Low-Rank-Modulated Functa: Exploring the Latent Space of Implicit Neural Representations for Interpretable Ultrasound Video Analysis

Julia Wolleb[1,2], Cristiana Baloescu[3], Alicia Durrer[4] Hemant D. Tagare[5]*, and Xenophon Papademetris[1,2,5]*

[1] Dept. of Biomedical Informatics & Data Science, Yale University, New Haven, USA
[2] Yale Biomedical Imaging Institute, Yale University, New Haven, USA
[3] Dept. of Emergency Medicine, Yale School of Medicine, New Haven, USA
[4] Dept. of Biomedical Engineering, University of Basel, Allschwil, Switzerland
[5] Dept. of Radiology & Biomedical Imaging, Yale University, New Haven, USA
julia.wolleb@yale.edu

Abstract. Implicit neural representations (INRs) have emerged as a powerful framework for continuous image representation learning. In *Functa*-based approaches, each image is encoded as a latent modulation vector that conditions a shared INR, enabling strong reconstruction performance. However, the structure and interpretability of the corresponding latent spaces remain largely unexplored. In this work, we investigate the latent space of *Functa*-based models for ultrasound videos and propose *Low-Rank-Modulated Functa* (*LRM-Functa*), a novel architecture that enforces a low-rank adaptation of modulation vectors in the time-resolved latent space. When applied to cardiac ultrasound, the resulting latent space exhibits clearly structured periodic trajectories, facilitating visualization and interpretability of temporal patterns. The latent space can be traversed to sample novel frames, revealing smooth transitions along the cardiac cycle, and enabling direct readout of end-diastolic (ED) and end-systolic (ES) frames without additional model training. We show that *LRM-Functa* outperforms prior methods in unsupervised ED and ES frame detection, while compressing each video frame to as low as rank $k = 2$ without sacrificing competitive downstream performance on ejection fraction prediction. Evaluations on out-of-distribution frame selection in a cardiac point-of-care dataset, as well as on lung ultrasound for B-line classification, demonstrate the generalizability of our approach. Overall, *LRM-Functa* provides a compact, interpretable, and generalizable framework for ultrasound video analysis. The code is available at `https://github.com/JuliaWolleb/LRM_Functa`.

Keywords: Implicit neural representations · Cardiac phase analysis · Ultrasound videos · Latent space trajectory · Interpretability

1 Introduction

Ultrasound imaging is widely used in clinical practice due to its portability, low cost, and real-time acquisition. Modern workflows rely on ultrasound videos

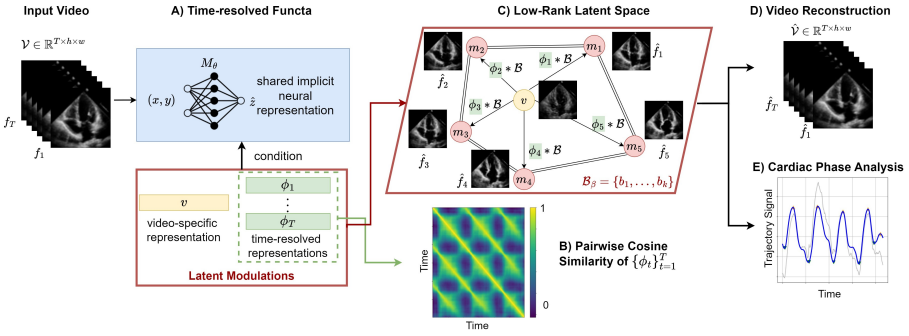

Fig. 1. Overview of our *LRM-Functa* framework. A) Our approach builds on *Vid-Functa*, where each ultrasound video is represented by a video-level modulation vector v and time-resolved modulation vectors $\phi = \{\phi_t\}_{t=1}^T$. (B) Pairwise cosine similarity of ϕ over time reveals a clear low-rank pattern. (C) Motivated by this observation, we constrain the latent modulation space to a low-rank subspace $\mathcal{B}_\beta$. (D/E) This structured latent space enables efficient compression-reconstruction and cardiac phase analysis.

