A Diffusion-Driven Temporal Super-Resolution and Spatial Consistency Enhancement Framework for 4D MRI imaging

Xuanru Zhou[1,2], Jiarun Liu[1,2,3], Shoujun Yu[1,2], Hao Yang[1,2,3], Cheng Li[1], Tao Tan[4], and Shanshan Wang[1(✉)]

[1] Paul C. Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
ss.wang@siat.ac.cn
[2] University of Chinese Academy of Sciences, Beijing, China
[3] Pengcheng Laboratory, Shenzhen, China
[4] Faculty of Applied Sciences, Macao Polytechnic University, Macao, China

Abstract. In medical imaging, 4D MRI enables dynamic 3D visualization, yet the trade-off between spatial and temporal resolution requires prolonged scan time that can compromise temporal fidelity—especially during rapid, large-amplitude motion. Traditional approaches typically rely on registration-based interpolation to generate intermediate frames. However, these methods struggle with large deformations, resulting in misregistration, artifacts, and diminished spatial consistency. To address these challenges, we propose TSSC-Net, a novel framework that generates intermediate frames while preserving spatial consistency. To improve temporal fidelity under fast motion, our diffusion-based temporal super-resolution network generates intermediate frames using the start and end frames as key references, achieving 6× temporal super-resolution in a single inference step. Additionally, we introduce a novel tri-directional Mamba-based module that leverages long-range contextual information to effectively resolve spatial inconsistencies arising from cross-slice misalignment, thereby enhancing volumetric coherence and correcting cross-slice errors. Extensive experiments were performed on the public ACDC cardiac MRI dataset and a real-world dynamic 4D knee joint dataset. The results demonstrate that TSSC-Net can generate high-resolution dynamic MRI from fast-motion data while preserving structural fidelity and spatial consistency. The code will be available at `https://github.com/Joker-ZXR/TSSC-Net`.

Keywords: Temporal super-resolution · Spatial consistency enhancement · Diffusion model · 4D MRI imaging.

1 Introduction

The evolution of four-dimensional magnetic resonance imaging (4D MRI) has revolutionized dynamic anatomical assessment, enabling volumetric visualization

of physiological motions across time—a capability that is critical for applications such as cardiac function analysis [28,21], tumor kinetics [13], and joint biomechanics investigations [8]. By acquiring time-resolved 3D image sequences, 4D MRI uniquely captures transient phenomena like myocardial contraction patterns and rapid synovial movements, providing clinicians with unprecedented insights into disease progression and treatment response. However, conventional 4D MRI acquisition protocols are hampered by inherent limitations [24,29]. The fundamental trade-off between spatial and temporal resolution dictated by MRI physics necessitates prolonged scan durations, which in turn compromise the temporal fidelity of the acquired data [6,25]. This limitation is more pronounced in cases involving rapid or large-amplitude anatomical motion. Consequently, the resulting sparse temporal sampling and potential spatial inconsistencies diminish the clinical utility of 4D MRI in time-sensitive diagnostic scenarios.

Recent advances in deep learning have spurred interest in data-driven approaches for 4D MRI generation. While generative adversarial networks (GANs) [9] and deformable registration techniques [1,2,5,15] have demonstrated preliminary success in temporal interpolation, they exhibit three critical shortcomings: (1) GANs often exhibit inherent limitations in spatiotemporal sequence modeling, as their adversarial training mechanism fails to capture smooth temporal evolution, resulting in progressive distortion on temporal trajectories [3]. (2) Conventional registration-based interpolation methods often lack explicit mechanisms to enforce smooth inter-phase motion continuity, which may introduce structural distortions during the interpolation process [1,2,5,15]. (3) Existing frameworks often lack explicit mechanisms to enforce cross-slice spatial consistency, which may lead to volumetric inconsistencies and reduce the reliability of subsequent 3D analyses [14,11]. In contrast to GANs and registration-based methods, diffusion models [22,20] demonstrate impressive generation performance in natural image synthesis but remain under-explored due to the unique coupled spatiotemporal challenge of MRI dynamics.

