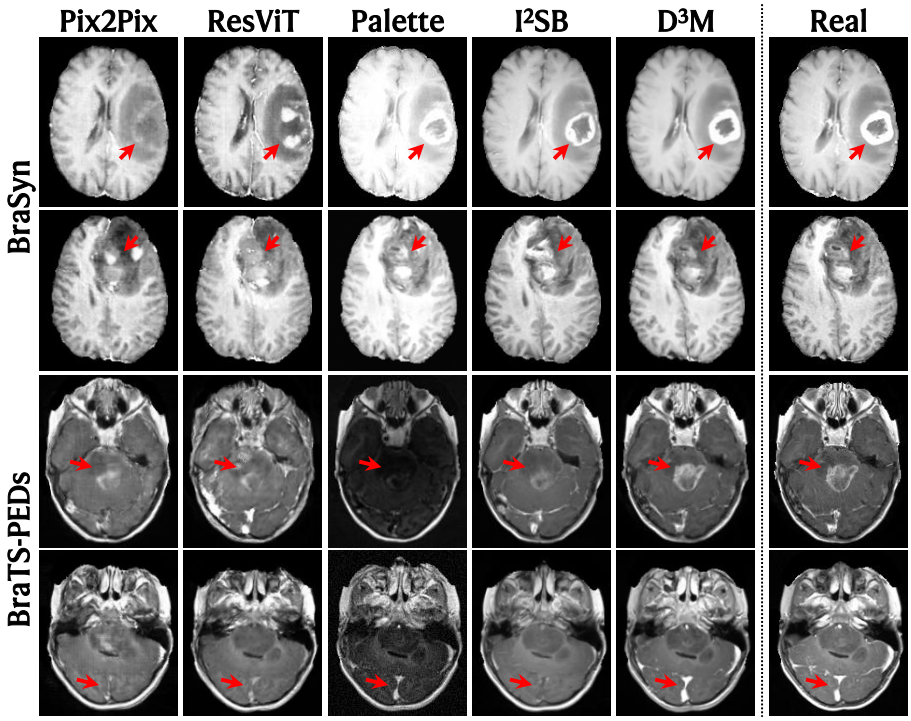

Fig. 2. Examples of synthesis results, shown together with the real CEMRI for reference. Note the tumor and vessel regions highlighted by arrows for comparison.

conventional diffusion models. All competing methods used the same training, validation, and test sets as our method. For fair comparison, the competing methods also used the enhanced tumor masks for training, where segmentation of the enhanced tumor was used as an auxiliary task like D^{3}M.

Examples of the synthesis results are shown in Fig. 2. The D^{3}M result is more consistent with the real image compared to those of the competing methods. In particular, in the tumor regions, the enhancement produced by D^{3}M agrees with the tumor in the real image, whereas the competing methods produce noticeable false positive and/or negative enhancement.

Next, we quantitatively compared D^{3}M with the competing methods by calculating the *Peak Signal-to-Noise Ratio* (PSNR) and *Structural Similarity Index Measure* (SSIM) between the synthesized and real images. Additionally, we analyzed the synthesis results specifically within the tumor regions. To obtain the tumor regions for the test images, we trained an nnU-Net segmentation model [9] based on the tumor annotations of training data, and the trained model was applied to the test images.

The PSNR and SSIM results are presented in Table 1. D^{3}M outperforms all other methods, achieving higher PSNR and SSIM, both for the whole image

Table 1. The means and standard deviations of the PSNR and SSIM between synthesized and real images, as well as the results within the tumor regions. The best results are highlighted in bold. Asterisks indicate statistically significant differences between the results of D^{3}M and each competing method based on the Wilcoxon signed-rank test: *** $p < 0.001$.

Model	BraSyn				BraTS-PEDs			
	Whole Image		Tumor Region		Whole Image		Tumor region	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Pix2Pix	$23.53^{***}_{\pm 2.89}$	$87.70^{***}_{\pm 4.25}$	$15.31^{***}_{\pm 4.40}$	$65.37^{***}_{\pm 18.38}$	$20.80^{***}_{\pm 2.68}$	$74.61^{***}_{\pm 8.59}$	$15.36^{***}_{\pm 4.78}$	$70.36^{***}_{\pm 16.49}$
ResViT	$23.54^{***}_{\pm 2.81}$	$87.43^{***}_{\pm 4.19}$	$15.21^{***}_{\pm 4.41}$	$65.19^{***}_{\pm 17.87}$	$20.96^{***}_{\pm 2.56}$	$75.23^{***}_{\pm 8.20}$	$17.48^{***}_{\pm 4.42}$	$77.82^{***}_{\pm 13.80}$
Palette	$18.95^{***}_{\pm 5.02}$	$39.97^{***}_{\pm 29.09}$	$14.00^{***}_{\pm 4.75}$	$67.86^{***}_{\pm 17.10}$	$13.61^{***}_{\pm 5.90}$	$31.92^{***}_{\pm 22.18}$	$9.22^{***}_{\pm 6.29}$	$60.55^{***}_{\pm 18.56}$
I^{2}SB	$25.01^{***}_{\pm 2.94}$	$89.11^{***}_{\pm 3.99}$	$17.05^{***}_{\pm 4.81}$	$71.59^{***}_{\pm 16.66}$	$21.98^{***}_{\pm 3.00}$	$77.05^{***}_{\pm 8.40}$	$17.91^{***}_{\pm 5.29}$	$80.76^{***}_{\pm 12.91}$
D^{3}M	$\mathbf{25.11}_{\pm 3.33}$	$\mathbf{90.95}_{\pm 3.86}$	$\mathbf{17.33}_{\pm 4.56}$	$\mathbf{73.21}_{\pm 16.22}$	$\mathbf{22.21}_{\pm 3.45}$	$\mathbf{79.29}_{\pm 9.02}$	$\mathbf{18.20}_{\pm 4.40}$	$\mathbf{82.07}_{\pm 13.33}$

