Multiscale Graph and Multi-Step Cross-Frame Mamba for Myocarditis Lesion Segmentation

Chengjin Yu[1], Hao Zhang[2], Yuanting Yan[2*], Dong Zhang[3*], Sangyin Lv[4], and Cailing Pu[5]

[1] School of Big Data and Statistics, Anhui University, Hefei, China,
[2] School of Computer Science and Technology, Anhui University, Hefei, China,
[3] School of Biomedical Engineering, Anhui Medical University, Hefei, China,
[4] Shaoxing People's Hospital, Shaoxing, China
[5] Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, China

Abstract. Myocarditis, an acute cardiac disorder progressing rapidly to life-threatening heart failure, requires precise lesion segmentation from Cine Magnetic Resonance Imaging (Cine-MRI) for timely intervention. Current segmentation accuracy is limited by two key challenges: 1) spatiotemporal discordance between myocardial motion patterns and evolving pathological features and 2) morphological complexity (irregular borders, scattered lesions). In this paper, we propose the MG-Mamba, a framework integrating deep state space models with graph-based spatiotemporal analysis. The architecture employs Mamba blocks to establish initial intra-/inter-frame dependencies in Cine-MRI sequences. For Challenge 1, we improve the detection of subtle abnormal motions through multi-step cross-frame analysis, extending beyond conventional adjacent-frame analysis. For Challenge 2, we further implement multi-scale patch division and constructs inter-patch graphs to concurrently capture global lesion distribution and local geometric patterns. Extensive evaluations on SYC-QC and SYC-SX clinical datasets demonstrate MG-Mamba's superior segmentation accuracy over ten state-of-the-art benchmarks, significantly advancing myocarditis diagnostic precision. The code is available at https://github.com/userZ-CY/MICCAI.

Keywords: Myocarditis · Cine-MRI · Graph · Mamba.

1 Introduction

Timely and accurate segmentation of myocarditis lesions is crucial for treatment and patient prognosis [1]. Myocarditis is a group of diseases characterized primarily by inflammation of the myocardium [2]. The myocarditis progresses rapidly and it might lead to severe complications, including arrhythmias, heart failure, and cardiogenic shock [3]. The Lake Louise criteria are commonly used in

*Corresponding authors: Yuanting Yan (ytyan2016@163.com) and Dong Zhang (dongzhang@ahmu.edu.cn)

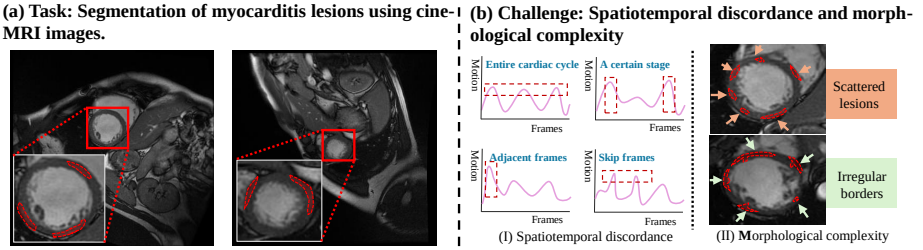

Fig. 1. (a): Segmenting myocarditis lesions using Cine-MRI images can determine the size, distribution, and severity of inflammation. This can help doctors provide targeted treatment. (b): (I) The abnormal myocardial motion caused by myocarditis can occur at different times. (II) The myocarditis lesions are unevenly distributed and irregularlyshaped.

clinical practice to diagnose myocarditis, including T2-weighted imaging, early gadolinium-enhanced imaging, and late gadolinium-enhanced imaging [4]. However, the nephrotoxicity of gadolinium contrast agents and the prolonged scan time limit their clinical application [5]. In contrast, Cine-MRI imaging technology eliminates the need of contrast agents and offers shorter scan time [6], providing a potentially safer and more effective approach for the segmentation of myocarditis lesions, as shown in Fig. 1 (a).

