From CT to Surgical Video: Robust Mitral Valve Segmentation across Multimodal Clinical Imaging

Jie Xu*, Bowen Guo*, Wutong Li, Shangkun Li, Yi Guo, Yuanyuan Wang†, and Zeju Li†

Fudan University, Shanghai, China

Abstract. Mitral valve segmentation is clinically important for quantitative valve assessment, intervention planning, and computer-assisted mitral valve repair. This challenge addresses mitral valve segmentation across three heterogeneous imaging settings: semi-supervised 3D cardiac CT segmentation, fully supervised 3D transesophageal echocardiography (TEE) segmentation, and semi-supervised frame-level surgical video segmentation. In this work, we present a practical competition solution tailored to these three tracks. For the semi-supervised tracks, we adopt an offline pseudo-labeling strategy to provide stable supervision for unlabeled data. Supervised models are first trained on limited annotations and then ensembled to generate pseudo labels. To reduce the effect of noisy predictions, we further apply uncertainty-based sample selection and task-specific pseudo-label refinement before semi-supervised retraining. For the fully supervised 3D TEE track, we build a robust Auto3DSeg-based pipeline with five-fold cross-validation, ensemble inference, test-time augmentation, and connected-component post-processing. Across all tracks, our design focuses on maximizing supervision from limited annotations while effectively exploiting unlabeled data when available.

1 Introduction

Mitral valve disease is a common and clinically significant form of valvular heart disease, with mitral regurgitation and mitral stenosis potentially leading to cardiac dysfunction, heart failure, and poor outcomes, particularly in older adults [4]. Imaging is central to diagnosis, severity assessment, and treatment planning [16,23]; however, quantitative evaluation remains challenging because the mitral valve is a dynamic three-dimensional apparatus involving the annulus, leaflets, chordae tendineae, papillary muscles, and the left ventricle [12]. Accurate segmentation is therefore essential for reproducible analysis of valve anatomy and function, supporting disease characterization, intervention planning, intraoperative decision-making, and patient-specific modeling [3,17,19].

Mitral valve segmentation spans diverse imaging scenarios across the clinical workflow. Cardiac CT supports high-resolution anatomical modeling and transcatheter intervention planning [6,24], 3D TEE provides real-time volumetric

*Equal contribution. †Corresponding authors.

information for lesion assessment and procedural guidance [13,14], and surgical videos capture the operative field for anatomical recognition and computer-assisted mitral valve repair [9]. Together, these modalities provide complementary perspectives for mitral valve analysis and establish a challenging benchmark for robust segmentation across heterogeneous clinical scenarios.

Deep learning has substantially advanced automated medical image segmentation, with methods such as nnU-Net [8], MedNeXt [18], and SAM [10] providing effective tools for segmenting complex anatomical structures. Their performance often depends on large-scale, high-quality annotations, which are difficult to obtain for mitral valve segmentation. For the mitral valve, high-quality annotation is particularly demanding: its thin leaflets, complex motion, and substantial deformation make the boundaries difficult to delineate consistently, while cardiac CT, 3D TEE, and surgical videos each present distinct visual characteristics. These data constraints make mitral valve segmentation a natural setting for both fully supervised and semi-supervised learning. The former aims to make the most of limited expert annotations, while the latter further seeks reliable training signals from unlabeled data through pseudo-labeling [11], consistency-based regularization [22], or related strategies. Together, they reflect a common clinical reality: scarce annotations coexist with abundant unlabeled data.

In this work, we present a unified competition solution for mitral valve segmentation across three heterogeneous imaging settings: semi-supervised 3D cardiac CT, fully supervised 3D TEE, and semi-supervised surgical video segmentation. Our central idea is to make supervision more reliable under different annotation regimes. In the semi-supervised tracks, we avoid unstable online pseudo-label updates by first training supervised models and then generating offline pseudo labels through model ensembling. We further use uncertainty estimation to select reliable pseudo-labeled samples, allowing unlabeled data to contribute additional training signal with reduced noise. In the fully supervised track, we exploit the limited annotations through a strong self-configuring Auto3DSeg pipeline, five-fold training, ensemble inference, test-time augmentation, and connected-component refinement. Together, these components provide a practical and effective framework for robust mitral valve segmentation, achieving the top validation leaderboard performance at the time of submission.

