MVAA-Seg: A Multi-Stage Segmentation Framework for Mitral Valve Anatomy Analysis

Liyang Zhang[1], Qing Xu[1,2†], Xiangjian He[1(✉)], and Zhen Chen[3(✉)]

[1] School of Computer Science, University of Nottingham Ningbo China, China
[2] School of Computer Science, University of Nottingham, UK
[3] Department of Data Science and Artificial Intelligence, The Hong Kong Polytechnic University, Hong Kong SAR
scxqx1@nottingham.ac.uk, sean.he@nottingham.edu.cn
z.chen@polyu.edu.hk

Abstract. Mitral valve anatomy analysis is essential for cardiac assessment and surgical planning, yet robust segmentation remains challenging because annotations are scarce and clinical modalities are heterogeneous. Existing methods are often designed for one modality and use uniform pseudo-labeling rules without anatomical priors. We propose MVAA-Seg, a multi-stage framework for the MICCAI 2026 Mitral Valve Anatomy Analysis (MVAA) Challenge, covering computed tomography (CT), three-dimensional transesophageal echocardiography (3D TEE), and surgical video. For CT, a confidence-guided semi-supervised nnUNetv2 pipeline ranks pseudo labels using prediction confidence and anatomical plausibility, exploiting unlabeled data while rejecting unreliable masks. For TEE, we use a 3D nnUNetv2 ResEnc M pipeline trained for 500 epochs on five folds and ensembled at inference. For video, we use UNet++ with an EfficientNet-B5 encoder. On the MVAA validation set, MVAA-Seg **(Team: NPM)** improves over the official baseline in all tasks, with DSC of 0.8585, 0.8543, and 0.8105 on CT, TEE, and video, respectively.

Keywords: Mitral valve anatomy analysis · medical image segmentation · semi-supervised learning

1 Introduction

Mitral valve disease is a prevalent valvular heart condition, and precise anatomical analysis is essential for quantitative assessment, surgical planning, and image-guided intervention. Robust segmentation remains difficult across heterogeneous modalities: CT provides high-resolution 3D anatomy but has limited annotations; 3D TEE suffers from speckle noise and weak boundaries; and surgical video shows large variations in illumination, viewpoint, and instrument occlusion. These differences demand tailored strategies.

Encoder–decoder architectures such as U-Net [20] and 3D U-Net [6] established the foundation for medical image segmentation, while nnU-Net [8, 9] showed

† Team Leader.

that systematic pipeline design can rival architectural novelty. For frame-wise segmentation, UNet++ [27] improves multi-scale fusion, and EfficientNet [23] provides scalable visual encoding. Semi-supervised pseudo-labeling, Mean Teacher, and confidence-based consistency learning [12, 24, 22] help exploit unlabeled data while limiting noisy supervision. Recent ultrasound studies further emphasize benchmarks [1, 3], boundary-aware modeling [5, 4], lightweight inference [19], and multi-architecture consistency [10], motivating our modality-specific design.

Despite these advancements, existing methods [16, 25, 26, 11] are typically designed and evaluated within a single imaging modality, which overlooks the clinical reality that mitral valve analysis requires integrated understanding across volumetric and frame-wise data. Moreover, standard pseudo-labeling strategies apply uniform confidence thresholds without considering anatomical plausibility, leading to noisy supervision in volumetric cardiac data. The MICCAI 2026 Mitral Valve Anatomy Analysis (MVAA) Challenge provides a comprehensive benchmark spanning CT, 3D TEE, and surgical video modalities, requiring robust segmentation across these heterogeneous settings. This motivates us to develop a modality-specific framework that pairs each imaging modality with its most suitable segmentation strategy and pseudo-label selection mechanism.

To this end, we propose MVAA-Seg, a multi-stage segmentation framework with three specialized branches. For CT segmentation, we design a confidence-guided semi-supervised nnUNetv2 pipeline that ranks pseudo labels using foreground confidence, low-percentile confidence, connected-component consistency, and anatomical volume plausibility, effectively exploiting the large unlabeled CT pool while filtering unreliable supervision. For TEE segmentation, we adopt the official nnUNetv2 ResEnc M preset [8, 9], train five folds for 500 epochs each, and ensemble their predictions to provide stable and robust volumetric segmentation. For video segmentation, we employ a UNet++ [27] model with an EfficientNet-B5 [23] encoder, optimized with a Dice–Focal objective and enhanced by conservative pseudo-label refinement and connected-component post-processing. Evaluated on the MVAA Challenge validation set, MVAA-Seg achieves substantial improvements over the official baseline across all three tasks, improving DSC from 0.6184 to 0.8585 on Task 1, from 0.7451 to 0.8543 on Task 2, and from 0.6440 to 0.8105 on Task 3.

2 Methodology

As illustrated in Figs. 1–3, MVAA-Seg uses three modality-specific branches: confidence-guided semi-supervised nnUNetv2 for CT, five-fold ResEnc M 3D nnUNetv2 for TEE, and UNet++ with an EfficientNet-B5 encoder for video.

2.1 Task 1: Confidence-Guided Semi-Supervised CT Segmentation

The CT segmentation task provides only 27 labeled scans, together with a substantially larger unlabeled pool. To exploit the unlabeled data while controlling pseudo-label noise, we design a two-stage nnUNetv2-based semi-supervised

pipeline. Let $\mathcal{D}_l = \{(x_i, y_i)\}_{i=1}^{N_l}$ denote the labeled CT set and $\mathcal{D}_u = \{x_j\}_{j=1}^{N_u}$ denote the unlabeled CT pool. We first train a fully supervised plain 3D nnUNetv2 model on $\mathcal{D}_l$ using five-fold cross-validation. The resulting five-fold ensemble is used as a seed model f_θ to predict all unlabeled CT volumes and export both hard pseudo masks and foreground probability maps. This fully supervised stage already provides a strong CT model, but its predictions on unlabeled scans are further filtered before being used as supervision.

