



Diffusion Kinetic Model for Breast Cancer Segmentation in Incomplete DCE-MRI

Tianxu Lv¹, Yuan Liu¹, Kai Miao³, Lihua Li², and Xiang Pan^{1,3}(✉)

¹ School of Artificial Intelligence and Computer Science, Jiangnan University,
Wuxi 214122, China

xiangpan@jiangnan.edu.cn

² Institute of Biomedical Engineering and Instrumentation,
Hangzhou Dianzi University, Hangzhou, China

³ Cancer Center, Faculty of Health Sciences, University of Macau,
Macau SAR, China

Abstract. Recent researches on cancer segmentation in dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) usually resort to the combination of temporal kinetic characteristics and deep learning to improve segmentation performance. However, the difficulty in accessing complete temporal sequences, especially post-contrast images, hinders segmentation performance, generalization ability and clinical application of existing methods. In this work, we propose a diffusion kinetic model (DKM) that implicitly exploits hemodynamic priors in DCE-MRI and effectively generates high-quality segmentation maps only requiring pre-contrast images. We specifically consider the underlying relation between hemodynamic response function (HRF) and denoising diffusion process (DDP), which displays remarkable results for realistic image generation. Our proposed DKM consists of a diffusion module (DM) and segmentation module (SM) so that DKM is able to learn cancer hemodynamic information and provide a latent kinetic code to facilitate segmentation performance. Once the DM is pretrained, the latent code estimated from the DM is simply incorporated into the SM, which enables DKM to automatically and accurately annotate cancers with pre-contrast images. To our best knowledge, this is the first work exploring the relationship between HRF and DDP for dynamic MRI segmentation. We evaluate the proposed method for tumor segmentation on public breast cancer DCE-MRI dataset. Compared to the existing state-of-the-art approaches with complete sequences, our method yields higher segmentation performance even with pre-contrast images. The source code will be available on <https://github.com/Medical-AI-Lab-of-JNU/DKM>.

Keywords: Deep learning · Kinetic representation · DCE-MRI · Cancer segmentation · Denoising Diffusion model

1 Introduction

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) revealing tumor hemodynamics information is often applied to early diagnosis and

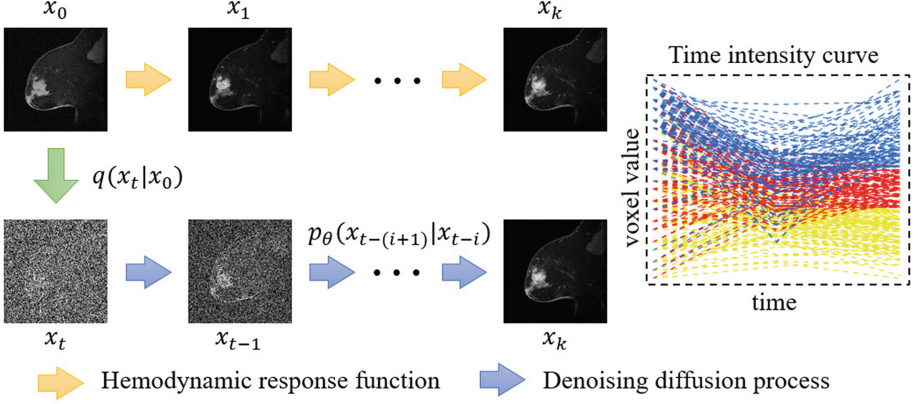


Fig. 1. Illustration of hemodynamic response function and denoising diffusion process as well as their underlying relation. Right is the time intensity curve (TIC). x_0 and x_k represent pre-contrast images and post-contrast images in DCE-MRI, respectively.

treatment of breast cancer [1]. In particular, automatically and accurately segmenting tumor regions in DCE-MRI is vital for computer-aided diagnosis (CAD) and various clinical tasks such as surgical planning. For the sake of promoting segmentation performance, recent methods utilize the dynamic MR sequence and exploit its temporal correlations to acquire powerful representations [2–4]. More recently, a handful of approaches take advantage of hemodynamic knowledge and time intensity curve (TIC) to improve segmentation accuracy [5, 6]. However, the aforementioned methods require the complete DCE-MRI sequences and overlook the difficulty in assessing complete temporal sequences and the missing time point problem, especially post-contrast phase, due to the privacy protection and patient conditions. Hence, these breast cancer segmentation models cannot be deployed directly in clinical practice.