However, due to the inherent technical limitations ingrained in Cine-MRI, it demonstrates insufficient contrast when imaging myocarditis lesions. In recent years, the researches on using deep learning to extract spatiotemporal information from Cine-MRI images for segmenting cardiomyopathy regions has achieved inspiring results [7]. Nevertheless, given the morphological complexity of the myocarditis lesions and spatiotemporal discordance of the myocarditis motion patterns, existing methods might not be directly applicable to the myocarditis segmentation. First, the motion patterns of myocarditis patients exhibit significant spatiotemporal inconsistency: the occurrence of abnormal motion is unpredictable; the compensatory mechanisms of healthy myocardium can mask subtle abnormalities in affected regions, as shown in Fig. 1 (b) I. Then, as shown in Fig. 1 (b) II, due to the infiltration patterns of inflammatory cells (either localized or diffuse) [8], myocardial inflammation lesions typically exhibit irregular shapes and scattered spatial distributions, which pose challenges in learning both the local structure information and global distribution information.

To address the aforementioned challenges, we propose a MG-Mamba model for myocarditis lesion segmentation. We first divides the image into different patches at multiple scales. Then, for Challenge 1, we employ a Multi-step Cross-frame Mamba Motion Analysis (MCMMA) module in each scale. This module designs multi-step cross-frame scanning sequences, utilizing the Mamba network's state space modeling capability to scan inter-frame relationships in varying orders. This approach effectively detects subtle pathological motion patterns caused by myocarditis. For Challenge 2, we propose a Multi-scale Graph

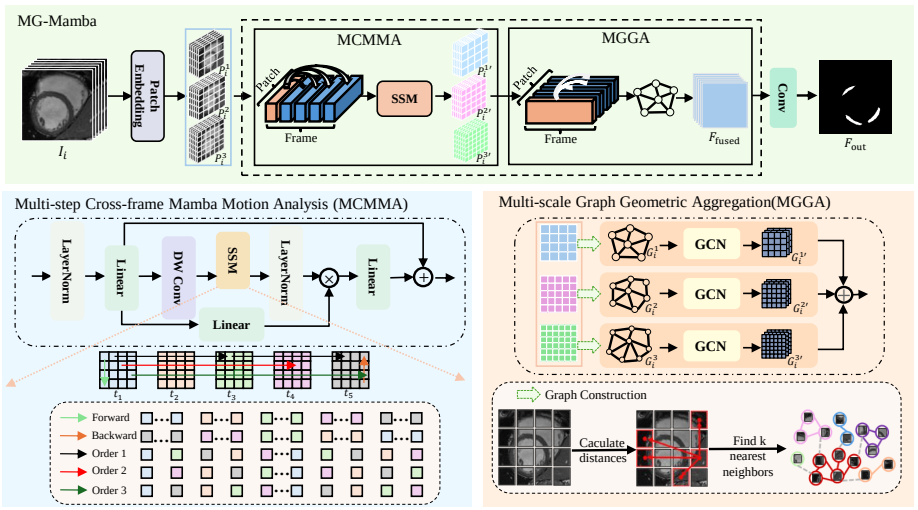

Fig. 2. The proposed MG-Mamba model consists of two main modules: MCMMA and MGGA. The MCMMA module captures abnormal myocardial motion features, while the MGGA module aggregates multi-scale geometric features of the lesion.

Geometric Aggregation (MGGA) module. This module first divides the image into different patches and constructs a graph by computing the Euclidean distances between intra-frame patches. Subsequently, we employ Graph Convolutional Networks (GCN) to aggregate geometric features across different scales, enabling the simultaneous capture of both local structural patterns and global morphological distributions of myocardial lesions.

Our contributions are summarized as follows: 1) We present a clinical method for segmentation of myocarditis lesions using Cine-MRI images. This method has significant clinical potential for myocarditis lesion segmentation without contrast agents and shorter scan times. 2) We propose a MCMMA module to address the spatiotemporal inconsistencies in myocardial motion through multi-step cross-frame state space modeling, and a MGGA module to address the morphology complexity by constructing multi-scale graphs that aggregate global lesion distribution and local geometric information. 3) Extensive experiments conducted on two clinical datasets (10,400 Cine-MRI images in total) demonstrate that the MG-Mamba is effective and superior to ten state-of-the-art methods.

2 Method

Overall Pipeline. Our MG-Mamba framework combines two key components: the Multi-step Cross-frame Mamba Motion Analysis (MCMMA) module (Section 2.1), designed to identify abnormal motion patterns via multi-step cross-frame state space modeling, and the Multi-scale Graph Geometric Aggregation

(MGGA) module (Section 2.2), which extracts lesion geometric features through multi-scale graph learning. The overall workflow is depicted in Fig. 2.

