Self-Supervised Spline Autoencoder for Smooth Latent Dynamics in Echocardiography

Leong Chit Jeff Kwan✉[1] [0009−0002−0645−150X] and Albert Chung[1] [0000−0003−4400−9261]

Department of Computer Science and Engineering,
The Hong Kong University of Science and Technology
lcjkwanaa@connect.ust.hk✉, achung@ust.hk

Abstract. Cardiac phase identification is important for assessing heart function in echocardiography, yet existing automatic approaches typically rely on supervised learning or strong structural assumptions. We show that self-supervised learning of smooth latent dynamics can enable phase-conditioned geometric analysis of cardiac motion. In this work, we present a Spline Autoencoder that learns temporally smooth and consistent evolutions of latent embeddings, capturing cardiac motion as low-rank dynamical structures. We discover a cardiac phase variable using a label-agnostic algorithm based on the spectral embedding of normalized latent trajectory velocities, enabling precise temporal localization of end-diastole (ED) and end-systole (ES) events. Evaluated on the EchoNet-Dynamic dataset, our method achieves a localization mean absolute error (MAE) of 2.3 frames (ED/ES MAEs of 2.7/3.6 frames after label-informed semantic assignment with 94.0% accuracy), which is competitive with supervised methods. Zero-shot evaluation on the EchoNet-Pediatric dataset demonstrates robust generalization across populations and imaging views, with a localization MAE of 3.0 frames. These results suggest that temporal smoothness and consistency are effective inductive priors for learning cardiac motion dynamics, yielding a latent geometry that supports precise phase-event localization, with lightweight calibration for semantic ED/ES assignment. Our code is available at https://github.com/LCJKwan/SplineAE-Cardiac.

Keywords: Echocardiography · Cardiac Phase Detection · Latent Dynamics · Spline · Autoencoder

1 Introduction

Cardiovascular diseases are among the leading causes of mortality worldwide [8]. Echocardiography offers a minimally invasive, non-ionizing, and accessible way to assess cardiac function. Accurate identification of key phases of the cardiac cycle, particularly end-diastole (ED) and end-systole (ES), underpins measurements of chamber size, ejection fraction, and myocardial function [5].

Recent deep learning methods perform cardiac phase analysis in echocardiography, including supervised CNN–RNN hybrids [13,7], transformer-based [11] and attention-based [16] models. In some pipelines, ED/ES events are derived from LV volume curves estimated by segmentation masks [9], which also facilitates analysis of beat-to-beat variability and cardiac phase. Although effective, these strategies often require extensive annotations, limiting scalability.

Unsupervised and weakly supervised alternatives reduce labeling by exploiting latent motion structure, from classical manifold learning [4,12] and LV deformation estimation [2] to representation learning [6,15]. However, these approaches often face a trade-off between smooth, interpretable phase progression and sharp ED/ES localization. For example, DeepHeartBeat explicitly parameterizes a circular phase variable by assuming periodicity and constant frequency, but produces temporally dispersed ED/ES estimates under natural variability [6]. On the other hand, Latent Motion Profiling (LMP) constrains motion to a 2D subspace, achieving accurate ED/ES localization but compressing dynamics into noisy trajectories without a recoverable cardiac phase variable [15].

To address this gap, we seek a representation that discovers dynamical phase and enables precise event localization. At the same time, we avoid label supervision or rigid structural assumptions during training, such as a predefined latent dimensionality or explicit periodicity. We propose that low-rank cardiac motion structure can be captured from *temporal smoothness and consistency* as effective inductive priors on the embeddings, by training an autoencoder with a spline-parameterized latent space trajectory. Moreover, we introduce an algorithm for explicit phase-conditional geometric disambiguation for reliable event localization in a high-dimensional latent space. Our contributions are as follows:

1. **Representation Learning of Cardiac Motion Dynamics:** By enforcing temporal smoothness and consistency, we train a self-supervised Spline Autoencoder that captures cardiac motion as low-rank latent trajectories.
2. **Dynamical Phase Discovery:** A cardiac phase variable is recovered from latent trajectory direction flow via spectral embedding of normalized velocities, allowing precise ED/ES localization.
3. **Cross-Population Robustness:** Competitive accuracy of localizing ED and ES (2.3 ± 2.5 MAE for A4CH view) on EchoNet-Dynamic and zero-shot generalization to EchoNet-Pediatric (3.0 ± 3.1 MAE for A4CH view, 3.0 ± 3.8 MAE for PSAX view), demonstrating stable latent trajectory geometry across populations and imaging views.

