

# MTCNet: Motion and Topology Consistency Guided Learning for Mitral Valve Segmentation in 4D Ultrasound

Rusi Chen<sup>1\*</sup>, Yuanting Yang<sup>2\*</sup>, Jiezhai Yao<sup>1\*</sup>, Hongning Song<sup>2\*</sup>, Ji Zhang<sup>2</sup>, Yongsong Zhou<sup>1</sup>, Yuhao Huang<sup>1</sup>, Ronghao Yang<sup>1</sup>, Dan Jia<sup>2</sup>, Yuhao Zhang<sup>1</sup>, Xing Tao<sup>1</sup>, Haoran Dou<sup>3</sup>, Qing Zhou<sup>2(✉)</sup>, Xin Yang<sup>1(✉)</sup>, and Dong Ni<sup>1,4,5,6</sup>

<sup>1</sup>Medical Ultrasound Image Computing (MUSIC) Lab, School of Biomedical Engineering, Medical School, Shenzhen University, Shenzhen, China  
xinyang@szu.edu.cn

<sup>2</sup>Renmin Hospital of Wuhan University, Wuhan, China  
qingzhou.wh.edu@hotmail.com

<sup>3</sup>Centre for CIMIM, Manchester University, Manchester, UK

<sup>4</sup>School of Artificial Intelligence, Shenzhen University, Shenzhen, China

<sup>5</sup>National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen, China

<sup>6</sup>School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China

**Abstract.** Mitral regurgitation is one of the most prevalent cardiac disorders. Four-dimensional (4D) ultrasound has emerged as the primary imaging modality for assessing dynamic valvular morphology. However, 4D mitral valve (MV) analysis remains challenging due to limited phase annotations, severe motion artifacts, and poor imaging quality. Yet, the absence of inter-phase dependency in existing methods hinders 4D MV analysis. To bridge this gap, we propose a Motion-Topology guided consistency network (**MTCNet**) for accurate 4D MV ultrasound segmentation in semi-supervised learning (SSL). MTCNet requires only sparse end-diastolic and end-systolic annotations. First, we design a cross-phase motion-guided consistency learning strategy, utilizing a bi-directional attention memory bank to propagate spatio-temporal features. This enables MTCNet to achieve excellent performance both per- and inter-phase. Second, we devise a novel topology-guided correlation regularization that explores physical prior knowledge to maintain anatomically plausible. Therefore, MTCNet can effectively leverage structural correspondence between labeled and unlabeled phases. Extensive evaluations on the first largest 4D MV dataset, with 1408 phases from 160 patients, show that MTCNet performs superior cross-phase consistency compared to other advanced methods (Dice: 87.30%, HD: 1.75mm). Both the code and the dataset are available at <https://github.com/crs524/MTCNet>.

**Keywords:** Mitral Valve Segmentation · 4D Ultrasound · Semi-supervised Learning · Consistency Guided Learning.

\* Rusi Chen, Yuanting Yang, Jiezhai Yao and Hongning Song contributed equally. Qing Zhou and Xin Yang are the corresponding authors of this work.

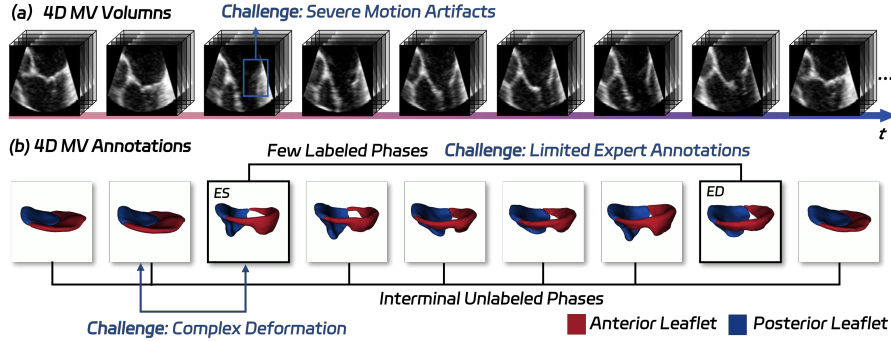


Fig. 1. Illustration of MV volumes and annotations in 4D ultrasound.

## 1 Introduction

Mitral regurgitation (MR) is a common cardiovascular disease with high morbidity and mortality [6,16]. Transesophageal Echocardiography (TEE) is the gold standard for diagnosing and quantifying MR. It offers a real-time view of the mitral valve (MV), providing both temporal and spatial perspectives [22]. Accurate 4D MV segmentation enables precise measurement of MV structure and functional analysis, as well as patient-specific 3D printing for surgical planning.

