


No Image, No Problem: End-to-End Multi-Task Cardiac Analysis Straight from Undersampled k-Space

Yundi Zhang^{1,2} 

YUNDI.ZHANG@TUM.DE

Sevgi Gokce Kafali^{1,2} 

S.KAFALI@TUM.DE

Niklas Bubeck^{1,4} 

NIKLAS.BUBECK@TUM.DE

Daniel Rueckert^{1,2,3,4} 

DANIEL.RUECKERT@TUM.DE

Jiazhen Pan^{1,2} 

JIAZHEN.PAN@TUM.DE

¹ Chair for AI in Healthcare and Medicine, Technical University of Munich, Germany

² TUM University Hospital, Munich, Germany

³ Biomedical Image Analysis Group, Department of Computing, Imperial College London

⁴ Munich Center for Machine Learning, Technical University of Munich, Germany

Editors: Under Review for MIDL 2026

Abstract

Before a cardiac MR image is reconstructed, the heart is represented in k-space, which encodes all information needed for analysis—including tissue structure, motion, and functional dynamics. However, information extraction (e.g., downstream analysis tasks such as segmentation or biomarker quantification) is usually performed in the image domain rather than in the k-space domain. This means that the quality of the information extraction is fundamentally limited by the image reconstruction quality. At the same time, the push toward unified models for diverse cardiac downstream tasks has accelerated, driven by advances in efficient representation learning. However, most work has focused on the image domain, overlooking the k-space potential as a direct, information-dense source for end-to-end, multi-task cardiac analysis. As a result, the development of robust and expressive k-space representations and their impact on downstream cardiac assessment remain significantly underexplored. To address this gap, we propose **k-space Multi-Task Representation (k-MTR)** learning, which enables solving different downstream tasks directly from undersampled k-space. By aligning the k-space and image-domain embeddings, k-MTR establishes a unified representation that simultaneously captures local anatomical detail, global spectral structure, and rich physiological signatures. We show that k-MTR matches or exceeds state-of-the-art image-based and k-space-based baselines across three clinically relevant tasks—disease classification, phenotype regression, and segmentation, providing the first systematic evidence that k-space alone can support comprehensive cardiac analysis. k-MTR represents a pivotal step toward scalable, reconstruction-free cardiac foundation models. The code will be made publicly available after the review process.

Keywords: Cardiac MRI, multi-task, k-space measurements, representation learning, foundation models

1. Introduction

Cardiac magnetic resonance imaging (MRI) is the gold standard for assessing heart structure and function (Ismail et al., 2022). Due to hardware limits, high costs, and requirements for breath-hold by patients, standard cardiac MRI protocols typically undersample k-space to accelerate image acquisitions (Blumen et al., 1997; Wang et al., 2001; Plein

and Kozerke, 2021). Thus, the conventional pipeline for cardiac analysis follows three steps—undersampled k-space acquisition, image reconstruction, and image-based downstream analysis—which has remained largely unchanged as the standard practice.

Deep learning has led to significant progress in both reconstruction (Schlemper et al., 2017; Pan et al., 2023b; Huang et al., 2025) and downstream tasks such as segmentation or analysis (Chen et al., 2020; Martin-Isla et al., 2020; Liu et al., 2024). However, the dominant "reconstruct-then-analyze" pipeline remains fundamentally sequential and introduces avoidable information bottlenecks. Whether reconstruction and downstream models are trained separately or jointly, optimizing for perceptual image quality inevitably introduces artifacts and suppresses subtle frequency-domain details that may carry essential diagnostic value. Task-aware or end-to-end approaches (e.g., motion estimation (Oksuz et al., 2019; Pan et al., 2023a) or segmentation (Machado et al., 2023; Xu and Oksuz, 2024; Wech et al., 2025)) mitigate some issues, but they still rely on supervision in the intensity image domain. This anchors optimization to fine-grained pixel fidelity rather than the true diagnostic semantics of cardiac MRI, leaving a fundamental gap between reconstruction-driven objectives and clinically meaningful prediction (Seitzer et al., 2018; Dohmen et al., 2025).

On the other hand, k-space intrinsically captures the critical information about the heart, including anatomical structure, motion, and physiological characteristics, which is essential for comprehensive analysis of cardiovascular diseases (CVD). However, directly leveraging k-space for downstream tasks remains largely unexplored. Recent studies have started to investigate end-to-end approaches, predicting segmentation maps or clinical labels directly from undersampled k-space (Schlemper et al., 2018; Li et al., 2024; Zhang et al., 2024b). While promising, these efforts remain in their early stages and largely limited to proof-of-concept or single-task applications.

In parallel, cardiac MRI foundation models for diverse downstream tasks are rapidly emerging, driven by advances in image-domain representation learning (Zhang et al., 2024a; Jacob et al., 2025; Zhang et al., 2025). Yet this progress has not extended to k-space. To the best of our knowledge, no prior work has explored end-to-end, multi-task cardiac analysis directly from k-space, despite it being the most information-dense representation available. More critically, the fundamental question of how to learn powerful, efficient, and generalizable k-space representations that could serve as a unified and scalable foundation for cardiac MRI analysis remains largely unexplored.

To address these gaps, we propose **k-MTR** (**k**-space **M**ulti-**T**ask **R**epresentation), an end-to-end framework that enables reconstruction-free comprehensive and diverse cardiac analysis tasks directly from undersampled k-space measurements, leveraging the representation alignment between k-space and image domains. Our key contributions are:

- **We unify k-space and image representations into a single, information-rich manifold.** k-MTR aligns undersampled k-space and image-domain embeddings into a shared latent space that jointly captures fine-grained anatomy, global spectral structure, and physiological dynamics, providing a foundation for robust, reconstruction-free, and versatile cardiac analysis.
- **We show that this representation can be used for a cardiac multi-task pipeline directly from k-space.** Across disease classification, phenotype regression, and segmentation, k-MTR consistently matches or outperforms state-of-the-art