2 Method

2.1 Learning Smooth Latent Trajectories via Video Reconstruction

Given a video sequence $X \in \mathbb{R}^{T \times C \times H \times W}$, each frame is encoded into a latent representation $\mathbf{z}_t \in \mathbb{R}^D$, which we interpret as samples from a continuous latent trajectory. Our objective is to learn this trajectory in a self-supervised manner that preserves temporal structure and captures the underlying cardiac dynamics. Rather than imposing explicit periodicity or dimensionality constraints, we only encourage the latent path to evolve coherently over time.

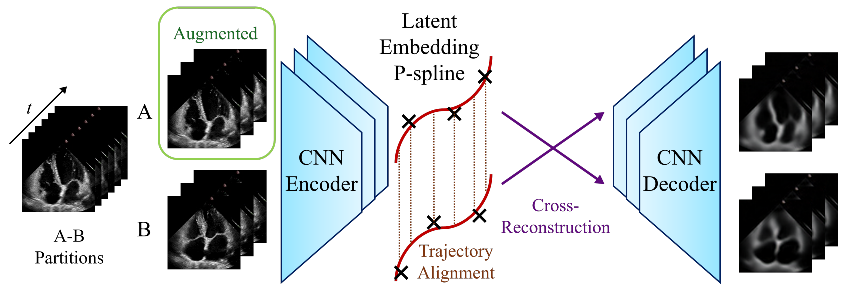

Fig. 1. Overview of the Spline Autoencoder self-supervised training framework. A video segment is split into A–B partitions, with augmentations applied to A. Two P-splines are fitted with temporal smoothness penalties on latent trajectories. The decoder uses cross-partition spline interpolations of latent embeddings, with an additional trajectory alignment regularization loss.

Fig. 1 illustrates our **Spline Autoencoder**. The spline fitted on the latent embeddings acts as a continuous-time model. This design enforces geometric smoothness while allowing the latent trajectory to assume flexible curve shapes. The encoder uses two residual blocks consisting of [3×3 convolution, GroupNorm, GELU, 3×3 convolution] before each down-sampling stage, with five stages followed by average pooling to obtain a latent embedding vector. The decoder mirrors this architecture with symmetric up-sampling stages.

Smooth P-spline Interpolation. For each video segment, timestamps are normalized to $[0, 1]$. A clamped uniform B-spline basis of degree $p = 3$ is chosen, with the cubic degree balancing expressivity and smoothness. We use $n_{\text{ctrl}} = 16$ control points to define the design matrix A, while latent vectors are stacked into Z. With 64-frame partitions, there is approximately 1 control point per 4 frames to capture curvature without fitting frame-level noise. We introduce a diagonal weighting matrix W based on local time intervals Δt estimated via the trapezoidal rule to mitigate non-uniform temporal sampling.

Temporal smoothness is enforced using a second-order finite-difference operator D acting on spline control points. The control points $P \in \mathbb{R}^{n_{\text{ctrl}} \times D}$ are estimated by solving

$$P^\star = \arg\min_P \; \|W^{1/2}(AP - Z)\|_F^2 + \lambda_{\text{spline}} \|DP\|_F^2.$$

This yields the closed-form normal equations

$$(A^\top W A + \lambda_{\text{spline}} D^\top D)\, P = A^\top W Z.$$

Spline fitting remains fully differentiable and can be integrated directly into end-to-end training. The spline can be evaluated at continuous query timestamps to obtain latent embedding vectors for cross-partition decoding.