However, automatic MV segmentation in 4D ultrasound faces challenges caused by a few phase annotations, severe motion artifacts, and complex deformations (See Fig. 1). To tackle these challenges, we aim to develop a 4D MV segmentation method with only end-diastolic (ED) and end-systolic (ES) phase annotations, while ensuring high accuracy and temporal coherence.

Despite the promising progress in deep learning [26,13,9,8], research on 4D MV segmentation remains limited, with most studies focusing on single-volume segmentation. Some studies [3,1] employed U-Net [19] for 3D MV segmentation, while others [5,17] utilized the nnU-Net [10] framework. Although these methods achieve promising accuracy in static 3D volumes, they struggle with MV motion due to a lack of information extraction from unannotated phases. Semi-supervised learning (SSL) [20,27,28] has emerged as a promising paradigm to enhance model performance with limited labeled data by utilizing the abundant information in unlabeled data. Ivantsits et al. [11] simplify MV as an idea tubular sheet to reconstruct 4D MV surface, which limits its use in simulating the common organic mitral regurgitation disease with valve abnormalities. Munaf'o et al. [18] proposed an SSL method for 4D MV segmentation, where ED and ES phases of all patients are treated as labeled, while intermediate phases are considered unlabeled ones. However, they treat each phase independently, neglecting patient-level phase consistency, leading to suboptimal segmentation.

To address these challenges, we leverage motion and topology-guided consistency learning to account for the relationships between patient-level labeled and

unlabeled phases while preserving anatomical constraints. In this paper, we propose a motion-topology guided consistency method termed MTCNet to regard all phases of a patient as input samples to mine cross-phase relevance. Firstly, to structurally enhance the motion coherence, we introduce motion-guided consistency learning (MCL). Through a well-designed bi-directional memory bank (BMB), MCL can encourage MTCNet to effectively learn from key labeled phases and then propagate the semantic features to the unlabeled ones. Secondly, considering the anatomical prior of the MV’s surface-volume invariance, we embed this property into a topology consistency regularization (TCR), optimizing inter-phase dissimilarity through an extra surface and volume continuity constraint. Extensive evaluations on the largest 4D MV dataset demonstrate that MTCNet outperforms existing SSL approaches. Our contributions are three-fold:

- We introduce a novel motion and topology-guided learning framework for 4D MV segmentation, which achieves superior per-phase and inter-phase performance with sparse ES and ED annotations.
- We propose a MCL strategy to effectively propagate semantic information across labeled and unlabeled phases, thereby enhancing motion consistency.
- We design a simple yet effective TCR with surface and volume variance that boosts segmentation accuracy and preserves topological coherence.

## 2 Methods

**Problem Setting.** Given a patient-level sequence  $\mathbf{D}$  comprising a labeled subset  $\mathbf{D}_l^i = \{(X_l, Y, t)\}$  and an unlabeled subset  $\mathbf{D}_u^i = \{(X_u, t')\}$  for patient  $i$ , where  $t$  and  $t'$  represent the indices of the ES and ED phases, and the intermediate phases, respectively. Our goal is to segment all phases within a cycle using a limited number of labeled phases. Specifically,  $X_l \in \mathbb{R}^{D \times H \times W}$  and  $X_u \in \mathbb{R}^{D \times H \times W}$  denote labeled phases and unlabeled phases, respectively.  $Y \in \{0, 1\}^{D \times H \times W}$  represents the MV labels of  $X_l$ .

### 2.1 Patient-level Cross-phase Learning Framework

The overall framework of our method is illustrated in Fig. 2. Inspired by the Mean Teacher architecture [21], a powerful SSL method, MTCNet is designed to enhance learning from both labeled and unlabeled data. For each training iteration, MTCNet takes triplet phases  $\mathcal{T}$  as input from the sequence  $\mathbf{D}_l^i$ . Specifically, one of the triplet phases is a labeled phase, while the other two phases are unlabeled volumes from intermediate phases. Thus, the input can be defined as  $\mathcal{T} = \{X_l, X_u^1, X_u^2\}$ , where  $X_l \in \{X_{ES}, X_{ED}\}$ . Simultaneously,  $\mathcal{T}$  is fed into the student and teacher models, which share the same architecture.