Recently, denoising diffusion probabilistic model (DDPM) [7, 8] has produced a tremendous impact on image generation field due to its impressive performance. Diffusion model is composed of a forward diffusion process that add noise to images, along with a reverse generation process that generates realistic images from the noisy input [8]. Based on this, several methods investigate the potential of DDPM for natural image segmentation [9] and medical image segmentation [10–12]. Specifically, Baranchuk *et al.* [9] explores the intermediate activations from the networks that perform the markov step of the reverse diffusion process and find these activations can capture semantic information for segmentation. However, the applicability of DDPM to medical image segmentation are still limited. In addition, existing DDPM-based segmentation networks are generic and are not optimized for specific applications. In particular, a core question for DCE-MRI segmentation is how to optimally exploit hemodynamic priors.

Based on the above observations, we innovatively consider the underlying relation between hemodynamic response function (HRF) and denoising diffusion

process (DDP). As shown in Fig. 1, during HRF process, only tumor lesions are enhanced and other non-tumor regions remain unchanged. By designing a network architecture to effectively transmute pre-contrast images into post-contrast images, the network should acquire hemodynamic inherent in HRF that can be used to improve segmentation performance. Inspired by the fact that DDPM generates images from noisy input provided by the parameterized Gaussian process, this work aims to exploit implicit hemodynamic information by a diffusion process that predict post-contrast images from noisy pre-contrast images. Specifically, given the pre-contrast and post-contrast images, the latent kinetic code is learned using a score function of DDPM, which contains sufficient hemodynamic characteristics to facilitate segmentation performance.

Once the diffusion module is pretrained, the latent kinetic code can be easily generated with only pre-contrast images, which is fed into a segmentation module to annotate cancers. To verify the effectiveness of the latent kinetic code, the SM adopts a simple U-Net-like structure, with an encoder to simultaneously conduct semantic feature encoding and kinetic code fusion, along with a decoder to obtain voxel-level classification. In this manner, our latent kinetic code can be interpreted to provide TIC information and hemodynamic characteristics for accurate cancer segmentation.

We verify the effectiveness of our proposed diffusion kinetic model (DKM) on DCE-MRI-based breast cancer segmentation using Breast-MRI-NACT-Pilot dataset [13]. Compared to the existing state-of-the-art approaches with complete sequences, our method yields higher segmentation performance even with pre-contrast images. In summary, the main contributions of this work are listed as follows:

- We propose a diffusion kinetic model that implicitly exploits hemodynamic priors in DCE-MRI and effectively generates high-quality segmentation maps only requiring pre-contrast images.
- We first consider the underlying relation between hemodynamic response function and denoising diffusion process and provide a DDPM-based solution to capture a latent kinetic code for hemodynamic knowledge.
- Compared to the existing approaches with complete sequences, the proposed method yields higher cancer segmentation performance even with pre-contrast images.

2 Methodology

The overall framework of the proposed diffusion kinetic model is illustrated in Fig. 2. It can be observed that the devised model consists of a diffusion module (DM) and a segmentation module (SM). Let $\{x_K, K = 0, 1, \dots, k\}$ be a sequence of images representing the DCE-MRI protocol, in which x_0 represents the pre-contrast image and x_k represents the late post-contrast image. The DM takes a noisy pre-contrast image x_t as input and generates post-contrast images to estimate the latent kinetic code. Once the DM is trained, the learned kinetic code

