A Semi-Supervised Knowledge Distillation Framework for Left Ventricle Segmentation and Landmark Detection in Echocardiograms

Haoyuan Chen[1], Yonghao Li[1], Long Yang[1], Han Wu[1], Lin Zhou[2(✉)],
Kaicong Sun[1], Dinggang Shen[1,3,4(✉)]

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, 201210, China
dgshen@shanghaitech.edu.cn
[2] The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, 325027, China
xyzyyxs@163.com
[3] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, 200230, China
[4] Shanghai Clinical Research and Trial Center, Shanghai, 201210, China

Abstract. Left ventricular segmentation and landmark detection from echocardiograms are routine practices in clinical settings for comprehensive evaluation of cardiovascular disease. Recently, deep learning-based models have been developed to interpret echocardiograms. However, existing methods face challenges in handling sparse annotations, limiting their clinical applicability. Additionally, their robustness can be significantly influenced by temporal inconsistency (*i.e.*, abrupt prediction fluctuations between consecutive frames) and inter-task conflict (*i.e.*, detected landmarks deviating from segmentation boundaries). To address these issues, we propose a novel semi-supervised framework that integrates: 1) a knowledge distillation method for generating pseudo labels of the numerous unlabeled frames to improve the performance; 2) a Task-aware Spatial-Temporal Network (TSTNet) along with consistency constraints that enhances robustness by enforcing temporal consistency across frames, and inter-task consistency between segmentation and landmark detection. Experimental results on two datasets (a public dataset with 500 subjects and a private dataset with 1,950 subjects) show that our proposed framework significantly outperforms the previous approaches. The source code and dataset are publicly available at https://github.com/chenhy-97/TSTNet.

Keywords: Ultrasound · Knowledge distillation · Semi-supervised learning · Left ventricle segmentation.

1 Introduction

Cardiovascular diseases represent a major clinical concern and remain the leading cause of death worldwide [5]. Due to unique advantages of echocardiography, such as low cost, real-time imaging, and the absence of ionizing radiation, it has become an indispensable tool for diagnosing and assessing heart conditions [7]. In clinical practice, the apical two-chamber (A2C) and four-chamber (A4C) views are particularly critical for diagnostic analysis [1]. In these views, sonographers first identify three key anatomical landmarks: the apex, as well as the left and right sides of the mitral valve. Subsequently, they manually delineate the left ventricular (LV) border to assess cardiac function [6]. However, speckle noise and acoustic shadows in ultrasound imaging often obscure the endocardial borders [13]. These challenges underscore the urgent need for accurate and efficient segmentation and landmark detection in echocardiograms.

Artificial intelligence (AI) demonstrates considerable promise in aiding sonographers with echocardiogram segmentation and landmark detection [15,8]. Traditional AI-based approaches focus on image-level tasks within keyframes [14,10]. Li et al. employed a multi-scale feature fusion mechanism for both segmentation and landmark detection on the end-systolic (ES) and end-diastolic (ED) frames [10]. Nevertheless, these methods fail on numerous intermediate frames, causing even worse performance when images are contaminated by speckle noise and acoustic shadows. To address these challenges, more recent methods incorporate temporal interaction across the unlabeled frames in echocardiogram videos [12,4,11]. Deng et al. utilized a memory network to fuse temporal features [21]. Zhang et al. leveraged inter-temporal interaction to enhance the performance of clinically applicable systems through temporal cross-attention [22]. Despite these advancements, current methods impose pixel-level constraints on the ED and ES frames, but they fail to address the issue of sparse annotations. Furthermore, temporal information is integrated primarily at the final layer of the network. This simple implicit interaction falls short in processing temporal inconsistency (*i.e.*, abrupt prediction fluctuations between consecutive frames) and leads to inter-task conflicts (*i.e.*, landmarks misaligned with segmentation boundaries).

To address these challenges, we propose a novel semi-supervised framework utilizing knowledge distillation for segmentation and landmark detection in echocardiograms. To the best of our knowledge, this is the first work to employ pseudo labels from knowledge distillation for segmentation and landmark detection in echocardiograms. The main contributions of our work are as follows: (1) We propose a teacher-student framework to address the issue of sparse annotations, where the pseudo labels generated by the teacher model possess robust performance on large-scale clinical datasets. (2) We introduce the Task-aware Spatial-Temporal Network (TSTNet) along with multi-consistency constraints for segmentation and landmark detection in echocardiograms, mitigating the robustness issues caused by temporal inconsistency and inter-task conflicts.