During training, the teacher network supervises the student network by generating high-confidence pseudo-labels. The teacher network’s parameters  $\theta_t$  are updated via exponential moving average (EMA) [20] from the student’s parameters  $\theta_s$ . It can be formulated as  $\theta_t = \alpha\theta_t + (1 - \alpha)\theta_s$ , where  $\alpha \in (0, 1)$  is the

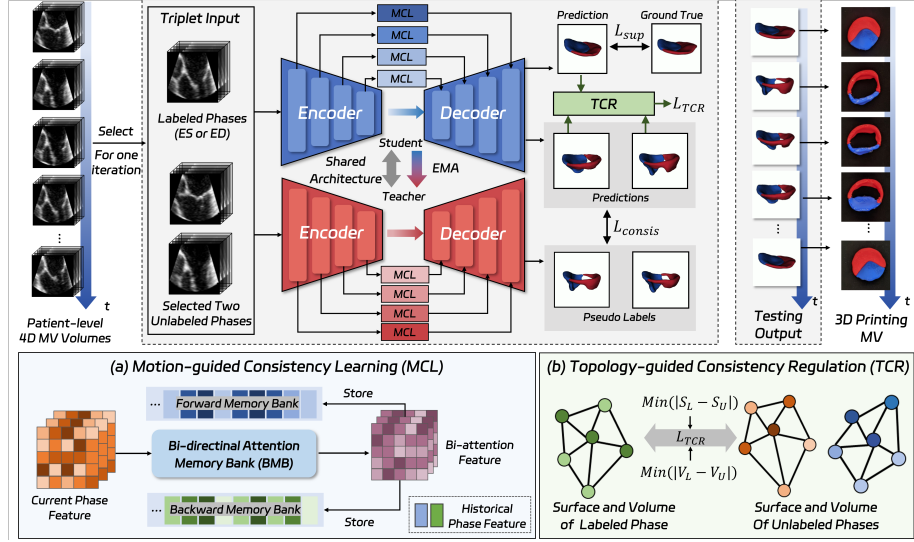


Fig. 2. Overall framework of our proposed MTCNet.

momentum mitigating the overfitting of the teacher network on limited labeled data. Ultimately, MTCNet generates the predicted segmentation masks for both the labeled and unlabeled volumes. The total training objective is:

$$\mathcal{L}_{seg} = \mathcal{L}_{sup}(f_s(\theta), Y) + \beta \cdot \mathcal{L}_{consis}(f_s(\theta), f_t(\theta)), \quad (1)$$

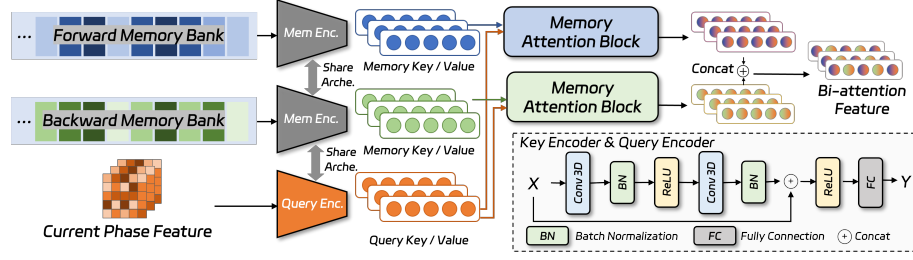
where  $\beta$  is the loss weight,  $L_{sup}$  and  $L_{consis}$  indicate the supervised loss and consistency loss, respectively.  $f_s(\theta)$  represents the segmentation model's output of labeled volume, and  $f_t(\theta)$  for unlabeled volumes. Both losses combine dice and binary cross-entropy loss with a 0.8 to 0.2 weight ratio. During the testing stage, MTCNet is able to predict masks for all phases in an end-to-end manner.

## 2.2 Motion-guided Consistency Learning

To effectively enhance the inter-phase motion coherence, we propose an MCL strategy through the BMB block. As shown in Fig. 2 (a), we propose forward memory bank  $M_f$  and backward memory bank  $M_b$  to store the temporal information for all phases of a patient.  $M_f$  captures systolic-diastolic deformation, while  $M_b$  encodes reverse motion, jointly modeling bidirectional patterns to reduce phase misalignment in cardiac cycles.

Fig. 3 illustrates the detailed design of the BMB block. Here, we take a memory bank to illustrate the mechanism, as both follow the same process. Given a memory bank  $M$  containing  $T$  memory phases, the memory encoder would generate memory key  $k^M \in \mathbb{R}^{C^k \times TDHW}$  and memory value  $v^M \in \mathbb{R}^{C^v \times TDHW}$ .