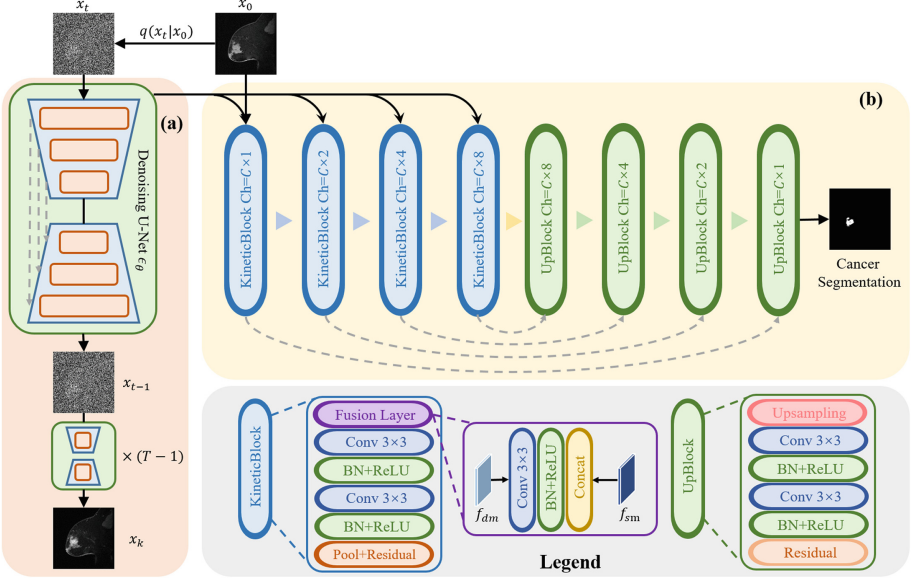


Fig. 2. Illustration of our method for implicitly exploiting hemodynamic information from pre-contrast images. The combination of learned kinetic code is an example.

is incorporated into the SM as hemodynamic priors to guide the segmentation process. Model details are shown as follows.

2.1 Diffusion Module

The diffusion module is following the denoising diffusion probabilistic model [8, 14]. Based on the consideration from nonequilibrium thermodynamics, DDPM approximates the data distribution by learning a Markov chain process which originates from the Gaussian distribution. The forward diffusion process gradually adds Gaussian noise to the data x_0 according to a variance schedule β_1, \dots, β_T [8]:

$$q(x_t|x_{t-1}) := \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t \mathbf{I}) \quad (1)$$

Particularly, a noisy image x_t can be directly obtained from the data x_0 :

$$q(x_t|x_0) := \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)\mathbf{I}) \quad (2)$$

where $\alpha_t := 1 - \beta_t$ and $\bar{\alpha}_t := \prod_{s=1}^t \alpha_s$. Afterwards, DDPM approximates the reverse diffusion process by the following parameterized Gaussian transitions:

$$p_\theta(x_{t-1}|x_t) := \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sum_\theta(x_t, t)) \quad (3)$$

where $\mu_\theta(x_t, t)$ is the learned posterior mean and $\sum_\theta(x_t, t)$ is a fixed set of scalar covariances. In particular, we employ a noise predictor network ($\epsilon_\theta(x_t, t)$) to predict the noise component at the step t (As shown in Fig. 2(a)).

Inspired by the property of DDPM [8], we devise the diffusion module by considering the pre-contrast images x_0 as source and regarding the post-contrast images x_k as target. Formally, a noisy sample can be acquired by:

$$x_t = \sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon, \epsilon \sim \mathcal{N}(0, \mathbf{I}) \quad (4)$$

where $\alpha_t := 1 - \beta_t$ and $\bar{\alpha}_t := \prod_{s=1}^t \alpha_s$. Next, we employ the reverse diffusion process to transform the noisy sample x_t to the post-contrast data x_k . As thus, the DM gradually exploits the latent kinetic code by comparing the pre-contrast and post-contrast images, which contains hemodynamic knowledge for segmentation.