2 Method

To address the performance degradation caused by sparse annotations in echocardiogram analysis [16], we propose a semi-supervised framework for left ventric-

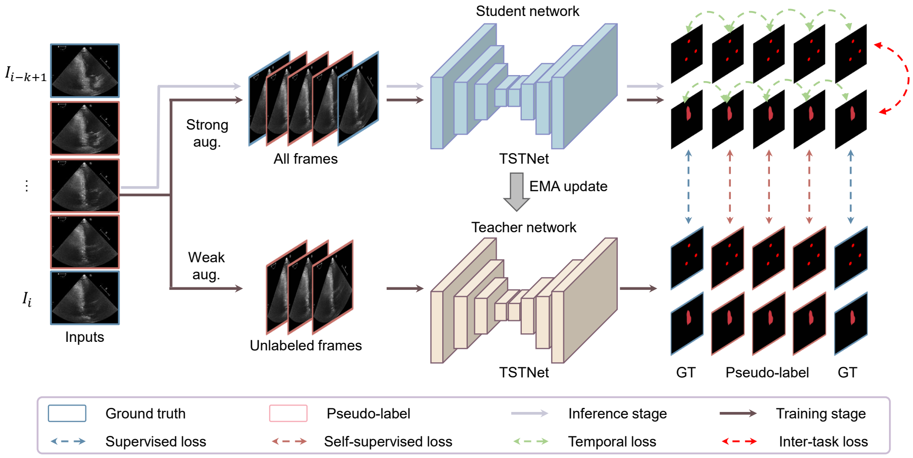


Fig. 1. Overview of our proposed semi-supervised echocardiogram segmentation and landmark detection framework.

ular segmentation and landmark detection. As shown in Fig. 1, our framework consists of a teacher-student architecture [20] where the teacher network generates pseudo labels to recover missing intermediate frame annotations. In both teacher and student networks, we use the proposed TSTNet, a novel backbone designed to jointly model cardiac motion patterns and anatomical structures. To further enhance learning robustness, we introduce consistency constraints that simultaneously enforce temporal coherence and anatomical plausibility. More detailed descriptions are given below.

2.1 Problem Formulation and Knowledge Distillation Network

Problem Formulation. In the semi-supervised echocardiogram segmentation and landmark detection tasks, the training set consists of echocardiography video sequences from multiple subjects. Given N frames of an echocardiogram video $V = \{I_t\}_{t=1}^{N}$ from training set, only a small number of frames are labeled as ED and ES frames $I_L \in \mathbb{R}^{2\times H\times W}$, while the majority of frames are unlabeled $I_U \in \mathbb{R}^{(N-2)\times H\times W}$, where $H \times W$ denotes the spatial resolution of each frame. Each labeled frame contains both the segmentation and landmark labels. The landmark label is represented by a heatmap constructed using a Gaussian function with a kernel size of five pixels. The objective is to train a multi-task model with both I_U and I_L for enhancing the segmentation and landmark detection performance.

Knowledge Distillation Network. The proposed framework integrates a dual-branch knowledge distillation network comprising a student network and a teacher network, designed to jointly leverage labeled and unlabeled frames. Both networks adopt the TSTNet backbone (Sec. 2.2). Within this dual-branch framework, the student network acts as the primary learner, actively exploring diverse imaging conditions. The student network processes all the frames including both I_{L} and I_{U} under strong augmentations, designed to mimic clinical imaging variations. This data perturbation encourages the student to learn robust representations against potential artifacts in clinical practice. From the augmented inputs, the student network generates segmentation probability map P_s and landmark heatmap prediction P_l.

To prevent the student from overfitting to noisy augmented data, the teacher network provides stabilized guidance by generating pseudo labels from anatomically consistent views of the same unlabeled frames. The teacher network operates on weakly augmented versions of the unlabeled frames I_{U}, applying only minimal perturbations to preserve anatomical coherence. This strategy ensures a stable generation of pseudo labels (S_p for segmentation and L_p for landmarks), which serve as surrogate supervision for the unlabeled data.