**Fig. 3.** Detailed design for bi-directional memory bank.

Similarly, the query encoder produces a query key  $k^Q \in \mathbb{R}^{C^k \times DHW}$  and a query value  $k^Q \in \mathbb{R}^{C^k \times DHW}$ , where  $D$ ,  $H$ , and  $W$  are the multi-scale feature dimensions. Current phase feature is the output feature of encoders both in the student and teacher models. Therefore, multi-scale memory aggregation minimizes information loss, especially for subtle boundary changes in low signal-to-noise ratio ultrasound phases. Then, we compute the normalized affinity matrix  $\mathbf{W}$  to effectively weigh the dependencies across phases:

$$\mathbf{W}_{ij} = \frac{\exp(c(\mathbf{k}_i^M, \mathbf{k}_j^Q))}{\sum_n \exp(c(\mathbf{k}_n^M, \mathbf{k}_j^Q))}, \quad (2)$$

where  $\mathbf{W} \in \mathbb{R}^{TDHW \times DHW}$ ,  $\mathbf{k}_i$  denotes the feature vector at the  $i$ -th position and  $c$  represents the dot product.

With the normalized affinity matrix  $\mathbf{W}$ , the aggregated readout feature  $v^Q \in \mathbb{R}^{C^v \times DHW}$  for the query phase is computed as a weighted sum of the memory features using a top-k operation. To be specific, the weighted sum of the top-k memory features is calculated as:  $v_r^Q = v_r^M \mathbf{W}_r$ , where  $r$  represents two forward and backward memory banks. Finally, the  $v_1^Q$  and  $v_2^Q$  are concatenated to obtain the bi-directional attention feature. This feature is stored in the memory banks  $M_f$  and  $M_b$  and passed to the decoder to produce the segmentation mask.

### 2.3 Topology-guided Consistency Regulation

To ensure structural integrity during complex deformations across phases, we propose the TCR mechanism. This approach is grounded in prior knowledge that surface area and volume should be stable throughout such deformations [15] (see Fig. 2 (b)). Specifically, for a predicted probability map  $P_t$  for  $t$ -th phase, we first obtain its binarized form  $B_t = \mathbb{I}(P_t > 0.5)$ , which represents the region of interest. Then, the normalized surface area  $S(P_t)$  of the  $t$ -th phase can be defined as:

$$S(P_t) = \frac{\int_S \nabla B_t(\mathbf{x}), d\mathcal{A}(\mathbf{x})}{\int_S \mathbb{I}(\nabla B_t(\mathbf{x}) > 0), d\mathcal{A}(\mathbf{x}) + \epsilon} \approx \frac{\sum_{v \in V} \nabla B_t(v), \Delta A(v)}{\sum_{v \in V} \mathbb{I}(\nabla B_t(v) > 0), \Delta A(v) + \epsilon}, \quad (3)$$

where  $\nabla = (\partial_x, \partial_y, \partial_z)$  denotes 3D Sobel operators so that  $\nabla B_t(\mathbf{x})$  represents the spatial gradient of the volumetric mask  $B_t$  at location  $\mathbf{x}$ .  $\Delta A(v)$  is the discrete voxel-level form and  $\epsilon$  prevents division by zero. Sobel operators ensure the computation is differentiable, enabling seamless integration into gradient-based optimization frameworks. Thus, the surface-area-based loss can be defined as:

$$\mathcal{L}_{\text{surf}} = \sum_{t=2}^3 \left( \left| 1 - \frac{S(P_t)}{S(P_1)} \right| + \lambda |S(P_t) - S(P_1)| \right). \quad (4)$$

The dual-term design overcomes key limitations of static shape constraints. The relative term adapts to patient-specific anatomy via the annotated  $S(P_1)$ , while the absolute term prevents error propagation through physical consistency. Follow a similar strategy to surface area consistency, the volume consistency loss can be formulated based on the voxel-wise summation of  $B_t$ :

$$\mathcal{L}_{\text{vol}} = \sum_{t=2}^3 \left( \left| 1 - \frac{V(P_t)}{V(P_1)} \right| + \lambda |V(P_t) - V(P_1)| \right), \quad (5)$$

where  $\sum_i B_t(i)$  is the voxel-wise summation of  $B_t$ . Finally, the total topological consistency regularization loss  $\mathcal{L}_{\text{tcp}}$  can be written as:  $\mathcal{L}_{\text{tcp}} = \mathcal{L}_{\text{surf}} + \mathcal{L}_{\text{vol}}$ . The  $\mathcal{L}_{\text{tcp}}$  acts as a physics-informed regularizer that enforces temporal plausibility. Combined with equation 1, the total loss for triplet input phases integrates both segmentation and physiological regularization  $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{seg}} + \sigma \cdot \mathcal{L}_{\text{tcp}}$ , where  $\sigma$  is a constant set to 0.1 by default based on empirical observations.

### 3 Experiments and Results

**Implementation Details.** For a fair comparison, all SSL experiments were conducted in the same setting, with only ES and ED annotated labels for training. Our model was implemented with PyTorch 1.11.0 on two NVIDIA GeForce RTX A40 GPUs. All input volumes were all resampled to  $128 \times 128 \times 128$  based on isotropic spacing resizing. During training, the Adam optimizer was used with an initial learning rate of  $10^{-4}$ , then reduced by 0.1 every 20 epochs.