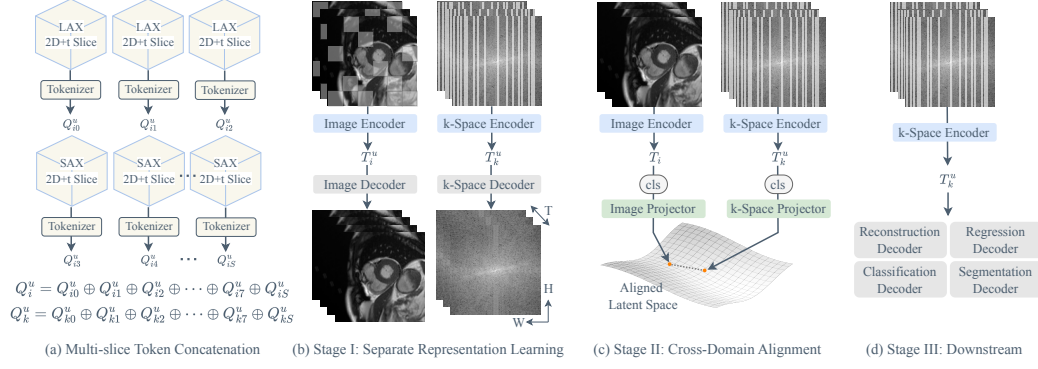


Figure 1: Overview of k-MTR. (a) Multi-slice tokenization prior to encoding. Each scan contains S multi-view $2D + t$ slices, with tokenizers applied to each slice. To enable multi-slice information exchange, input image tokens Q_i^u and k-space tokens Q_k^u are concatenated separately over all S undersampled slices for their respective encoders. (b–d) Training pipeline illustrated with a single slice for clarity. (b) Stage I: Learn separate representations via unsupervised reconstruction of undersampled k-space measurements and multi-view masked $2D + t$ slices. (c) Stage II: Align undersampled k-space representations T_k^u and fully-sampled image representations T_i using contrastive learning. (d) Stage III: Fine-tune the pretrained k-space encoder with lightweight decoders for various downstream tasks.

image-based and k-space-based baselines, demonstrating that undersampled k-space can serve as the basis for high-performance, end-to-end cardiac analysis.

- **We show that k-MTR learns powerful, interpretable, and clinically meaningful k-space representations.** k-MTR produces structured latent clusters aligned with clinical traits and supports reliable reconstructions, establishing a transparent, trustworthy representational backbone for comprehensive cardiac downstream tasks.

2. k-MTR

Unlike conventional imaging workflows, which perform downstream analysis only after reconstructing images from fully sampled measurements, k-MTR enables multiple downstream tasks directly from undersampled k-space data in three training stages, as illustrated in Figure 1. Let fully-sampled complex-valued cardiac k-space measurements be $X_k \in \mathbb{C}^{S \times T \times H \times W}$, where S , T , H , and W represent the number of slices, time frames, height, and width, respectively. The slices include both $2D + t$ long-axis (LAX) and short-axis (SAX) views, which are concatenated along the slice dimension. The corresponding image stack is written as $X_i \in \mathbb{C}^{S \times T \times H \times W}$. For undersampling, we apply a $4 \times$ acceleration mask $M \in \mathbb{Z}^{S \times T \times W}$ on the phase-encoding (W) direction, where each entry $M_{ijp} \in \{0, 1\}$ indicates whether the position is sampled (1) or omitted (0), following standard undersampling schemes for clinical MRI acquisitions. The mask is repeated along the H dimension to

obtain \tilde{M} , and the undersampled k-space data is generated via element-wise multiplication $X_k^u = \tilde{M} \odot X_k$.

Stage I: Domain-specific representation learning in the image and k-space domains. In the first stage, we independently learn robust representations from masked cardiac MR images and from undersampled k-space measurements, following the masked autoencoder (MAE) paradigm (He et al., 2022). This enables each domain (i.e., k-space and image) to develop high-level, semantically rich feature representations. For the image domain, multi-view $2D + t$ images are randomly masked at the patch level to obtain X_i^u . For the frequency domain, the k-space data is masked using a predefined acceleration mask \tilde{M} to obtain X_k^u . Each domain employs its own tokenizer \mathcal{T} , encoder \mathcal{E} , decoder \mathcal{D} , mask tokens T^m . With \oplus denoting concatenation, the reconstruction objectives are

$$\hat{X}_i = \mathcal{D}_i(\mathcal{E}_i(Q_i^u) \oplus T_i^m), \quad \hat{X}_k = \mathcal{D}_k(\mathcal{E}_k(Q_k^u) \oplus T_k^m), \quad (1)$$

where $Q_i^u = \mathcal{T}_i(X_i^u)$ and $Q_k^u = \mathcal{T}_k(X_k^u)$ denote the tokenized inputs. For both tokenizers, real and imaginary components are treated as two input channels, and tokens from different slices are concatenated along the sequence-length dimension (Fig. 1(a)). By reconstructing fully-sampled images and k-space measurements from masked inputs, both encoders learn domain-specific yet semantically rich features, capturing structures in the image domain and acquisition data characteristics in the frequency domain.

Stage II: Cross-domain alignment through contrastive learning. After pre-training the encoders, we establish a shared latent space that aligns multi-slice image and k-space representations from the same subject. By aligning high-level representations from k-space and the image domain, it is possible to construct a unified manifold where anatomical localization, global frequency structure, and physiological patterns coexist. Such a manifold allows models to learn directly from measurements, preserving information that reconstruction might obscure, while enabling scalable, task-agnostic representations. A critical design choice is that the image representations $T_i = \mathcal{E}_i(\mathcal{T}_i(X_i))$ are extracted from fully sampled images to preserve complete semantic content, whereas the k-space representations $T_k^u = \mathcal{E}_k(\mathcal{T}_k(X_k^u))$ are derived **solely from undersampled data**. This configuration reinforces the representational capacity of the k-space encoder, which is essential for Stage III, where downstream analysis must operate under the clinically realistic constraint of using only undersampled k-space data. By aligning the representations from two domains, we explicitly encourage the k-space encoder to embed all necessary anatomical cues—even those missing due to undersampling—into its latent representation.