rather than single frames, particularly in cardiac and lung imaging, where temporal dynamics are critical for diagnosis and quantification [1,10,25]. Precise identification of end-systolic (ES) and end-diastolic (ED) frames is essential in echocardiography because key measures, such as ejection fraction and ventricular hypertrophy, are defined at specific cardiac phases [15,18]. However, ultrasound videos present challenges for storage, transmission, and analysis due to their size, noise, and high temporal redundancy. This has motivated the development of low-dimensional and task-specific video representations [12,13]. In this work, we build on implicit neural representations (INRs) [24] to learn compact, continuous representations for ultrasound video compression and analysis. We adopt the *Functa* framework [4,7], which encodes entire datasets using a shared INR backbone modulated by per-image latent vectors. While this approach achieves efficient reconstruction, the structure and interpretability of the latent space remains largely unexplored. To address this gap, building on time-resolved *Vid-Functa* [28] and inspired by [29], we propose *Low-Rank-Modulated Functa (LRM-Functa)*, shown in Figure 1. This novel architecture enforces a structured, low-rank latent space shown in Figure 1.C, producing compact video representations at high compression rates. The latent trajectories are directly interpretable and can be visualized for unsupervised cardiac phase analysis, enabling extraction of ED and ES frame indices without additional model training.

Related Work INRs model signals as continuous neural functions and have been applied to 2D, 3D, and spatiotemporal data [30], where a neural network learns coordinate-to-intensity mappings for self-supervised reconstruction. They are widely used in medical image analysis [17], including ultrasound 3D reconstruction [8] or speed-of-sound estimation [3]. The *Functa* framework [4,7] extends the INR paradigm by combining a shared network M_θ with per-image

latent modulation vectors, enabling entire datasets to be encoded within a single model and allowing efficient reconstruction [2,4,5,7]. *VidFuncta* [28] further extends this approach to videos through time-resolved modulation vectors, while Latent-INRs use frame-specific latent codes learned jointly with hypernetworks for video compression [16]. However, these methods focus on reconstruction and downstream tasks, leaving the interpretability of the latent space largely unexplored. Structured latent representations have been explored in dynamic human view synthesis [20]. Low-rank adaptation methods (LoRA), originally developed for parameter-efficient fine-tuning of large language models [11], have been applied to INRs for compact video compression [26]. In medical image analysis, learned video representations support tasks such as cardiac phase detection, functional quantification, and pathology classification [6,12]. Prior work has investigated interpretable latent trajectories using variational and convolutional autoencoders to analyze cardiac motion [22,29]. However, structured and interpretable latent representations remain unexplored in INR-based architectures.

Contributions To the best of our knowledge, we are the first to analyze and structure the latent space of *Functa*-based frameworks. Our novel architecture, *LRM-Functa*, enforces a low-rank latent representation, enabling interpretable visualization and analysis of temporal trajectories in cardiac ultrasound. This enables direct identification of ED and ES frames from the latent trajectories without additional model training, outperforming state-of-the-art unsupervised methods, and generalizing to out-of-distribution (OOD) cardiac views. Furthermore, we explore the compression–reconstruction trade-off, demonstrating that accurate ejection fraction estimation is preserved even under high compression down to $k = 2$ components per frame. We additionally validate the generalizability of our approach on B-line detection in lung ultrasound (LUS) videos.

2 Method

2.1 Exploration of the Latent Space of VidFuncta

We explore the *VidFuncta* architecture as our baseline INR for ultrasound video analysis [28], illustrated in Figure 1.A. *VidFuncta* employs a shared neural network M_θ, implemented as a multilayer perceptron (MLP), whose learnable parameters θ are shared across the whole dataset. Each video is conditioned on a learnable video-level latent vector v using FiLM-style modulation [21], and temporal dynamics are captured by frame-specific modulation vectors ϕ_t, one per frame f_t for $t = 1, \ldots, T$. To explore the structure of this latent space, we visualize pairwise cosine similarities between all frame vectors $\phi = \{\phi_t\}_{t=1}^{T}$. The resulting $T \times T$ matrices exhibit low-rank patterns with clear periodicity along the side diagonals, as shown in Figure 1.B, indicating that the latent space encodes rich temporal dynamics. This observation motivates the low-rank constraint on the latent space we propose in this work, as described in Section 2.2.