2 Methods

2.1 Overview and Datasets

We used the datasets provided by the MVAA 2026 challenge, *Mitral Valve Anatomy Analysis Using Multimodal Imaging Data* [5]. The benchmark consists of three mitral valve segmentation tracks with different imaging modalities and annotation settings. Track 1 is a semi-supervised 3D cardiac CT segmentation task with 27 labeled cases and 1040 unlabeled cases. Track 2 is a fully supervised 3D transesophageal echocardiography (TEE) segmentation task with 105 labeled training cases. Track 3 is a semi-supervised frame-level surgical video segmentation task with 120 labeled frames and 1379 unlabeled frames.

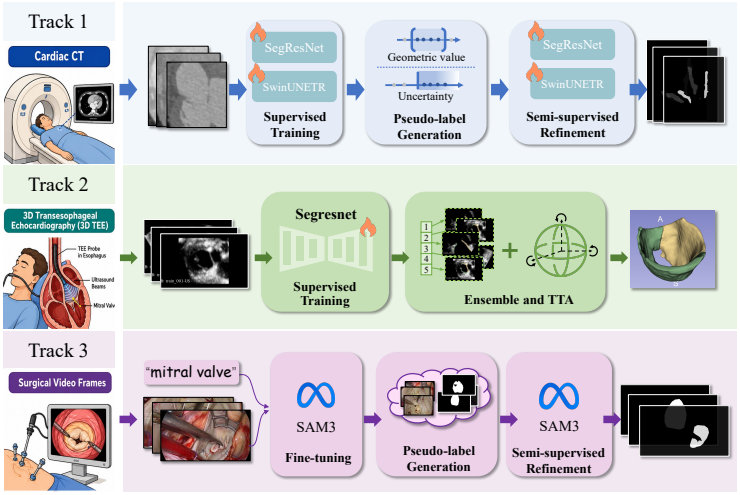

Fig. 1. Overview of our three-track mitral valve segmentation framework, covering semi-supervised 3D CT segmentation, fully supervised 3D TEE segmentation, and semi-supervised surgical video segmentation.

For the semi-supervised tracks, i.e., Track 1 and Track 3, we denote the track index by $t \in \{1, 3\}$. The labeled and unlabeled training sets are defined as

$$\mathcal{D}_t^l = \left\{ \left(\mathbf{x}_{t,i}^l, \mathbf{y}_{t,i}^l \right) \right\}_{i=1}^{N_t^l}, \quad \mathcal{D}_t^u = \left\{ \mathbf{x}_{t,i}^u \right\}_{i=1}^{N_t^u}, \quad t \in \{1, 3\}, \tag{1}$$

where $\left(N_1^l, N_1^u\right) = (27, 1040)$ and $\left(N_3^l, N_3^u\right) = (120, 1379)$. Here, $\mathbf{x}$ denotes an input volume or video frame, and $\mathbf{y}$ denotes the corresponding mask.

For Track 2, which follows a fully supervised setting, the labeled training set is defined as

$$\mathcal{D}_2^l = \left\{ \left(\mathbf{x}_{2,i}^l, \mathbf{y}_{2,i}^l \right) \right\}_{i=1}^{N_2^l}, \quad N_2^l = 105. \tag{2}$$