Before alignment, class tokens from each encoder are projected via domain-specific projectors \mathcal{P}_i and \mathcal{P}_k :

$$\hat{Z}_i = \mathcal{P}_i(T_i), \quad \hat{Z}_k = \mathcal{P}_k(T_k^u). \quad (2)$$

Contrastive learning is then applied across all subjects \mathcal{N} in a batch, encouraging embeddings from the same subject to be closer while pushing embeddings from different subjects further apart (Hager et al., 2023; Zhang et al., 2025). Both k-space and image encoders are fine-tuned in this stage. For subject m with image and k-space embeddings z_{m_i} and z_{m_k} , the image-to-k-space loss $l_{i,k}$ and the total loss are

$$l_{i,k} = - \sum_{m \in \mathcal{N}} \log \frac{\exp(\cos(z_{m_i}, z_{m_k})/\tau)}{\sum_{n \in \mathcal{N}, n \neq m} \exp(\cos(z_{m_i}, z_{n_k})/\tau)}, \quad \mathcal{L} = \lambda l_{i,k} + (1 - \lambda) l_{k,i}, \quad (3)$$

where k-space-to-image loss $l_{k,i}$ is calculated analogously, λ is the loss weight, and τ is a small constant. This alignment stage allows the model to unify two fundamentally different data sources—image space and k-space—into a coherent, semantically meaningful manifold.

Stage III: Downstream learning directly from undersampled k-space. In the final stage, we leverage the pretrained k-space encoder to directly perform a range of downstream clinical tasks using only undersampled k-space measurements X_k^u . Lightweight task-specific decoders are attached to handle phenotype regression, disease classification, and anatomical segmentation. For each task, both k-space encoder and task-specific decoder are fine-tuned end-to-end in a supervised fashion. To promote trustworthiness and interpretability, we also evaluate k-MTR’s reconstruction ability directly from undersampled k-space without an explicit Fourier transform. Similar to AUTOMAP (Zhu et al., 2018), we learn a mapping from undersampled k-space to the image domain by integrating an adaptive reconstruction decoder with the pretrained k-space encoder. This enables the model to simultaneously support image reconstruction and clinical analysis within a unified representation space.

3. Dataset and Implementation

Datasets. For all 3 stages, we pretrain and fine-tune k-MTR using multi-view $2D + t$ cine MRI scans from 42,000 subjects that are part of the UK Biobank study (Petersen et al., 2015). Each scan comprises 6 SAX and 3 LAX views, with a spatial resolution of 128×128 and 50 time frames. As the UK Biobank study provides images only, corresponding k-space measurements need to be generated. To this end, we first introduce a synthetic phase to the real-valued images using a Gaussian-smoothed B0 field variation (Brown and Semelka, 1999). The resulting complex-valued images are then transformed into k-space via Fourier transform. Cartesian undersampling (Ahmad et al., 2015) with an acceleration factor of 4 is applied consistently to k-space measurements during both pretraining and fine-tuning through all 3 stages. For each downstream task, model evaluation is performed on a held-out set of cine MRI scans of 1,000 subjects from UK Biobank. For downstream targets, segmentation maps are generated using a convolution-based model (Bai et al., 2018) and subsequently quality-controlled. For phenotype prediction, we derive 18 phenotypes from both SAX and LAX segmentation maps (full details are provided in A). For disease classification, we focus on three cardiac conditions: coronary artery disease (CAD), high blood pressure, and hypertension (labels are derived following Zhang et al. (2025)).

Implementation. In stage I, imaging and k-space MAEs use 6 encoder and 2 decoder layers with a token size of 1024. The image patch size is (5, 8, 8) in (T, H, W) for each slice and the masking ratio is 70%. k-space data is undersampled by a factor of 4. The batch size is set to 2. All experiments are implemented in PyTorch and executed on a single NVIDIA A100 GPU.

In stage II, we apply gradient checkpointing to lower the computational cost of training, allowing a batch size of 256 for contrastive learning. The projectors consist of two-layer multi-layer perceptrons (MLPs) that map the token embeddings from image and k-space pipeline from 1025 to 128 dimensions. τ is set to 0.1 and the loss weight λ is 0.5.

In the final stage, the k-space encoder pretrained in Stages I and II is fully fine-tuned in Stage III for various downstream tasks. The task-specific decoders are designed as follows:

Table 1: Mean absolute error for phenotype prediction across methods: ResNet-50 (upper bound, trained on fully sampled images), ResNet-50^u (trained on undersampled images), ResNet-50^k (trained on undersampled zero-filled k-space data), MAE (k-MTR counterpart without cross-domain alignment), and the proposed k-MTR. Best: bold, second: green, third: underlined.

