

Latent Motion Profiling for Annotation-free Cardiac Phase Detection in Adult and Fetal Echocardiography Videos

Yingyu Yang¹, Qianye Yang¹, Kangning Cui^{1,2}, Can Peng¹, Elena D’Alberti^{3,4}, Netzahualcoyotl Hernandez-Cruz¹, Olga Patey⁴, Aris T. Papageorgiou⁴, and J. Alison Noble¹

¹ Department of Engineering Science, University of Oxford, Oxford, UK
`firstname.lastname@eng.ox.ac.uk`

² Hong Kong Centre for Cerebro-Cardiovascular Health Engineering, Hong Kong, China

³ Department of Maternal and Child Health and Urological Sciences, Sapienza University of Rome, Rome, Italy

⁴ Nuffield Department of Women’s and Reproductive Health, University of Oxford, Oxford, UK

Abstract. The identification of cardiac phase is an essential step for analysis and diagnosis of cardiac function. Automatic methods, especially data-driven methods for cardiac phase detection, typically require extensive annotations, which is time-consuming and labour-intensive. In this paper, we present an unsupervised framework for end-diastole (ED) and end-systole (ES) detection through self-supervised learning of latent cardiac motion trajectories from 4-chamber-view echocardiography videos. Our method eliminates the need for manual annotations—including ED ES indices, segmentation, or volumetric measurements—by training a reconstruction model to encode interpretable spatiotemporal motion patterns. Evaluated on the EchoNet-Dynamic benchmark, the approach achieves mean absolute error (MAE) of 3.0 frames (58.3 ms) for ED and 2.0 frames (38.8 ms) for ES detection, matching state-of-the-art supervised methods. Extended to fetal echocardiography, the model demonstrates robust performance with MAE 1.5 frames (20.7ms) for ED and 1.7 frames (25.3ms) for ES, despite the fact that the fetal heart model is built using non-standardized heart views due to fetal heart positioning variability. Our results demonstrate the potential of the proposed latent motion trajectory strategy for cardiac phase detection in adult and fetal echocardiography. This work advances unsupervised cardiac motion analysis, offering a scalable solution for clinical populations lacking annotated data. Code is released at <https://github.com/YingyuYy/CardiacPhase>.

Keywords: Cardiac Phase Detection · Latent Motion Model · Adult heart · Fetal heart.

1 Introduction

Echocardiography is a widely used clinical modality for assessing cardiac function due to its low cost and absence of ionizing radiation. Precise identification of cardiac phases, particularly two important key-points: end-diastole (ED) and end-systole (ES), is critical to derive quantitative functional parameters such as the ejection fraction and ventricular longitudinal strain in adults [6], as well as to perform various measurements for fetal heart assessment [2]. ED and ES also enable motion pattern alignment for spatiotemporal group analysis [11]. ED is defined as the frame before mitral valve closure and ES marks aortic valve closure [10]. In the 4-chamber (4CH) view, one of the most informative echocardiographic planes, the current clinical practice relies on visual assessment by sonographers or cardiologists to identify these key frames, a subjective process that introduces inter-observer variability and depends heavily on operator experience.

Automated cardiac phase detection in 4CH echocardiography has employed direct and indirect approaches. Direct data-driven methods use deep learning architectures to predict ED/ES end-to-end. Spatio-temporal information is processed through hybrid CNN-RNN frameworks [3, 8], Vision Transformers [14] to estimate temporal ED/ES probability. Authors of [13] apply a multi-task CNN network to perform fetal heart localization, phase detection and standard plane classification. Indirect methods derive ED/ES from intermediate prediction such as left ventricle (LV) volume curves [9] or temporal segmentation masks [19], extracting phases from extrema. Both strategies require extensive manual annotations (e.g., ED/ES indices, LV segmentation), limiting their applicability in annotation-scarce situations.

