DINO-CMR: DINOv3-Driven View-Hybrid Segmentation with Decoupled Measurement for Multi-Sequence CMR

Xuan Zu[1], Qing Xu[1,2†], Xiangjian He[1(✉)], and Zhen Chen[3(✉)]

[1] School of Computer Science, University of Nottingham Ningbo China, China
[2] School of Computer Science, University of Nottingham, UK
[3] Department of Data Science and Artificial Intelligence, The Hong Kong Polytechnic University, Hong Kong SAR
scxqx1@nottingham.ac.uk, sean.he@nottingham.edu.cn
z.chen@polyu.edu.hk

Abstract. Cardiac magnetic resonance (CMR) segmentation is critical for quantitative assessment of ventricular function and myocardial tissue characterization, yet robust segmentation remains challenging due to variations across sequences, imaging planes, and acquisition centers. The MICCAI 2026 Universal Multi-Sequence, Multi-Center and Multi-View CMR Segmentation Challenge further requires segmentation masks and clinical measurements to be produced across cine (CINE) and late gadolinium enhancement (LGE) data with different view-specific label spaces. To address this challenge, we propose DINO-CMR, a task-specific framework that combines volumetric CINE segmentation, view-hybrid LGE segmentation, and decoupled measurement estimation. The CINE branch employs a masked-autoencoder-pretrained MedNeXt-S for volumetric segmentation with phase-aware ejection fraction estimation. The LGE branch selects the strongest segmentation source for each view and retains scar-mass estimates from a separate DINOv3-based path when segmentation and measurement objectives favor different models. Five-fold ensembling with test-time augmentation is applied before generating final masks and clinical measurements. On the challenge validation set, DINO-CMR **(Team: NPM)** achieves task scores of 0.6477 for CINE and 0.6937 for LGE, with an overall score of 0.6707.

Keywords: Cardiac MRI segmentation · multi-sequence multi-view · view-hybrid assignment · decoupled measurement estimation

1 Introduction

Cardiac magnetic resonance (CMR) is fundamental to the quantitative assessment of ventricular function and myocardial tissue characterization, providing critical measurements such as ejection fraction and scar burden for clinical

† Team Leader.

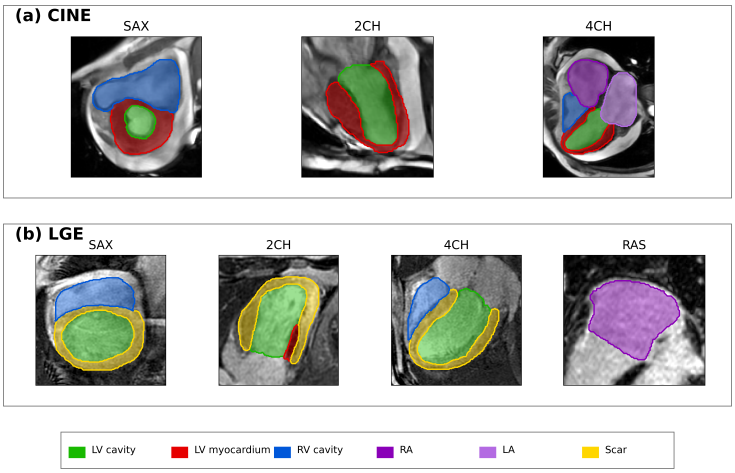

Fig. 1. Representative Challenge data. CINE includes short-axis (SAX), two-chamber (2CH), and four-chamber (4CH) views for cardiac structure segmentation and EF estimation. LGE includes SAX, 2CH, 4CH, and RAS views for chamber and scar assessment.

decision-making [22,27,2,1]. However, robust CMR segmentation remains inherently challenging due to the substantial variability in cardiac appearance across scanners, acquisition centers, imaging sequences, cardiac phases, and imaging planes. Earlier benchmarks have demonstrated that even high-performing methods can degrade substantially when evaluated across multiple anatomical structures, acquisition settings, centers, and disease groups [3,4,13].

Encoder-decoder architectures such as 3D U-Net [6], and V-Net [17] have established the foundation for medical image segmentation in both two-dimensional and volumetric settings. Building upon this paradigm, UNet++ [26], ResUNet++ [11], and nnU-Net [10] improved skip-connection design, residual decoding, and task-specific pipeline configuration. More recently, MedNeXt [20] adapted ConvNeXt-style blocks [15] to volumetric medical images, and masked autoencoder (MAE) pretraining [9] has emerged as a practical approach to learn representations from unlabeled images before supervised fine-tuning. Transformer-based [12,25,16] and self-supervised models have also been explored for medical segmentation and representation learning [5,8,24]. In parallel, DINOv3 [23] provides transferable dense features through large-scale self-supervised pretraining, offering strong visual representations for downstream segmentation tasks.