2.2 Track 1: Semi-supervised 3D Cardiac CT Segmentation

Framework and Supervised Training. For Track 1, we used MONAI Auto3DSeg [1], which automatically generates preprocessing pipelines, model configurations, and training settings for 3D medical image segmentation. We adopted two Auto3DSeg backbones, SegResNet [15] and SwinUNETR [21]. Both models were trained using five-fold cross-validation on the labeled data. SegResNet was optimized with a deep-supervision Dice-plus-cross-entropy loss:

$$\mathcal{L}_{\text{SegResNet}} = \sum_{s=0}^{3} \alpha_s \left[\mathcal{L}_{\text{Dice}} \left(\mathbf{p}^{(s)}, \mathbf{y}^{(s)} \right) + \mathcal{L}_{\text{CE}} \left(\mathbf{p}^{(s)}, \mathbf{y}^{(s)} \right) \right], \tag{3}$$

where $\mathbf{p}^{(s)}$ and $\mathbf{y}^{(s)}$ are the prediction and ground truth at the s-th output scale, and $\alpha_s = 2^{-s}$ for $s = 0, 1, 2, 3$. SwinUNETR was trained with the standard Dice-plus-cross-entropy loss on the final prediction. This stage produced 10 supervised models, including five SegResNet models and five SwinUNETR models.

Pseudo-label Generation. Before generating pseudo labels, we filtered unlabeled cases according to the geometric distribution of the 27 labeled cases. Let $\mathcal{G}$ denote the set of geometric attributes, including image size, voxel spacing, and physical extent, and let $g(\mathbf{x})$ be the value of attribute g for case $\mathbf{x}$. For each $g \in \mathcal{G}$, we computed the first quartile Q_1^g, the third quartile Q_3^g, and the interquartile range $\mathrm{IQR}^g = Q_3^g - Q_1^g$ from the labeled cases. An unlabeled case $\mathbf{x}_{1,i}^u$ was retained only if $g(\mathbf{x}_{1,i}^u) \in [Q_1^g - 3.0 \times \mathrm{IQR}^g, Q_3^g + 3.0 \times \mathrm{IQR}^g]$ for all $g \in \mathcal{G}$. This retained 906 out of 1040 unlabeled cases, denoted as $\widetilde{\mathcal{D}}_1^u = \{\widetilde{\mathbf{x}}_{1,i}^u\}_{i=1}^{906}$. For each retained case, we averaged the predictions from the 10 supervised models:

$$\bar{\mathbf{p}}_{1,i}^u = \frac{1}{10} \sum_{j=1}^{10} M_j \left(\widetilde{\mathbf{x}}_{1,i}^u \right). \tag{4}$$

The initial binary pseudo label was obtained by thresholding at 0.5, followed by 26-connected component refinement: $\hat{\mathbf{y}}_{1,i}^u = \mathrm{LCC}_{26}\left(\mathbf{1}(\bar{\mathbf{p}}_{1,i}^u \geq 0.5)\right)$, where $\mathbf{1}(\cdot)$ is the indicator function and LCC_{26} keeps the largest 3D connected component.

We estimated pseudo-label uncertainty using the entropy of the ensemble probability map. Let $h(p) = -p \log p - (1-p)\log(1-p)$ denote the binary entropy. The sample-level uncertainty was computed as

$$u_{1,i} = \frac{1}{|\mathcal{R}_{1,i}^u|} \sum_{v \in \mathcal{R}_{1,i}^u} h\left(\bar{\mathbf{p}}_{1,i}^u(v)\right), \tag{5}$$

where $\mathcal{R}_{1,i}^u$ denotes the local region around the predicted target. We ranked all pseudo-labeled cases by $u_{1,i}$ and selected the lowest-uncertainty 50%, yielding 453 pseudo-labeled cases $\mathcal{D}_1^p = \{(\mathbf{x}_{1,i}^p, \hat{\mathbf{y}}_{1,i}^p)\}_{i=1}^{453}$.

Semi-supervised Training and Inference. The selected pseudo-labeled cases were combined with the labeled data, forming $\mathcal{D}_1^{\mathrm{ssl}} = \mathcal{D}_1^l \cup \mathcal{D}_1^p$. We used the same five-fold split as in supervised training, where the split was determined only by the labeled cases and all pseudo-labeled cases were added to each training fold. Since hard pseudo labels were used, the same supervised objectives were applied to both labeled and pseudo-labeled data. For final submission, we ensembled the 10 semi-supervised models, thresholded the averaged probability map with a validation-selected foreground threshold τ_1, and retained the largest 26-connected component as the final prediction.