2.1 Multi-step Cross-frame Mamba Motion Analysis (MCMMA)

To detect subtle pathological motion patterns caused by myocarditis, we propose the MCMMA module, which captures motion abnormalities by multi-step cross-frame state space modeling. We first create multi-step inter-frame scanning sequences by spanning varying numbers of frames. Then, we employ Mamba [9] model to scan created sequences extract motion features.

Specifically, for each frame $I_i \in \mathbb{R}^{H\times W}$ (with $i = 1, 2, \ldots, T$), H denotes the height, W denotes the width and T denotes the frames(in this experiment, T = 20.). Then we divide it into image patches at different scales of 8×8, 10×10, and 16×16. Through Patch Embedding (PE) operations, we generate patch embeddings $P_i^1 \in \mathbb{R}^{n_1\times d_1}$, $P_i^2 \in \mathbb{R}^{n_2\times d_2}$, and $P_i^3 \in \mathbb{R}^{n_3\times d_3}$, where n_1, n_2, and n_3 denote the number of patches, and d_1, d_2, and d_3 denote the embedding dimensions.

$$P_i^1 = PE^{8\times 8}(I_i), \quad P_i^2 = PE^{10\times 10}(I_i), \quad P_i^3 = PE^{16\times 16}(I_i) \tag{1}$$

Next, for the multi-scale patches P_i^1, P_i^2, P_i^3, we sequentially process the features $P \in \mathbb{R}^{T\times n\times d}$ obtained from the T frames of data at each scale, where T denotes the number of frames, n denotes the number of patches, d denotes the embedding dimensions.

Afterward, at each scale, we employ the Mamba model to learn inter-frame dependencies of the P_i^j=$\{P_1^j, P_2^j, P_3^j, \ldots, P_T^j\}$. To enhance the motion modeling capability of our block, we introduce a triple-branch architecture that operates in parallel: a forward SSM branch, a backward SSM branch, and an inter-frame SSM branch.

At j-th scale, we first perform forward SSM to scan from the first frame to the T-th frame, while backward SSM to scan from the T-th frame to the first frame.

$$\widetilde{P}_f = SSM(f) : P_1^j \to P_2^j \to \ldots \to P_{T-1}^j \to P_T^j \tag{2}$$

$$\widetilde{P}_b = SSM(b) : P_T^j \to P_{T-1}^j \to \ldots \to P_2^j \to P_1^j \tag{3}$$

For inter-frame SSM, we scan across frames by varying the frame interval ($\Delta_t = 2, 3, 5$):

$$\widetilde{P}_{\Delta_t} = SSM(\Delta_t) : P_{[1][x]}^j \to P_{[1+\Delta_t][x]}^j \to P_{[1+2*\Delta_t][x]}^j \to \ldots P_T^j \tag{4}$$

Where $[a][b]$ indicate the b-th patch embedding in a-th frame.

After, we fuse the output of the forward, backward and inter-frame SSM:

$$X = \mathrm{LN}(\widetilde{P}_f \odot \widetilde{P}_b \odot \widetilde{P}_{\Delta_{t=2}} \odot \widetilde{P}_{\Delta_{t=3}} \odot \widetilde{P}_{\Delta_{t=5}}) \tag{5}$$

where $\odot$ denotes the Hadamard product. Given three different scales, we can get ($X^1 \in \mathbb{R}^{T\times n_1\times d_1}, X^2 \in \mathbb{R}^{T\times n_2\times d_2}, X^3 \in \mathbb{R}^{T\times n_3\times d_3}$).