**Datasets and Evaluation Metrics.** We collected 4D TEE MV data from 160 patients with 1408 phases from cooperating hospitals. The in-house dataset comprised 147 cases with functional mitral regurgitation, and 13 normal cases. The dataset was randomly split into training (112 cases), validation (16 cases), and testing (32 cases) subsets. The number of phases varied per case due to heart rate differences, with an average of nine phases per cycle. All imaging data were acquired using the Philips X5-1 transthoracic volume probe. Two sonographers with more than 10 years of experience manually annotated the MV volumes, consisting anterior leaflets (AL) and posterior leaflets (PL). We also validated on the mid-diastolic (MD) phase and its adjacent transitional phase (MD-1) between MD and ED. For evaluation, Dice coefficient, 95th-percentile Hausdorff Distance (HD) and Conformity (Conf) [4], were adopted for quantitative comparison.

**Table 1.** Comparison results for different methods. All Phases stands for total phases in a cardiac cycle. \* denotes statistically significant differences (t-test) between MTCNet and compared SOTA methods.

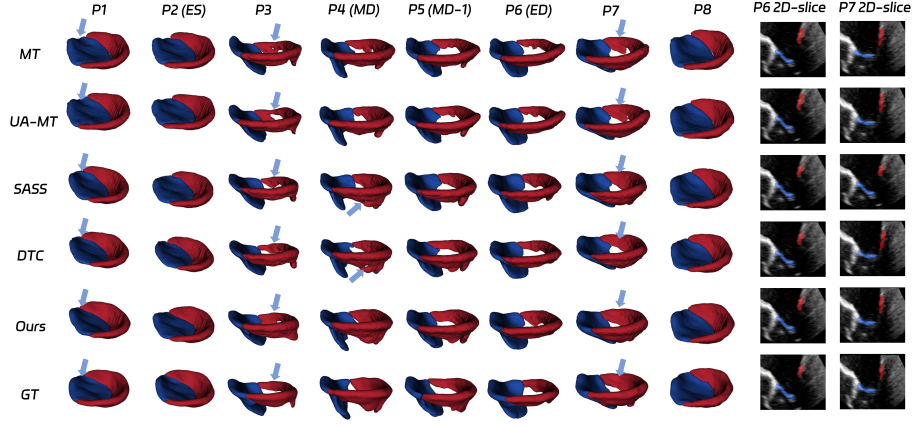
| Method        | ES           |             |              | MD           |             |              | MD-1         |             |              | ED           |             |              | All Phases   |             |              |
|---------------|--------------|-------------|--------------|--------------|-------------|--------------|--------------|-------------|--------------|--------------|-------------|--------------|--------------|-------------|--------------|
|               | Dice         | HD          | Conf         | Dice         | HD          | Conf         | Dice         | HD          | Conf         | Dice         | HD          | Conf         | Dice         | HD          | Conf         |
| MT[20]        | 86.13        | 1.74        | 66.96        | 79.25        | 2.94        | 32.62        | 85.63        | 1.89        | 64.42        | 87.50        | 1.41        | 70.58        | 85.61*       | 1.78        | 63.27*       |
| UA-MT[27]     | 86.44        | <b>1.68</b> | 67.85        | 79.76        | 2.75        | 37.50        | 86.22        | 1.85        | 66.53        | 87.85        | 1.35        | 71.62        | 86.08*       | <b>1.69</b> | 65.11*       |
| SASS[12]      | 85.90        | 1.85        | 66.45        | 79.67        | 2.74        | 35.21        | 86.64        | <b>1.69</b> | 68.03        | 87.76        | 1.45        | 71.29        | 86.07*       | 1.71        | 64.75*       |
| DTC[14]       | 84.71        | 2.07        | 61.04        | 79.14        | 3.15        | 25.81        | 86.26        | 1.99        | 66.39        | 87.50        | 1.54        | 70.10        | 85.31*       | 1.92*       | 60.69*       |
| ICT[23]       | 86.57        | 1.79        | 68.01        | 80.17        | 2.83        | 40.29        | 85.79        | 2.03        | 58.71        | 88.17        | 1.33        | 72.42        | 86.23*       | 1.77        | 64.81*       |
| MCF[24]       | 85.64        | 2.09        | 64.77        | 78.53        | 3.29        | 16.00        | 86.33        | 1.97        | 65.92        | 88.20        | 1.52        | 72.35        | 85.54*       | 2.06*       | 60.01*       |
| CC-Net[7]     | 85.23        | 1.88        | 63.38        | 78.63        | 2.89        | 33.67        | 86.14        | 1.75        | 66.53        | 87.17        | 1.39        | 69.61        | 85.16*       | 1.79        | 64.81*       |
| VM[2]         | -            | -           | -            | 73.99        | 4.03        | 6.65         | 83.89        | 2.37        | 57.24        | 87.79        | 1.69        | 70.35        | 83.01*       | 2.46*       | 42.33*       |
| T-VOS [25]    | -            | -           | -            | 80.29        | 2.81        | <b>44.54</b> | 83.97        | 2.66        | 60.63        | 85.15        | 2.05        | 63.98        | 84.56*       | 2.12*       | 61.36*       |
| <b>MTCNet</b> | <b>87.42</b> | 1.83        | <b>70.55</b> | <b>82.69</b> | <b>2.45</b> | 34.96        | <b>87.14</b> | 2.03        | <b>69.14</b> | <b>89.20</b> | <b>1.33</b> | <b>75.10</b> | <b>87.30</b> | 1.75        | <b>66.71</b> |