2.3 Track 2: Fully Supervised 3D TEE Segmentation

For Track 2, we again used MONAI Auto3DSeg [1] and selected the SegResNet-based configuration [15]. Since Auto3DSeg does not provide a dedicated 3D

ultrasound preprocessing template, we used the MRI-style volumetric preprocessing preset, including foreground cropping, spacing resampling, and intensity normalization. The model takes a single-channel 3D TEE volume as input and predicts three classes, including the background.

The model was trained under a fully supervised five-fold cross-validation setting using the same deep-supervision Dice-plus-cross-entropy loss as Eq. 3. For inference, we combined five-fold probability ensembling with 8-view TTA, covering all flip combinations along the three spatial axes. The averaged probability map was converted to a preliminary segmentation by voxel-wise arg max label assignment. Finally, for each foreground class, only the largest 26-connected component was retained to suppress isolated false positives.

2.4 Track 3: Semi-supervised Surgical Video Segmentation

Framework and Supervised Fine-tuning. For Track 3, we adopted SAM3 [2] as the base model due to its concept-conditioned segmentation ability. Given an image frame and the text prompt "mitral valve", SAM3 can directly produce the corresponding mask, avoiding manual point or box prompts during inference. To efficiently adapt SAM3 to the surgical video domain, we fine-tuned it with LoRA [7] while keeping backbone parameters frozen. We trained SAM3 using five-fold cross-validation on labeled frames, yielding five supervised SAM3 models. Following the official setting, each model was optimized with a weighted combination of concept recognition, localization, and mask supervision losses:

$$\begin{aligned}\mathcal{L}_{\text{SAM3}} = &\sum_{s\in\mathcal{S}}\Big[20\mathcal{L}_{\text{cls}}^{(s)} + 20\mathcal{L}_{\text{pres}}^{(s)} + 5\mathcal{L}_{\text{box}}^{(s)} + 2\mathcal{L}_{\text{GIoU}}^{(s)}\Big] \\ &+ 200\mathcal{L}_{\text{focal}}^{\text{main}} + 10\mathcal{L}_{\text{Dice}}^{\text{main}},\end{aligned} \tag{6}$$

where $\mathcal{S}$ denotes the set of decoder outputs, and "main" denotes the final mask.

Pseudo-label Generation. The five supervised SAM3 models were used to generate pseudo labels for 1379 unlabeled frames. During inference, we applied four-view TTA, including the original image, horizontal flip, vertical flip, and horizontal-vertical flip, and averaged predictions at the probability level after mapping them back to the original image space. As in Track 1, we estimated sample-level uncertainty using the entropy of the ensemble probability map and retained the lowest-uncertainty 50%, keeping 690 frames and discarding 689 highly uncertain frames. For the retained frames, we used a FixMatch-style high-confidence pixel selection strategy [20]. Pixels with probabilities larger than 0.95 were assigned to the foreground, pixels with probabilities smaller than 0.05 were assigned to the background, and the remaining ambiguous pixels were ignored:

$$\hat{\mathbf{y}}_{3,i}^{u}(v) = \begin{cases} 1, & \bar{\mathbf{p}}_{3,i}^{u}(v) > 0.95, \\ 0, & \bar{\mathbf{p}}_{3,i}^{u}(v) < 0.05, \\ \text{ignore}, & \text{otherwise}. \end{cases} \tag{7}$$

Frames without valid high-confidence pixels were discarded, resulting in 580 pseudo-labeled frames $\mathcal{D}_3^p = \{(\mathbf{x}_{3,i}^p, \hat{\mathbf{y}}_{3,i}^p)\}_{i=1}^{580}$.