Despite these advancements, existing methods still face two fundamental challenges in multi-sequence CMR analysis. The difficulty is particularly evident for myocardial pathology analysis, where scar regions are small, heterogeneous, and sensitive to sequence alignment and measurement rules [13,18,7]. First, different views and sequences require distinct label spaces and segmentation strategies, making a one-size-fits-all architecture suboptimal. Second, the model that produces the best segmentation masks does not necessarily yield the most reliable clinical measurements, as segmentation quality and measurement sta-

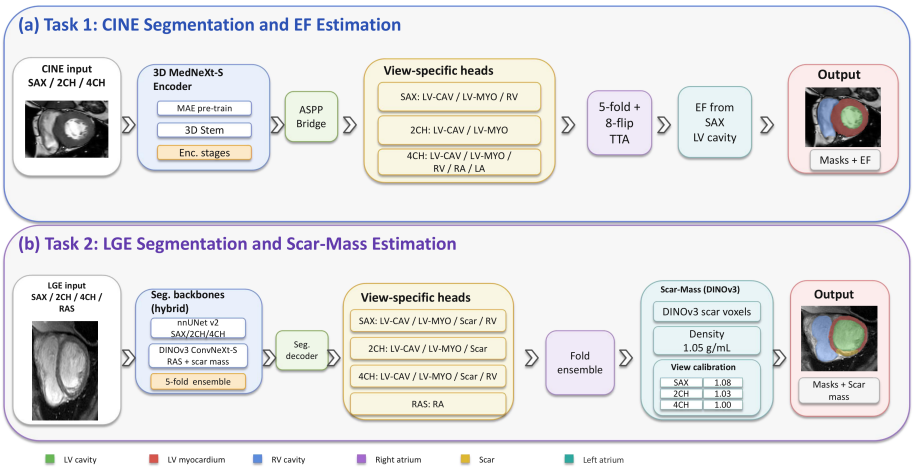

Fig. 2. Overview of DINO-CMR. The CINE branch uses MAE-pretrained MedNeXt-S for volumetric segmentation, followed by five-fold and eight-flip test-time ensembling and EF estimation from SAX LV-cavity volumes. The LGE branch uses a view-hybrid segmentation strategy, where SAX, 2CH, and 4CH masks are produced by nnU-Net and the RAS mask is retained from the DINOv3 ConvNeXt-S branch. Scar mass is estimated through a separate DINOv3-based conversion path with fixed view-wise calibration.

bility may favor different sources. The MICCAI 2026 Universal Multi-Sequence, Multi-Center and Multi-View CMR Segmentation Challenge [19] formalizes these requirements by evaluating both cine (CINE) segmentation with ejection fraction (EF) estimation and late gadolinium enhancement (LGE) segmentation with scar-mass estimation across multiple cardiac views. Representative CINE and LGE examples are shown in Fig. 1. This motivates us to develop a task-specific framework that assigns modality-appropriate models to each sequence and decouples segmentation from clinical measurement estimation when the two objectives diverge.

To this end, we propose DINO-CMR, a task-specific framework for multi-sequence and multi-view CMR segmentation. The CINE branch employs an MAE-pretrained MedNeXt-S for volumetric segmentation with phase-aware EF estimation. The LGE branch adopts a view-hybrid strategy that selects the strongest segmentation source per view, while retaining scar-mass estimates from a separate DINOv3 ConvNeXt-S path that provides more stable measurement statistics. Five-fold ensembling with test-time augmentation is applied before generating final masks and clinical measurements. Evaluated on the challenge validation set, DINO-CMR achieves task scores of 0.6477 for CINE and 0.6937 for LGE, with an overall score of 0.6707.

2 Methodology

As illustrated in Fig. 2, the proposed DINO-CMR comprises two task-specific branches: a CINE branch for volumetric cardiac structure segmentation and EF estimation, and an LGE branch for view-hybrid chamber and scar segmentation with decoupled scar-mass estimation. The two branches share a unified inference logic—view-specific prediction, fold ensembling, and deterministic measurement conversion—but maintain independent segmentation backbones to accommodate their distinct label spaces and clinical output requirements.