**Table 2.** Ablation study of MTCNet. M and T represent MCL and TCR, respectively. Four key phases denote the four key phases (ES, MD, MD-1, ED). \* denotes statistically significant differences (t-test) between best model and other configurations.

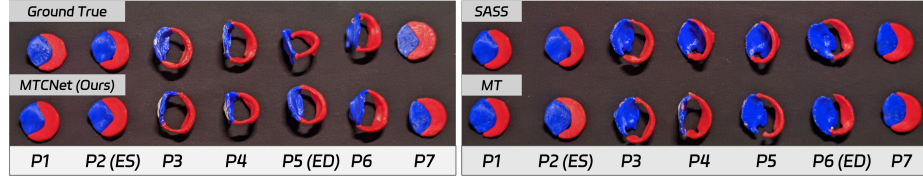
| Method    | Four Key Phases |             |              |              |             |              |              |             |              | All Phases   |             |              |              |             |              |              |             |              |
|-----------|-----------------|-------------|--------------|--------------|-------------|--------------|--------------|-------------|--------------|--------------|-------------|--------------|--------------|-------------|--------------|--------------|-------------|--------------|
|           | AL              |             |              | PL           |             |              | Mean         |             |              | AL           |             |              | PL           |             |              | Mean         |             |              |
|           | Dice            | HD          | Conf         | Dice         | HD          | Conf         | Dice         | HD          | Conf         | Dice         | HD          | Conf         | Dice         | HD          | Conf         | Dice         | HD          | Conf         |
| Based     | 85.81           | 1.93        | <b>62.16</b> | 83.21        | 2.29        | 55.55        | 84.51        | 2.11        | 58.85        | 86.95        | <b>1.74</b> | <b>66.98</b> | 84.08        | 1.98        | 59.53        | 85.51*       | 1.86        | 63.25*       |
| Based+M   | 86.90           | 1.92        | 61.63        | 85.15        | 2.23        | 60.42        | 86.02        | 2.08        | 61.03        | 87.75        | 1.78        | 66.71        | 85.57        | 2.00        | 62.12        | 86.66*       | 1.89        | 64.41*       |
| Based+M+T | <b>87.64</b>    | <b>1.90</b> | 60.35        | <b>85.58</b> | <b>1.92</b> | <b>64.52</b> | <b>86.61</b> | <b>1.91</b> | <b>62.43</b> | <b>88.41</b> | 1.75        | 66.82        | <b>86.19</b> | <b>1.76</b> | <b>66.61</b> | <b>87.30</b> | <b>1.75</b> | <b>66.71</b> |

**Comparison Study.** We evaluated MTCNet against multiple SSL methods [20,27,12,14,23,24,7], the registration-based approach VoxelMorph (VM) [2], and the two-shot video object segmentation method (T-VOS) [25]. VM and T-VOS both require one first phase as reference. As depicted in Table 1, MTCNet achieves state-of-the-art (SOTA) performance, surpassing compared methods in Dice (87.30%) and Conf (66.71%) in all phases segmentation. Specifically, our method surpasses the best baseline ICT (86.23%) by 1.07% ( $p < 0.05$ ). Notably, although MD is the most challenging phase due to the leaflets are maximally separated and often near the left ventricular wall, MTCNet surpasses the sub-optimal T-VOS method by 2.4% (82.69% vs. 80.29%) in Dice, highlighting MTCNet’s ability in handling complex deformation. Additionally, compared to reference-phase-driven methods VM and T-VOS, MTCNet shows more potential results in capturing continuous motion more effectively. Moreover, MTCNet achieves robust performance for both labeled phases (ES and ED) and unlabeled phases (MD and MD-1). This highlights MTCNet’s superior ability in per-phase and inter-phase performance.