The training objective combines supervised and self-supervised losses. For labeled frames, the supervised loss $\mathcal{L}_{sup}$ computes the discrepancy between the student's predictions and ground-truth (GT) annotations. The segmentation loss employs the Dice coefficient to measure overlap between P_s and the GT masks S, while the landmark loss uses mean squared error (MSE) to align P_l with the Gaussian-smoothed heatmaps L:

$$\mathcal{L}_{sup} = \underbrace{\left(1 - \frac{2\sum_{i \in M} S^i P_s^i}{\sum_{i \in M} S^{i2} + \sum_{i \in M} P_s^{i2}}\right)}_{\text{Segmentation Loss (Dice)}} + \underbrace{\frac{1}{3}\sum_{j=1}^{3}\left(L^j - P_l^j\right)^2}_{\text{Landmark Detection Loss (MSE)}} \quad , \tag{1}$$

where M denotes the set of all pixel locations in the image, and j indexes the three anatomical landmarks. For unlabeled frames, the self-supervised loss $\mathcal{L}_{self}$ applies an identical formulation but substitutes S and L with the teacher-generated pseudo labels S_p and L_p.

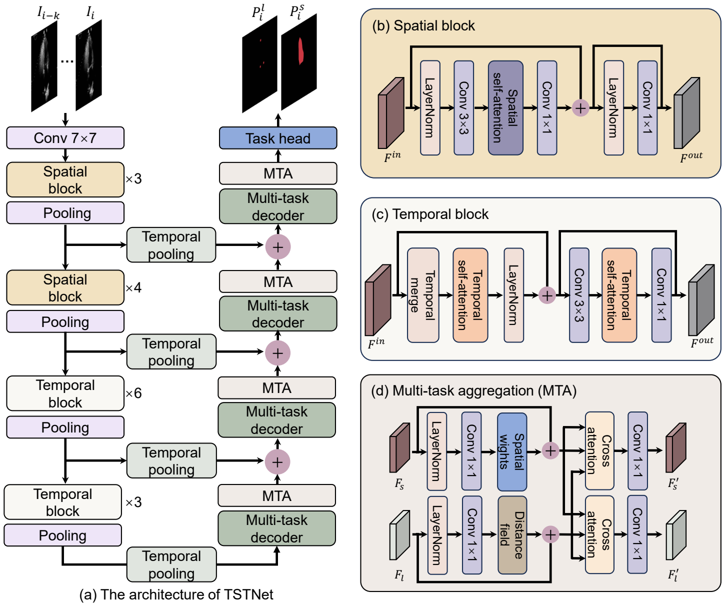
During optimization, gradients are computed and backpropagated solely through the student network. The teacher's parameters are updated via an exponential moving average (EMA) of the student's weights, ensuring gradual refinement of the pseudo label quality while preventing abrupt changes. This update rule is defined in Eq. 2.

$$\theta_{\text{teacher}}^{t+1} = \phi \cdot \theta_{\text{teacher}}^{t} + (1 - \phi) \cdot \theta_{\text{student}}^{t}, \tag{2}$$

where $\theta_{\text{teacher}}^{t}$ represents the parameters of the teacher network at time step t, $\theta_{\text{student}}^{t}$ denotes the parameters of the student network at the same time step, and ϕ is a smoothing coefficient.

2.2 TSTNet

TSTNet is built on the U-shaped architecture to resolve temporal inconsistencies and task conflicts in echocardiogram analysis by integrating spatio-temporal



(a) The architecture of TSTNet

(b) Spatial block

(c) Temporal block

(d) Multi-task aggregation (MTA)

Fig. 2. Overview of the proposed TSTNet for per-frame echocardiogram prediction. (a) The architecture of TSTNet. (b) The spatial block for spatial feature extraction. (c) The temporal block for temporal feature modeling. (d) The multi-task aggregation for combining segmentation and landmark detection tasks.

coherence and anatomical constraints. As shown in Fig. 2 (a), the network processes k consecutive frames $\{I_{i-k+1}, ..., I_i\}$ through a four-stage encoder-decoder structure, generating the segmentation mask P_s^i and landmark heatmap P_l^i of the frame i. The encoder consists of four stages: first applies two stages with spatial blocks (Fig. 2 (b)) that align myocardial boundaries across adjacent frames using spatial self-attention, followed by two stages with temporal blocks (Fig. 2 (c)) that model cardiac motion continuity through dynamically weighted temporal attention. This design explicitly decouples structural alignment from motion dynamics, progressively extracting hierarchical features at four spatial resolutions (1/2 to 1/16 scale).