Unsupervised ED/ES detection methods offer promising solutions to reduce annotation dependency. Early approaches like those by Gifani and Shalbaf [5, 17] mapped apical 4CH echocardiography videos to low-dimensional manifolds via locally linear embedding (LLE), detecting phases through density analysis. However, their case-specific embeddings lack interpretability and generalizability. Laumer et al. [7] advanced this by learning a circular latent trajectory from cardiac videos, with ED/ES phases semantically aligned to specific arc regions in the learned latent circle. While innovative, this method cannot pinpoint exact ED/ES timings, limiting potential clinical utility. More recently, a training-free unsupervised approach based on left ventricular expansion-contraction dynamics, called DDSB, was introduced [1]. Though computationally efficient, it struggles with videos containing short ED/ES intervals, hindering broad applicability.

We propose LMP (Latent Motion Profiling), an annotation-free framework for unsupervised cardiac phase detection in apical 4CH echocardiography. Our key contributions are as follows.

- **Unsupervised Interpretability:** A novel latent motion subspace where cardiac dynamics are encoded into two orthogonal physiologically meaningful directions (septal and lateral movement). ED and ES occupy distinct trajectory regions, enabling annotation-free detection.

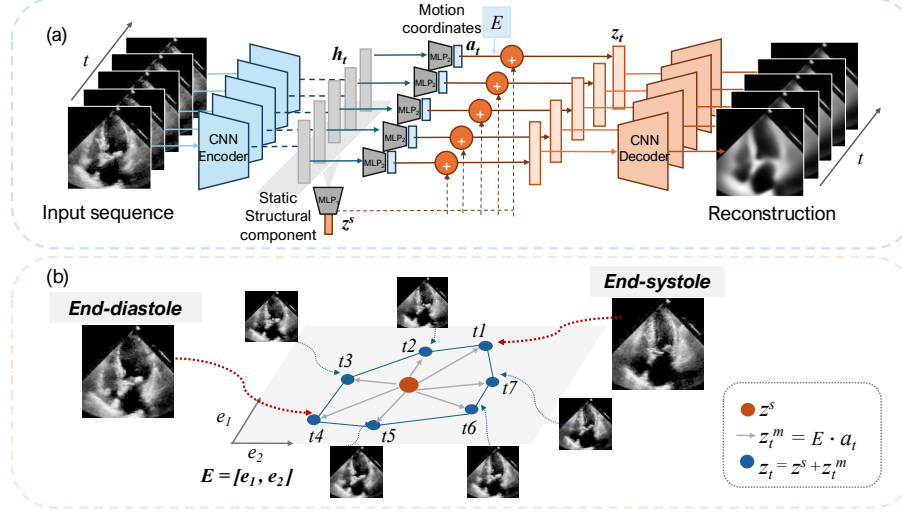


Fig. 1. Latent Cardiac Motion Profiling via Frame-wise Reconstruction. (a) Structure-motion decomposition: Input frame x_t is encoded as $z_t = z^s + \mathbf{E}a_t$, where $z^s \in \mathbb{R}^D$ represents static anatomy and $\mathbf{E}a_t$ models motion in orthogonal subspace $\mathbf{E} = [e_1, \dots, e_K]$. (b) Motion trajectory: Temporal evolution of $\{a_t\}$ in $\mathbb{R}^K$ ($K = 2$ in our case) reveals ED/ES phases as geometric landmarks, enabling unsupervised detection through trajectory analysis.

- **Label-Free Generalization:** By training solely on video reconstruction, LMP eliminates dependence on ED/ES indices, segmentations, or clinical measurements, overcoming a critical bottleneck in fetal and paediatric applications where annotations are scarce.
- **Cross-Population Robustness:** Consistent performance in adult and fetal echocardiography, achieving: supervised-level accuracy on adult data benchmarks (MAE: 3.0 ED / 2.0 ES frames), and on fetal results (1.5 ED / 1.7 ES frames). This demonstrates the first unified framework for cardiac phase detection in both populations.

2 Method

2.1 Echocardiography Motion Profiling via Video Reconstruction

Given a video sequence $X \in \mathbf{R}^{T \times H \times W}$, where T , H , and W denote the temporal, spatial height, and spatial width dimensions, respectively. Our objective is to learn a temporally coherent sequence of low-dimensional latent vectors $\{z_t\}_{t=1}^T \in \mathbf{R}^D$ for video reconstruction through an autoencoder framework. Inspired by the low-dimensional representation analysis of cardiac deformation[15, 16], we propose a *structure-motion decomposition* approach. Specifically, for each frame

x_t , the latent representation z_t is decomposed into two components:

$$z_t = z^s + z_t^m \quad (1)$$

Static structural component: $z^s \in \mathbf{R}^D$ encodes subject-specific anatomical features and accounts for structural variations between distinct videos. It is computed as the mapping of the mean latent embedding through a MLP $z^s = \mathbf{MLP}_1(\frac{1}{T} \sum_{t=1}^T h_t)$, where h_t is the frame-wise embedding via a convolutional encoder (Fig. 1(a)).