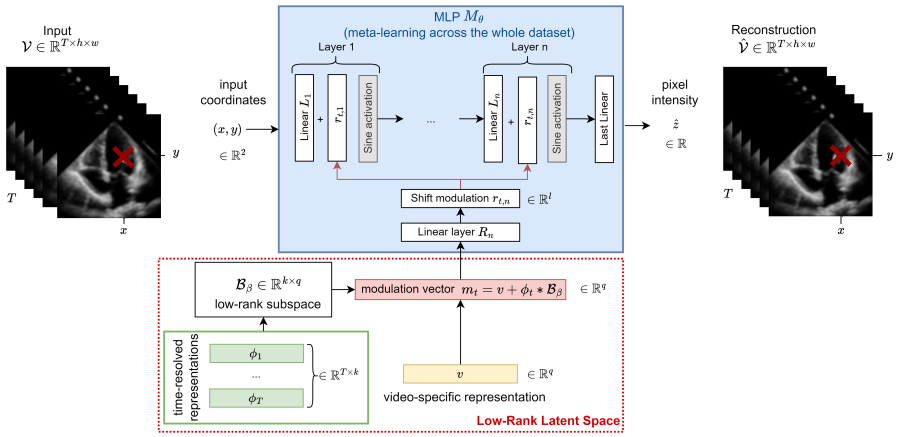

Fig. 2. Proposed architecture for low-rank adaptation of the modulation vectors. Each video $\mathcal{V}$ is encoded into a video-level representation v and time-resolved frame representations ϕ_t, which are projected onto a rank-k subspace $\mathcal{B}_\beta$ to produce modulation vectors $m_t = v + \mathcal{B}_\beta \phi_t$. These vectors shift the activations of a shared MLP M_θ, trained to reconstruct pixel intensities $\hat{z}$ at each coordinate (x, y), shown as a red cross.

2.2 Low-Rank Adaptation of the Latent Modulation Vectors

We present *LRM-Functa*, which defines a low-rank latent space within the *Vid-Functa* framework for video representation. Building on LoRA [11,26] and latent motion profiling in convolutional autoencoders [29], each frame-specific modulation vector is constrained to $m_t = v + \mathcal{B}_\beta \, \phi_t$, as illustrated in Figure 1.C. Here, $v \in \mathbb{R}^q$ denotes the video-level latent vector, $\mathcal{B}_\beta = \{b_1, \ldots, b_k\} \in \mathbb{R}^{k \times q}$ defines a learnable rank-k subspace of $\mathbb{R}^q$, and $\phi_t \in \mathbb{R}^k$ represents the learnable frame-specific update coefficients. This ensures that all updates from v to individual frames f_t are constrained to the low-rank subspace spanned by $\mathcal{B}_\beta$. If rank $k = 2$, $\mathcal{B}_\beta$ can be visualized as a 2D plane. The architecture is shown in Figure 2.

Orthogonal Constraint: To explore the structure of the low-rank subspace $\mathcal{B}_\beta \in \mathbb{R}^{k \times q}$ with learnable parameters β, we study two variants: The baseline *LRM-Functa basic* (*LRM-Functa*$_b$) initializes β randomly, whereas *LRM-Functa ortho* (*LRM-Functa*$_o$) uses an orthogonal basis initialization. During training, the modulation vectors v and ϕ, and the shared parameters θ and β, are updated following the meta-learning schedule used in [28]: The inner loop performs $G = 10$ optimization steps over v and $\{\phi_t\}_{t=1}^b$ for each batch of b video frames, followed by an outer-loop update of θ and β. Modulation vectors are initialized to zero before each inner-loop optimization. The loss function is given by

$$\mathcal{L}_t = \frac{1}{N} \sum_{i=1}^{N} \| M_{\theta, v, \phi_t, \beta}(x_i, y_i) - z_i \|_2^2 + \lambda_{ortho} \left\| \mathcal{B}_\beta^\top \mathcal{B}_\beta - I \right\|_F, \tag{1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, I is the identity matrix, and $\lambda_{ortho} = 0$ in the basic case, and $\lambda_{ortho} = 1$ in the orthogonal variant.

2.3 Inference and Downstream Task Integration

During inference, we follow Algorithm 1 of *VidFuncta* [28], freezing the shared parameters θ and β while fitting the modulation vectors v and temporal modulations $\phi = \{\phi_t\}_{t=1}^{T}$ to each video. The fitted v and ϕ are stored as a compact representation of the video, and the reconstructed video $\hat{\mathcal{V}}$ is computed. To extract signals from the raw trajectories ϕ, we apply Algorithm 1 of [29], as shown in Figure 3. We estimate the principal motion direction p using PCA [27] on the motion directions $d_t = \frac{\phi_{t+1} - \phi_t}{\|\phi_{t+1} - \phi_t\|}$. The raw signal s_t is obtained by projecting ϕ_t onto p, which is corrected for baseline wander and smoothed via a Savitzky–Golay filter [23] to obtain s_{filt}. ED and ES indices are identified as valleys and peaks of s_{filt} using a prominence-based threshold.