Table 2. Ablation studies for investigating the individual benefit of MSSDM and DSIMD in D^{3}M. The means and standard deviations of PSNR and SSIM are presented and the best results are in bold. Asterisks indicate statistically significant differences between the results of D^{3}M and each setting based on the Wilcoxon signed-rank test: *** $p < 0.001$.

Model	Whole Image		Tumor Region	
	PSNR	SSIM	PSNR	SSIM
D^{3}M	$\mathbf{25.11}_{\pm 3.33}$	$\mathbf{90.95}_{\pm 3.86}$	$\mathbf{17.33}_{\pm 4.56}$	$\mathbf{73.21}_{\pm 16.22}$
Without MSSDM	$24.76^{***}_{\pm 3.29}$	$90.40^{***}_{\pm 3.96}$	$16.07^{***}_{\pm 4.63}$	$70.76^{***}_{\pm 17.19}$
Without MSSDM and DSIMD	$24.05^{***}_{\pm 3.31}$	$90.00^{***}_{\pm 4.01}$	$15.82^{***}_{\pm 4.63}$	$70.41^{***}_{\pm 17.10}$

and in the tumor regions. This highlights the superior performance of D^{3}M in synthesizing tumor regions. The Wilcoxon signed-rank test was also performed for statistical comparison, and its results (also shown in Table 1) indicate that the improvements achieved by D^{3}M are highly statistically significant ($p < 0.001$).

3.3 Ablation Study

To confirm the individual benefit of the major components MSSDM and DSIMD of D^{3}M, we performed ablation studies on the BraSyn dataset. In the ablation studies, the overall pipeline of D^{3}M was revised and retrained with the same data split as described in Section 3.1. The detailed results are shown in Table 2. First, we removed MSSDM from D^{3}M, where the output of DSIMD was directly used to obtain the final velocity terms at each step. The PSNR and SSIM decrease with the removal of MSSDM, especially in the tumor area. This result shows the benefit of applying geometric correction. In addition, we further removed DSIMD and used a single decoder instead. The PSNR and SSIM decrease even more, demonstrating the benefit of using a dual-stream decoder to separately handle image and mask information.

4 Conclusion

We have proposed D^{3}M to improve the quality of CEMRI synthesis with brain tumors from NCMRIs. D^{3}M models synthesis errors as misinterpretation of enhanced regions and thus incorporates geometric correction in the diffusion model. The geometric correction is achieved with two modules, MSSDM and DSIMD. MSSDM deforms the synthesized image to adjust the enhancement, where the deformation field is estimated by MSSDM from the output of DSIMD that encodes the knowledge about tumor subcomponents. Experimental results on two publicly available datasets show that our approach outperforms existing state-of-the-art methods in synthesizing high-quality CEMRIs, particularly in tumor regions.

Acknowledgments. This work was supported by the Beijing Municipal Natural Science Foundation (7242273) and the Key Program of National Natural Science Foundation of China (82330057).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: VoxelMorph: a learning framework for deformable medical image registration. IEEE Transactions on Medical Imaging **38**(8), 1788–1800 (2019)
2. Brünjes, R., Hofmann, T.: Anthropogenic gadolinium in freshwater and drinking water systems. Water Research **182**, 115966 (2020)
3. Chen, T.: On the importance of noise scheduling for diffusion models. arXiv preprint arXiv:2301.10972 (2023)
4. Dalmaz, O., Yurt, M., Çukur, T.: ResViT: Residual vision transformers for multimodal medical image synthesis. IEEE Transactions on Medical Imaging **41**(10), 2598–2614 (2022)
5. Ghaffari, M., Sowmya, A., Oliver, R.: Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the BraTS 2012–2018 challenges. IEEE Reviews in Biomedical Engineering **13**, 156–168 (2019)
6. Gui, L., Ye, C., Yan, T.: CAVM: Conditional autoregressive vision model for contrast-enhanced brain tumor MRI synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 161–170. Springer (2024)
7. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)
8. Inoue, K., Fukushi, M., Furukawa, A., Sahoo, S.K., Veerasamy, N., Ichimura, K., Kasahara, S., Ichihara, M., Tsukada, M., Torii, M., Mizoguchi, M., Taguchi, Y., Nakazawa, S.: Impact on gadolinium anomaly in river waters in Tokyo related to the increased number of MRI devices in use. Marine Pollution Bulletin **154**, 111148 (2020)

9. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021)
10. 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)
11. Kazerooni, A.F., Khalili, N., Liu, X., Gandhi, D., Jiang, Z., Anwar, S.M., Albrecht, J., Adewole, M., Anazodo, U., Anderson, H., et al.: The brain tumor segmentation in pediatrics (BraTS-PEDs) challenge: Focus on pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs). arXiv preprint arXiv:2404.15009 (2024)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
13. Li, H.B., Conte, G.M., Anwar, S.M., Kofler, F., Ezhov, I., van Leemput, K., Piraud, M., Diaz, M., Cole, B., Calabrese, E., et al.: The brain tumor segmentation (BraTS) challenge 2023: Brain MR image synthesis for tumor segmentation (BraSyn). ArXiv (2023)
14. Liu, G.H., Vahdat, A., Huang, D.A., Theodorou, E.A., Nie, W., Anandkumar, A.: I^2SB: Image-to-image schrödinger bridge. In: Proceedings of the 40th International Conference on Machine Learning. pp. 22042–22062 (2023)
15. Mallio, C.A., Radbruch, A., Deike-Hofmann, K., van der Molen, A.J., Dekkers, I.A., Zaharchuk, G., Parizel, P.M., Zobel, B.B., Quattrocchi, C.C.: Artificial intelligence to reduce or eliminate the need for gadolinium-based contrast agents in brain and cardiac MRI: A literature review. Investigative Radiology **58**(10), 746–753 (2023)
16. Meng, X., Sun, K., Xu, J., He, X., Shen, D.: Multi-modal modality-masked diffusion network for brain MRI synthesis with random modality missing. IEEE Transactions on Medical Imaging (2024)
17. Preetha, C.J., Meredig, H., Brugnara, G., Mahmutoglu, M.A., Foltyn, M., Isensee, F., Kessler, T., Pflüger, I., Schell, M., Neuberger, U., Petersen, J., Wick, A., Heiland, S., Debus, J., Platten, M., Idbaih, A., Brandes, A.A., Winkler, F., van den Bent, M.J., Nabors, B., Stupp, R., Maier-Hein, K.H., Gorlia, T., Tonn, J.C., Weller, M., Wick, W., Bendszus, M., Vollmuth, P.: 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)
18. Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., Norouzi, M.: Palette: Image-to-image diffusion models. In: ACM SIGGRAPH 2022 Conference Proceedings. pp. 1–10 (2022)
19. Salimans, T., Ho, J.: Progressive distillation for fast sampling of diffusion models. arXiv preprint arXiv:2202.00512 (2022)
20. Salimans, T., Karpathy, A., Chen, X., Kingma, D.P.: PixelCNN++: Improving the PixelCNN with discretized logistic mixture likelihood and other modifications. arXiv preprint arXiv:1701.05517 (2017)
21. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
22. Villanueva-Meyer, J.E., Mabray, M.C., Cha, S.: Current clinical brain tumor imaging. Neurosurgery **81**(3), 397 (2017)
23. Wamelink, I.J., Azizova, A., Booth, T.C., Mutsaerts, H.J., Ogunleye, A., Mankad, K., Petr, J., Barkhof, F., Keil, V.C.: Brain tumor imaging without gadolinium-based contrast agents: Feasible or fantasy? Radiology **310**(2), e230793 (2024)

24. Wang, Z., Yang, Y., Chen, Y., Yuan, T., Sermesant, M., Delingette, H., Wu, O.: Mutual information guided diffusion for zero-shot cross-modality medical image translation. IEEE Transactions on Medical Imaging (2024)
25. Woolen, S.A., Shankar, P.R., Gagnier, J.J., MacEachern, M.P., Singer, L., Davenport, M.S.: Risk of nephrogenic systemic fibrosis in patients with stage 4 or 5 chronic kidney disease receiving a group II gadolinium-based contrast agent: a systematic review and meta-analysis. JAMA Internal Medicine **180**(2), 223–230 (2020)
26. Xu, C., Tian, S., Wang, B., Zhang, J., Polat, K., Alhudhaif, A., Li, S.: Common-unique decomposition driven diffusion model for contrast-enhanced liver MR images multi-phase interconversion. IEEE Journal of Biomedical and Health Informatics (2024)
27. Xu, C., Zhang, D., Chong, J., Chen, B., Li, S.: Synthesis of gadolinium-enhanced liver tumors on nonenhanced liver MR images using pixel-level graph reinforcement learning. Medical Image Analysis **69**, 101976 (2021)
28. Zagoruyko, S.: Wide residual networks. arXiv preprint arXiv:1605.07146 (2016)
29. Zhou, Y., Chen, T., Hou, J., Xie, H., Dvornek, N.C., Zhou, S.K., Wilson, D.L., Duncan, J.S., Liu, C., Zhou, B.: Cascaded multi-path shortcut diffusion model for medical image translation. Medical Image Analysis **98**, 103300 (2024)