Dynamic motion component: $z_t^m \in \mathbf{R}^D$ captures temporally evolving cardiac motion relative to z^s (Fig. 1(b)). z_t^m lies in a low-rank subspace spanned by an orthogonal basis $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_K] \in \mathbf{R}^{D \times K}$, where $K \ll D$. The basis vectors $\{\mathbf{e}_k\}_{k=1}^K$, optimized during training, define a motion subspace that preserves trajectory smoothness and physiological plausibility. Temporal dynamics within this subspace are governed by learnable coefficients $\mathbf{a}_t \in \mathbf{R}^K$ through another MLP: $\mathbf{a}_t = \mathbf{MLP}_2(h_t)$ [18]. The motion component is computed $z_t^m = \mathbf{E}\mathbf{a}_t$.

Finally, the composite latent z_t is decoded by the decoder $\mathcal{D}$ to reconstruct the input frame. To enforce consistency between the decomposed latent space and pixel-level video dynamics, we apply the following loss function for optimization:

$$\mathcal{L}(X) = \frac{1}{T} \sum_{t=1}^T \|\mathcal{D}(z^s) - x_t\|_2^2 + \sum_{t=1}^T \|\mathcal{D}(z_t) - x_t\|_2^2 \quad (2)$$

where the first term aims to find the z^s that reconstructs the Fréchet mean image of a given sequence and the second term focuses on the reconstruction of each temporal frame.

2.2 Cardiac Phase Identification in Learned Motion Space

Cardiac motion during one cardiac cycle, as captured by the 4CH view, demonstrates the contraction of the ventricle (with the valve moving toward the apex and the ventricular volume decreasing) and the relaxation of the ventricle (with the valve moving away from the apex and the ventricular volume increasing). In this work, we choose to set the motion subspace with $K = 2$ due to its simplicity and the interpretability of the disentanglement of cardiac motion in the 4CH view. We find that through the reconstruction task, the latent motion subspace learns, in an unsupervised manner, the movement of the septal and lateral heart wall, respectively (Fig. 2). This results in a latent motion trajectory following a back-and-forth motion pattern, with the extremities of the trace representing the points of motion phase change, which are closely related to ED and ES (Fig. 3(b) and Fig. 3(d)). We apply Algorithm 1 to detect ED and ES indices of a given motion trajectory.

3 Experiments

We conducted unsupervised training experiments on two independent echocardiography video datasets: (1) a public adult dataset and (2) a private fetal

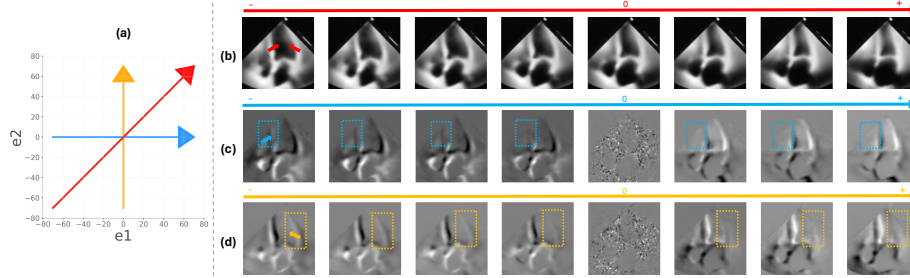


Fig. 2. Latent motion disentanglement visualization in the motion subspace spanned by $\mathbf{E} = [e_1, e_2]$. (a) Three latent motion trajectories: two axes-specific motions (orange and blue lines) and their combination (red line). (b) Reconstruction of the red trajectory showing cardiac contraction with left ventricle volume decreasing and the movement of the septum and lateral wall (red arrows). (c) Reconstruction difference between red and blue trajectories, with minimal differences in the dashed region indicating e_1 axis correlation with septal movement (blue arrow). (d) Reconstruction difference between red and orange trajectories, with minimal differences in the dashed region indicating e_2 axis correlation with lateral movement (orange arrow).

dataset. Both experiments shared core architectures but employed distinct pre-processing and evaluation protocols to address domain-specific challenges.