3 Experiments

We consider three datasets. EchoNet-Dynamic [19] contains 10,030 four-chamber cardiac videos labeled with ejection fraction values and ES/ED frames. For OOD evaluation, we use 17 point-of-care cardiac ultrasound (POCUS) videos showing a two-chamber parasternal long axis view, collected and labeled for ES/ED frames at *anonymous hospital*. We further considered 200 lung ultrasound video clips from adult emergency department patients at *anonymous hospital*, each annotated for the presence of B-lines. We reserved 40 cases for testing. Ethics approval was obtained for the creation of these retrospective, anonymized datasets. All videos are downsampled to a spatial resolution of 112×112, and normalized to values between 0 and 1. We use Python version 3.11.13 and PyTorch version 2.4.1. The vector v has dimension $q = 256$, and we vary k between 2 and 512. All remaining hyperparameters follow the configurations of [28]. All *LRM-Functa* models are trained for 80,000 iterations on a 24GB NVIDIA RTX A5000 GPU, which takes 20 hours per model. We compare *LRM-Functa* against frame-wise *COIN++* [5] and *MedFuncta* [7], time-resolved *VidFuncta* [28], and a convolutional *Autoencoder* [29]. We consider three tasks for quantitative anaylsis:

Compression-Reconstruction Task: We evaluate reconstructions using Peak Signal-to-Noise Ratio (PSNR) and 3D Structural Similarity Index (SSIM3D) between the original video $\mathcal{V}$ and its reconstruction $\hat{\mathcal{V}}$ on the predefined EchoNet-Dynamic test set of 1277 samples. Different compression rates are tested by varying the rank k of the subspace $\mathcal{B}_\beta$, defining the length of the vectors ϕ_t.

ED and ES Frame Detection: We apply the pipeline described in Section 2.3 to detect ED and ES frames on the trajectories ϕ on the EchoNet-Dynamic test set. We compute the mean absolute error (MAE) between the labeled frame and the closest detected frames. For OOD performance on the POCUS set, we use models pretrained on EchoNet-Dynamic train set without fine-tuning.

Downstream Tasks on the Reconstructions: We first reconstruct the videos

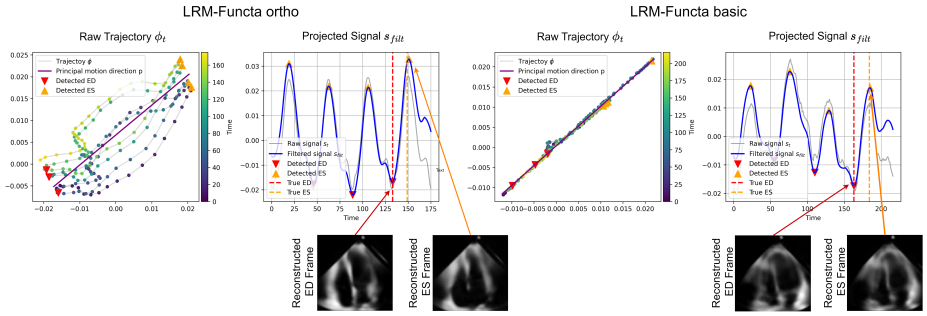

Fig. 3. For $k = 2$, we visualize the raw trajectories ϕ_t over time. *LRM-Functa*$_o$ exhibits spiral-like trajectories, whereas *LRM-Functa*$_b$ collapses the trajectory to a line. Projecting ϕ_t onto the principal motion direction p and filtering yields the 1D signal s_{filt}, which enables direct interpretable identification of ED and ES frames.