2.2 Segmentation Module

Once pretrained, the DM outputs multi-scale latent kinetic code f_{dm} from intermediate layers, which is fed into the SM to guide cancer segmentation. As shown in Fig. 2(b), the SM consists of four kinetic blocks and four up blocks. Each kinetic block is composed of a fusion layer, two convolutional layers, two batch normalization layers, two ReLU activation functions, a max pooling layer and a residual addition. Specifically, to obtain sufficient expressive power to transform the learned kinetic code into higher-level features, at least one learnable linear transformation is required. To this end, a linear transformation, parametrized by a weight matrix W , is applied to the latent code f_{dm} , followed by a batch normalization, ReLU activation layer and concatenation, which can be represented as follows:

$$\hat{f} = \mathbb{C}(\phi(\text{BN}(W * f_{dm}); f_{sm})) \quad (5)$$

where $*$ represents 1×1 based convolution operation, W is the weight matrix, BN represents batch normalization, ϕ represents ReLU activation function and \mathbb{C} is concatenation operation. In this way, the hemodynamic knowledge can be incorporated into the SM to capture more expressive representations to improve segmentation performance.

2.3 Model Training

To maintain training stability, the proposed DKM adopts a two-step training procedure for cancer annotation. In the first step, the DM is trained to transform pre-contrast images into post-contrast images for a latent space where hemodynamic priors are exploited. In particular, the diffusion loss for the reverse diffusion process can be formulated as follows:

$$\mathcal{L}_{\text{DM}} = \mathbb{E}_{t, \epsilon, x} \|\epsilon_\theta(x_t, t; x_0, x_k) - \epsilon\|^2 \quad (6)$$

where ϵ_θ represents the denoising model that employs an U-Net structure, x_0 and x_k are the pre-contrast and post-contrast images, respectively, ϵ is Gaussian distribution data $\sim \mathcal{N}(0, \mathbf{I})$, and t is a timestep.

For a second step, we train the SM that integrates the previously learned latent kinetic code to provide tumor hemodynamic information for voxel-level

prediction. Considering the varying sizes, shapes and appearances of tumors that results from intratumor heterogeneity and results in difficulties of accurate cancer annotation, we design the segmentation loss as follows:

$$\begin{aligned}\mathcal{L}_{\text{SM}} &= \mathcal{L}_{\text{seg}} + \lambda \mathcal{L}_{\text{SSIM}} \\ &= \mathcal{L}_{\text{CE}}(S, G) + \mathcal{L}_{\text{Dice}}(S, G) + \lambda \left(1 - \frac{(2\mu_S\mu_G + C_1)(2\varphi_{SG} + C_2)}{(\mu_S^2 + \mu_G^2 + C_1)(\varphi_S^2 + \varphi_G^2 + C_2)}\right)\end{aligned}\quad (7)$$

where $\mathcal{L}_{\text{SSIM}}$ is used to evaluate tumor structural characteristics, S and G represents segmentation map and ground truth, respectively; μ_S is the mean of S and μ_G is the mean of G ; φ_S represents the variance of S and φ_G represents the variance of G ; C_1 and C_2 denote the constant to hold training stable [15], and φ_{SG} is the covariance between S and G . The λ is set as 0.5 empirically. Following [16], $C_1 = (k_1 L)^2$ and $C_2 = (k_2 L)^2$, where k_1 is set as 0.01, k_2 is set as 0.03 and L is set as the range of voxel values.

3 Experiments

Dataset: To demonstrate the effectiveness of our proposed DKM, we evaluate our method on 4D DCE-MRI breast cancer segmentation using the Breast-MRI-NACT-Pilot dataset [13], which contains a total of 64 patients with the contrast-enhanced MRI protocol: a pre-contrast scan, followed by 2 consecutive post-contrast time points (As shown in Fig. 3). Each MR volume consists of 60 slices and the size of each slice is 256×256 . Regarding preprocessing, we conduct zero-mean unit-variance intensity normalization for the whole volume. We divided the original dataset into training (70%) and test set (30%) based on the scans. Ground truth segmentations of the data are provided in the dataset for tumor annotation. No data augmentation techniques are used to ensure fairness.