In this work, we present TSSC-Net, a novel framework that overcomes key challenges in 4D MRI by bridging temporal gaps and enhancing spatial consistency. The main contributions of our work are as follows:

- We propose a diffusion-driven temporal super-resolution network with cross-frame attention to generate continuous and coherent intermediate frames, addressing sparse sampling and ensuring smooth temporal interpolation in dynamic MRI.
- We design a spatial consistency enhancement network using residual tri-directional Mamba blocks for multi-directional scanning, capturing long-range spatial correlations and correcting cross-slice inconsistencies.
- Our method delivers outstanding performance: achieving competitive results on ACDC cardiac MRI with minor deformations and significantly improving knee joint MRI with large motions. Notably, TSSC-Net can produce 6× temporal super-resolution in a single inference step.

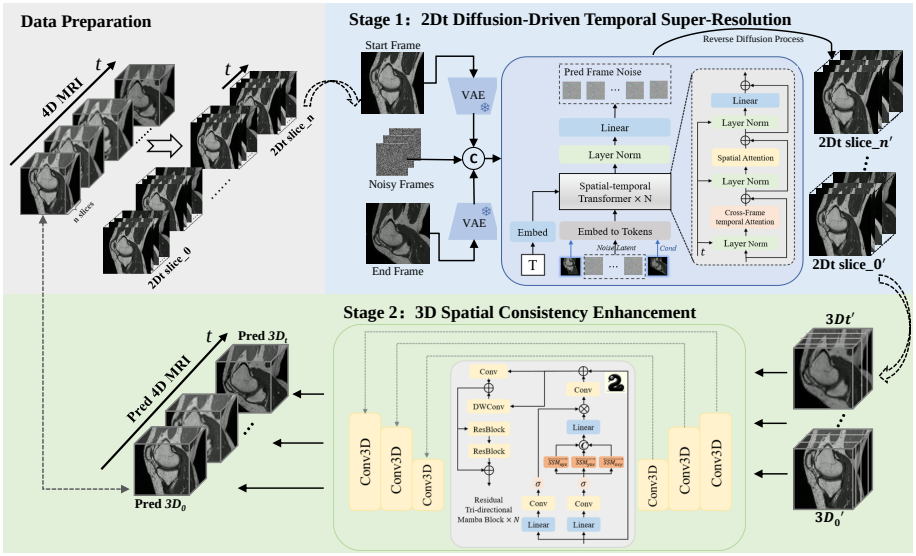

Fig. 1. The overall framework of the proposed method. The framework operates in two stages. In Stage 1, diffusion-driven temporal super-resolution is performed on 2Dt data to generate intermediate frames. In Stage 2, spatial consistency enhancement is applied to the generated 3D volumes to ensure volumetric coherence.

2 Proposed Method

2.1 Framework Overview

The overall framework of the proposed diffusion-driven temporal super-resolution and spatial consistency enhancement network (TSSC-Net) is illustrated in Fig. 1. The framework begins by preprocessing 4D MRI data into multiple 2Dt sequences. In Stage 1, a diffusion-driven temporal super-resolution network improves the temporal resolution by generating intermediate frames. These enhanced slices are then reassembled into 3D volumes and further refined in Stage 2 using a spatial consistency enhancement network. This stage leverages residual tri-directional Mamba blocks to enforce volumetric coherence across orthogonal planes. Further details are provided below.

2.2 Diffusion-driven temporal super-resolution

In order to address the challenges posed by sparse key frame information and complex inter-frame motion in video temporal super-resolution, we adopt a diffusion-based generative model owing to its ability to generate high-quality, detail-rich outputs with stable training dynamics. The diffusion model's inherent capability to progressively denoise complex signals enables it to capture intricate spatiotemporal features, making it particularly well-suited for bridging