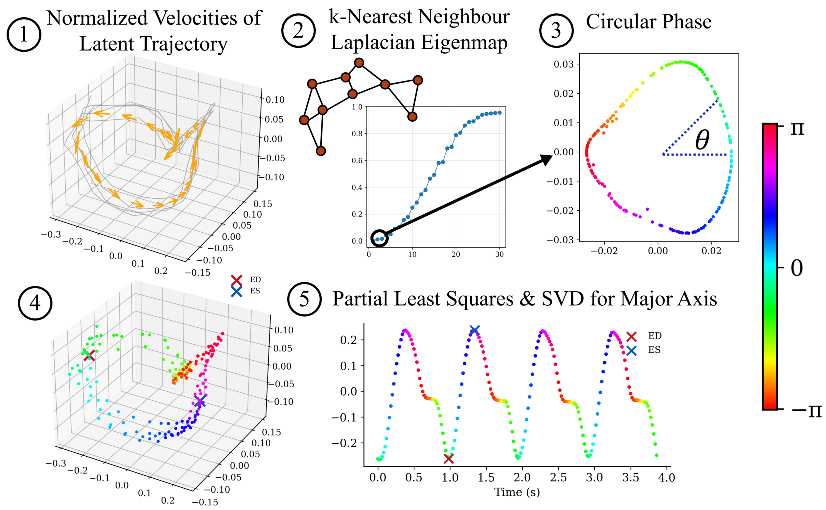

Fig. 2. Pipeline for dynamical phase discovery. (1) Obtain normalized velocities from the trajectory of latent embeddings. (2) Compute the Laplacian eigenmap of the normalized velocities from a k-nearest neighbours affinity graph. (3) Assign circular phase using the two smallest nontrivial components. (4) Assign phase to the original latent trajectory. (5) Use Partial Least Squares to find the 2D plane best describing the loop, and the corresponding major axis, obtained via SVD, for extrema localization.

Sampling Cross-Consistency. To encourage trajectory consistency beyond reconstruction alone, we introduce a cross-consistency training strategy.

Two frame subsets, A and B, are randomly sampled from the same video segment over the same temporal interval, with different frames and timestamps. Splines S_A and S_B are independently fitted to their respective latent embeddings.

Cross-partition reconstruction. Each spline is evaluated at the timestamps of the other subset. The interpolated latents are decoded to reconstruct the corresponding frames using an L_2 loss. Partitions are mutually exclusive if the video has more than 127 frames, but may randomly overlap otherwise.

Latent trajectory alignment. In addition to reconstruction, spline predicted latents are aligned with encoder-derived latents from the opposite subset using an L_2 consistency penalty. The final objective is

$$\mathcal{L} = \mathcal{L}_{\text{cross-MSE}} + \lambda_{\text{align}} \left(\|\mathbf{z}^A - S_B(t^A)\|_2^2 + \|\mathbf{z}^B - S_A(t^B)\|_2^2 \right).$$

This strategy encourages the encoder to learn latent representations that are consistent with the fitted splines across partitions.

2.2 Dynamical Phase Discovery

Algorithm 1 and Fig. 2 summarize the proposed label-agnostic dynamical phase discovery framework. At inference, no fitted spline is used. We interpret encoder

Algorithm 1 Dynamical phase discovery from latent trajectory

Input: Latent trajectory of encoder embeddings $\{\mathbf{z}_t\}_{t=1}^{T}$
Output: Dynamical phase $\{\theta_t\}$ and projected signal $\{s_t\}$
1: **Preprocess to normalized tangent space:**
– Center and linearly detrend $\mathbf{z}_t$ over time
– Compute Savitzky–Golay temporal derivative $\dot{\mathbf{z}}_t$, window $w = 11$, order $p = 3$
– Normalize velocities $\tilde{\mathbf{z}}_t = \dot{\mathbf{z}}_t / \|\dot{\mathbf{z}}_t\|_2$
2: **Spectral embedding and phase assignment:**
– Construct a k-nearest-neighbour affinity graph ($k = 15$)
– Compute graph Laplacian eigenvectors and keep the two smallest nontrivial components $\mathbf{v}_t = (v_t^{(1)}, v_t^{(2)})$
– Assign phase $\theta_t = \mathrm{atan2}(v_t^{(2)}, v_t^{(1)})$
3: **Phase-aligned latent projection:**
– Fit a PLS model using the circular phase encoding $[\cos(\theta_t), \sin(\theta_t)]$ and $\tilde{\mathbf{z}}_t$
– Compute SVD of the coefficient matrix $\mathbf{C} = U\Sigma V^\top$
– Select the leading left singular vector $\mathbf{u}_0$
4: **Projection:**
– Project latent states $s_t = \mathbf{z}_t^\top \mathbf{u}_0$

embedding vectors as tracing out a trajectory over time, analogous to an evolving dynamical system. We use normalized latent velocities for computational stability in high-dimensional spaces. We assume that latent trajectories exhibit a dominant cyclic geometry, which is empirically verified. We choose $k = 15$ neighbours to construct the affinity graph, balancing the preservation of semantically similar samples across cycles against over-connecting temporally adjacent samples. We estimate a phase-conditioned subspace, yielding a projection that captures the contraction-relaxation loop while suppressing non-cyclic fluctuations. Event localization is performed on the projection, where phase-aligned oscillations yield stable extrema.