The decoder integrates four parallel multi-task branches, each processing features at distinct scales through dedicated decoder-MTA pairs. Temporal skip connections deliver multi-scale encoder features to corresponding decoders, where each multi-task decoder is immediately followed by a Multi-Task Aggregation (MTA) module (Fig. 2 (d)). Every MTA enforces anatomical consistency through two operations: First, spatial weights derived from segmentation predictions guide landmark localization toward valid chamber boundaries. Second, landmark distance fields refine segmentation edges to match expected anatomical topology. In the end of the decoder, task-specific prediction heads generate the segmentation map P_s^i and landmark heatmap P_l^i.

2.3 Multi-consistency Constraints

Temporal Motion Smoothness Constraint. Cardiac motion in echocardiography exhibits inherent smoothness due to myocardial elasticity [19]. To enforce this physiological prior, we propose a second-order temporal constraint that penalizes abrupt acceleration changes in both segmentation and landmark predictions. The temporal loss $\mathcal{L}_{temp}$ is calculated in Eq. 3.

$$\mathcal{L}_{temp} = \sum_{t=2}^{N-1} \left[\lambda_S \left\| P_s^{t+1} - 2P_s^t + P_s^{t-1} \right\|_F^2 + \lambda_L \sum_{k=1}^{3} \left\| P_{lk}^{t+1} - 2P_{lk}^t + P_{lk}^{t-1} \right\|_F^2 \right], \tag{3}$$

where t and k index the time frames and landmarks, $\lambda_S = 0.8$ and $\lambda_L = 0.2$ denote the weighting factors balancing segmentation and landmark smoothness, calibrated on validation data, $\| \cdot \|_F^2$ presents Frobenius norm squared, measuring the total magnitude of acceleration changes.

Inter-task Anatomical Constraint. To enforce precise anatomical relationships between landmarks and segmentation boundary, we introduce a boundary-aware constraint mechanism. Given segmentation label S^t at frame t, the boundary coordinates $\mathcal{B}^t$ are derived through morphological gradient operation [17]. The inter-task anatomical constraint loss is formulated in Eq. 4.

$$\mathcal{L}_{task} = \frac{1}{3N} \sum_{t=1}^{N} \sum_{k=1}^{3} \left(1 - \exp\left(-\frac{d_{k,t}^2}{2\sigma^2} \right) \right), \tag{4}$$

where $d_{k,t} = \min_{(u,v)\in\mathcal{B}^t} \|(x_k, y_k) - (u, v)\|_2$ calculates minimum Euclidean distance for each landmark, with $\sigma = 5$ mm controlling the constraint tightness.

Integrated Optimization. The overall loss is formulated as the integration of the four loss functions: $\mathcal{L}_{sup}$, $\mathcal{L}_{self}$, $\mathcal{L}_{temp}$, and $\mathcal{L}_{task}$, defined by $\mathcal{L}_{total} = \mathcal{L}_{sup} + \alpha\mathcal{L}_{self} + \beta\mathcal{L}_{temp} + \gamma\mathcal{L}_{task}$.

3 Experiments

3.1 Dataset and Implementation Details

We validate the effectiveness of our proposed framework using two datasets: (1) **CAMUS** consists of 1,000 echocardiography videos from 500 patients, including A2C and A4C views, each covering half cardiac cycle with 15 frames [9]. (2) **The Second Affiliated Hospital of Wenzhou Medical University (SAH-WMU) Dataset** comprises 1,950 patients, including A2C and A4C views, each encompassing a full cardiac cycle with 30 frames. For both datasets, we partition the data into training, validation, and testing sets in a 7:1:2 ratio. For data augmentations, strong augmentation includes adding Gaussian noise, high-confidence contrast, and brightness enhancement. Weak augmentation involves performing low contrast and brightness enhancement. The inputs are resized to 256×256. We employ Dice Similarity Score (DSC), Intersection over Union

Table 1. The performance of comparative experiments in segmentation and landmark detection tasks.