Phenotype	Image-based		k-space-based		
	ResNet-50	ResNet-50 ^u	ResNet-50 ^k	MAE	k-MTR
LVEDV (mL)	6.58 \pm 9.10	7.35 \pm 9.30	9.57 \pm 10.75	12.88 \pm 11.95	<u>8.15</u> \pm 11.11
LVESV (mL)	5.19 \pm 5.15	5.75 \pm 5.48	7.05 \pm 7.29	7.38 \pm 7.26	<u>6.02</u> \pm 6.46
LVSV (mL)	5.68 \pm 6.29	6.55 \pm 6.44	7.03 \pm 6.57	8.42 \pm 7.72	6.50 \pm 6.83
LVEF (%)	2.95 \pm 2.39	3.30 \pm 2.52	3.58 \pm 2.68	6.09 \pm 3.73	3.14 \pm 2.40
LVCO (L/min)	0.59 \pm 0.58	0.61 \pm 0.61	0.68 \pm 0.66	0.70 \pm 0.71	<u>0.65</u> \pm 0.65
LVM (g)	5.51 \pm 5.67	7.01 \pm 6.67	8.00 \pm 7.36	10.09 \pm 8.85	<u>7.20</u> \pm 7.42
RVEDV (mL)	9.24 \pm 10.64	10.49 \pm 10.83	11.33 \pm 9.99	14.76 \pm 12.88	10.60 \pm 11.99
RVESV (mL)	6.15 \pm 6.28	7.35 \pm 6.76	7.63 \pm 6.91	8.71 \pm 7.78	6.98 \pm 7.07
RVSV (mL)	7.58 \pm 7.43	8.06 \pm 7.25	8.41 \pm 6.96	9.64 \pm 8.28	7.86 \pm 7.63
RVEF (%)	3.30 \pm 2.97	3.56 \pm 3.06	3.66 \pm 3.08	5.97 \pm 3.71	3.28 \pm 2.84
LAV max (mL)	6.65 \pm 6.22	8.74 \pm 7.49	10.86 \pm 10.05	8.81 \pm 8.47	8.51 \pm 8.21
LAV min (mL)	4.61 \pm 4.95	5.60 \pm 6.05	7.24 \pm 8.33	5.29 \pm 6.82	5.49 \pm 6.39
LASV (mL)	4.66 \pm 3.93	5.41 \pm 4.18	5.84 \pm 4.99	5.78 \pm 4.96	5.34 \pm 4.49
LAEF (%)	4.28 \pm 4.01	4.82 \pm 4.77	5.74 \pm 5.66	6.71 \pm 5.22	4.72 \pm 4.80
RAV max (mL)	7.67 \pm 6.60	10.75 \pm 9.57	12.89 \pm 10.38	12.07 \pm 10.49	10.93 \pm 9.66
RAV min (mL)	5.84 \pm 4.98	7.18 \pm 6.52	8.77 \pm 7.59	8.31 \pm 7.43	7.36 \pm 7.18
RASV (mL)	5.80 \pm 4.80	7.08 \pm 6.44	8.01 \pm 6.78	7.40 \pm 6.58	7.07 \pm 6.30
RAEF (%)	5.09 \pm 4.48	5.55 \pm 4.97	6.53 \pm 5.55	6.56 \pm 5.49	<u>5.74</u> \pm 5.10

the phenotype regression and classification decoders both use two-layer MLPs with 256-dimensional embeddings, while the segmentation decoder is a U-Net-based model (Zhou et al., 2023) with an embedding dimension of 576. The batch sizes are set to 16 for regression, 32 for classification, and 2 for segmentation. For reconstruction, we adopt the same decoder architecture as the k-space decoder used in Stage I, but train it from scratch. The batch size is set to 2. More implementation details are provided in Appendix C.

Baselines. For phenotype regression, we use ResNet-50 (He et al., 2016) and a MAE-based model (Zhang et al., 2024a) as baseline methods. For ResNet-50, we evaluate three input settings: (1) fully-sampled images as an upper bound (ResNet-50), (2) undersampled images (ResNet-50^u), and (3) undersampled zero-filled k-space measurements (ResNet-50^k). These baselines enable us to assess: the performance degradation caused by 4-fold under-sampling ((1) vs. (2)); the effectiveness of k-MTR in predicting clinical phenotypes directly from k-space ((2) vs. k-MTR); and the necessity of the k-MTR design ((3) vs. k-MTR).

Table 2: Classification performance across three cardiac diseases for ResNet-50 (trained on fully-sampled images), ResNet-50^u (undersampled images), and the proposed k-MTR. Positive class ratios are shown as percentages of the cohort. AP: average precision. Top: green, second: underlined.

Disease	Method	AUC-ROC↑	F1 Score↑	Recall↑	Precision↑	AP↑
CAD (7.4%)	ResNet-50	<u>0.608</u>	0.000	0.000	0.000	<u>0.139</u>
	ResNet-50 ^u	0.590	0.000	0.000	0.000	0.088
	k-MTR	0.737	0.282	0.500	0.197	0.234
High Blood Pressure (25.8%)	ResNet-50	<u>0.685</u>	<u>0.095</u>	<u>0.054</u>	0.359	0.397
	ResNet-50 ^u	0.550	0.000	0.000	0.000	0.273
	k-MTR	0.697	0.474	0.781	<u>0.340</u>	<u>0.386</u>
Hypertension (20.8%)	ResNet-50	<u>0.607</u>	<u>0.047</u>	<u>0.024</u>	0.714	<u>0.296</u>
	ResNet-50 ^u	0.525	0.000	0.000	0.000	0.219
	k-MTR	0.710	0.417	0.603	<u>0.319</u>	0.356

The MAE-based model uses the same architecture as k-MTR in Stage III but does not include the Stage II pretraining, isolating the contribution of cross-domain representation alignment. For classification, we use ResNet-50 trained on fully-sampled images and its undersampled counterpart ResNet-50^u as image-based baselines. For segmentation, we adopt nnU-Net (Isensee et al., 2021) trained on fully-sampled images as the upper bound (nnU-Net) and trained on undersampled images (nnU-Net^u). Additionally, we compare against LI-Net (Schlemper et al., 2018), a model specifically developed for segmentation from undersampled cine MR images. For reconstruction, MAE-based k-GIN (Pan et al., 2023b) is our baseline.

4. Results

In this section, we show that k-MTR provides efficient and comprehensive representations directly from undersampled k-space, preserving rich cardiac structural and functional information. These representations enable accurate performance across a wide range of downstream cardiac analysis tasks without requiring explicit image reconstruction.

4.1. k-Space Representations Enable Strong Downstream Performance

We evaluate k-MTR on several clinically relevant tasks, including SAX and LAX phenotype prediction, cardiac disease classification, and comprehensive image segmentation across different planes and time points.

Phenotype prediction. As shown in Tab. 1, k-MTR achieves phenotype prediction accuracy competitive with image-based ResNet-50^u, despite operating only on undersampled k-space. This demonstrates that downstream cardiac analysis does not inherently require

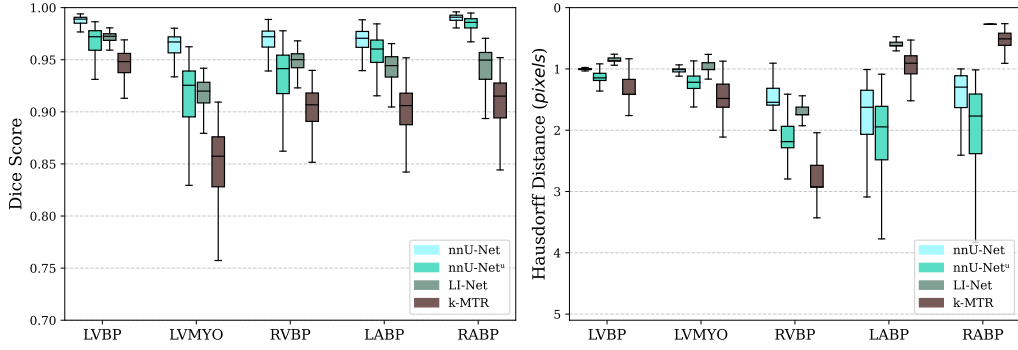


Figure 2: Segmentation Dice scores and Hausdorff distances of the proposed k-MTR method, compared with the upper bound nnU-Net, nnU-Net^u, and LI-Net.