2.3 Event Localization and Label Assignment

ED/ES *event localization* can be estimated from the geometry of dynamics. However, *assigning semantic labels* (ED versus ES) is nontrivial, because the phase variable is invariant up to a global rotation. Prior methods resolve this ambiguity through additional information. LMP [15] constrains the latent space to 2D and uses a single sample label to orient the global motion axis. DeepHeartBeat [6] plots a histogram to identify the absolute phase distributions of ED/ES. Other approaches incorporate physiological cues, such as LV volume or motion information [2,4,12], which implicitly provide semantic orientation.

Because we relax structural constraints apart from temporal smoothness, we introduce a lightweight label-informed calibration to perform semantic assignment. We compute a global ED-ES alignment axis by averaging the difference vector of ED and ES embeddings over the training set. During inference, we enforce signed orientation using the dot product between this axis and the difference axis between found extrema, so that ED/ES correspond to valleys/peaks.

3 Experiments

We perform self-supervised training on EchoNet-Dynamic [9], which contains 10,030 apical-4-chamber (A4CH) view videos. We use the same training, validation, and test splits as in [9]. We further evaluate the generalization of our method without any fine-tuning on EchoNet-Pediatric [10]. It contains 3,176 A4CH view videos and 4,424 parasternal short axis (PSAX) videos. Both EchoNet datasets have only one pair of ground-truth ED/ES annotations for each video, so we present the mean absolute error (MAE) in frames between the ground truth and the temporally closest prediction.

3.1 Implementation Details

The experiments were accelerated by four RTX 5090 GPUs with an AMD EPYC 9B14 96-core processor, with `PyTorch 2.10.0` with `CUDA 12.8`. We use the provided video resolution of 112×112, normalizing pixel intensities to $[-1, 1]$. From each video, we randomly sample 64-frame partitions A and B, allowing replacement if the video is short. We use the AdamW optimizer with a learning rate of 0.0001, and a weight decay of 0.01, and a batch size of 32 for 200 epochs. We use $\lambda_{align} = 0.001$ for regularization and a spline smoothness penalty of $\lambda_{spline} = 1.0$. We use ultrasound-specific augmentations from `usaugment` [14], adapted for temporally consistent application across video frames. Depth attenuation, Gaussian shadow, haze artifacts, and speckle reduction are stochastically applied to partition A, while partition B remains unaltered for each sampled video segment. This encourages the model to learn representations robust to ultrasound-specific artifacts. The full video is used at inference time.

4 Results

4.1 Properties of the Latent Trajectory

On the EchoNet-Dynamic test set, the average participation ratio of the latent trajectories is 3.85 ± 0.98 for the 512-dimensional frame embeddings. This demonstrates that the trajectories span low-rank linear subspaces. Furthermore, we performed a lightweight topological verification using persistent homology [3] with Ripser [1]. The normalized maximum H_1 persistence (ranging from 0 to 1) is 0.42 ± 0.11, with approximately 97% of trajectories exceeding a persistence threshold of 0.2, indicating consistent nontrivial 1-cycle features.

4.2 Cardiac Phase Detection

We evaluate whether the learned dynamical representation enables accurate localization of cardiac phase events. Table 1 compares our approach with supervised and unsupervised baselines. Our method achieves 2.3 ± 2.5 localization MAE, outperforming unsupervised methods and approaching supervised approaches that rely on phase labels, LV segmentation, or volume estimation.

Table 1. Cardiac phase detection on EchoNet-Dynamic test set. Localization error statistics are computed by pooling ED/ES prediction errors for other methods. The global axis ED/ES assignment, with 94.0% accuracy, was used for our method.