Method	SAHWMU				CAMUS			
	DSC↑	IoU↑	HD95↓	ADE↓	DSC↑	IoU↑	HD95↓	ADE↓
U-Net [18]	87.4±7.3	78.3±9.3	5.8±3.7	-	91.4±4.9	84.5±7.4	2.6±1.7	-
TranUnet [3]	89.3±7.2	81.3±9.9	5.1±4.6	-	91.8±4.5	85.2±7.0	2.3±1.1	-
SwinUnet [2]	87.4±0.6	78.1±9.2	6.2±4.3	-	90.4±5.4	83.0±7.9	2.7±2.1	-
CLAS [21]	88.5±6.3	79.9±9.2	5.4±4.2	-	91.6±5.5	84.9±8.2	2.9±3.3	-
SAMUS [12]	91.0±5.3	83.9±8.0	3.9±2.5	-	92.2±4.8	85.9±7.6	2.5±1.9	-
MemSAM [4]	91.2±4.2	84.5±6.8	3.9±1.8	-	92.0±4.7	85.5±7.5	2.5±2.1	-
EchoGLAD [14]	-	-	-	3.7±2.2	-	-	-	4.2±2.6
Echo-STG [11]	-	-	-	4.6±2.4	-	-	-	4.6±3.1
EchoEFNet [10]	89.5±5.6	81.4±8.5	4.8±3.6	4.5±3.0	91.8±4.1	85.9±6.7	3.0±1.6	4.3±3.4
CSC [22]	90.4±6.5	83.5±8.9	4.3±3.1	3.9±2.2	92.7±4.8	86.8±7.4	2.5±1.8	4.9±2.9
Ours	**92.7±4.0**	**86.6±6.6**	**3.3±1.9**	**2.7±2.1**	**93.8±3.4**	**88.4±5.8**	**2.0±1.0**	**3.6±2.5**

(IoU), Hausdorff Distance 95% (HD95), and Average Distance Error (ADE) to assess the segmentation and landmark detection tasks.

The experiments were conducted on an Ubuntu 18.04 system utilizing an NVIDIA A100 80G HPC cluster. Python 3.11 along with PyTorch 2.4.0 was employed for the implementation. The training configuration includes 100 epochs, a batch size of 4, and an initial learning rate of 0.0001. The optimization was performed using the AdamW optimizer. The hyper-parameters were set as follows: α, β, γ, ϕ are 1, 5, 0.1, and 0.99, respectively.

3.2 Comparison with State-of-the-Art Methods

We have evaluated the proposed framework against state-of-the-art methods across three categories: segmentation (U-Net [18], TransUnet [3], Swin-Unet [2], CLAS [21], SAMUS [12], MemSAM [4]), landmark detection (EchoGLAD [14], Echo-STG [11]), and multi-task models (CSC [22], EchoEFNet [10]).

Quantitative Comparison. The results of the echocardiogram segmentation and landmark detection tasks are depicted in Table 1. For the segmentation task, the proposed framework improves DSC, IoU, and HD95, by 1.5%, 2.1%, and 0.6 mm, respectively, in the SAHWMU dataset. In the CAMUS dataset, the improvements are 1.1%, 1.6%, and 0.5 mm, respectively. For the landmark detection task, the proposed framework achieves a reduction of 1.0 mm and 0.6 mm in ADE on the SAHWMU and CAMUS datasets, respectively. The quantitative comparisons conclusively show superior performance of our framework, with statistically improvements over existing methods.

Qualitative Comparison. A qualitative analysis using segmentation and landmark detection results is performed on two patients, as shown in Fig. 3. The ED and ES frames from these patients exhibit blurred endocardial and left ventricular landmarks, primarily due to speckle noise around the ventricle. Other baseline and multi-task methods fail to recognize the blurred endocardial boundaries accurately and left ventricular landmarks. In contrast, our method demonstrates