Semi-supervised Training and Inference. The pseudo-labeled frames were combined with the labeled set under the same five-fold split, forming $\mathcal{D}_3^{\text{ssl}} = \mathcal{D}_3^l \cup \mathcal{D}_3^p$. The semi-supervised objective was

$$\mathcal{L}_{\text{SSL}} = \mathcal{L}_{\text{SAM3}}(\mathcal{D}_3^l) + \lambda \mathcal{L}_{\text{SAM3}}(\mathcal{D}_3^p), \tag{8}$$

where losses on pseudo-labeled frames were computed only over valid foreground/background pixels and λ was set to 0.5. For final inference, we ensembled the five semi-supervised SAM3 models with same four-view TTA strategy and applied a validation-selected foreground threshold τ_3 to obtain the final masks.

3 Experiments and Results

3.1 Implementation Details

All experiments were implemented in PyTorch and trained on NVIDIA RTX 4090D GPUs with 48GB memory. AdamW was used for optimization, and automatic mixed precision was enabled when applicable.

For Track 1, SegResNetDS and SwinUNETR were trained with five-fold cross-validation. SegResNetDS followed the Auto3DSeg-generated configuration with a patch size of $160 \times 128 \times 96$, batch size 3, 1250 epochs, learning rate 2×10^{-4}, and weight decay 1×10^{-5}. SwinUNETR used pretrained weights, a patch size of $96 \times 96 \times 96$, feature size 48, up to 1000 epochs, learning rate 4×10^{-4}, and weight decay 1×10^{-5}. Both models used RAS reorientation, foreground cropping, CT intensity normalization or clipping, and standard 3D spatial/intensity augmentations; SwinUNETR additionally resampled volumes to $0.3574 \times 0.3574 \times 0.5$ mm.

For Track 2, SegResNetDS was trained with five-fold cross-validation using the Auto3DSeg MRI-style volumetric preprocessing preset. The patch size was $224 \times 160 \times 96$, batch size was 1, and training lasted 1500 epochs with learning rate 2×10^{-4}, weight decay 1×10^{-5}, and a 3-epoch warmup strategy. The same family of 3D augmentations as Track 1 was used.

For Track 3, frames were resized to 1008×1008. SAM3 was fine-tuned with LoRA using rank $r = 16$ and scaling factor $\alpha = 32$. Models were trained for 100 epochs with batch size 4 and learning rate 2×10^{-4}. Semi-supervised training used mixed batches of labeled and pseudo-labeled frames. Augmentations included flipping, rotation within $\pm 10°$, and brightness/contrast perturbation.

All three tracks were evaluated using Dice similarity coefficient (Dice), Hausdorff distance (HD), and average surface distance (ASD).

3.2 Results

Track 1 Table 1 reports the validation performance of different Track 1 settings. The SegResNet–SwinUNETR ensemble outperformed both single backbones, improving DSC from 0.8586 and 0.8506 to 0.8616. Semi-supervised training further improved the combined ensemble to 0.8620 at the default threshold

Table 1. Performance of Track 1 under different model and training settings. All methods use 5-fold ensemble inference unless otherwise specified and apply largest connected component post-processing.

Method	τ	**DSC** ↑	**HD** ↓	**ASD** ↓
SegResNet	0.50	0.8586	4.70	0.2722
SwinUNETR	0.50	0.8506	4.66	0.2821
SegResNet + SwinUNETR	0.50	0.8616	4.49	**0.2666**
Semi-supervised SegResNet	0.50	0.8607	4.56	0.2731
Semi-supervised SwinUNETR	0.50	0.8607	4.41	0.2695
Semi-supervised SegResNet + SwinUNETR	0.50	0.8620	4.51	0.2706
	0.45	0.8631	4.45	0.2677
	0.40	0.8632	4.44	0.2674
Semi-supervised SegResNet + SwinUNETR	0.35	0.8639	4.41	0.2673
	0.30	**0.8642**	4.36	0.2674
	0.25	0.8636	**4.31**	0.2700

Table 2. Performance of Track 2 under different model and inference settings.