Method	**Supervision**	**MAE (frames)**		
		Localization	**ED**	**ES**
R3D-50 [7]	LV volume	1.9 ± 2.6	2.1 ± 2.5	1.7 ± 2.7
MAEF [16]	LV segmentation	2.4 ± 2.3	2.3 ± 2.3	2.4 ± 2.2
UVT [11]	EF, ED, ES	5.3 ± 10.5	7.2 ± 12.9	3.4 ± 6.8
DDSB† [2]	None	6.6 ± 7.9	4.3 ± 4.8	8.9 ± 9.5
LMP [15]	None	$\underline{2.5 \pm 2.9}$	$\underline{3.0 \pm 3.3}$	$\mathbf{2.0 \pm 2.2}$
Ours	ED/ES Axis§	$\mathbf{2.3 \pm 2.5}$	$\mathbf{2.7 \pm 4.0}$	$\underline{3.6 \pm 5.1}$

† Reported from the LMP [15] re-run on the test set for fair comparison.
§ Label-agnostic localization, label-informed global-axis calibration for ED/ES.

Table 2. Zero-shot cardiac phase detection on EchoNet-Pediatric. Localization detects ED/ES indiscriminately, and the separate ED/ES errors reported are obtained using the global axis found from the adult dataset.

View	**Labeling Acc. (%)**	**MAE (frames)**		
		Localization	**ED**	**ES**
A4CH	93.8	3.0 ± 3.1	3.0 ± 4.3	4.4 ± 5.4
PSAX	83.2	3.0 ± 3.8	4.5 ± 6.8	6.2 ± 8.5

Semantic assignment is evaluated separately using the global axis alignment. After alignment, labeling accuracy reaches 94%, yielding ED and ES errors of 2.7 ± 4.0 and 3.6 ± 5.1 frames, respectively. Compared with LMP [15], which constrains latent dynamics to a low-dimensional space, our method maintains smooth phase organization while preserving competitive localization accuracy.

4.3 Zero-shot Generalization to Pediatric Echocardiography

We evaluate zero-shot generalization on the EchoNet-Pediatric dataset. This introduces a domain shift due to differences in age, anatomy, and imaging views. As shown in Table 2, localization remains stable across adult and pediatric data, with 3.0 MAE for both A4CH and PSAX views. This suggests that cardiac event localization performance depends primarily on the geometry of the latent space trajectory. Using the global ED-ES alignment axis obtained from adult training data, semantic assignment reaches 93.8% accuracy on pediatric A4CH videos. Labeling accuracy drops to 83.2% for PSAX videos. This suggests that changes in imaging view affect the semantic direction of ED/ES in latent space. Overall, these results demonstrate that the learned representation generalizes across populations and imaging views without retraining, with geometric features more robustly preserved than the semantic ED/ES orientation in latent space.

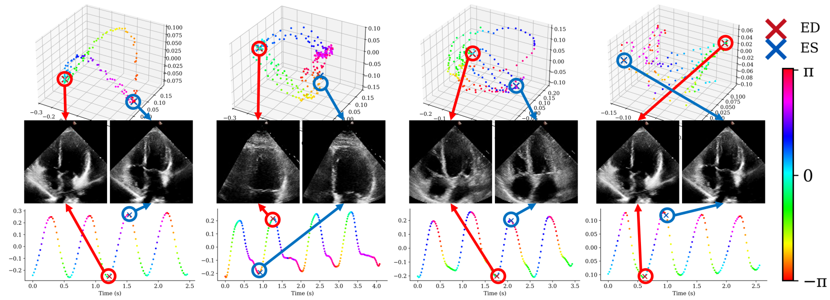

Fig. 3. Rows from top to bottom show the 3D PCA of latent trajectories, ED (left) and ES (right) frames, and the phase-conditioned major axis projection. Ground-truth ED frames are assigned phase = 0 for visual comparison.

4.4 Qualitative Results

Fig. 3 provides qualitative examples of our results, visualizing the latent trajectory, ED/ES key frames, and the phase-conditioned major axis projection. There are four test cases shown from left to right: (1) A cyclic structure with precise ED/ES detection. (2) The cycle is evident, but there is a label assignment error where ED corresponds to peaks rather than valleys. (3) The loop drifts nonlinearly across cycles. (4) 3D PCA fails to reveal interpretable geometry. Across all cases, the major axis projection demonstrates that the dynamical phase varies smoothly across time, with a stable oscillation. Conditioning on the phase variable is also important in extracting the periodic component that localizes ED/ES effectively, across geometric variations. In general, we find that localization is accurate when a coherent latent loop can be found. Conversely, we find localization failure when a loop cannot be constructed from the data, for example, in low-quality videos with the heart chambers moving out of view.