**Ablation Study.** To demonstrate the impact of different components, we performed an ablation study in Table 2. Based-MTCNet means directly setting labeled and unlabeled data from the same patient based on Mean Teacher learning strategy. When integrating MCL, the mean Dice improves by 1.51% (84.51%  $\rightarrow$  86.02%) for all four key phases segmentation. Specifically, for AL and PL Dice, notably increasing +1.09% and +1.94%, respectively. This validates MCL’s ability to propagate motion cues across phases. TCR further refines



**Fig. 4.** Examples in consecutive phases with 3D volumes (Left seven columns) and 2D slices (Right two columns). Blue and red represent the AL and PL, respectively.



**Fig. 5.** Examples of 3D printing models of MV among different methods.

anatomical plausibility, reducing PL HD by 0.31 mm ( $2.23 \rightarrow 1.92$ ) and achieving a mean HD of 1.91 mm. Under full-phase evaluation, applying TCR improves PL Conf by 2.3% ( $64.41\% \rightarrow 66.71\%$ ), which significantly improved the consistency among different phases. The best model, Based+MCL+TCR, achieves a 1.79% improvement in Dice, a 0.11 mm reduction in HD and a 3.46% improvement in Conf compared to the baseline. The progressive improvements highlight their complementary roles: MCL captures temporal dependencies through BMB, while TCR enforces surface-volume continuity via physical regularization.

**Qualitative Results.** We further perform a qualitative comparison. Fig. 4 clearly shows that MTCNet produces smoother and more complete segmentations. Particularly in the MD phase where severe motion and deformation often cause holes (blue arrows). Artifacts in this region are most prominent due to the valve’s motion, indicating MTCNet’s superior ability to handle complex motion and structure changes. To further validate the clinical applicability of our approach, we compared the 3D printing results (See Fig. 5). While other methods produce incomplete shapes and missing details, our model achieves greater realism and completeness, showcasing its potential for patient-specific planning.

## 4 Conclusion

In this study, we present MTCNet, a SSL framework that leverages motion and topology-guided consistency learning for 4D MV segmentation. MTCNet performs better than the SOTA with sparse annotation, especially in motion and structure coherence. We contribute this gain in performance in two key contributions. First, the proposed MCL strategy can learn motion information across phases through BMB mechanism, improving segmentation performance in per-phase and inter-phase. Additionally, a robust TCR strategy with surface and volume variance as prior knowledge boosts segmentation accuracy while preserving topological coherence. MTCNet can also facilitate MV dynamics analysis and support personalized hemodynamic assessment and treatment planning.

**Acknowledgments.** This work was supported by the grant from National Natural Science Foundation of China (82201851, 82102046, 12326619, 62171290); Science and Technology Planning Project of Guangdong Province (2023A0505020002); Frontier Technology Development Program of Jiangsu Province (BF2024078); Guangxi Province Science Program (2024AB17023); Yunnan Major Science and Technology Special Project Program (202402AA310052); Yunnan Key Research and Development Program(202503AP140014).

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

## References

1. Aly, A.H., Khandelwal, P., Aly, A.H., Kawashima, T., Mori, K., Saito, Y., Hung, J., Gorman III, J.H., Pouch, A.M., Gorman, R.C., et al.: Fully automated 3d segmentation and diffeomorphic medial modeling of the left ventricle mitral valve complex in ischemic mitral regurgitation. *Medical Image Analysis* **80**, 102513 (2022)
2. Balakrishnan, G., Zhao, A., Sabuncu, M., Guttag, J., Dalca, A.V.: Voxelmorph: A learning framework for deformable medical image registration. *IEEE TMI: Transactions on Medical Imaging* **38**, 1788–1800 (2019)
3. Carnahan, P., Moore, J., Bainbridge, D., Eskandari, M., Chen, E.C., Peters, T.M.: Deepmitral: Fully automatic 3d echocardiography segmentation for patient specific mitral valve modelling. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V* 24. pp. 459–468. Springer (2021)
4. Chang, H.H., Zhuang, A.H., Valentino, D.J., Chu, W.C.: Performance measure characterization for evaluating neuroimage segmentation algorithms. *NeuroImage* **47**(1), 122–135 (2009)
5. Chen, J., Li, H., He, G., Yao, F., Lai, L., Yao, J., Xie, L.: Automatic 3d mitral valve leaflet segmentation and validation of quantitative measurement. *Biomedical Signal Processing and Control* **79**, 104166 (2023)
6. El Sabbagh, A., Reddy, Y.N., Nishimura, R.A.: Mitral valve regurgitation in the contemporary era: insights into diagnosis, management, and future directions. *JACC: Cardiovascular Imaging* **11**(4), 628–643 (2018)