Category	**Setting**	**DSC** ↑	**HD** ↓	**ASD** ↓
	Fold 0	0.8488	11.16	0.6275
	Fold 1	0.8513	11.29	0.5940
Single fold	Fold 2	0.8481	11.44	0.6292
	Fold 3	0.8559	12.30	0.5949
	Fold 4	0.8537	11.44	0.5939
	5-fold ensemble	0.8612	10.57	0.5618
Inference Method	5-fold ensemble + TTA	0.8624	10.40	0.5507
	5-fold ensemble + TTA + LCC	**0.8635**	**9.70**	**0.5285**

Table 3. Performance comparison of SAM3-based methods under different training settings and threshold values. All results are obtained with 5-fold ensemble inference and four-view test-time augmentation, including the original view, horizontal flip, vertical flip, and horizontal–vertical flip. Largest connected component post-processing is applied to all methods.

Method	τ	**Dice** ↑	**HD** ↓	**ASD** ↓
SAM3	0.50	0.8119	74.55	12.37
Semi-supervised SAM3	0.50	0.8189	71.30	11.72
	0.45	0.8233	67.68	10.98
	0.40	0.8270	65.68	10.43
Semi-supervised SAM3	0.35	0.8302	64.74	10.19
	0.30	**0.8317**	63.57	**10.14**
	0.25	0.8303	**63.35**	10.36

of 0.50. Threshold tuning brought additional gains, with the best DSC of 0.8642 at $\tau = 0.30$ and the best HD of 4.31 at $\tau = 0.25$. These results show that model ensembling, offline pseudo-labeling, uncertainty-based filtering, and threshold tuning were all beneficial.

Track 2 Table 2 summarizes the Track 2 validation results. Single-fold models achieved DSC scores between 0.8481 and 0.8559, while five-fold ensembling improved DSC to 0.8612. Adding TTA and largest connected component refinement further improved the final performance to 0.8635 DSC, 9.70 HD, and 0.5285 ASD. This indicates that the inference pipeline effectively improved robustness and reduced isolated false positives.

Track 3 Table 3 shows the SAM3-based results for surgical video segmentation. Semi-supervised training improved the default-threshold DSC from 0.8119 to 0.8189, confirming the value of selected pseudo-labeled frames. Threshold tuning further increased DSC to 0.8317 at $\tau = 0.30$, while the best HD was obtained at $\tau = 0.25$. Compared with the supervised baseline, the final setting improved DSC by 1.98 percentage points and reduced HD from 74.55 to 63.57 and ASD from 12.37 to 10.14, demonstrating the effectiveness of LoRA adaptation, ensemble pseudo-labeling, uncertainty filtering, and high-confidence pixel selection.

4 Conclusion

In this work, we presented a practical solution for the MVAA 2026 challenge across semi-supervised 3D cardiac CT segmentation, fully supervised 3D TEE segmentation, and semi-supervised surgical video segmentation. For the semi-supervised tracks, we used offline pseudo-labeling with model ensembling, uncertainty-based filtering, and task-specific refinement to exploit unlabeled data more reliably. For the fully supervised TEE track, we built an Auto3DSeg-based pipeline with five-fold training, ensemble inference, TTA, and connected-component refinement. The validation results show consistent gains from these designs, demonstrating the effectiveness of reliable pseudo-labeling and robust inference for multimodal mitral valve segmentation under limited annotations.

Acknowledgments. This work was supported by the National Key R&D Program of China (2024YFF0507300, 2024YFF0507303), and the National Natural Science Foundation of China (Grant No. 62531004).