2.2 Multi-scale Graph Geometric Aggregation(MGGA)

To capture the morphological complexity of myocarditis lesions, we propose the MGGA module, which learns geometric features through multi-scale graph learning. Specifically, we further model the geometric features of the multi-scale embeddings (X^1, X^2, X^3) in each frame. Take t-th frame $X_t^1 = \{x_{t1}^1, x_{t2}^1, ...x_{tn_1}^1\}$ as an example, each patch embedding x_{t1}^1 can be seen as an vertex. Then we compute the Euclidean distance between embeddings within the same frame. Then, we employ the k-nearest neighbor (k-NN) algorithm to identify the k nearest neighbors for each vertex. Based on these neighbors, we construct an intra-patch graph $G_i^j = (X_t^1, E)$, where E denotes the corresponding set of edges, and each vertex $x_{ti}^j \in P$ only connects to its k-nearest neighbors. Then, we use a Graph Convolution Network [10] on the intra-patch graph G_i^j to aggregate geometric features of the lesion regions. It aggregates small local features of the lesion into global distribution features. For three scale space, we can get:

$$G_t^{1'} = GCN(G_t^1), \quad G_t^{2'} = GCN(G_t^2), \quad G_t^{3'} = GCN(G_t^3) \tag{6}$$

where G_i^1, G_i^2, G_i^3 represent the graphs constructed at different scales. $G_i^{1'}$, $G_i^{2'}$, $G_i^{3'}$ denote the outputs of graph convolution.

Finally, we aggregate the feature maps of different scales to obtain the final multi-scale fused feature map F_{fused}. This is achieved by performing a weighted summation of the normalized feature maps $G_i^{1'}$, $G_i^{2'}$, and $G_i^{3'}$, where the weights are learned during training.

$$F_{fused} = w_1 \cdot G_i^{1'} + w_2 \cdot G_i^{2'} + w_3 \cdot G_i^{3'} \tag{7}$$

where w_1, w_2, and w_3 are the learnable weights for each scale.

2.3 Loss function

For the fused feature map F_{fused}, we generate the lesion segmentation map F_{out} through a conv operation with a kernel size of $20 \times 1 \times 1$. Then, we employ combining Cross-entropy loss and Dice loss to optimize our model.

$$L = \alpha \cdot L_{CE} + (1 - \alpha) \cdot L_{Dice} \tag{8}$$

where α is a weighting parameter that balances the influence of the two losses, α is 0.6.

3 Experiments

3.1 Dataset and Implementation Details

Dataset. Our datasets consist of Cine-MRI images from short-axis views of the left ventricle, collected from two hospitals in China and referred to as SYC-QC

and SYC-SX. These datasets include 320 and 200 myocarditis cases, respectively, totaling of 10,400 Cine-MRI images. Due to variations in MRI scanners, the number of frames in cine-sequences varies between 20 and 25. To standardize the data, all sequences were uniformly downsampled to 20 frames at equal intervals. This ensures consistency across different scanners and patients, facilitating feature extraction and model training.

Evaluation Metrics. We employ the Dice score(Dice)[%], Precision(Pre)[%], Recall(Rec)[%] and Hausdorff Distance(HD)[mm] as metrics.

Implementation Details. Our model MG-Mamba is implemented using PyTorch 2.2.1 with CUDA 12.0. All experiments were run on an NVIDIA GeForce RTX 4090 GPU. During training, we used a batch size of 8 per GPU for 200 epochs. Optimization is performed using Adam with a Cosine Annealing with Warm Restarts scheduler(initial learning rate of 1e-3).

Table 1. Comparison with other methods on the SYC-QC and SYC-SX. The best-performing results are highlighted in bold. Natural denotes natural image segmentation methods. Medical denotes medical image segmentation methods.

	Methods	SYC-QC				SYC-SX			
		Dice↑	**Pre↑**	**Rec↑**	**HD↓**	**Dice↑**	**Pre↑**	**Rec↑**	**HD↓**
Natural	OW-VISF(IJCV'24) [11]	74.74	68.62	84.39	5.16	73.56	65.10	86.96	5.85
	Video-kMaX(WACV'24) [12]	73.81	65.94	87.03	5.49	72.15	65.79	82.61	5.49
	TarViS(CVPR'23) [13]	77.28	73.98	82.77	4.66	73.90	66.35	86.39	5.78
	TDSNet(INS'24) [14]	71.01	59.15	**91.97**	7.02	70.32	61.28	86.52	7.83
	CFFM++(TPAMI'24) [15]	78.92	75.42	84.90	5.70	75.47	68.90	86.31	5.01
Medical	CAS-Net(MedIA'23) [16]	78.48	71.64	88.51	6.06	76.62	67.94	**90.23**	5.34
	PolypNext(IJCARS'23) [17]	75.16	70.81	82.87	4.94	74.96	67.03	88.03	6.56
	MemSAM(CVPR'24) [18]	80.74	75.93	87.84	5.64	80.26	75.37	87.62	5.84
	ZePT(CVPR'24) [19]	81.50	79.28	85.77	3.99	80.09	76.31	86.02	4.01
	Vivim(arXiv'24) [20]	83.68	80.97	88.21	4.68	82.41	77.73	89.32	5.01
	Ours	**86.92**	**88.51**	86.54	**3.84**	**86.35**	**87.56**	86.45	**3.99**