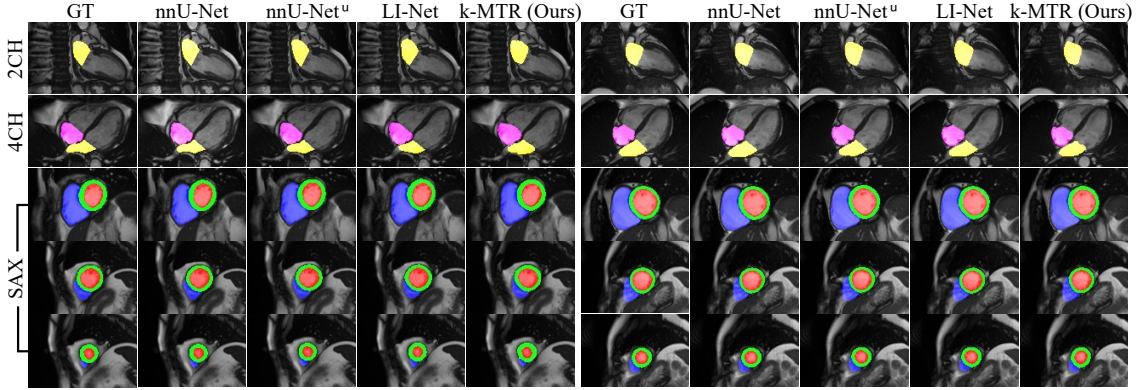


Figure 3: Segmentation examples for the 2-chamber, 4-chamber, and SAX views, comparing predictions from nnU-Net, nnU-Net^u, LI-Net, and the proposed k-MTR.

working in the image domain, challenging the conventional reliance on reconstructed images. Moreover, k-MTR consistently surpasses both the k-space-based ResNet-50^k and the MAE baseline (which shares the same architecture but lacks cross-domain alignment). These gains emphasize the rich physiological information captured by k-space representations learned by k-MTR and underscore the critical role of the proposed image-k-space alignment mechanism in enabling robust phenotype prediction. The image-based ResNet-50 serves as an upper bound and further illustrates the performance drop caused by undersampling.

Disease classification. As shown in Tab. 2, k-MTR consistently outperforms both ResNet-50^u trained on undersampled images derived from the same 4-fold k-space data and the standard ResNet-50 trained on fully sampled images across all three cardiac disease categories. These gains arise from k-MTR’s ability to learn robust k-space representations and align them with informative image-domain features. The resulting latent space cap-

tures both cardiac anatomy and temporal dynamics, enabling more accurate identification of disease-related patterns and greater robustness to noise, aliasing, and other artifacts introduced by undersampling.

Image segmentation. As shown in Fig. 2 and Fig. 3, k-MTR achieves competitive Dice scores and Hausdorff distances, with only a slight reduction compared to nnU-Net-based baselines and LI-Net, a model specifically designed for segmentation of undersampled images. This difference is expected, as segmentation relies on fine-grained spatial details, which are more naturally represented in the image domain, making the task inherently more challenging for k-MTR. Nonetheless, these results demonstrate that coherent anatomical representations can be learned directly from raw k-space, even without explicit image reconstruction.

4.2. Clustering in Latent Space

To further demonstrate the meaningful k-space latent representations learned by our model, we visualize the k-space embeddings using 3D t-SNE in Fig. 4, with points color-coded according to key cardiac phenotypes, LVEDV and LVM (more visualizations are provided in B). Since both spatial and temporal information are critical for accurate phenotype estimation, the presence of clearly separated clusters suggests that k-MTR captures semantically and temporally meaningful structures directly from the undersampled k-space domain, reflecting a robust understanding of cardiac variability across subjects.

4.3. Reconstruction Ability for Interpretability

To demonstrate the trustworthiness and interpretability of k-MTR, we evaluate its adaptive reconstruction decoder. Although reconstruction is not the primary goal of k-MTR, the model is able to perform end-to-end reconstruction from undersampled k-space to images

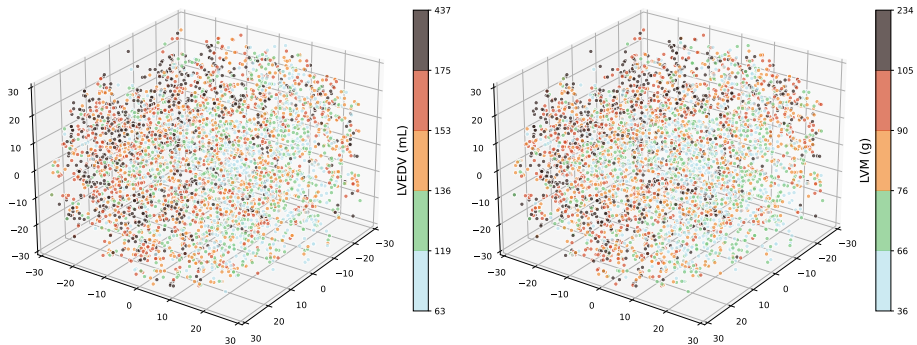


Figure 4: The 3D t-SNE visualization of the k-space and image aligned latent representations. Latent embeddings are labeled with different phenotypes, categorized into 5 groups according to the ground truth, and shown in different colors.