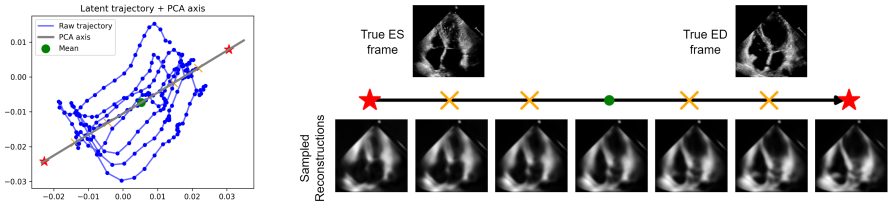

Fig. 4. We plot the raw trajectory ϕ, its mean, and and first principal component for using *LRM-Functa*$_o$. Walking along this axis generates latent encodings, from which reconstructed frames show a smooth progression from ES to ED. Overshooting, indicated by red stars, produces exaggerated contraction and relaxation.

$\hat{\mathcal{V}}$ for EchoNet-Dynamic and the LUS dataset. On these reconstructions, a ResNet 18-3D [9] is trained for 20 epochs using 5-fold cross-validation to predict ejection fraction and classify B-lines, respectively. Mean squared error (MSE) is used for ejection fraction prediction and binary cross-entropy loss for B-line classification. Performance is evaluated using MAE, R^2, and root mean squared error (RMSE) for ejection fraction prediction, and accuracy, F1-score, and area under the receiver operating characteristic curve (AUROC) for B-line classification.

4 Results and Discussion

4.1 Trajectory Visualization and Walk in the Latent Space

As shown in Section 4.2, we can compress ϕ_t to a length of $k = 2$ and visualize it directly. In Figure 3, we show the raw latent trajectories $\phi = \{\phi_t\}_{t=1}^{T}$ over time for two EchoNethDynamic test examples. *LRM-Functa*$_o$ organizes the latent space into a periodic spiral, while in *LRM-Functa*$_b$ the trajectories move back and forth along a line, suggesting that even $k = 1$ may be sufficient to

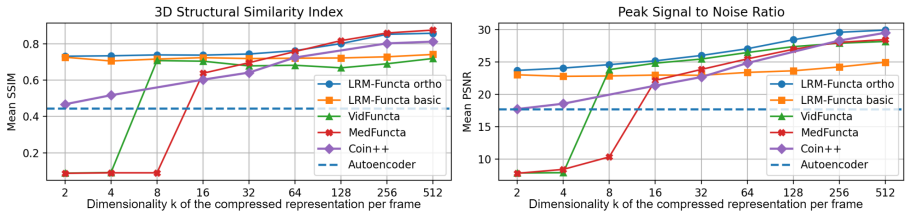

Fig. 5. Mean SSIM and PSNR scores on the EchoNet-Dynamic test set for all comparing methods across varying latent dimensions k (x-axis shown on a log2 scale). For small k, our *LRM-Functa* variants are the only methods to produce stable reconstructions, while for higher ranks, *LRM-Functa*$_o$ achieves among the highest performance.

capture cardiac motion. We project ϕ onto the principal motion direction p and filter it to obtain s_{filt}, as described in Section 2.3. Peaks and valleys of s_{filt} allow unsupervised readout of ED and ES indices. Figure 4 shows the raw latent trajectory of ϕ for another test example using *LRM-Functa*$_o$, with the first principal component of ϕ computed via PCA plotted as a gray axis. Traversing this latent axis and reconstructing frames from sampled points reveals a smooth ES-to-ED progression, demonstrating that the latent space encodes cardiac motion and supports continuous-time frame sampling.

4.2 Quantitative Results

Compression-Reconstruction Results: In Figure 5, we show the PSNR and SSIM scores on the EchoNet-Dynamic test set as a function of the rank k, which controls the compression rate. For all comparing methods, k denotes the dimensionality of the compressed representation per frame. Our *LRM-Functa* approaches (blue and orange curves) is the only method that maintains stable reconstruction quality at very low ranks ($k = 2$ and $k = 4$). At higher ranks, *LRM-Functa*$_o$ aligns with or even outperforms all comparing methods. Encoding a new video at test time requires 2.28 GB of GPU memory and 27.5 seconds for a video of length $T = 200$.