2.1 CINE Segmentation and EF Estimation

Ejection fraction estimation depends on phase-wise left ventricular (LV) cavity volumes, which are most naturally derived from volumetric short-axis (SAX) predictions. This motivates the use of a three-dimensional backbone for the CINE branch. Each cardiac phase is processed as a volume $x \in \mathbb{R}^{1\times D\times H\times W}$ with z-score normalization. We adopt MedNeXt-S [20] as the encoder-decoder backbone, whose encoder is initialized via MAE pretraining [9] on the Challenge SAX CINE training images to learn structural representations without additional labeled data.

The output space is view-dependent: SAX predicts LV cavity, LV myocardium, and RV cavity; two-chamber (2CH) predicts LV cavity and LV myocardium; four-chamber (4CH) additionally predicts RV, right atrium (RA), and left atrium (LA). Five-fold predictions are combined with eight-flip test-time augmentation by averaging class logits:

$$\hat{y} = \arg\max_{c} \frac{1}{KT} \sum_{k=1}^{K} \sum_{t=1}^{T} \mathcal{T}_t^{-1}\left(z_{k,t}^{(c)}\right), \qquad K = 5,\ T = 8, \tag{1}$$

where $z_{k,t}^{(c)}$ is the logit for class c from fold k under augmentation t, and $\mathcal{T}_t^{-1}$ maps the augmented prediction back to the original space. A connected-component step removes small isolated regions from the ensembled mask.

EF is derived from the ensembled SAX LV-cavity predictions. Voxel counts are converted to physical volume using image spacing, and end-diastolic volume (EDV) and end-systolic volume (ESV) are identified from the valid phase sequence:

$$\mathrm{EF} = \frac{\mathrm{EDV} - \mathrm{ESV}}{\mathrm{EDV}} \times 100. \tag{2}$$

EF is computed once from the final ensembled mask rather than aggregated from fold-specific values, ensuring measurement consistency.

2.2 LGE Segmentation and Scar-Mass Estimation

LGE segmentation requires handling multiple views with distinct label definitions, and development experiments revealed that the model producing the best

Table 1. Official validation results of DINO-CMR on the challenge validation set.

Task	Score↑	DSC↑	HD↓	ASD↓	EF PCC↑	Mass vPCC↑	Mass RAE↓
Overall	0.6707	-	-	-	-	-	-
CINE	0.6477	0.8509	9.5104	0.8458	0.9188	-	-
LGE	0.6937	0.6699	-	-	-	0.8385	0.5161

segmentation masks does not necessarily yield the most stable scar-mass statistics. To address both requirements, we adopt a view-hybrid segmentation strategy with decoupled mass estimation.

For segmentation, SAX, 2CH, and 4CH masks are produced by view-specific 2D nnU-Net models [10], which provide stronger overlap-based accuracy for the main cardiac views. The RAS mask is retained from the DINOv3 ConvNeXt-S branch [23,15], which offers more reliable atrial predictions. SAX and 4CH contain LV cavity, LV myocardium, scar, and RV cavity; 2CH contains LV cavity, LV myocardium, and scar; RAS contains RA. Fold logits are combined before producing the final label map.

Scar mass is deliberately not recomputed from the view-hybrid masks. Instead, mass estimates are retained from the DINOv3 ConvNeXt-S LGE branch, which exhibited more stable mass correlation during development. For a scar voxel count N_{scar} and spacing (s_x, s_y, s_z), the uncorrected mass is:

$$M_{\text{scar}} = N_{\text{scar}} s_x s_y s_z \, \rho / 1000, \tag{3}$$

with myocardial tissue density $\rho = 1.05$ g/mL. Fixed view-wise calibration factors of 1.08, 1.03, and 1.00 are applied for SAX, 2CH, and 4CH, respectively. These constants were determined from view-level mass bias observed during development and are not learned parameters. In this way, segmentation and clinical measurement estimation are decoupled, allowing each to use its most reliable source.

2.3 Training Optimization

The CINE model is optimized with cross-entropy and soft Dice losses:

$$\mathcal{L}_{\text{cine}} = \mathcal{L}_{\text{CE}} + (1 - \text{Dice}). \tag{4}$$

We use Adam with a learning rate of 10^{-4} and weight decay of 10^{-4}, followed by ReduceLROnPlateau scheduling. Training runs for 60 epochs with batch size 1.