7. Huang, H., Chen, Z., Chen, C., Lu, M., Zou, Y.: Complementary consistency semi-supervised learning for 3d left atrial image segmentation. *Computers in Biology and Medicine* **165**, 107368 (2023)
8. Huang, Y., Chang, A., Dou, H., Tao, X., Zhou, X., Cao, Y., Huang, R., Frangi, A.F., Bao, L., Yang, X., et al.: Flip learning: Weakly supervised erase to segment nodules in breast ultrasound. *Medical Image Analysis* **102**, 103552 (2025)
9. Huang, Y., Yang, X., Liu, L., Zhou, H., Chang, A., Zhou, X., Chen, R., Yu, J., Chen, J., Chen, C., et al.: Segment anything model for medical images? *Medical Image Analysis* **92**, 103061 (2024)
10. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods* **18**(2), 203–211 (2021)
11. Ivantsits, M., Tautz, L., Huellebrand, M., Walczak, L., Akansel, S., Khasyanova, I., Kempfert, J., Sündermann, S., Falk, V., Hennemuth, A.: Mv-gnn: Generation of continuous geometric representations of mitral valve motion from 3d+ t echocardiography. *Computers in Biology and Medicine* **182**, 109154 (2024)
12. Li, S., Zhang, C., He, X.: Shape-Aware Semi-supervised 3D Semantic Segmentation for Medical Images, p. 552–561. Springer International Publishing (2020)
13. Luo, G., Xu, M., Chen, H., Liang, X., Tao, X., Ni, D., Jeong, H., Kim, C., Stock, R., Baumgartner, M., et al.: Tumor detection, segmentation and classification challenge on automated 3d breast ultrasound: The tdsc-abus challenge. *arXiv preprint arXiv:2501.15588* (2025)
14. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 35, pp. 8801–8809 (2021)
15. May-Newman, K., Yin, F.: A constitutive law for mitral valve tissue (1998)
16. McDonagh, T.A., Metra, M., Adamo, M., Gardner, R.S., Baumbach, A., Böhm, M., Burri, H., Butler, J., Čelutkienė, J., Chioncel, O., et al.: 2021 esc guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the task force for the diagnosis and treatment of acute and chronic heart failure of the european society of cardiology (esc) with the special contribution of the heart failure association (hfa) of the esc. *European heart journal* **42**(36), 3599–3726 (2021)
17. Munafò, R., Saitta, S., Ingallina, G., Denti, P., Maisano, F., Agricola, E., Redaelli, A., Votta, E.: A deep learning-based fully automated pipeline for regurgitant mitral valve anatomy analysis from 3d echocardiography. *IEEE Access* (2024)
18. Munafò, R., Saitta, S., Tondi, D., Ingallina, G., Denti, P., Maisano, F., Votta, E., et al.: Automatic 4d mitral valve segmentation from transesophageal echocardiography: a semi-supervised learning approach. *Medical & Biological Engineering & Computing* pp. 1–16 (2025)
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III* 18. pp. 234–241. Springer (2015)
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. 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)

22. Vahanian, A., Beyersdorf, F., Praz, F., Milojevic, M., Baldus, S., Bauersachs, J., Capodanno, D., Conradi, L., De Bonis, M., De Paulis, R., et al.: 2021 esc/eacts guidelines for the management of valvular heart disease: developed by the task force for the management of valvular heart disease of the european society of cardiology (esc) and the european association for cardio-thoracic surgery (eacts). *European heart journal* **43**(7), 561–632 (2022)
23. Verma, V., Kawaguchi, K., Lamb, A., Kannala, J., Solin, A., Bengio, Y., Lopez-Paz, D.: Interpolation consistency training for semi-supervised learning. *Neural Networks* **145**, 90–106 (2022)
24. Wang, Y., Xiao, B., Bi, X., Li, W., Gao, X.: Mcf: Mutual correction framework for semi-supervised medical image segmentation. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15651–15660 (2023)
25. Yan, K., Li, X., Wei, F., Wang, J., Zhang, C., Wang, P., Lu, Y.: Two-shot video object segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2257–2267 (2023)
26. Yan, Z., Han, T., Huang, Y., Liu, L., Zhou, H., Chen, J., Shi, W., Cao, Y., Yang, X., Ni, D.: A foundation model for general moving object segmentation in medical images. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2024)
27. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: Medical image computing and computer assisted intervention–MICCAI 2019: 22nd international conference, Shenzhen, China, October 13–17, 2019, proceedings, part II 22. pp. 605–613. Springer (2019)
28. Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D.P., Chen, D.Z.: Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20. pp. 408–416. Springer (2017)