4.5 Ablation Studies

We investigate the effect of the smoothness regularization term λ_{spline} when fitting the cubic spline. The selected setting, $\lambda_{\text{spline}} = 1.0$, achieved the lowest localization error of 2.3 ± 2.5 MAE. Decreasing the smoothness penalty led to a consistent increase in error, with localization errors of 2.9 ± 3.4, 3.4 ± 3.8, 3.4 ± 3.8, and 3.4 ± 3.9 frames for $\lambda_{\text{spline}} = 0.1$, 0.01, 0.001, and 0.0001, respectively. Moreover, we find that removing cross-partition consistency training degrades localization accuracy and increases variability. Enabling cross-partition reconstruction alone with $\lambda_{align} = 0$ improves localization from 2.5 ± 2.7 to 2.4 ± 2.4 frames. To test whether the phase-conditioned subspace improves localization, we implemented an LMP-style baseline that detects the extrema along a PCA-derived principal motion axis [15]. This increases the localization error to 4.5 ± 4.1 MAE, indicating that maximal-variance geometry without using loop phase information is insufficient for accurate localization.

5 Conclusion and Discussion

We present a self-supervised Spline Autoencoder for cardiac motion representation learning, using temporal smoothness and consistency as inductive priors, together with an algorithm that extracts a dynamical phase variable for cardiac phase detection. Experiments on EchoNet-Dynamic demonstrate competitive localization accuracy compared with supervised approaches, while the learned representation maintains a coherent phase structure. Furthermore, zero-shot evaluation on EchoNet-Pediatric shows strong generalization across populations and imaging views, suggesting that learned geometry of dynamics remains robust even when semantic latent directions may shift.

We anticipate future work will investigate connections between learned dynamics and clinical pathologies for downstream tasks. Moreover, our framework requires few domain-specific assumptions, suggesting potential applicability to other cyclic or quasi-periodic temporal dynamics.

Acknowledgments. Leong Chit Jeff Kwan is supported by the Hong Kong PhD Fellowship Scheme. The authors gratefully acknowledge this support.

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

References

1. Bauer, U.: Ripser: Efficient computation of Vietoris–Rips persistence barcodes. Journal of Applied and Computational Topology **5**(3), 391–423 (2021). https://doi.org/10.1007/s41468-021-00071-5
2. Bu, Z., Liu, Y., Huo, J., Peng, J., Wang, K., Zhou, G., Sparks, R., Dasgupta, P., Granados, A., Ourselin, S.: DDSB: An unsupervised and training-free method for phase detection in echocardiography. In: Xu, X., Cui, Z., Rekik, I., Ouyang, X., Sun, K. (eds.) Machine Learning in Medical Imaging. Lecture Notes in Computer Science, vol. 15241, pp. 42–51. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-73284-3_5
3. Ghrist, R.: Barcodes: The persistent topology of data. Bulletin of the American Mathematical Society **45**(1), 61–75 (2008). https://doi.org/10.1090/S0273-0979-07-01191-3
4. Gifani, P., Behnam, H., Shalbaf, A., Sani, Z.A.: Automatic detection of end-diastole and end-systole from echocardiography images using manifold learning. Physiological Measurement **31**(9), 1091–1103 (2010). https://doi.org/10.1088/0967-3334/31/9/002
5. Lang, R.M., Badano, L.P., Mor-Avi, V., Afilalo, J., Armstrong, A., Ernande, L., Flachskampf, F.A., Foster, E., Goldstein, S.A., Kuznetsova, T., Lancellotti, P., Muraru, D., Picard, M.H., Rietzschel, E.R., Rudski, L., Spencer, K.T., Tsang, W., Voigt, J.U.: Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. Journal of the American Society of Echocardiography **28**(1), 1–39.e14 (2015). https://doi.org/10.1016/j.echo.2014.10.003