3.1 Adult Echocardiography Experiment

The adult analysis used the EchoNet-Dynamic dataset [12], containing 10,030 apical 4CH view videos. They have annotated one single ED and ES frames for each video. We followed the train/validation/test splits in [12].

Evaluation Since only one ground truth ED/ES was provided, we calculated the mean absolute error (MAE) between the ground truth and the temporally closest prediction.

3.2 Fetal Echocardiography Experiment

Our private fetal dataset comprised 449 four-chamber view B-mode videos of normal fetal hearts from 313 healthy participants imaged during the second trimester of gestation at the John Radcliffe Hospital, Oxford University Hospitals NHS Foundation Trust, UK. Each participant contributed between 1 to 4 videos. The videos were acquired using GE Voluson E10 and E8 ultrasound machines, with frame rates ranging from 23 to 103 Hz (mean 65 Hz). The video frames have dimensions of 1136×786 pixels. We manually cropped the heart area and aligned it with apical 4CH view orientation (Fig. 3. (a)). A fetal cardiologist annotated 44 test videos generally containing 4 cardiac cycles, providing continuous ED/ES labels. The remaining 405 videos were split 85:15 for training and validation.

Training strategies Two training approaches were implemented: 1) *Standard Training*: Model initialization as for the adult echocardiography experiment.

Algorithm 1 ED/ES detection from latent motion trajectory

Input: Trajectory point coordinates $\{\mathbf{a}_t = (a_t^1, a_t^2)\}_{t=1}^T$ **Output:** End-diastole indices Group_{ED} , End-systole indices Group_{ES} **1: Principal orientation detection:**

- Compute normalized displacement vectors: $d(t) = \frac{\mathbf{a}_{t+1} - \mathbf{a}_t}{\|\mathbf{a}_{t+1} - \mathbf{a}_t\|}$
- Apply RANSAC [4] to determine inlier displacement vectors and compute the main orientation axis $\mathbf{v} = (v_1, v_2)$ using PCA

2: Trajectory projection to the principal orientation axis:

- Normalize the main orientation direction $\bar{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$ and compute the mean position of the trajectory μ
- Projected trajectory: $\mathbf{s} = \{s_t\}_{t=1}^T, s_t = (a_t - \mu) \cdot \bar{\mathbf{v}}$

3: Projected trajectory preprocessing:

- Smoothing: Apply Savitzky-Golay filter to $\mathbf{s}$ with window $w = 8$, order $p = 2$
- Remove baseline wander: Apply high-pass filter (cutoff 0.5 Hz) if low-frequency power ratio exceeds 0.1
- Get filtered signal $\mathbf{s}_{filt}$

4: ED/ES indices detection:

- Find peaks and valleys in $\mathbf{s}_{filt}$ using prominence threshold $0.3(\mathbf{s}_{filt}^{max} - \mathbf{s}_{filt}^{min})$
 - Assign indices: $\text{Group}_{ES} = \text{Peaks}, \text{Group}_{ED} = \text{Valleys}$
-

2) *Transfer Learning*: Fine-tuning the best-performing adult echocardiography model on fetal data, preserving learned motion priors.

Evaluation We adopted three complementary metrics: (1) *GT-centric MAE* focuses on temporal alignment completeness by measuring the error between each ground truth (GT) frame and its closest prediction. It penalizes failures to detect true events. (2) *Prediction-centric MAE* evaluates detection precision by computing the error between each prediction and its nearest GT annotation. It penalizes extraneous or false-positive detections. (3) *Matched-pair MAE* assesses localization accuracy for temporally aligned pairs: GT and prediction pairs were matched if their temporal offset was less than 50% of the mean cardiac cycle length (derived from test data). This metric isolates errors only for confident matches, avoiding noise from mismatches.