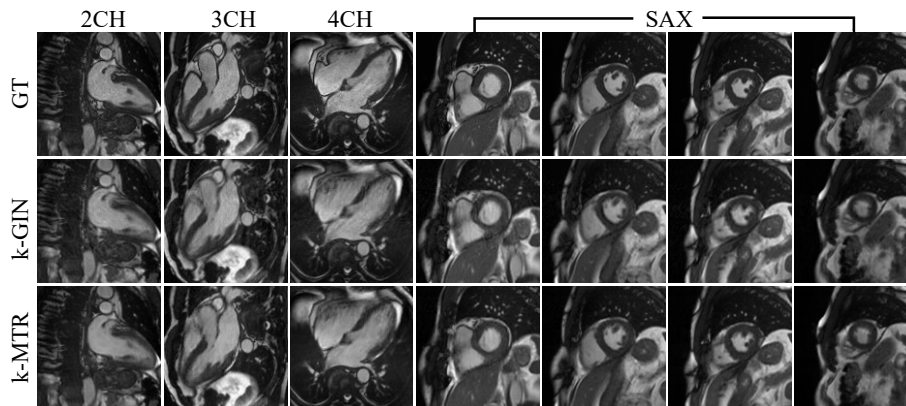


Figure 5: Representative examples of end-to-end image reconstructions by the proposed k-MTR from undersampled k-space measurements, compared with baseline k-GIN.

without an explicit Fourier transform—highlighting its strong, intrinsically interpretable latent representations. Remarkably, k-MTR achieves a PSNR of 38.18 ± 2.48 dB, matching the performance of the cardiac reconstruction-specific model k-GIN (38.30 ± 2.98 dB). Example images in Fig. 5 further illustrate the k-MTR’s ability to produce high-fidelity reconstructions that remain strongly consistent with the reference images.

5. Discussion and Conclusion

In this work, we introduced k-MTR, a framework that learns compact and expressive representations directly from undersampled k-space measurements, enabling accurate multi-task cardiac analysis without explicit image reconstruction. By aligning k-space and image-domain embeddings, k-MTR captures both frequency-domain structure and spatially localized anatomy, yielding semantically consistent and information-rich representations. k-MTR’s superior performance on phenotype prediction, disease classification, and segmentation reveals the largely overlooked diagnostic value embedded in k-space.

A limitation, however, is that our evaluation currently relies on simulated undersampling. Due to the scarcity of large, publicly available real multi-coil CMR k-space datasets, we have not yet validated k-MTR on genuine clinical acquisitions. Therefore, a critical next step is to test on prospectively acquired k-space data to establish real-world reliability. Furthermore, we will also focus on systematically probing k-MTR’s robustness under different undersampling ratios and sampling patterns to understand how far acceleration can be pushed without degrading performance, providing insights for more efficient and task-aware CMR acquisition protocols.

In conclusion, k-MTR paves a scalable, reconstruction-free path to k-space-driven cardiac foundation models, unlocking rich physiological information for more efficient, generalizable, and clinically robust cardiac analysis.

Acknowledgments

This research has been conducted using the UK Biobank Resource under Application Number 87802. This work is funded by the European Research Council (ERC) project Deep4MI (884622). Dr. Sevgi Gokce Kafali has been sponsored by the Alexander von Humboldt Foundation.

References

- Rizwan Ahmad, Hui Xue, Shivraman Giri, Yu Ding, Jason Craft, and Orlando P Simonetti. Variable density incoherent spatiotemporal acquisition (vista) for highly accelerated cardiac mri. *Magnetic resonance in medicine*, 74(5):1266–1278, 2015.
- Wenjia Bai, Matthew Sinclair, Giacomo Tarroni, Ozan Oktay, Martin Rajchl, Ghislain Vialant, Aaron M Lee, Nay Aung, Elena Lukaschuk, Mihir M Sanghvi, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of cardiovascular magnetic resonance*, 20(1):65, 2018.
- David A Bluemke, Jerrold L Boxerman, Ergin Atalar, and Elliot R McVeigh. Segmented k-space cine breath-hold cardiovascular mr imaging: Part 1. principles and technique. *AJR. American journal of roentgenology*, 169(2):395–400, 1997.
- Mark A Brown and Richard C Semelka. *MRI: basic principles and applications*. Willey-Liss, 1999.
- Chen Chen, Chen Qin, Huaqi Qiu, Giacomo Tarroni, Jinming Duan, Wenjia Bai, and Daniel Rueckert. Deep learning for cardiac image segmentation: a review. *Frontiers in cardiovascular medicine*, 7:25, 2020.
- Melanie Dohmen, Mark A Klemens, Ivo M Baltruschat, Tuan Truong, and Matthias Lenga. Similarity and quality metrics for mr image-to-image translation. *Scientific Reports*, 15(1):3853, 2025.
- Paul Hager, Martin J Menten, and Daniel Rueckert. Best of both worlds: Multimodal contrastive learning with tabular and imaging data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23924–23935, 2023.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.
- Wenqi Huang, Veronika Spieker, Siying Xu, Gastao Cruz, Claudia Prieto, Julia A Schnabel, Kerstin Hammernik, Thomas Kuestner, and Daniel Rueckert. Subspace implicit neural representations for real-time cardiac cine mr imaging. In *International Conference on Information Processing in Medical Imaging*, pages 168–183. Springer, 2025.

- Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.
- Tevfik F Ismail, Wendy Strugnell, Chiara Coletti, Maša Božić-Iven, Sebastian Weingaertner, Kerstin Hammernik, Teresa Correia, and Thomas Kuestner. Cardiac mr: from theory to practice. *Frontiers in cardiovascular medicine*, 9:826283, 2022.
- Athira J Jacob, Indraneel Borgohain, Teodora Chitiboi, Puneet Sharma, Dorin Comaniciu, and Daniel Rueckert. Towards a cmr foundation model for multi-task cardiac image analysis. *Journal of Cardiovascular Magnetic Resonance*, page 101967, 2025.
- Ruochen Li, Jiazhen Pan, Youxiang Zhu, Juncheng Ni, and Daniel Rueckert. Classification, regression and segmentation directly from k-space in cardiac mri. In *International Workshop on Machine Learning in Medical Imaging*, pages 31–41. Springer, 2024.
- Zelong Liu, Komal Kainth, Alexander Zhou, Timothy W Deyer, Zahi A Fayad, Hayit Greenspan, and Xueyan Mei. A review of self-supervised, generative, and few-shot deep learning methods for data-limited magnetic resonance imaging segmentation. *NMR in Biomedicine*, 37(8):e5143, 2024.
- Inês Machado, Esther Puyol-Antón, Kerstin Hammernik, Gastao Cruz, Devran Ugurlu, Ihsane Olakorede, Ilkay Oksuz, Bram Ruijsink, Miguel Castelo-Branco, Alistair Young, et al. A deep learning-based integrated framework for quality-aware undersampled cine cardiac mri reconstruction and analysis. *IEEE Transactions on Biomedical Engineering*, 71(3):855–865, 2023.
- Carlos Martin-Isla, Victor M Campello, Cristian Izquierdo, Zahra Raisi-Estabragh, Bettina Baeßler, Steffen E Petersen, and Karim Lekadir. Image-based cardiac diagnosis with machine learning: a review. *Frontiers in cardiovascular medicine*, 7:1, 2020.
- Ilkay Oksuz, James Clough, Bram Ruijsink, Esther Puyol-Antón, Aurelien Bustin, Gastao Cruz, Claudia Prieto, Daniel Rueckert, Andrew P King, and Julia A Schnabel. Detection and correction of cardiac mri motion artefacts during reconstruction from k-space. In *International conference on medical image computing and computer-assisted intervention*, pages 695–703. Springer, 2019.
- Jiazhen Pan, Wenqi Huang, Daniel Rueckert, Thomas Küstner, and Kerstin Hammernik. Motion-compensated mr cine reconstruction with reconstruction-driven motion estimation. *arXiv preprint arXiv:2302.02504*, 2023a.
- Jiazhen Pan, Suprosanna Shit, Özgün Turgut, Wenqi Huang, Hongwei Bran Li, Nil Stolt-Ansó, Thomas Küstner, Kerstin Hammernik, and Daniel Rueckert. Global k-space interpolation for dynamic mri reconstruction using masked image modeling. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 228–238. Springer, 2023b.
- S. E. Petersen, P. M. Matthews, J. M. Francis, M. D. Robson, and et al. UK Biobank’s cardiovascular magnetic resonance protocol. *JCMR*, pages 1–7, 2015.

- Sven Plein and Sebastian Kozerke. Are we there yet? the road to routine rapid cmr imaging, 2021.
- Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony N Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE transactions on Medical Imaging*, 37(2):491–503, 2017.
- Jo Schlemper, Ozan Oktay, Wenjia Bai, Daniel C Castro, Jinming Duan, Chen Qin, Jo V Hajnal, and Daniel Rueckert. Cardiac mr segmentation from undersampled k-space using deep latent representation learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 259–267. Springer, 2018.
- Maximilian Seitzer, Guang Yang, Jo Schlemper, Ozan Oktay, Tobias Würfl, Vincent Christlein, Tom Wong, Raad Mohiaddin, David Firmin, Jennifer Keegan, et al. Adversarial and perceptual refinement for compressed sensing mri reconstruction. In *International conference on medical image computing and computer-assisted intervention*, pages 232–240. Springer, 2018.
- YI Wang, Richard Watts, Ian R Mitchell, Thanh D Nguyen, Jeffrey W Bezanson, Geoffrey W Bergman, and Martin R Prince. Coronary mr angiography: selection of acquisition window of minimal cardiac motion with electrocardiography-triggered navigator cardiac motion prescanning—initial results. *Radiology*, 218(2):580–585, 2001.
- Tobias Wech, Oliver Schad, Simon Sauer, Jonas Kleineisel, Nils Petri, Peter Nordbeck, Thorsten A Bley, Bettina Baeßler, Bernhard Petritsch, and Julius F Heidenreich. Joint image reconstruction and segmentation of real-time cardiovascular magnetic resonance imaging in free-breathing using a model based on disentangled representation learning. *Journal of Cardiovascular Magnetic Resonance*, 27(1):101844, 2025.
- Ruru Xu and Ilkay Oksuz. Segmentation-aware mri subsampling for efficient cardiac mri reconstruction with reinforcement learning. *Image and Vision Computing*, 150:105200, 2024.
- Yundi Zhang, Chen Chen, Suprosanna Shit, Sophie Starck, Daniel Rueckert, and Jiazhen Pan. Whole heart 3d+ t representation learning through sparse 2d cardiac mr images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 359–369. Springer, 2024a.
- Yundi Zhang, Nil Stolt-Ansó, Jiazhen Pan, Wenqi Huang, Kerstin Hammernik, and Daniel Rueckert. Direct cardiac segmentation from undersampled k-space using transformers. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–4. IEEE, 2024b.
- Yundi Zhang, Paul Hager, Che Liu, Suprosanna Shit, Chen Chen, Daniel Rueckert, and Jiazhen Pan. Towards cardiac mri foundation models: Comprehensive visual-tabular representations for whole-heart assessment and beyond. *Medical Image Analysis*, 106:103756, 2025. ISSN 1361-8415. doi: <https://doi.org/10.1016/j.media.2025.103756>. URL <https://www.sciencedirect.com/science/article/pii/S1361841525003032>.

Lei Zhou, Huidong Liu, Joseph Bae, Junjun He, Dimitris Samaras, and Prateek Prasanna. Self pre-training with masked autoencoders for medical image classification and segmentation. In *2023 IEEE 20th international symposium on biomedical imaging (ISBI)*, pages 1–6. IEEE, 2023.

Bo Zhu, Jeremiah Z Liu, Stephen F Cauley, Bruce R Rosen, and Matthew S Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492, 2018.