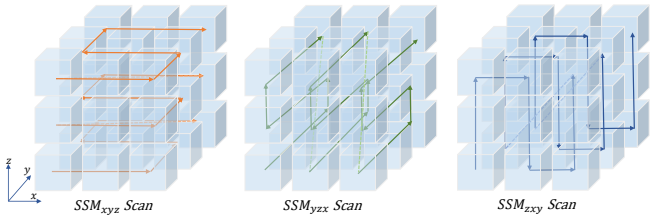

Fig. 2. Illustration of the three distinct selective scan orders in the residual tri-directional Mamba blocks. Each order processes volumetric data along a different sequence of x, y, and z axes, effectively capturing multi-directional dependencies for enhanced spatial consistency.

the temporal gap between key frames. Specifically, our approach conditions the diffusion process on the known start and end frames, I_0 and I_1. These frames serve as fixed boundary conditions that guide the diffusion model throughout the generation process [18,7].

The forward diffusion process defines a Markov chain [12] that progressively adds noises to a clean image over T steps. In TSSC-Net, we add noises to intermediate frames(x_0) and train the model to predict the noises, conditioned on I_0 (start frame) and I_1 (end frame). This allows the model to learn the noise distribution of intermediate dynamics. Starting from x_0, the forward process generates a sequence $x_1, x_2, \ldots, x_T$. For each step $t = 1, \ldots, T$,

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

with $\beta_t \epsilon (0, 1)$ being the noises variance at step t. This forward process is fixed (non-learnable) and is used to provide training pairs of noisy images and their corresponding clean target. In experiments, T is set to 1000.

The reverse diffusion process is the generative stage of TSSC-Net. Starting from a random noise image $x_T \sim$N(0,I), the model iteratively denoises it to produce a coherent image. At each reverse step t, the goal is to sample x_{t-1} from the conditional distribution.

$$p_\theta(x_{t-1}|x_t, I_0, I_1) = N(x_{t-1}; \mu_\theta(x_t, t, I_0, I_1), {\sigma_t}^2 I), \tag{2}$$

where μ_θ is a learned mean function and ϵ_t is a fixed variance at step t.

Furthermore, as shown in Stage 1 of Fig. 1, we employ a transformer-based [19] architecture within our diffusion model that partitions the noisy image into a sequence of tokens. By integrating a spatiotemporal attention mechanism, the model effectively captures cross-frame correlations in both spatial and temporal dimensions. This design not only enhances the denoising process but also ensures that the generated frames exhibit consistent structural details and realistic motion patterns. Finally, leveraging this cross-frame temporal attention enables robust 6× frame interpolation, yielding synthesized outputs with high visual quality and temporal coherence.

2.3 Spatial Consistency Enhancement

After the temporal super-resolution stage, the independently generated 2D slices are concatenated into initial 3D volumes, which may lead to cross-slice inconsistencies and discontinuities. To address these issues, we adopt the Mamba architecture, a framework built on state space models (SSMs) [10,27] that overcomes limitations of conventional methods such as the quadratic complexity of transformers and the restricted receptive fields of CNNs. The Mamba architecture is particularly effective for processing 3D images and long sequences, as it efficiently captures long-range dependencies [17]. Specifically, the hidden state is updated and the output is generated via:

$$h_t = \bar{A}_t h_{t-1} + \bar{B}_t x_t \tag{3}$$

$$y_t = \bar{C}_t h_t \tag{4}$$

where h_t denotes the hidden state at time t, x_t is the input, and $\bar{A}_t$, $\bar{B}_t$ and $\bar{C}_t$ are learned parameters. The framework is designed to capture long-range dependencies while maintaining computational efficiency. To address spatial inconsistencies in the initial 3D volumes, we introduce a spatial consistency enhancement network that employs residual tri-directional Mamba blocks. Each block processes the 3D volume along three orthogonal directions, denoted as $\overleftrightarrow{SSM_{xyz}}$, $\overleftrightarrow{SSM_{yzx}}$, and $\overleftrightarrow{SSM_{zxy}}$, to capture horizontal, vertical, and inter-slice dependencies, respectively. The bidirectional structure of each SSM, capturing both forward and backward dependencies along its respective axis, enables robust integration of multi-directional contextual information.