3.2 Comparison with State-of-the-art Methods

We validated our model on the SYC-QC and SYC-SX datasets, comparing it with ten state-of-the-art methods, including five natural image segmentation methods and five medical image segmentation methods. We compared our model with OW-VISF [11], Video-kMaX [12], TarViS [13], TDSNet [14] and CFFM++ [15] for natural image segmentation, and with CAS-Net [16], PolypNext [17], MemSAM [18], ZePT [19] and Vivim [20] for medical image segmentation. As shown in Table 1, our method outperforms others on both datasets, achieving a Dice of 86.92% and HD of 3.84 mm on SYC-QC, and a Dice of 86.35% and HD of 3.99 mm on SYC-SX. For visualization, we selected three top-performing methods from each category for comparison, including CFFM++, TarViS, OW-VISF,

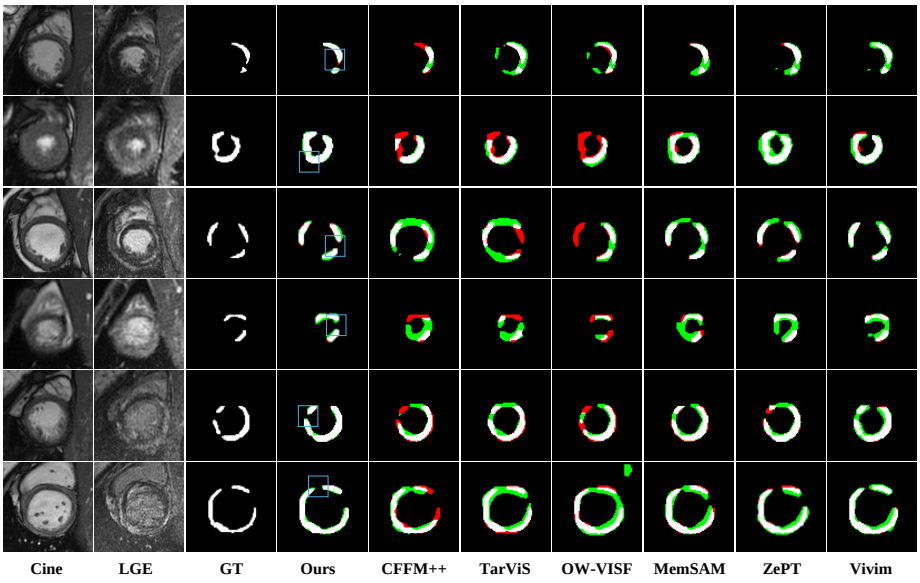

Fig. 3. Qualitative comparison of six different methods on SYC-QC. The green area in the figure indicates false positives, and the red area indicates false negatives. The blue boxed area indicates that we segmented the lesion details better.

MemSAM, ZePT and Vivim. As illustrated in Fig. 3, our model demonstrates superior segmentation performance, particularly in capturing fine details of myocarditis lesions.

3.3 Ablation Studies and Analysis

Effectiveness of different components. We conducted ablation studies on the SYC-QC dataset to validate the contributions of the MCMMA and MGGA modules in MG-Mamba. Three model versions were compared: M1 (backbone), M2 (backbone + MCMMA), and M3 (backbone + MGGA). Results in Table 2 show that both M2 and M3 outperform M1, while the MG-Mamba model (MCMMA + MGGA) achieves the best performance across all metrics. This demonstrates the complementary and synergistic effects of MCMMA and MGGA in enhancing lesion segmentation accuracy and robustness.