3.3 Implementation Details

We preprocessed videos by resizing frames to 128×128 pixels and normalizing intensities to $[0,1]$. We randomly extracted 50-frame clips from each video for training efficiency and temporally downsampled them to 25 frames. The reconstruction model was trained for 500 epochs using the Adam optimizer (learning rate = 0.001 for adult data and 0.0001 for fetal data, batch size = 32) with extensive data augmentation: random brightness/contrast adjustments ($\pm 20\%$), Gaussian blurring ($\sigma \in [0.25, 1.5]$), additive noise ($\sigma = 0.01$), rotation ($\pm 60^\circ$), translation (10% offset range), scaling (0.85–1.15), and horizontal flipping. Model selection was guided by validation split performance. Inference used full-length original videos without temporal sampling.

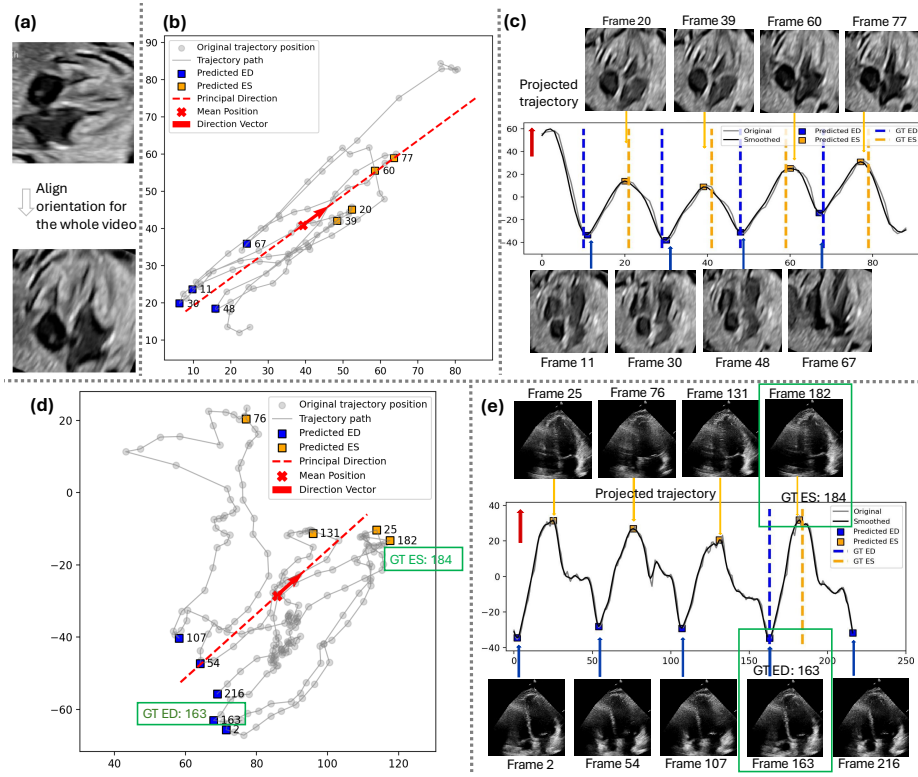


Fig. 3. Latent motion trajectory for ED/ES indices detection. (a-c) A fetal test example. (a) Align fetal 4CH view to canonical apical orientation. (b) 2D latent motion trajectory. (c) Following Algorithm 1, the motion trajectory is projected to the principal direction. Peaks of the projected trajectory indicate ES and valleys indicate ED. Predicted ED/ES image frames are shown. (d-e) An adult test example.

4 Results

4.1 Cardiac Phase Detection in Adult Echocardiography

Table 1 summarizes cardiac phase detection results for the EchoNet-Dynamic test split, comparing our method with existing approaches from [9, 19, 14]. Notably, MAEF and UVT did not report MAE in milliseconds. To enable cross-method comparison, we converted frame-based MAE to temporal error (ms) using a dataset’s mean frame rate of 51.52 fps (19.42 ms per frame), as 80% of test samples were recorded at 50 fps. For DDSB [1], which originally reported results on a subset of EchoNet-Dynamic, we re-evaluated their official implementation on the test split to ensure fair comparison - this accounts for discrepancies from their published results. Our unsupervised method achieves MAEs of 3 frames (58.3 ms) for end-diastole (ED) and 2 frames (39.8 ms) for end-systole (ES) detection. This performance not only surpasses the unsupervised approach, but

Table 1. Cardiac phase detection in adult echocardiography (1276 test videos of [12]).