The DINOv3 LGE branch is trained in two stages. Phase A uses a one-vs-rest combination of binary cross-entropy, focal, and Dice losses:

$$\mathcal{L}^{A}_{\text{lge}} = 0.5\mathcal{L}_{\text{BCE}} + 0.4\mathcal{L}_{\text{Focal}}(\gamma = 2.0) + 0.5\mathcal{L}_{\text{Dice}}. \tag{5}$$

Phase B replaces the Dice term with Tversky loss [21] ($\alpha = 0.3$, $\beta = 0.7$) to emphasize recall of missed scar regions:

$$\mathcal{L}^{B}_{\text{lge}} = 0.5\mathcal{L}_{\text{BCE}} + 0.4\mathcal{L}_{\text{Focal}}(\gamma = 2.0) + 0.5\mathcal{L}_{\text{Tversky}}(\alpha = 0.3, \beta = 0.7). \tag{6}$$

Table 2. Selected entries from the final validation leaderboard. Scores are rounded as displayed by the challenge server.

Rank	Team	Overall	T1	CINE DSC	CINE HD	CINE ASD	EF PCC	T2	LGE DSC	Mass vPCC	Mass RAE
1	TsingHK MedVision	0.91	0.84	0.99	2.72	0.03	1.00	0.98	0.97	1.00	0.00
2	xiaoji	0.88	0.77	0.95	5.58	0.20	0.97	0.99	0.98	1.00	0.00
3	cemrg	0.88	0.76	0.95	6.06	0.24	1.00	0.99	0.98	1.00	0.00
4	kateneuro	0.77	0.72	0.91	8.05	0.46	1.00	0.81	0.74	1.00	0.00
5	adinath	0.74	0.70	0.90	9.64	0.49	0.97	0.79	0.77	0.91	0.37
				...							
16	hnaqvi	0.67	0.70	0.91	8.11	0.47	0.94	0.65	0.73	0.52	1.83
17	**NPM (Ours)**	**0.67**	**0.65**	**0.85**	**9.51**	**0.85**	**0.92**	**0.69**	**0.67**	**0.84**	**0.52**
18	tobi_hao	0.67	0.65	0.87	11.14	0.79	0.91	0.68	0.75	0.74	2.25

Table 3. Component development analysis of DINO-CMR. $\mathcal{M}$: MAE-pretrained CINE; $\mathcal{T}$: TTA and post-processing; $\mathcal{N}$: nnU-Net LGE masks; $\mathcal{D}$: decoupled DINOv3 mass; $\mathcal{V}$: view-hybrid selection. Each row corresponds to a server-evaluated development configuration rather than a strict controlled ablation.

	Components							Overall		CINE (Task 1)				LGE (Task 2)			
	$\mathcal{M}$	$\mathcal{T}$	$\mathcal{N}$	$\mathcal{D}$	$\mathcal{V}$					CINE			EF		LGE	Mass	Mass
							Overall↑	T1↑	DSC↑	HD↓	ASD↓	PCC↑	T2↑	DSC↑	vPCC↑	RAE↓
$\mathcal{C}_1$							0.6577	0.6279	0.8484	11.14	0.9788	0.8898	0.6875	0.6611	0.8385	0.5161
$\mathcal{C}_2$	✓						0.6661	0.6446	0.8505	9.95	0.8434	0.9114	0.6875	0.6611	0.8385	0.5161
$\mathcal{C}_3$	✓	✓					0.6676	0.6477	0.8509	9.51	0.8458	0.9188	0.6875	0.6611	0.8385	0.5161
$\mathcal{C}_4$	✓	✓	✓				0.6596	0.6477	0.8509	9.51	0.8458	0.9188	0.6715	0.6690	0.7462	0.6441
$\mathcal{C}_5$	✓	✓	✓	✓			0.6703	0.6477	0.8509	9.51	0.8458	0.9188	0.6930	0.6690	0.8385	0.5161
$\mathcal{C}_6$	✓	✓	✓	✓	✓		**0.6707**	0.6477	0.8509	9.51	0.8458	0.9188	**0.6937**	**0.6699**	0.8385	0.5161

Focal loss [14] addresses class imbalance across cardiac structures. The nnU-Net models follow view-specific 2D configurations with the standard training protocol [10]. No additional medical dataset is used: MAE pretraining relies only on Challenge CINE images, and the DINOv3 branch starts from publicly available non-medical pretrained weights.