Feature Visualization Analysis. To further validate the effectiveness of our proposed MG-Mamba model, we perform feature visualization analysis on this model and several comparison models with better performance. In the experiment, we use the t-SNE technique to reduce the dimensions of normal and lesion patches in the features for visualization. The experimental results are shown in Fig. 4 (a). We can see that our MG-Mamba model exhibits clearer class separation and tighter intra-class clustering compared to other models, demonstrating its superior discriminative capability.

Table 2. Ablation Studies of different components of our method.

Methods	Modules		SYC-QC			
	MCMMA	MGGA	**Dice↑**	**Pre↑**	**Rec↑**	**HD↓**
M1	-	-	82.64	81.58	85.23	4.48
M2	✓	-	$83.61_{\uparrow 0.97}$	$82.56_{\uparrow 0.98}$	$86.28_{\uparrow 1.05}$	$4.03_{\downarrow 0.45}$
M3	-	✓	$84.49_{\uparrow 1.85}$	$84.90_{\uparrow 3.32}$	$85.61_{\uparrow 0.38}$	$3.92_{\downarrow 0.56}$
Ours	✓	✓	$\mathbf{86.92}_{\uparrow 4.28}$	$\mathbf{88.51}_{\uparrow 6.93}$	$\mathbf{86.54}_{\uparrow 1.31}$	$\mathbf{3.84}_{\downarrow 0.64}$

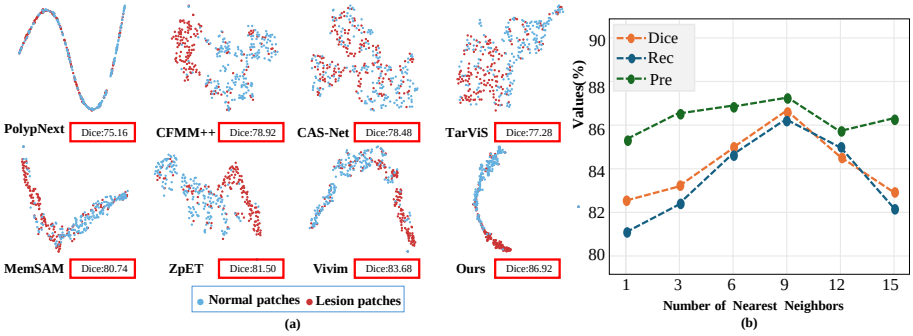

Fig. 4. (a) shows the t-SNE results of our MG-Mamba model and other comparative models. (b) shows the segmentation results of MG-Mamba with respect to different numbers k of nearest neighbors in k-NN graph.

Stability Analysis. In the MG-Mamba model, the neighborhood size K of the k-NN algorithm influences the learning of geometric features of myocarditis lesions. To evaluate the effect of different values of K on the model performance, we trained the model on the SYC-QC dataset with K values set to {1, 3, 6, 9, 12, 15}. As shown in Fig. 4 (b), the model's performance initially improves and then declines as K increases, peaking at K=9. This value provides an optimal local space for capturing detailed geometric information.

4 Conclusion.

In our work, we propose the MG-Mamba framework for myocarditis lesion segmentation in Cine-MRI images. The framework integrates two key modules: MCMMA for multi-step cross-frame motion analysis to address spatiotemporal inconsistencies, and MGGA for multi-scale graph geometric aggregation to capture both local structural patterns and global lesion distributions. Experimental results demonstrate significant performance gains over state-of-the-art methods, highlighting the effectiveness of our approach.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (No.62376002, No.62101491, No.62401016), in part by the Foundation of Sichuan Provincial People's Hospital (Grant Number 2023QN25).