Method	Supervision	Annotation	MAE (frames)		MAE (ms)	
			mean (<i>std</i>)		mean (<i>std</i>)	
			ED	ES	ED	ES
R3D-50 [9]	Yes	LV volume	2.1 (2.5)	1.7 (2.7)	40.2 (48.9)	32.6 (51.6)
MAEF [19]	Yes	LV segmentation	2.3 (2.3)	2.4 (2.2)	44.5 (44.2)	46.1 (42.3)
UVT [14]	Yes	EF,ED,ES	7.2 (12.9)	3.4 (6.8)	139.2 (250.4)	65.0 (131.9)
DDSB [1] ^a	No	None	4.3 (4.8)	8.9 (9.5)	83.4 (93.3)	175.4 (185.4)
Ours	No	None	3.0 (3.3)	2.0 (2.2)	58.3 (66.4)	39.8 (43.0)

^a Rerun using official implementation**Table 2.** Cardiac phase detection in fetal echocardiography (44 test videos)

Training	Metric	#Samples		MAE (frames)		MAE (ms)	
				mean (<i>std</i>)		mean (<i>std</i>)	
		ED	ES	ED	ES	ED	ES
Standard	GT-centric	165	161	3.2 (7.2)	3.8 (9.5)	46.3 (103.0)	53.7 (126.3)
	Prediction-centric	157	154	2.2 (4.7)	2.4 (4.1)	31.3 (69.6)	34.5 (57.8)
	Match-pair	155	152	1.5 (1.4)	1.8 (1.4)	21.0 (19.7)	25.7 (19.4)
Transfer	GT-centric	165	161	3.2 (8.9)	2.7 (5.2)	48.5 (142.6)	39.6 (77.5)
	Prediction-centric	159	157	2.2 (4.8)	2.2 (3.8)	31.3 (71.9)	31.8 (52.8)
	Match-pair	157	155	1.5 (1.3)	1.7 (1.5)	20.7 (17.9)	25.3 (21.7)

also matches the accuracy of state-of-the-art supervised methods. An example of detection shown in Fig. 3(e).

4.2 Cardiac Phase Detection in Fetal Echocardiography

Table 2 presents fetal cardiac phase detection results using our unsupervised method. Transfer learning from adult echocardiography preserved learned ED/ES motion patterns, achieving a reduction of 1.1 frames in ES MAE compared to standard fetal training for GT-centric metrics. Both approaches learned interpretable cardiac motion trajectories without any annotation from real-world fetal data. Our method has correctly detected 95% ED frames and 96% ES frames. Match-pair MAE (ED: 1.5 frames, ES: 1.7 frames) shows strong alignment with the ground truth. An example of detection shown in Fig. 3(c). In addition, the inference speed of the proposed method is efficient. Taking a video of 184 frames as an example, the prediction of a latent motion trajectory takes approximately 36.8 ms on an NVIDIA RTX 5000 GPU. The extraction of ED/ES frame indices from the latent motion trajectory takes around 6 ms on a Dell Inc. Precision 5860 Tower PC with a 4.8 GHz (max) Intel Xeon(R) w5-2565X processor.

5 Conclusion and Discussion

In this work, we propose an unsupervised cardiac phase detection framework that learns interpretable latent motion trajectories from echocardiography videos without requiring annotations (ED/ES indices, LV segmentation, or volume measurements). Validated for both adult (apical 4CH views) and fetal echocardiography, our approach achieves performance comparable to supervised methods. Notably, this work establishes a foundation for annotation-free cardiac motion analysis for diverse clinical populations. Finally, in this work, we mainly focus on apical 4CH view videos (where we have aligned the fetal 4CH view videos to apical orientation), while in reality, positional variability often results in non-apical 4CH views (e.g., basal/transverse) in the fetal cardiac examination, requiring future work to address view-invariant motion modelling. Beyond phase detection, the learned latent space may enable novel applications, including motion artifact disentanglement and pathology-sensitive trajectory clustering, which will be investigated in future work.