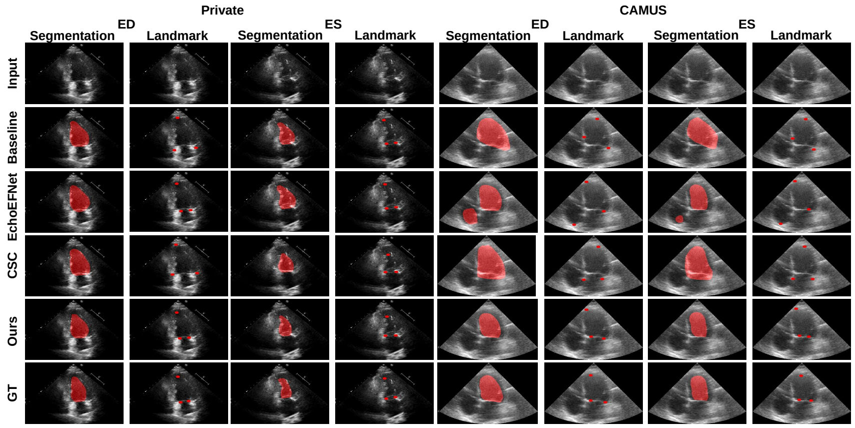


Fig. 3. Comparison of segmentation and landmark detection results for different methods on the SAHWMU and CAMUS datasets. The baseline method is the best-performing task-specific method in Table 1.

Table 2. Ablation studies of the proposed framework. **Sup.** represents supervised using only ED and ES frames. **Semi.** refers to the semi-supervised using intermediate frames. **Cons.** indicates the inclusion of temporal and inter-task constraints.

Method	SAHWMU				CAMUS			
	DSC↑	IoU↑	HD95↓	ADE↓	DSC↑	IoU↑	HD95↓	ADE↓
Sup.	91.6±4.5	84.9±7.2	3.7±2.1	2.9±2.1	92.4±3.7	87.8±6.3	2.6±1.8	3.7±2.6
Semi.	92.6±3.9	86.4±6.4	3.4±2.0	2.8±2.1	93.4±4.0	88.0±6.5	2.5±1.4	3.7±3.5
Semi.+ Cons.	**92.7±4.0**	**86.6±6.6**	**3.3±1.9**	**2.7±2.1**	**93.8±3.4**	**88.4±5.8**	**2.0±1.0**	**3.6±2.5**

superior performance in accurately identifying these landmarks, highlighting its robustness against speckle noise and achieving precise landmark detection in challenging scenarios.

3.3 Ablation Studies

Ablation studies validate the contributions of key components (Table 2). The semi-supervised strategy improves performance by propagating pseudo labels to unlabeled frames. Multi-consistency constraints further enhance anatomical plausibility, reducing landmark errors. Optimal results emerge when synergistically combining these innovations, demonstrating complementary benefits between architectural design and physiological priors.

4 Conclusion

In this study, we propose a semi-supervised framework for echocardiogram segmentation and landmark detection. Different from the existing works, our framework has three key contributions. First, we use a teacher-student network with knowledge distillation to generate anatomically consistent pseudo labels from unlabeled frames, which effectively addresses the challenge of sparse annotation. Second, we introduce TSTNet, a novel network combining hierarchical attention mechanisms, to jointly model cardiac motion patterns and anatomical structures. Last, we incorporate temporal consistency regularization and inter-task consistency constraints to enhance learning robustness. Experimental results on two datasets containing 2,450 subjects show that our framework significantly outperforms existing methods in echocardiogram segmentation and landmark detection.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (grant numbers 82441023, U23A20295, 62131015, 82394432), the China Ministry of Science and Technology (S20240085, STI2030-Major Projects-2022ZD0209000, STI2030-Major Projects-2022ZD0213100), Shanghai Municipal Central Guided Local Science and Technology Development Fund (No. YDZX20233100001001), The Key R&D Program of Guangdong Province, China (grant number 2023B0303040001), and HPC Platform of ShanghaiTech University.

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

References

1. Boyd, A.C., Schiller, N.B., Thomas, L.: Principles of transthoracic echocardiographic evaluation. Nature Reviews Cardiology **12**(7), 426–440 (2015)
2. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision. pp. 205–218. Springer (2022)
3. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
4. Deng, X., Wu, H., Zeng, R., Qin, J.: Memsam: Taming segment anything model for echocardiography video segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9622–9631 (2024)
5. Koskinas, K.C., Van Craenenbroeck, E.M., Antoniades, C., Blüher, M., Gorter, T.M., Hanssen, H., Marx, N., McDonagh, T.A., Mingrone, G., Rosengren, A., et al.: Obesity and cardiovascular disease: an esc clinical consensus statement. European Heart Journal **45**(38), 4063–4098 (2024)
6. Lancellotti, P., Pellikka, P.A., Budts, W., Chaudhry, F.A., Donal, E., Dulgheru, R., Edvardsen, T., Garbi, M., Ha, J.W., Kane, G.C., et al.: The clinical use of stress echocardiography in non-ischaemic heart disease: recommendations from the european association of cardiovascular imaging and the american society of echocardiography. European Heart Journal–Cardiovascular Imaging **17**(11), 1191–1229 (2016)