The spatial consistency enhancement network is trained using a composite loss function that integrates three components: an MSE loss for voxel-wise fidelity, a Wavelet Transform loss [26] to capture multi-scale feature consistency, and Total Variation (TV) regularization [23] to encourage spatial smoothness. Formally, the overall loss is defined as:

$$\mathcal{L}_{\mathrm{SC}} = \lambda_{\mathrm{MSE}}\mathcal{L}_{\mathrm{MSE}} + \lambda_{\mathrm{WT}}\mathcal{L}_{\mathrm{Wavelet}} + \lambda_{\mathrm{TV}}\mathcal{L}_{\mathrm{TV}} \tag{5}$$

with the weighting factors λ_{MSE}, λ_{WT} and λ_{TV} all set to 1 based on experimental validation.

3 Experiments

3.1 Dataset

To verify the proposed method for 4D image generation, two 4D image datasets are used, each for the heart and knee. The publicly available ACDC [4] dataset consists of 150 4D cardiac MRI scans, each spanning a full cardiac cycle from end-diastole (ED) to end-systole (ES). All scans were resampled in the x–y plane using bilinear interpolation, and zero-padding was applied along the z-axis to obtain a standardized volume size of 256×256×32. The dataset was split into

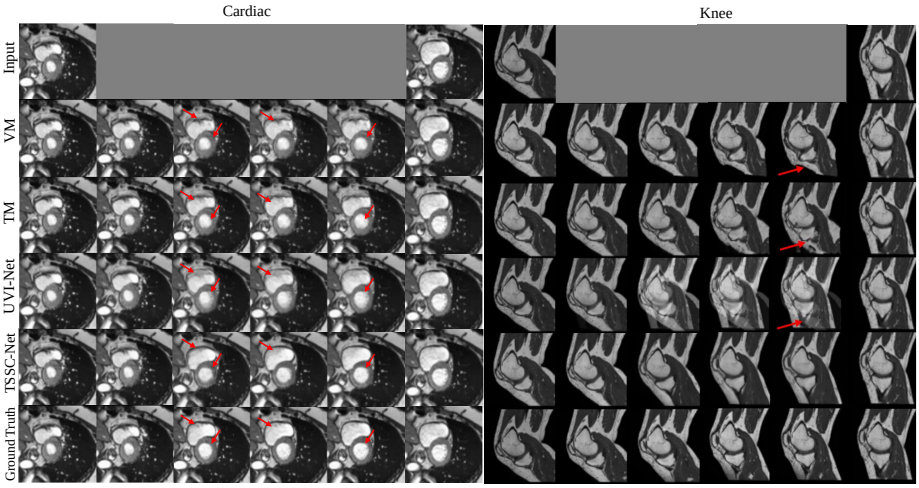

Fig. 3. Qualitative comparison between synthesized cardiac (left) and knee (right) MRI sequences, displaying six evenly-spaced frames from each 12-frame series. Red arrows highlight regions where TSSC-Net better preserves anatomical details.

100 training samples and 50 testing samples. To assess generalizability, we also collected an in-house 4D dynamic knee MRI dataset (12 cases) using a United Imaging 3T scanner. Each scan consists of 12 3D frames covering a full flexion-extension cycle, acquired with a coronal T1-weighted sequence. The volumes were resampled to $256\times256\times96$ to ensure consistent spatial resolution. In total, 1,152 2Dt slices were used for the first-stage temporal super-resolution task, and 144 3D volumes were used for the second-stage spatial consistency enhancement.

3.2 Implementation Details

Experiments were conducted using PyTorch 11.7 on an NVIDIA A100 80GB GPU. In the first stage, the diffusion-driven temporal super-resolution network was trained for 100,000 iterations using the Adam optimizer with a fixed learning rate of 1×10^{-4}. A linear noise schedule was applied, with β_t increasing from 10^{-6} to 10^{-2}, and DDIM acceleration was employed during inference to expedite the sampling process. This stage focused on achieving temporal super-resolution by generating high-quality intermediate frames. Once the diffusion-driven temporal super-resolution network was fully trained, its parameters were frozen. In the second stage, the spatial consistency enhancement network was then trained for 100 epochs using the Adam optimizer. The training began with a learning rate of 1×10^{-4} that was linearly decayed over time, with the aim of refining the 3D voxel coherence and eliminating cross-slice inconsistencies.