References

1. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: MONAI: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
2. Carion, N., Gustafson, L., Hu, Y.T., Debnath, S., Hu, R., Suris, D., Ryali, C., Alwala, K.V., Khedr, H., Huang, A., et al.: Sam 3: Segment anything with concepts. arXiv preprint arXiv:2511.16719 (2025)
3. Corinzia, L., Laumer, F., Candreva, A., Taramasso, M., Maisano, F., Buhmann, J.M.: Neural collaborative filtering for unsupervised mitral valve segmentation in echocardiography. Artificial intelligence in medicine **110**, 101975 (2020)
4. Figlioli, G., Sticchi, A., Christodoulou, M.N., Hadjidemetriou, A., Amorim Moreira Alves, G., De Carlo, M., Praz, F., Caterina, R.D., Nikolopoulos, G.K., Bonovas, S., et al.: Global prevalence of mitral regurgitation: a systematic review and meta-analysis of population-based studies. Journal of Clinical Medicine **14**(8), 2749 (2025)
5. Hammad, M., Bai, J., Zhao, J., Jia, T., Zhang, X., Xu, X., Ma, J., Li, S.: Mitral valve anatomy analysis using multimodal imaging data (2026). https://doi.org/10.5281/zenodo.19726755, https://doi.org/10.5281/zenodo.19726755
6. Hashimoto, G., Lopes, B.B., Sato, H., Fukui, M., Garcia, S., Gössl, M., Enriquez-Sarano, M., Sorajja, P., Bapat, V.N., Lesser, J., et al.: Computed tomography planning for transcatheter mitral valve replacement. Structural Heart **6**(1), 100012 (2022)
7. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. Iclr **1**(2), 3 (2022)
8. 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)
9. Ivantsits, M., Tautz, L., Sündermann, S., Wamala, I., Kempfert, J., Kuehne, T., Falk, V., Hennemuth, A.: Dl-based segmentation of endoscopic scenes for mitral valve repair. In: Current Directions in Biomedical Engineering. vol. 6, p. 20200017. De Gruyter (2020)
10. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026 (2023)
11. Lee, D.H., et al.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, ICML. vol. 3, p. 896. Atlanta (2013)
12. McCarthy, K.P., Ring, L., Rana, B.S.: Anatomy of the mitral valve: understanding the mitral valve complex in mitral regurgitation. European Journal of echocardiography **11**(10), i3–i9 (2010)
13. Munafo, 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 **12**, 5295–5308 (2024)
14. Munafò, R., Saitta, S., Tondi, D., Ingallina, G., Denti, P., Maisano, F., Agricola, E., Votta, E.: Automatic 4d mitral valve segmentation from transesophageal echocardiography: a semi-supervised learning approach. Medical & Biological Engineering & Computing pp. 1–16 (2025)

15. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization. pp. 311–320. Springer International Publishing, Cham (2019)
16. Otto, C.M., Nishimura, R.A., Bonow, R.O., Carabello, B.A., Erwin III, J.P., Gentile, F., Jneid, H., Krieger, E.V., Mack, M., McLeod, C., et al.: 2020 acc/aha guideline for the management of patients with valvular heart disease: a report of the american college of cardiology/american heart association joint committee on clinical practice guidelines. Circulation **143**(5), e72–e227 (2021)
17. Pouch, A.M., Wang, H., Takabe, M., Jackson, B.M., Gorman III, J.H., Gorman, R.C., Yushkevich, P.A., Sehgal, C.M.: Fully automatic segmentation of the mitral leaflets in 3d transesophageal echocardiographic images using multi-atlas joint label fusion and deformable medial modeling. Medical image analysis **18**(1), 118–129 (2014)
18. Roy, S., Koehler, G., Ulrich, C., Baumgartner, M., Petersen, J., Isensee, F., Jaeger, P.F., Maier-Hein, K.H.: Mednext: transformer-driven scaling of convnets for medical image segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 405–415. Springer (2023)
19. Schneider, R.J., Perrin, D.P., Vasilyev, N.V., Marx, G.R., Del Nido, P.J., Howe, R.D.: Mitral annulus segmentation from 3d ultrasound using graph cuts. IEEE transactions on medical imaging **29**(9), 1676–1687 (2010)
20. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems **33**, 596–608 (2020)
21. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-supervised pre-training of swin transformers for 3d medical image analysis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20730–20740 (2022)
22. 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)
23. 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)
24. Weir-McCall, J.R., Blanke, P., Naoum, C., Delgado, V., Bax, J.J., Leipsic, J.: Mitral valve imaging with ct: relationship with transcatheter mitral valve interventions. Radiology **288**(3), 638–655 (2018)