7. Lancellotti, P., Price, S., Edvardsen, T., Cosyns, B., Neskovic, A.N., Dulgheru, R., Flachskampf, F.A., Hassager, C., Pasquet, A., Gargani, L., et al.: The use of echocardiography in acute cardiovascular care: recommendations of the european association of cardiovascular imaging and the acute cardiovascular care association. European Heart Journal-Cardiovascular Imaging **16**(2), 119–146 (2015)
8. Laumer, F., Di Vece, D., Cammann, V.L., Würdinger, M., Petkova, V., Schönberger, M., Schönberger, A., Mercier, J.C., Niederseer, D., Seifert, B., et al.: Assessment of artificial intelligence in echocardiography diagnostics in differentiating takotsubo syndrome from myocardial infarction. JAMA cardiology **7**(5), 494–503 (2022)
9. Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Jodoin, P.M., Grenier, T., et al.: Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE transactions on medical imaging **38**(9), 2198–2210 (2019)
10. Li, H., Wang, Y., Qu, M., Cao, P., Feng, C., Yang, J.: Echoefnet: multi-task deep learning network for automatic calculation of left ventricular ejection fraction in 2d echocardiography. Computers in Biology and Medicine **156**, 106705 (2023)
11. Li, H., Yang, J., Xuan, Z., Qu, M., Wang, Y., Feng, C.: A spatio-temporal graph convolutional network for ultrasound echocardiographic landmark detection. Medical Image Analysis **97**, 103272 (2024)
12. Lin, X., Xiang, Y., Zhang, L., Yang, X., Yan, Z., Yu, L.: Samus: Adapting segment anything model for clinically-friendly and generalizable ultrasound image segmentation. arXiv preprint arXiv:2309.06824 (2023)
13. Madani, A., Ong, J.R., Tibrewal, A., Mofrad, M.R.: Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. NPJ digital medicine **1**(1), 1–11 (2018)
14. Mokhtari, M., Mahdavi, M., Vaseli, H., Luong, C., Abolmaesumi, P., Tsang, T.S., Liao, R.: Echoglad: Hierarchical graph neural networks for left ventricle landmark detection on echocardiograms. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 227–237. Springer (2023)
15. Nedadur, R., Wang, B., Tsang, W.: Artificial intelligence for the echocardiographic assessment of valvular heart disease. Heart **108**(20), 1592–1599 (2022)
16. Ouzir, N., Basarab, A., Liebgott, H., Harbaoui, B., Tourneret, J.Y.: Motion estimation in echocardiography using sparse representation and dictionary learning. IEEE Transactions on Image Processing **27**(1), 64–77 (2017)
17. Rong, W., Li, Z., Zhang, W., Sun, L.: An improved canny edge detection algorithm. In: 2014 IEEE international conference on mechatronics and automation. pp. 577–582. IEEE (2014)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Porc. of the MICCAI. pp. 234–241. Springer (2015)
19. Saric, M., Armour, A.C., Arnaout, M.S., Chaudhry, F.A., Grimm, R.A., Kronzon, I., Landeck, B.F., Maganti, K., Michelena, H.I., Tolstrup, K.: Guidelines for the use of echocardiography in the evaluation of a cardiac source of embolism. Journal of the American Society of Echocardiography **29**(1), 1–42 (2016)
20. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems **30** (2017)
21. Wei, H., Cao, H., Cao, Y., Zhou, Y., Xue, W., Ni, D., Li, S.: Temporal-consistent segmentation of echocardiography with co-learning from appearance and shape. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020:

23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23. pp. 623–632. Springer (2020)

22. Zhang, Z., Yu, C., Zhang, H., Gao, Z.: Embedding tasks into the latent space: cross-space consistency for multi-dimensional analysis in echocardiography. IEEE Transactions on Medical Imaging (2024)