3 Experiments

3.1 Experimental Setup

We evaluate DINO-CMR on the MICCAI 2026 Universal Multi-Sequence, Multi-Center and Multi-View CMR Segmentation Challenge validation set. All experiments are conducted on an NVIDIA RTX A5000 GPU. Task 1 provides 105 training and 15 validation cases per CINE view (SAX, 2CH, 4CH). Task 2 provides 40 training and 7 validation cases per LGE view (SAX, 2CH, 4CH), with 80 training and 14 validation cases for RAS. CINE is evaluated by Dice similarity coefficient (DSC), Hausdorff distance (HD), average surface distance (ASD), and EF Pearson correlation coefficient (PCC). LGE is evaluated by DSC, mass-vector PCC (vPCC), and mass relative absolute error (RAE). The official

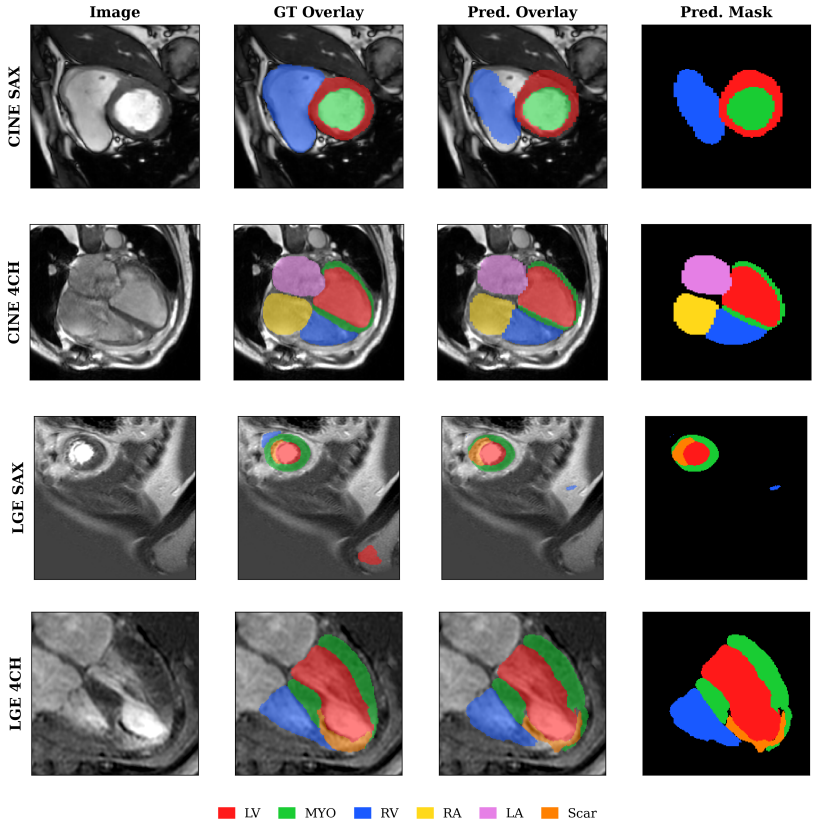

Fig. 3. Qualitative comparison of CINE and LGE results. Columns: input image, ground-truth overlay, predicted overlay, and predicted mask.

task scores are:

$$S_1 = 0.7\left(0.4\,\mathrm{DSC} + \frac{0.3}{1+\mathrm{HD}} + \frac{0.3}{1+\mathrm{ASD}}\right) + 0.3\max(0, \mathrm{EF\ PCC}), \tag{7}$$

$$S_2 = 0.7\,\mathrm{DSC} + 0.3\left(0.5\max(0, \mathrm{vPCC}) + \frac{0.5}{1+\mathrm{RAE}}\right), \tag{8}$$

and the overall score is $(S_1 + S_2)/2$. All reported scores are from the official challenge server.

3.2 Validation Results

As shown in Table 1, DINO-CMR achieves an overall score of 0.6707. The CINE branch obtains a task score of 0.6477 with a DSC of 0.8509 and an EF PCC of 0.9188, demonstrating balanced improvement in both segmentation accuracy and functional estimation. The LGE branch achieves a task score of 0.6937 with a DSC of 0.6699 and a mass vPCC of 0.8385. Although LGE DSC remains

lower than CINE DSC, the decoupled DINOv3 mass path preserves strong mass correlation and avoids the degradation observed when mass is recomputed from alternative masks.