Frame Detection Results: In Table 1, we report frame detection accuracy on EchoNet-Dynamic for ranks $k = 2$ and $k = 512$, corresponding to compression rates of roughly 3000× and 25×, respectively. *LRM-Functa*$_b$ achieves results close to the state-of-the-art fully supervised approach [14], which reports an ED MAE of 2.1 and an ES MAE of 1.7 on the same test set [14,29]. As shown in Figure 3, *LRM-Functa*$_b$ collapses the latent trajectory to an almost one-dimensional line, suggesting that cardiac motion can largely be represented with $k = 1$ value per frame. This produces a clean projection onto the principal motion direction p, resulting in a clear filtered 1D signal s_{filt} and facilitating ED/ES frame detection. This explains the stronger performance over *LRM-Functa*$_o$ in Table 1. For OOD evaluation on the POCUS test set for $k = 512$, all INR-based

Table 1. Mean ± standard deviation for the ED and ES MAE on the EchoNet-Dynamic and the OOD POCUS test sets. All methods are unsupervised.

	EchoNet, $k = 2$		EchoNet, $k = 512$		OOD POCUS, $k = 512$	
Method	ED MAE	ES MAE	ED MAE	ES MAE	ED MAE	ES MAE
Coin++	8.29 ± 12.3	5.82 ± 10.8	4.46 ± 6.9	3.26 ± 5.5	1.83 ± 1.1	2.32 ± 1.4
MedFuncta	3.72 ± 3.8	3.48 ± 3.4	2.90 ± 3.3	2.17 ± 2.4	1.61 ± 1.3	1.27 ± 1.5
VidFuncta	3.06 ± 2.6	2.90 ± 2.8	3.08 ± 3.3	2.37 ± 2.4	$\mathbf{1.33 \pm 1.2}$	1.89 ± 3.8
Autoencoder	2.81 ± 3.5	2.26 ± 2.7	3.82 ± 4.7	2.44 ± 2.6	3.06 ± 3.5	4.24 ± 2.6
LRM-Functa$_b$	$\mathbf{2.26 \pm 2.5}$	$\mathbf{1.98 \pm 2.1}$	$\mathbf{2.48 \pm 2.9}$	2.21 ± 2.2	1.40 ± 1.5	$\mathbf{1.13 \pm 1.4}$
LRM-Functa$_o$	2.42 ± 2.7	2.15 ± 2.2	3.04 ± 3.5	$\mathbf{2.16 \pm 2.2}$	$\mathbf{1.33 \pm 1.3}$	1.39 ± 2.5

methods maintain strong frame detection performance, while the convolutional *Autoencoder* fails to generalize. Overall, the MAE scores on the POCUS dataset are lower than on EchoNet-Dynamic, likely due to the lower frame rate.

Downstream Task on the Reconstructions: Table 2 summarizes the performance of downstream models trained and tested on reconstructed videos $\hat{\mathcal{V}}$. The bottom row shows scores for the original videos $\mathcal{V}$, serving as an upper bound. For ejection fraction, even at a compression level of $k = 2$, *LRM-Functa*$_o$ largely preserves downstream performance. This highlights the strong compression capability of our approach while preserving clinically relevant information. Compared to *LRM-Functa*$_b$, *LRM-Functa*$_o$ maintains greater latent space variability through the orthogonal constraint, resulting in slightly improved reconstruction fidelity, as shown in Figure 5. This is beneficial for downstream tasks on reconstructed videos, where preserving fine-grained visual information is important, explaining the improved performance of *LRM-Functa*$_o$ in Table 2. For the LUS dataset, finer spatial details are needed to reliably detect B-lines, therefore we focus on $k = 512$. Reconstructions of all *Functa*-based approaches achieve competitive AUROC performance, preserving diagnostically relevant features, whereas the *Autoencoder* fails to retain details required for this task.

5 Conclusion

We explore the latent space of time-resolved INR architectures for ultrasound video analysis and introduce *LRM-Functa*, which applies a low-rank constraint to the modulated latent space. This improves reconstruction quality even at extreme compression rates of $k = 2$ values per frame and allows direct visualization of the latent space for model interpretability. Reconstructions and downstream tasks demonstrate that diagnostically relevant information for estimating ejection fraction and classifying B-lines is preserved. The resulting latent space is clearly structured and interpretable: the cardiac cycle forms a continuous trajectory that can be directly visualized, and ED and ES frames can be identified

Table 2. Mean ± std of downstream performance on $\hat{\mathcal{V}}$, evaluated on the reconstructed EchoNet-Dynamic and LUS test sets using 5-fold cross-validation.