Table 1. Quantitative results of comparison models on the ACDC cardiac dataset and in-house knee joint dataset. The best results are highlighted in bold while the second best results are marked with an underline.

Dataset	Method	MAE↓	PSNR (dB)↑	SSIM↑
Cardiac	VM[2]	0.011±0.004	30.920±2.559	0.960±0.021
	TM[5]	0.010±0.004	31.264±2.509	0.963±0.020
	UVI-Net[16]	<u>0.007±0.003</u>	**33.256±2.583**	<u>0.975±0.012</u>
	TSSC-Net	**0.007±0.002**	<u>32.971±1.584</u>	**0.977±0.010**
Knee	VM[2]	0.074±0.007	16.506±0.597	0.579±0.007
	TM[5]	<u>0.063±0.005</u>	<u>17.580±0.399</u>	<u>0.635±0.007</u>
	UVI-Net[16]	0.069±0.007	17.360±0.595	0.598±0.013
	TSSC-Net	**0.047±0.006**	**20.041±0.934**	**0.730±0.024**

3.3 Results and Discussion

In our experiments, TSSC-Net was benchmarked against three registration-based baselines: VM [2], TM [5], and UVI-Net [16]. These methods estimate flow fields from I_0 to I_1 and use linear interpolation to generate intermediate frames. VM (VoxelMorph) is a registration-based method that aligns images at the voxel level. TM (TransMorph) builds upon this by integrating transformer modules to capture global context and enhance registration accuracy. Conversely, UVI-Net is an unsupervised bidirectional registration interpolation approach that synthesizes intermediate frames through bidirectional flow estimation.

Table 1 presents a detailed account of the quantitative results obtained on the ACDC cardiac dataset and the in-house knee joint dataset. Regarding the cardiac dataset, TSSC-Net achieved a PSNR of 32.971 dB and an SSIM of 0.977. Although UVI-Net registered a slightly higher PSNR (33.256 dB), the comparable SSIM values suggest that both methods effectively preserve structural details when motion is relatively minor. In contrast, on the knee dataset, characterized by substantial inter-frame motion, TSSC-Net markedly outperformed the baseline methods, attaining a PSNR of 20.041 dB and an SSIM of 0.730, while VM and TM produced considerably lower scores.

Qualitative comparisons from Fig. 3 further corroborate these findings. In the cardiac dataset, the regions indicated by red arrows highlight areas where TSSC-Net preserves fine anatomical details and yields smoother transitions compared to the other methods. In the knee dataset, the differences are even more pronounced: while baseline methods exhibit noticeable discontinuities and loss of structural integrity, TSSC-Net consistently maintains coherent representations across frames. This performance gap underscores TSSC-Net's effectiveness in addressing large inter-frame motion—an issue that remains challenging for existing methods.

The ablation study in Table 2, supported by the qualitative results in Fig. 4, highlights the importance of the spatial consistency enhancement network in maintaining volumetric coherence. Without this network (w/o SC), PSNR and SSIM drop significantly on both datasets, with noticeable slice-wise inconsistencies in the generated volumes. As shown in Fig. 4, the absence of spatial

Table 2. Quantitative results of the ablation study comparing the TSSC-Net and its variant without spatial consistency enhancement network (w/o SC).