3.3 Challenge Leaderboard

Table 2 presents selected entries from the final validation leaderboard. DINO-CMR ranks 17th among all participating teams with an overall score of 0.67.

Figure 3 presents representative qualitative results. The CINE predictions preserve principal chamber boundaries in SAX and 4CH views. The LGE examples reveal remaining challenges: small scar regions and weak myocardial boundaries can still be missed or fragmented.

3.4 Component Development Analysis

To validate the contribution of each component, Table 3 presents the progressive development configurations evaluated on the challenge server. Starting from the DINOv3 baseline ($\mathcal{C}_1$), replacing the CINE branch with MAE-pretrained MedNeXt-S ($\mathcal{C}_2$) raises the overall score from 0.6577 to 0.6661, with HD decreasing from 11.14 to 9.95 and EF PCC improving from 0.8898 to 0.9114. This confirms that volumetric pretraining better preserves phase-wise structural consistency, which directly benefits functional estimation. Adding test-time augmentation and post-processing ($\mathcal{C}_3$) further improves the CINE task score to 0.6477, primarily through reduced boundary deviation.

For the LGE branch, introducing nnU-Net masks without decoupling mass estimation ($\mathcal{C}_4$) actually reduces the overall score from 0.6676 to 0.6596. Although LGE DSC improves from 0.6611 to 0.6690, mass vPCC drops sharply from 0.8385 to 0.7462, indicating that models optimized for voxel-level overlap do not necessarily produce masks whose scar statistics are well-calibrated for clinical measurement. This confirms our core observation: the best segmentation source does not necessarily yield the most reliable clinical measurements. Restoring the decoupled DINOv3 mass path ($\mathcal{C}_5$) immediately recovers mass vPCC to 0.8385, raising the overall score to 0.6703. The final view-level mask selection ($\mathcal{C}_6$) further increases LGE DSC from 0.6690 to 0.6699, producing the reported score of 0.6707. These results collectively demonstrate that CINE benefits from stronger volumetric modeling, while LGE benefits from allowing segmentation and measurement estimation to draw on different sources.

4 Conclusion

In this work, we present DINO-CMR, a DINOv3-driven task-specific framework for multi-sequence and multi-view CMR segmentation. The CINE branch employs MAE-pretrained MedNeXt-S for volumetric segmentation with phase-aware ejection fraction estimation. The LGE branch adopts a view-hybrid strategy that selects the strongest segmentation source per view, while decoupling scar-mass

estimation into a separate DINOv3-based path that preserves more stable measurement statistics. The component development analysis confirms that the best segmentation source and the best clinical measurement source are not always the same: nnU-Net masks improved overlap-based metrics, whereas the DINOv3 branch maintained stronger mass correlation. This observation supports the principle of task-specific model selection over a single unified architecture. Evaluated on the challenge validation set, DINO-CMR achieves an overall score of 0.6707, demonstrating that modality-specific modeling with decoupled measurement estimation provides a practical strategy for multi-sequence CMR analysis.

References

1. Bai, J., Qiu, R., Chen, J., Wang, L., Li, L., Tian, Y., Wang, H., Lu, Y., Zhao, J.: A two-stage method with a shared 3d u-net for left atrial segmentation of late gadolinium-enhanced mri images. Cardiovascular Innovations and Applications **8**(1), 976 (2023)
2. Bai, J., Zhu, J., Chen, Z., Yang, Z., Lu, Y., Li, L., Li, Q., Wang, W., Zhang, H., Wang, K., et al.: A benchmark framework for the right atrium cavity segmentation from lge-mris. IEEE Transactions on Medical Imaging (2025)
3. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Gonzalez Ballester, M.A., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE Trans. Med. Imaging **37**(11), 2514–2525 (2018)
4. Campello, V.M., Gkontra, P., Izquierdo, C., Martín-Isla, C., Sojoudi, A., Full, P.M., Maier-Hein, K., Zhang, Y., He, Z., Ma, J., et al.: Multi-centre, multi-vendor and multi-disease cardiac segmentation: The M&Ms challenge. IEEE Trans. Med. Imaging **40**(12), 3543–3554 (2021)
5. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. Med. Image Anal. **97**, 103280 (2024)
6. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: MICCAI. pp. 424–432 (2016)
7. Ding, W., Li, L., Qiu, J., Wang, S., Huang, L., Chen, Y., Yang, S., Zhuang, X.: Aligning multi-sequence CMR towards fully automated myocardial pathology segmentation. IEEE Trans. Med. Imaging **42**(12), 3474–3486 (2023)
8. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: Brainlesion. pp. 272–284 (2022)
9. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: CVPR. pp. 15979–15988 (2022)
10. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2021)
11. Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., de Lange, T., Halvorsen, P., Johansen, H.D.: ResUNet++: An advanced architecture for medical image segmentation. In: ISM. pp. 225–230 (2019)