	Ejection fraction, $k = 2$			B-line classifier [$\times 10^{-2}$], $k = 512$		
Method	MAE	RMSE	R^2 score	Accuracy	F1 score	AUROC
Coin++	9.15 ± 0.1	12.17 ± 0.2	0.03 ± 0.03	77.5 ± 5.0	67.8 ± 8.7	80.9 ± 9.0
MedFuncta	9.42 ± 0.1	12.94 ± 0.1	-0.02 ± 0.02	81.0 ± 6.3	73.1 ± 7.7	90.1 ± 1.9
VidFuncta	9.14 ± 0.1	13.09 ± 0.2	-0.02 ± 0.03	**83.5 ± 5.2**	**76.6 ± 7.3**	**91.9 ± 3.6**
Autoencoder	7.16 ± 0.3	9.79 ± 0.5	0.47 ± 0.06	68.0 ± 2.1	50.2 ± 6.5	66.6 ± 5.7
LRM-Functa$_b$	5.70 ± 0.2	7.63 ± 0.3	$\mathbf{0.67 \pm 0.03}$	77.5 ± 6.4	70.9 ± 3.9	89.5 ± 2.2
LRM-Functa$_o$	$\mathbf{5.29 \pm 0.1}$	$\mathbf{7.14 \pm 0.1}$	$\mathbf{0.67 \pm 0.01}$	82.0 ± 5.4	75.3 ± 5.6	91.7 ± 3.7
Original	4.93 ± 0.3	6.71 ± 0.4	0.70 ± 0.03	87.5 ± 5.6	82.8 ± 7.2	91.5 ± 5.9

without additional model training. Our approach outperforms state-of-the-art methods and generalizes to OOD cardiac views. Traversing trajectories in the latent space enables sampling of novel frames, revealing smooth transitions from ED to ES. In future work, we will leverage this latent representation for further downstream tasks, such as anomaly detection and congenital heart disease classification. We also aim to further improve reconstruction fidelity to preserve fine structural details necessary for assessing left ventricular hypertrophy. Overall, our results highlight the potential of *LRM-Functa* latent spaces to provide compact and explainable representations for medical video analysis.

Acknowledgments. This study was funded by the Swiss National Science Foundation (grant number P500PT_222349).

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

References

1. Al Mukaddim, R., MacKay, E., Gessert, N., Erkamp, R., Sethuraman, S., Sutton, J., Bharat, S., Jutras, M., Baloescu, C., Moore, C.L., et al.: Spatiotemporal deep learning-based cine loop quality filter for handheld point-of-care echocardiography. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **71**(11), 1577–1587 (2024)
2. Bauer, M., Dupont, E., Brock, A., Rosenbaum, D., Schwarz, J.R., Kim, H.: Spatial functa: Scaling functa to imagenet classification and generation. arXiv preprint arXiv:2302.03130 (2023)
3. Byra, M., Jarosik, P., Karwat, P., Klimonda, Z., Lewandowski, M.: Implicit neural representations for speed-of-sound estimation in ultrasound. In: 2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS). pp. 1–4. IEEE (2024)
4. Dupont, E., Kim, H., Eslami, S., Rezende, D., Rosenbaum, D.: From data to functa: Your data point is a function and you can treat it like one. arXiv preprint arXiv:2201.12204 (2022)