6. Laumer, F., Fringeli, G., Dubatovka, A., Manduchi, L., Buhmann, J.M.: DeepHeartBeat: Latent trajectory learning of cardiac cycles using cardiac ultrasounds. In: Alsentzer, E., McDermott, M.B.A., Falck, F., Sarkar, S.K., Roy, S., Hyland, S.L. (eds.) Proceedings of the Machine Learning for Health NeurIPS Workshop. Proceedings of Machine Learning Research, vol. 136, pp. 194–212. PMLR (2020), https://proceedings.mlr.press/v136/laumer20a.html
7. Li, Y., Li, H., Wu, F., Luo, J.: Semi-supervised learning improves the performance of cardiac event detection in echocardiography. Ultrasonics **134**, 107058 (2023). https://doi.org/10.1016/j.ultras.2023.107058
8. Martin, S.S., Aday, A.W., Allen, N.B., Almarzooq, Z.I., Anderson, C.A.M., Arora, P., Avery, C.L., Baker-Smith, C.M., Bansal, N., Beaton, A.Z., Commodore-Mensah, Y., Currie, M.E., Elkind, M.S.V., Fan, W., Generoso, G., Gibbs, B.B., Heard, D.G., Hiremath, S., Johansen, M.C., Kazi, D.S., Ko, D., Leppert, M.H., Magnani, J.W., Michos, E.D., Mussolino, M.E., Parikh, N.I., Perman, S.M., Rezk-Hanna, M., Roth, G.A., Shah, N.S., Springer, M.V., St-Onge, M.P., Thacker, E.L., Urbut, S.M., Van Spall, H.G.C., Voeks, J.H., Whelton, S.P., Wong, N.D., Wong, S.S., Yaffe, K., Palaniappan, L.P.: 2025 heart disease and stroke statistics: A report of US and global data from the American Heart Association. Circulation **151**(8), e41–e660 (2025). https://doi.org/10.1161/CIR.0000000000001303
9. Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C.P., Heidenreich, P.A., Harrington, R.A., Liang, D.H., Ashley, E.A., Zou, J.Y.: Video-based AI for beat-to-beat assessment of cardiac function. Nature **580**(7802), 252–256 (2020). https://doi.org/10.1038/s41586-020-2145-8
10. Reddy, C.D., Lopez, L., Ouyang, D., Zou, J.Y., He, B.: Video-based deep learning for automated assessment of left ventricular ejection fraction in pediatric patients. Journal of the American Society of Echocardiography **36**(5), 482–489 (2023). https://doi.org/10.1016/j.echo.2023.01.015
11. Reynaud, H., Vlontzos, A., Hou, B., Beqiri, A., Leeson, P., Kainz, B.: Ultrasound video transformers for cardiac ejection fraction estimation. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Lecture Notes in Computer Science, vol. 12906, pp. 495–505. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87231-1_48
12. Shalbaf, A., AlizadehSani, Z., Behnam, H.: Echocardiography without electrocardiogram using nonlinear dimensionality reduction methods. Journal of Medical Ultrasonics **42**(2), 137–149 (2015). https://doi.org/10.1007/s10396-014-0588-y
13. Taheri Dezaki, F., Liao, Z., Luong, C., Girgis, H., Dhungel, N., Abdi, A.H., Behnami, D., Gin, K., Rohling, R., Abolmaesumi, P., Tsang, T.: Cardiac phase detection in echocardiograms with densely gated recurrent neural networks and global extrema loss. IEEE Transactions on Medical Imaging **38**(8), 1821–1832 (2019). https://doi.org/10.1109/TMI.2018.2888807
14. Tupper, A., Gagné, C.: Revisiting data augmentation for ultrasound images. Transactions on Machine Learning Research (2025), https://openreview.net/forum?id=iGcxlTLIL5
15. 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: Gee, J.C., et al. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. Lecture Notes in Computer Science, vol. 15973, pp. 316–325. Springer, Cham (2026). https://doi.org/10.1007/978-3-032-05185-1_31

16. Zeng, Y., Tsui, P.H., Pang, K., Bin, G., Li, J., Lv, K., Wu, X., Wu, S., Zhou, Z.: MAEF-Net: Multi-attention efficient feature fusion network for left ventricular segmentation and quantitative analysis in two-dimensional echocardiography. Ultrasonics **127**, 106855 (2023). https://doi.org/10.1016/j.ultras.2022.106855