12. Jiao, Y., Xu, Q., Luo, Y., He, X., Chen, Z., Duan, W.: Tm-unet: Token-memory enhanced sequential modeling for efficient medical image segmentation. In: 2026 IEEE 23rd International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2026)
13. Li, L., Wu, F., Wang, S., Luo, X., Martin-Isla, C., Zhai, S., Zhang, J., Liu, Y., Zhang, Z., Ankenbrand, M.J., et al.: MyoPS: A benchmark of myocardial pathology segmentation combining three-sequence cardiac magnetic resonance images. Med. Image Anal. **87**, 102808 (2023)
14. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: ICCV. pp. 2980–2988 (2017)
15. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A ConvNet for the 2020s. In: CVPR. pp. 11976–11986 (2022)
16. Luo, Y., Xu, Q., Huang, H., Ouyang, Y., Chen, Z., Duan, W.: Msm-seg: A modality-and-slice memory framework with category-agnostic prompting for multi-modal brain tumor segmentation. arXiv preprint arXiv:2510.10679 (2025)
17. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3DV. pp. 565–571 (2016)
18. Qiu, J., Li, L., Wang, S., Zhang, K., Chen, Y., Yang, S., Zhuang, X.: MyoPS-Net: Myocardial pathology segmentation with flexible combination of multi-sequence CMR images. Med. Image Anal. **84**, 102694 (2023)
19. Qu, T., Zhang, H., Zhang, L., Zhao, C., Zhang, N., Wang, H., Zhou, Z., Zou, M., Bo, K., Zhao, P., Jin, X., Su, Z., Jiang, K., Liu, H., Du, Y., Wang, M., Yan, R., Wang, Z., Huang, T., Xu, L., Zhang, H.: BAAI Cardiac Agent: An intelligent multimodal agent for automated reasoning and diagnosis of cardiovascular diseases from cardiac magnetic resonance imaging. arXiv preprint arXiv:2604.04078 (2026)
20. Roy, S., Koehler, G., Ulrich, C., Baumgartner, M., Petersen, J., Isensee, F., Jaeger, P.F., Maier-Hein, K.H.: MedNeXt: Transformer-driven scaling of ConvNets for medical image segmentation. In: MICCAI. pp. 405–415 (2023)
21. Salehi, S.S.M., Erdogmus, D., Gholipour, A.: Tversky loss function for image segmentation using 3D fully convolutional deep networks. In: MLMI. pp. 379–387 (2017)
22. Schulz-Menger, J., Bluemke, D.A., Bremerich, J., Flamm, S.D., Fogel, M.A., Friedrich, M.G., Kim, R.J., von Knobelsdorff-Brenkenhoff, F., Kramer, C.M., Pennell, D.J., Plein, S., Nagel, E.: Standardized image interpretation and post-processing in cardiovascular magnetic resonance – 2020 update. J. Cardiovasc. Magn. Reson. **22**(1), 19 (2020)
23. Simeoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: DINOv3. arXiv preprint arXiv:2508.10104 (2025)
24. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-supervised pre-training of swin transformers for 3D medical image analysis. In: CVPR. pp. 20730–20740 (2022)
25. Xu, Q., Yuan, K., Luo, Y., Zhai, Y., Duan, W., Navab, N., Chen, Z.: Lapfm: A laparoscopic segmentation foundation model via hierarchical concept evolving pre-training. arXiv preprint arXiv:2512.08439 (2025)
26. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: UNet++: A nested U-Net architecture for medical image segmentation. In: DLMIA/ML-CDS. pp. 3–11 (2018)
27. Zhu, J., Bai, J., Zhou, Z., Liang, Y., Chen, Z., Chen, X., Zhang, X.: Ras dataset: a 3d cardiac lge-mri dataset for segmentation of right atrial cavity. Scientific Data **11**(1), 401 (2024)