5. Dupont, E., Loya, H., Alizadeh, M., Goliński, A., Teh, Y.W., Doucet, A.: Coin++: Neural compression across modalities. arXiv preprint arXiv:2201.12904 (2022)
6. Farhad, M., Masud, M.M., Beg, A., Ahmad, A., Ahmed, L.A., Memon, S.: Cardiac phase detection in echocardiography using convolutional neural networks. Scientific reports **13**(1), 8908 (2023)
7. Friedrich, P., Bieder, F., McGinnis, J., Wolleb, J., Rueckert, D., Cattin, P.C.: Medfuncta: A unified framework for learning efficient medical neural fields. In: Medical Imaging with Deep Learning (2026)
8. Grutman, T., Bismuth, M., Glickstein, B., Ilovitsh, T.: Implicit neural representation for scalable 3d reconstruction from sparse ultrasound images. npj Acoustics **1**(1), 14 (2025)
9. Hara, K., Kataoka, H., Satoh, Y.: Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 6546–6555 (2018)
10. Hernandez, K.A.L., Rienmüller, T., Baumgartner, D., Baumgartner, C.: Deep learning in spatiotemporal cardiac imaging: A review of methodologies and clinical usability. Computers in Biology and Medicine **130**, 104200 (2021)
11. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. ICLR **1**(2), 3 (2022)
12. Jiao, J., Droste, R., Drukker, L., Papageorghiou, A.T., Noble, J.A.: Self-supervised representation learning for ultrasound video. In: 2020 IEEE 17th international symposium on biomedical imaging (ISBI). pp. 1847–1850. IEEE (2020)
13. Jin, T., Yu, X., Ai, Z., Guo, H.: Artificial intelligence in ultrasound imaging: A review of progress from machine learning to large language model. Advanced Ultrasound in Diagnosis and Therapy **9**(4), 483–496 (2025)
14. Li, Y., Li, H., Wu, F., Luo, J.: Semi-supervised learning improves the performance of cardiac event detection in echocardiography. Ultrasonics **134**, 107058 (2023)
15. Mada, R.O., Lysyansky, P., Daraban, A.M., Duchenne, J., Voigt, J.U.: How to define end-diastole and end-systole? impact of timing on strain measurements. JACC: Cardiovascular Imaging **8**(2), 148–157 (2015)
16. Maiya, S.R., Gupta, A., Gwilliam, M., Ehrlich, M., Shrivastava, A.: Latent-inr: A flexible framework for implicit representations of videos with discriminative semantics. In: European Conference on Computer Vision. pp. 285–302. Springer (2024)
17. Molaei, A., Aminimehr, A., Tavakoli, A., Kazerouni, A., Azad, B., Azad, R., Merhof, D.: Implicit neural representation in medical imaging: A comparative survey. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2381–2391 (2023)
18. Oikonomou, E., Holste, G., Coppi, A., Baloescu, C., McNamara, R., Khera, R.: Artificial intelligence-enabled detection and phenotyping of left ventricular hypertrophy on real-world point-of-care cardiac ultrasonography and its implications for patient outcomes. Circulation **150**(Suppl_1), A4139043–A4139043 (2024)
19. Ouyang, D., He, B., Ghorbani, A., Lungren, M.P., Ashley, E.A., Liang, D.H., Zou, J.Y.: Echonet-dynamic: a large new cardiac motion video data resource for medical machine learning. In: NeurIPS ML4H Workshop: Vancouver, BC, Canada. vol. 5 (2019)
20. Peng, S., Zhang, Y., Xu, Y., Wang, Q., Shuai, Q., Bao, H., Zhou, X.: Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9054–9063 (2021)

21. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
22. Ryser, A., Manduchi, L., Laumer, F., Michel, H., Wellmann, S., Vogt, J.E.: Anomaly detection in echocardiograms with dynamic variational trajectory models. In: Machine Learning for Healthcare Conference. pp. 425–458. PMLR (2022)
23. Schafer, R.W.: What is a savitzky-golay filter?[lecture notes]. IEEE Signal processing magazine **28**(4), 111–117 (2011)
24. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. Advances in neural information processing systems **33**, 7462–7473 (2020)
25. Takizawa, R., Kodera, S., Kabayama, T., Matsuoka, R., Ando, Y., Nakamura, Y., Settai, H., Takeda, N.: Video clip model for multi-view echocardiography interpretation (2025), https://arxiv.org/abs/2504.18800
26. Truong, A., Mahmoud, A.H., Konaković Luković, M., Solomon, J.: Low-rank adaptation of neural fields. In: Proceedings of the SIGGRAPH Asia 2025 Conference Papers. pp. 1–12 (2025)
27. Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. Chemometrics and intelligent laboratory systems **2**(1-3), 37–52 (1987)
28. Wolleb, J., Bieder, F., Friedrich, P., Tagare, H.D., Papademetris, X.: Vidfuncta: Towards generalizable neural representations for ultrasound videos. In: International Workshop on Advances in Simplifying Medical Ultrasound. pp. 109–119. Springer (2025)
29. Yang, Y., Yang, Q., Cui, K., Peng, C., D'Alberti, E., Hernandez-Cruz, N., Patey, O., Papageorghiou, A.T., Noble, J.A.: Latent motion profiling for annotation-free cardiac phase detection in adult and fetal echocardiography videos. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 316–325. Springer (2025)
30. Zhu, Y., Liu, Y., Zhang, Y., Liang, D.: Implicit neural representation for medical image reconstruction. Physics in Medicine & Biology **70**(12), 12TR01 (2025)