Dataset	Method	MAE↓	PSNR (dB)↑	SSIM↑
Cardiac	w/o SC	0.016±0.004	29.343±1.552	0.942±0.015
	TSSC-Net	**0.007±0.002**	**32.971±1.584**	**0.977±0.010**
Knee	w/o SC	0.049±0.006	18.128±0.908	0.679±0.028
	TSSC-Net	**0.047±0.006**	**20.041±0.934**	**0.730±0.024**

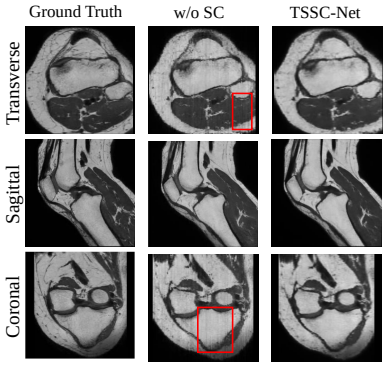

Fig. 4. Qualitative comparisons of generated images from the ablation study. The red boxes indicate areas of noticeable improvement in structural coherence when the spatial consistency network is included.

consistency enhancement leads to misalignment across adjacent slices. By incorporating residual tri-directional Mamba blocks, the spatial consistency enhancement network effectively captures long-range spatial dependencies, enabling the model to learn and correct spatial inconsistencies, thereby ensuring a more coherent and anatomically faithful generation. Overall, the experimental results demonstrate that TSSC-Net effectively generates high-fidelity 4D medical images, achieving competitive performance on cardiac data and notable improvements in large-motion knee imaging. Qualitative comparisons confirm its ability to reduce cross-slice inconsistencies and enhance structural coherence, while the ablation study highlights the critical role of the spatial consistency enhancement network in maintaining volumetric consistency. Ultimately, TSSC-Net generates a complete 12-frame 4D dynamic MRI sequence from only two input frames, achieving 6× temporal super-resolution in a single inference step.

4 Conclusion

In this paper, we introduced TSSC-Net, a novel 4D MRI imaging framework that combines a diffusion-driven temporal super-resolution network with a spatial consistency enhancement network based on residual tri-directional Mamba blocks. Experiments on both the ACDC cardiac dataset and an in-house knee

joint dataset demonstrate that TSSC-Net achieves competitive performance on low-motion data and significantly outperforms conventional registration-based methods in high-motion scenarios by mitigating cross-slice inconsistencies and preserving anatomical integrity. Notably, TSSC-Net can generate a complete 12-frame 4D dynamic MRI sequence from only two input frames in a single inference step, achieving 6× temporal super-resolution. These results underscore its potential for reliable 4D MRI imaging, and future work will focus on further enhancing robustness and adapting the framework to diverse anatomical regions and motion patterns.

Acknowledgments. This research was partly supported by the National Natural Science Foundation of China (No. 62222118, No. U22A2040), Shenzhen Science and Technology Program (No. JCYJ20220531100213029), Shenzhen Medical Research Fund (No. B2402047), National Key R&D Program of China (No. 2023YFA1011400), Key Laboratory for Magnetic Resonance and Multimodality Imaging of Guangdong Province (No. 2023B1212060052), and Youth lnnovation Promotion Association CAS.

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.: An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9252–9260 (2018)
2. 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)
3. Baniya, A.A., Lee, T.K., Eklund, P.W., Aryal, S.: A survey of deep learning video super-resolution. IEEE Transactions on Emerging Topics in Computational Intelligence **8**(4), 2655–2676 (2024)
4. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE transactions on medical imaging **37**(11), 2514–2525 (2018)
5. Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y.: Transmorph: Transformer for unsupervised medical image registration. Medical image analysis **82**, 102615 (2022)
6. Cohen, J.D., Noll, D.C., Schneider, W.: Functional magnetic resonance imaging: Overview and methods for psychological research. Behavior Research Methods, Instruments, & Computers **25**(2), 101–113 (1993)
7. Danier, D., Zhang, F., Bull, D.: Ldmvfi: Video frame interpolation with latent diffusion models. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 1472–1480 (2024)
8. Eck, B.L., Yang, M., Elias, J.J., Winalski, C.S., Altahawi, F., Subhas, N., Li, X.: Quantitative mri for evaluation of musculoskeletal disease: cartilage and muscle composition, joint inflammation, and biomechanics in osteoarthritis. Investigative radiology **58**(1), 60–75 (2023)

9. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)
10. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
11. Han, K., Xiong, Y., You, C., Khosravi, P., Sun, S., Yan, X., Duncan, J.S., Xie, X.: Medgen3d: A deep generative framework for paired 3d image and mask generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 759–769. Springer (2023)
12. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
13. Ingrisch, M., Sourbron, S.: Tracer-kinetic modeling of dynamic contrast-enhanced mri and ct: a primer. Journal of pharmacokinetics and pharmacodynamics **40**, 281–300 (2013)
14. Jeong, J., Kim, K.D., Nam, Y., Cho, K., Kang, J., Hong, G.S., Kim, N.: Generating high-resolution 3d ct with 12-bit depth using a diffusion model with adjacent slice and intensity calibration network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 366–375. Springer (2023)
15. Kim, B., Ye, J.C.: Diffusion deformable model for 4d temporal medical image generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 539–548. Springer (2022)
16. Kim, J., Yoon, H., Park, G., Kim, K., Yang, E.: Data-efficient unsupervised interpolation without any intermediate frame for 4d medical images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11353–11364 (2024)
17. Liu, J., Yang, H., Zhou, H.Y., Yu, L., Liang, Y., Yu, Y., Zhang, S., Zheng, H., Wang, S.: Swin-umamba†: Adapting mamba-based vision foundation models for medical image segmentation. IEEE Transactions on Medical Imaging pp. 1–1 (2024)
18. Lu, H., Yang, G., Fei, N., Huo, Y., Lu, Z., Luo, P., Ding, M.: Vdt: General-purpose video diffusion transformers via mask modeling. arXiv preprint arXiv:2305.13311 (2023)
19. Peebles, W., Xie, S.: Scalable diffusion models with transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4195–4205 (2023)
20. Po, R., Yifan, W., Golyanik, V., Aberman, K., Barron, J.T., Bermano, A., Chan, E., Dekel, T., Holynski, A., Kanazawa, A., et al.: State of the art on diffusion models for visual computing. In: Computer Graphics Forum. vol. 43, p. e15063. Wiley Online Library (2024)
21. Rizk, J.: 4d flow mri applications in congenital heart disease. European radiology **31**(2), 1160–1174 (2021)
22. 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)
23. Strong, D., Chan, T.: Edge-preserving and scale-dependent properties of total variation regularization. Inverse problems **19**(6), S165 (2003)
24. Wang, S., Ke, Z., Cheng, H., Jia, S., Ying, L., Zheng, H., Liang, D.: DIMENSION: dynamic mr imaging with both k-space and spatial prior knowledge obtained via multi-supervised network training. NMR in Biomedicine **35**(4), e4131 (2022)
25. Wang, S., Wu, R., Jia, S., Diakite, A., Li, C., Liu, Q., Zheng, H., Ying, L.: Knowledge-driven deep learning for fast mr imaging: Undersampled mr image reconstruction from supervised to un-supervised learning. Magnetic Resonance in Medicine **92**(2), 496–518 (2024)

26. Yang, Q., Tang, B., Deng, L., Zhu, P., Ming, Z.: Wtformer: Rul prediction method guided by trainable wavelet transform embedding and lagged penalty loss. Advanced Engineering Informatics **62**, 102710 (2024)
27. Yang, Y., Xing, Z., Yu, L., Huang, C., Fu, H., Zhu, L.: Vivim: A video vision mamba for medical video segmentation. arXiv preprint arXiv:2401.14168 (2024)
28. Zhao, F., Zhang, H., Wahle, A., Thomas, M.T., Stolpen, A.H., Scholz, T.D., Sonka, M.: Congenital aortic disease: 4d magnetic resonance segmentation and quantitative analysis. Medical image analysis **13**(3), 483–493 (2009)
29. Zou, J., Li, C., Jia, S., Wu, R., Pei, T., Zheng, H., Wang, S.: Selfcolearn: Self-supervised collaborative learning for accelerating dynamic mr imaging. Bioengineering **9**(11), 650 (2022)