Appendix A. Phenotype Values

The cardiac phenotypes shown in the work include Left Ventricular End-Diastolic Volume (LVEDV), Left Ventricular End-Systolic Volume (LVESV), Left Ventricular Stroke Volume (LVSV), Left Ventricular Ejection Fraction (LVEF), Left Ventricular Cardiac Output (LVCO), Left Ventricular Mass (LVM), Right Ventricular End-Diastolic Volume (RVEDV), Right Ventricular End-Systolic Volume (RVESV), Right Ventricular Stroke Volume (RVSV), Right Ventricular Ejection Fraction (RVEF), Left Atrium Volume Maximum (LAV max), Left Atrium Volume Minimum (LAV min), Left Atrium Stroke Volume (LASV), Left Atrium Ejection Fraction (LAEF), Right Atrium Volume Maximum (RAV max), Right Atrium

Table 3: Mean and standard deviation using cohort-level mean estimates for all physiological and anthropometric features, as well as SAX and LAX phenotype features.

Phenotype	Mean-guess
LVEDV (mL)	25.01 \pm 20.49
LVESV (mL)	13.79 \pm 11.59
LVSV (mL)	14.58 \pm 12.13
LVEF (%)	4.52 \pm 3.81
LVCO (L/min)	0.95 \pm 0.84
LVM (g)	17.71 \pm 12.84
RVEDV (mL)	28.40 \pm 20.63
RVESV (mL)	16.37 \pm 12.18
RVSV (mL)	15.40 \pm 11.61
RVEF (%)	4.81 \pm 3.89
LAV max (mL)	16.57 \pm 13.40
LAV min (mL)	9.74 \pm 9.45
LASV (mL)	8.79 \pm 7.03
LAEF (%)	6.51 \pm 6.15
RAV max (mL)	19.19 \pm 15.42
RAV min (mL)	12.68 \pm 10.75
RASV (mL)	9.90 \pm 8.05
RAEF (%)	7.17 \pm 5.97

Volume Minimum (RAV min), Right Atrium Stroke Volume (RASV), and Right Atrium Ejection Fraction (RAEF).

Tab. 3 shows the cohort-level mean and standard deviations of all phenotypes of the test dataset.

Appendix B. t-SNE Visualization Examples

Fig. 6 shows the t-SNE visualization of the k-space latent representations of k-MTR.

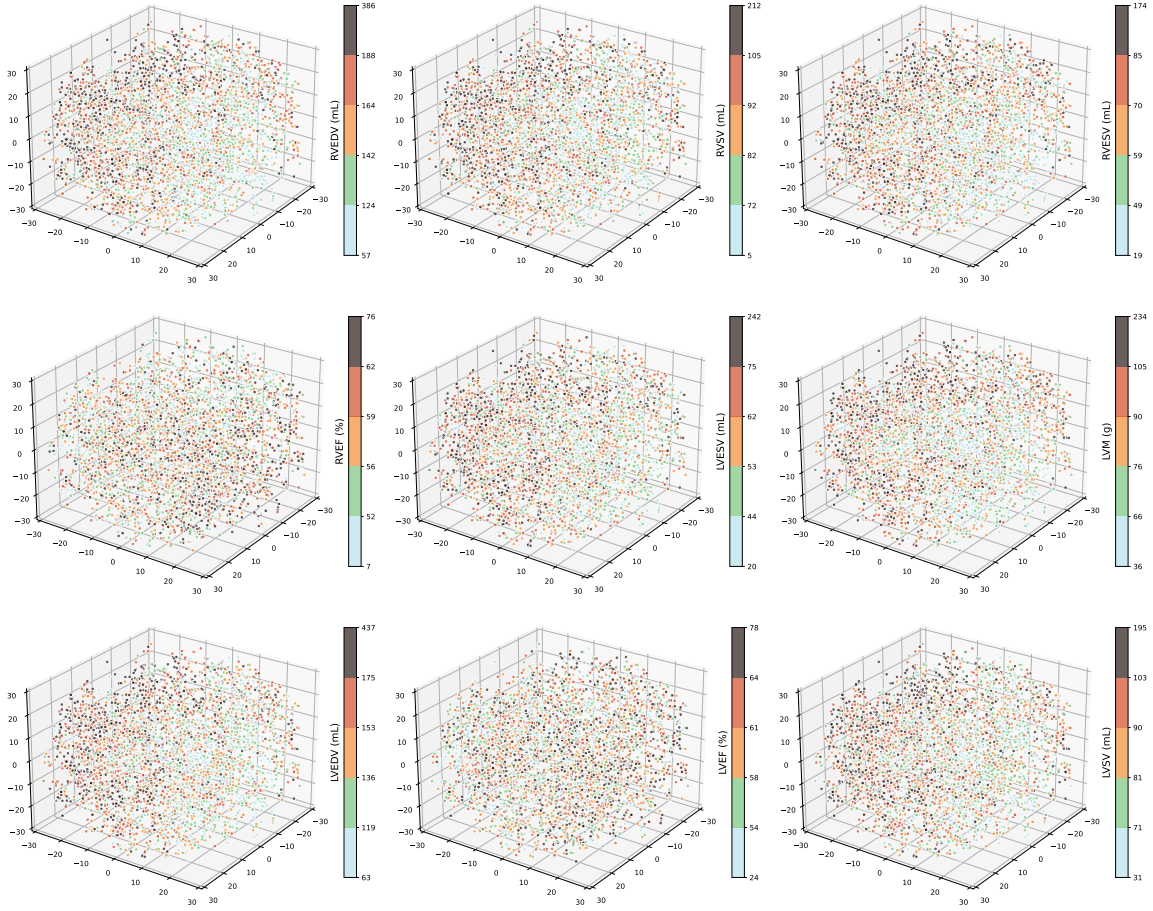


Figure 6: The 3D t-SNE visualization of the k-space latent representations after cross-domain alignment. Latent embeddings are labeled with different phenotypes, categorized into 5 groups according to the ground truth, and shown in different colors.

Appendix C. Implementation Details

For stage I, the initial learning rate for image pipeline is 3×10^{-4} and for k-space pipeline is 10^{-5} . We use a cosine annealing scheduler with warmup of 10 epochs and a weight decay of 10^{-4} . The same scheduler is applied to all 3 stages. For stage II, the initial learning rate is set to 10^{-5} . For stage III, the learning rates are 10^{-5} , 10^{-6} , 10^{-5} , and 10^{-5} for regression, classification, segmentation, and reconstruction, respectively.