Acknowledgments. This work was partly supported by the InnoHK-funded Hong Kong Centre for Cerebro-cardiovascular Health Engineering (COCHE) Project 2.1 (Cardiovascular risks in early life and fetal echocardiography). Co-authors J. Alison Noble and Aris Papageorgiou were supported by the Oxford Partnership Comprehensive Biomedical Research Centre with funding from the NIHR Biomedical Research Centre (BRC) funding scheme.

Disclosure of Interests. The authors have no competing interests.

References

1. 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*. pp. 42–51. Springer Nature Switzerland, Cham (2025)
2. Crispi, F., Valenzuela-Alcaraz, B., Cruz-Lemini, M., Gratacós, E.: Ultrasound assessment of fetal cardiac function. *Australasian journal of ultrasound in medicine* **16**(4), 158–167 (2013)
3. Dezaki, F.T., Liao, Z., Luong, C., Girgis, H., Dhungel, N., Abdi, A.H., Behnami, D., Gin, K., Rohling, R., Abolmaesumi, P., et al.: Cardiac phase detection in echocardiograms with densely gated recurrent neural networks and global extrema loss. *IEEE transactions on medical imaging* **38**(8), 1821–1832 (2018)
4. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* **24**(6), 381–395 (1981)
5. Gifani, P., Behnam, H., Shalbf, A., Sani, Z.A.: Automatic detection of end-diastole and end-systole from echocardiography images using manifold learning. *Physiological measurement* **31**(9), 1091 (2010)

6. 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., et al.: Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the american society of echocardiography and the european association of cardiovascular imaging. *European Heart Journal-Cardiovascular Imaging* **16**(3), 233–271 (2015)
7. Laumer, F., Fringeli, G., Dubatovka, A., Manduchi, L., Buhmann, J.M.: Deep-heartbeat: Latent trajectory learning of cardiac cycles using cardiac ultrasounds. In: *Machine Learning for Health*. pp. 194–212. PMLR (2020)
8. Lee, L.H., Noble, J.A.: Automatic determination of the fetal cardiac cycle in ultrasound using spatio-temporal neural networks. In: *2020 IEEE 17th international symposium on biomedical imaging (ISBI)*. pp. 1937–1940. IEEE (2020)
9. Li, Y., Li, H., Wu, F., Luo, J.: Semi-supervised learning improves the performance of cardiac event detection in echocardiography. *Ultrasonics* **134**, 107058 (2023)
10. 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)
11. McLeod, K., Sermesant, M., Beerbaum, P., Pennec, X.: Spatio-temporal tensor decomposition of a polyaffine motion model for a better analysis of pathological left ventricular dynamics. *IEEE transactions on medical imaging* **34**(7), 1562–1575 (2015)
12. 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., et al.: Video-based ai for beat-to-beat assessment of cardiac function. *Nature* **580**(7802), 252–256 (2020)
13. Pu, B., Li, K., Chen, J., Lu, Y., Zeng, Q., Yang, J., Li, S.: Hfscdd: a hybrid neural network for fetal standard cardiac cycle detection in ultrasound videos. *IEEE Journal of Biomedical and Health Informatics* (2024)
14. Reynaud, H., Vlontzos, A., Hou, B., Beqiri, A., Leeson, P., Kainz, B.: Ultrasound video transformers for cardiac ejection fraction estimation. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI* 24. pp. 495–505. Springer (2021)
15. Rohé, M.M., Sermesant, M., Pennec, X.: Low-dimensional representation of cardiac motion using barycentric subspaces: a new group-wise paradigm for estimation, analysis, and reconstruction. *Medical image analysis* **45**, 1–12 (2018)
16. 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)
17. Shalbaf, A., AlizadehSani, Z., Behnam, H.: Echocardiography without electrocardiogram using nonlinear dimensionality reduction methods. *Journal of Medical Ultrasonics* **42**, 137–149 (2015)
18. Wang, Y., Yang, D., Bremond, F., Dantcheva, A.: Latent image animator: Learning to animate images via latent space navigation. In: *ICLR 2022-The International Conference on Learning Representations* (2022)
19. 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)