Competing Methods and Evaluation Metrics: To comprehensively evaluate the proposed method, We compare it with 3D segmentation methods, including Dual Attention Net (DANet) [17], MultiResUNet [18] and multi-task learning network (MTLN) [19], and 4D segmentation methods, including LNet [20], 3D patch U-Net [21], and HybridNet [5]. All approaches are evaluated using 1) Dice Similarity Coefficient (DSC) and 2) Jaccard Index (JI).

Implementation Details: We implement our proposed framework with PyTorch using two NVIDIA RTX 2080Ti GPUs to accelerate model training. Following DDPM [8], we set 128, 256, 256, 256 channels for each stage in the DM and set the noise level from 10^{-4} to 10^{-2} using a linear schedule with $T = 1000$. Once the DM is trained, we extract intermediate feature maps from four resolutions for further segmentation task. Similar to DM, the SM also consists of four resolution blocks. However, unlike channel settings of DM, we set 128, 256,

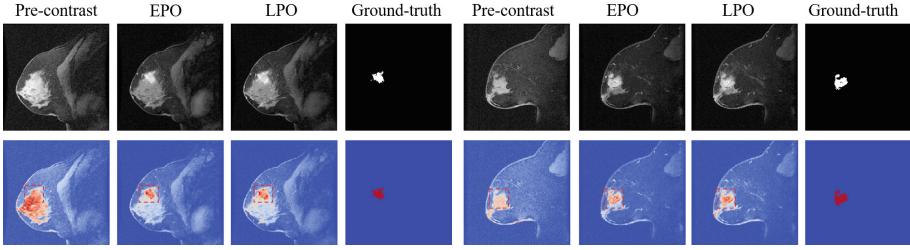


Fig. 3. Examples of breast DCE-MRI sequences. The bottom is heatmaps to observe the intensity change of cancers. EPO: early post-contrast, and LPO: late post-contrast.

Table 1. Cancer segmentation comparison between our method and previous models (Mean \pm Std). The scans are employed for testing.

Method	Scans	Dice (%) \uparrow	JI (%) \uparrow
DANet [17]	post-contrast	52.3 ± 3.1	40.5 ± 3.1
MultiResUNet [18]	post-contrast	55.6 ± 2.8	43.2 ± 3.0
MTLN [19]	post-contrast	54.2 ± 2.5	41.6 ± 2.8
LNet [20]	complete sequence	52.3 ± 2.9	40.4 ± 3.2
3D patch U-Net [21]	complete sequence	53.8 ± 2.8	41.1 ± 3.0
HybridNet [5]	complete sequence	64.4 ± 2.4	51.5 ± 2.6
DKM (ours)	pre-contrast	71.5 ± 2.5	58.5 ± 2.6

512, 1024 channels for each stage in the SM to capture expressive and sufficient semantic information. The SM is optimized by Adam with a learning rate 2×10^{-5} and a weight decay 10^{-6} . The model is trained for 500 epochs with the batch size to 1. No data augmentation techniques are used to ensure fairness.

Comparison with SOTA Methods: The quantitative comparison of the proposed method to recent state-of-the-art methods is reported in Table 1. Experimental results demonstrate that the proposed method comprehensively outperforms other models with less scans (i.e., pre-contrast) in testing. We attribute it to the ability of the diffusion module to exploit hemodynamic priors to guide the segmentation task. Specifically, in comparison with 3D segmentation models (e.g. MTLN), our method yields higher segmentation scores. The possible reason is that our method is able to exploit the time intensity curve, which contains richer information compared to post-contrast scan. Besides, we can observe that our method achieves improvements when compared to 4D segmentation models using complete sequence. Our method outperforms the HybridNet by 7.1% and 7.0% in DSC and JI, respectively. It is probably due to two aspects: 1) The hemodynamic knowledge is implicitly exploited by the diffusion module from pre-contrast images, which is useful for cancer segmentation. 2) The intermediate activations from