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

References

1. Ratko Lasica, Lazar Djukanovic, Lidija Savic, Gordana Krljanac, Marija Zdravkovic, Marko Ristic, Andjelka Lasica, Milika Asanin, and Arsen Ristic. Update on myocarditis: From etiology and clinical picture to modern diagnostics and methods of treatment. *Diagnostics*, 13(19):3073, 2023.
2. Michael A Seidman and Bruce McManus. Myocarditis. *Cardiovascular pathology*, pages 553–575, 2022.
3. Emil Brociek, Agata Tymińska, Andrea Silvio Giordani, Alida Linda Patrizia Caforio, Romuald Wojnicz, Marcin Grabowski, and Krzysztof Ozierański. Myocarditis: etiology, pathogenesis, and their implications in clinical practice. *Biology*, 12(6):874, 2023.
4. Julian A Luetkens, Anton Faron, Alexander Isaak, Darius Dabir, Daniel Kuetting, Andreas Feisst, Frederic C Schmeel, Alois M Sprinkart, and Daniel Thomas. Comparison of original and 2018 lake louise criteria for diagnosis of acute myocarditis: results of a validation cohort. *Radiology: Cardiothoracic Imaging*, 1(3):e190010, 2019.
5. Susana Coimbra, Susana Rocha, Nícia Reis Sousa, Cristina Catarino, Luís Belo, Elsa Bronze-da Rocha, Maria João Valente, and Alice Santos-Silva. Toxicity mechanisms of gadolinium and gadolinium-based contrast agents—a review. *International Journal of Molecular Sciences*, 25(7):4071, 2024.
6. David C Wendell and Robert M Judd. Cardiac cine imaging. *Basic Principles of Cardiovascular MRI: Physics and Imaging Technique*, pages 145–159, 2015.
7. Fabian Isensee, Paul F Jaeger, Peter M Full, Ivo Wolf, Sandy Engelhardt, and Klaus H Maier-Hein. Automatic cardiac disease assessment on cine-mri via time-series segmentation and domain specific features. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8*, pages 120–129. Springer, 2018.
8. Loïc Bière, Nicolas Piriou, Laura Ernande, François Rouzet, and Olivier Lairez. Imaging of myocarditis and inflammatory cardiomyopathies. *Archives of cardiovascular diseases*, 112(10):630–641, 2019.
9. Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.
10. Si Zhang, Hanghang Tong, Jiejun Xu, and Ross Maciejewski. Graph convolutional networks: a comprehensive review. *Computational Social Networks*, 6(1):1–23, 2019.
11. Omkar Thawakar, Sanath Narayan, Hisham Cholakkal, Rao Muhammad Anwer, Salman Khan, Jorma Laaksonen, Mubarak Shah, and Fahad Shahbaz Khan. Video instance segmentation in an open-world. *International Journal of Computer Vision*, pages 1–12, 2024.
12. Inkyu Shin, Dahun Kim, Qihang Yu, Jun Xie, Hong-Seok Kim, Bradley Green, In So Kweon, Kuk-Jin Yoon, and Liang-Chieh Chen. Video-kmax: A simple unified approach for online and near-online video panoptic segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 229–239, 2024.

13. Ali Athar, Alexander Hermans, Jonathon Luiten, Deva Ramanan, and Bastian Leibe. Tarvis: A unified approach for target-based video segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18738–18748, 2023.
14. Haochen Yuan, Junjie Peng, and Zesu Cai. Tdsnet: A temporal difference based network for video semantic segmentation. *Information Sciences*, 686:121335, 2024.
15. Guolei Sun, Yun Liu, Henghui Ding, Min Wu, and Luc Van Gool. Learning local and global temporal contexts for video semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
16. Caixia Dong, Songhua Xu, Duwei Dai, Yizhi Zhang, Chunyan Zhang, and Zongfang Li. A novel multi-attention, multi-scale 3d deep network for coronary artery segmentation. *Medical Image Analysis*, 85:102745, 2023.
17. Debayan Bhattacharya, Konrad Reuter, Finn Behrendt, Lennart Maack, Sarah Grube, and Alexander Schlaefer. Polypnextlstm: a lightweight and fast polyp video segmentation network using convnext and convlstm. *International journal of computer assisted radiology and surgery*, 19(10):2111–2119, 2024.
18. Xiaolong Deng, Huisi Wu, Runhao Zeng, and Jing Qin. Memsam: Taming segment anything model for echocardiography video segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9622–9631, 2024.
19. Yankai Jiang, Zhongzhen Huang, Rongzhao Zhang, Xiaofan Zhang, and Shaoting Zhang. Zept: Zero-shot pan-tumor segmentation via query-disentangling and self-prompting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11386–11397, 2024.
20. Yijun Yang, Zhaohu Xing, and Lei Zhu. Vivim: a video vision mamba for medical video object segmentation. *arXiv preprint arXiv:2401.14168*, 2024.