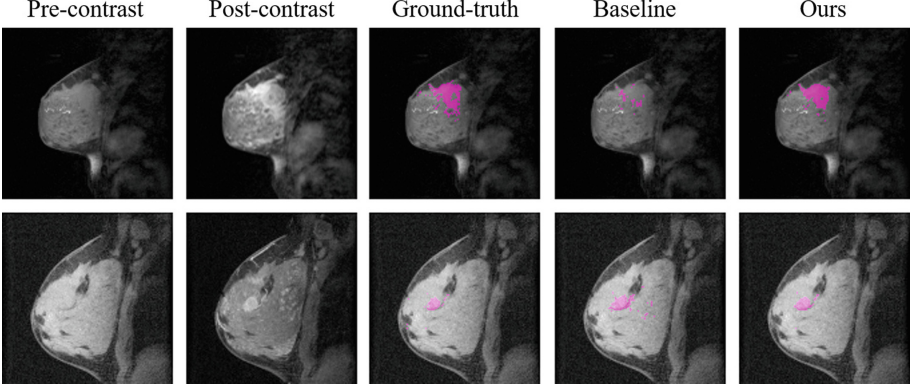


Fig. 4. Visual comparison of segmentation performance. The baseline is implemented without the incorporation of kinetic code.

Table 2. Ablation study for the incorporation of intermediate kinetic code.

f_1	f_2	f_3	f_4	Dice (%) \uparrow	JI (%) \uparrow
\checkmark				67.9 ± 2.4	54.4 ± 2.4
	\checkmark			68.6 ± 2.3	55.0 ± 2.2
		\checkmark		68.3 ± 2.4	54.8 ± 2.5
			\checkmark	69.3 ± 2.0	55.5 ± 2.1
	\checkmark	\checkmark		67.6 ± 2.3	53.2 ± 2.4
	\checkmark		\checkmark	70.1 ± 2.1	56.2 ± 2.3
		\checkmark	\checkmark	71.5 ± 2.5	58.5 ± 2.6
	\checkmark	\checkmark	\checkmark	70.2 ± 2.3	56.4 ± 2.3
\checkmark	\checkmark	\checkmark	\checkmark	69.5 ± 2.1	55.9 ± 2.4

diffusion models effectively capture the semantic information and are excellent pixel-level representations for the segmentation problem [9]. Thus, combining the intermediate features can further promote the segmentation performance. In a word, the proposed framework can produce accurate prediction masks only requiring pre-contrast images. This is useful when post-contrast data is limited.

Ablation Study: To explore the effectiveness of the latent kinetic code, we first conduct ablation studies to select the optimal setting. We denote the intermediate features extracted from each stage in the DM as f_1 , f_2 , f_3 , and f_4 , respectively, where f_i represents the feature map of i -th stage. Table 2 reports the segmentation performance with different incorporations of intermediate kinetic codes. It can be observed that the latent kinetic code is able to guide the network training for better segmentation results. Specifically, we note that the incorporation of f_3 and f_4 achieves the highest scores among these combinations, and

outperforms the integration of all features by 2.0% and 2.6% in DSC and JI, respectively. We attribute it to the denoising diffusion model that receives the noisy input, leading to the noise of shallow features. In contrast, the deep features capture essential characteristics to reveal the structural information and hemodynamic changes of tumors. Figure 4 shows visual comparison of segmentation performance. The above results reveal that incorporation of kinetic code comfortably outperform the baseline without hemodynamic information.

4 Conclusion

We propose a diffusion kinetic model by exploiting hemodynamic priors in DCE-MRI to effectively generate high-quality segmentation results only requiring pre-contrast images. Our models learns the hemodynamic response function based on the denoising diffusion process and estimates the latent kinetic code to guide the segmentation task. Experiments demonstrate that our proposed framework has the potential to be a promising tool in clinical applications to annotate cancers.

Acknowledgements. This work is supported in part by the National Key R&D Program of China under Grants 2021YFE0203700 and 2018YFA0701700, the Postgraduate Research & Practice Innovation Program of Jiangsu Province KYCX23_2524, and is supported by National Natural Science Foundation of China grants 61602007, U21A20521 and 61731008, Zhejiang Provincial Natural Science Foundation of China (LZ15F010001), Jiangsu Provincial Maternal and Child Health Research Project (F202034), Wuxi Health Commission Precision Medicine Project (J202106), Jiangsu Provincial Six Talent Peaks Project (YY-124), and the Science and Technology Development Fund, Macau SAR (File no. 0004/2019/AFJ and 0011/2019/AKP).

References

1. Fan, M., Xia, P., Clarke, R., Wang, Y., Li, L.: Radiogenomic signatures reveal multiscale intratumour heterogeneity associated with biological functions and survival in breast cancer. *Nat. Commun.* **11**(1), 4861 (2020)
2. Vidal, J., Vilanova, J.C., Martí, R., et al.: A U-net ensemble for breast lesion segmentation in DCE MRI. *Comput. Biol. Med.* **140**, 105093 (2022)
3. Qiao, M., et al.: Three-dimensional breast tumor segmentation on DCE-MRI with a multilabel attention-guided joint-phase-learning network. *Comput. Med. Imaging Graph.* **90**, 101909 (2021)
4. Nalepa, J., et al.: Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors. *Artif. Intell. Med.* **102**, 101769 (2020)
5. Lv, T., et al.: A hybrid hemodynamic knowledge-powered and feature reconstruction-guided scheme for breast cancer segmentation based on dce-mri. *Med. Image Anal.* **82**, 102572 (2022)
6. Wang, S., et al.: Breast tumor segmentation in DCE-MRI with tumor sensitive synthesis. *IEEE Trans. Neural Netw. Learn. Syst.* (2021)
7. 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)

8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. *Adv. Neural. Inf. Process. Syst.* **33**, 6840–6851 (2020)
9. Baranchuk, D., Rubachev, I., Voynov, A., Khrulkov, V., Babenko, A.: Label-efficient semantic segmentation with diffusion models. In: *International Conference on Learning Representation (ICLR)* (2022)
10. Fernandez, V., et al.: Can segmentation models be trained with fully synthetically generated data? In: Zhao, C., Svoboda, D., Wolterink, J.M., Escobar, M. (eds.) *SASHIMI 2022. LNCS*, vol. 13570, pp. 79–90. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16980-9_8
11. Wolleb, J., Sandkühler, R., Bieder, F., Valmaggia, P., Cattin, P.C.: Diffusion models for implicit image segmentation ensembles. In: *International Conference on Medical Imaging with Deep Learning*, pp. 1336–1348. PMLR (2022)
12. Junde, W., et al.: Medsegdiff: medical image segmentation with diffusion probabilistic model (2023)
13. Newitt, D., Hylton, N.: Single site breast DCE-MRI data and segmentations from patients undergoing neoadjuvant chemotherapy. *Cancer Imaging Arch.* **2** (2016)
14. Kim, B., Ye, J.C.: Diffusion deformable model for 4D temporal medical image generation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *MICCAI 2022. LNCS*, vol. 13431, pp. 539–548. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16431-6_51
15. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
16. Lin, J., Lin, H., Zhang, Z., Yiwen, X., Zhao, T.: SSIM-variation-based complexity optimization for versatile video coding. *IEEE Signal Process. Lett.* **29**, 2617–2621 (2022)
17. Fu, J., et al.: Dual attention network for scene segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3146–3154 (2019)
18. Ibtehaz, N., Rahman, M.S.: Multiresunet: rethinking the U-net architecture for multimodal biomedical image segmentation. *Neural Netw.* **121**, 74–87 (2020)
19. Zhou, Y., et al.: Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images. *Med. Image Anal.* **70**, 101918 (2021)
20. Denner, S., et al.: Spatio-temporal learning from longitudinal data for multiple sclerosis lesion segmentation. In: Crimi, A., Bakas, S. (eds.) *BrainLes 2020. LNCS*, vol. 12658, pp. 111–121. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72084-1_11
21. Khaled, R., Vidal, J., Martí, R.: Deep learning based segmentation of breast lesions in DCE-MRI. In: Del Bimbo, A., et al. (eds.) *ICPR 2021. LNCS*, vol. 12661, pp. 417–430. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-68763-2_32