Given an input volume x, the seed ensemble produces a probability map $p = f_\theta(x)$ and a hard segmentation mask $\hat{y} = \arg\max_c p_c$. For each pseudo mask, we compute foreground confidence and anatomical plausibility statistics. For the predicted foreground region $\Omega_f = \{v|\hat{y}(v) > 0\}$, the mean foreground confidence is defined as:

$$s_{\text{mean}} = \frac{1}{|\Omega_f|} \sum_{v \in \Omega_f} p_{\hat{y}(v)}(v). \tag{1}$$

To avoid selecting masks that are confident only in central regions but uncertain near boundaries, we also compute low-percentile foreground confidence:

$$s_{10} = P_{10}\left(\{p_{\hat{y}(v)}(v) \mid v \in \Omega_f\}\right), \quad s_{50} = P_{50}\left(\{p_{\hat{y}(v)}(v) \mid v \in \Omega_f\}\right), \tag{2}$$

where P_{10} and P_{50} denote the 10th and 50th percentiles. Unlike standard confidence-thresholded approaches [22] that mainly rely on prediction confidence, our criterion additionally incorporates anatomical volume and connected-component statistics.

Let $V(\hat{y})$ be the physical foreground volume of the pseudo mask and V_{med} be the median foreground volume estimated from the labeled masks. We define a volume deviation term:

$$s_{\text{vol}} = \left|\log \frac{V(\hat{y}) + \epsilon}{V_{\text{med}} + \epsilon}\right|. \tag{3}$$

We further compute the largest connected-component ratio r_{lcc}, the number of foreground connected components n_{cc}, and the mean foreground entropy h_{fg}. The final confidence ranking score is:

$$S(\hat{y}) = s_{\text{mean}} + 0.50 s_{10} + 0.25 s_{50} + 0.10 r_{\text{lcc}} - 0.15 s_{\text{vol}} - 0.05 \min(n_{\text{cc}} - 1, 5) - 0.10 h_{\text{fg}}. \tag{4}$$

Cases with empty foreground predictions are assigned the lowest priority. We then sort all unlabeled cases according to $S(\hat{y})$ and select the top 627 pseudo labels, corresponding to the high-confidence subset before the ranking score begins to drop noticeably.

The final semi-supervised CT dataset therefore contains the 27 real labeled CT volumes and 627 confidence-ranked pseudo-labeled volumes. We train an nnUNetv2 ResEnc M model on this combined set using five-fold cross-validation for 500 epochs per fold. To avoid optimistic model selection, pseudo-labeled cases are used only for training, while each validation split contains real labeled cases only. During inference, predictions from all five ResEnc M folds are ensemble with the standard nnUNetv2 sliding-window strategy. This confidence-guided semi-supervised ResEnc M pipeline improves Task 1 from the fully supervised

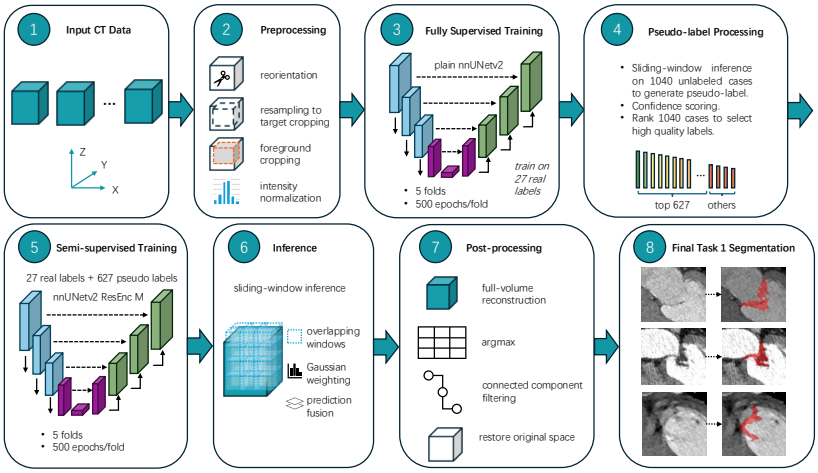

Fig. 1. Task 1 pipeline for 3D CT segmentation. After preprocessing, a fully supervised plain nnUNetv2 five-fold ensemble generates pseudo labels for 1040 unlabeled CT scans. Confidence scoring ranks these cases and selects the top 627 pseudo labels, which are combined with 27 real labels to train the five-fold ResEnc M nnUNetv2 model. Final inference uses sliding-window prediction followed by reconstruction, argmax, connected-component filtering, and restoration to the original space.

plain nnUNetv2 result of 0.8583 DSC, 4.6146 HD, and 0.2769 ASD to 0.8585 DSC, 4.5231 HD, and 0.2724 ASD.

2.2 Task 2: ResEnc M 3D nnUNetv2 for TEE Segmentation

3D TEE data differs substantially from CT in image contrast, noise distribution, and anatomical visibility. To address these characteristics, we adopt the official ResEnc M preset of 3D nnUNetv2 [8, 9] for volumetric TEE segmentation. The model follows the nnUNetv2 protocol, including preprocessing, dataset fingerprinting, patch-based optimization, and sliding-window inference. As illustrated in Fig. 2, the workflow converts 3D TEE volumes into the nnUNetv2 dataset format, applies automatic preprocessing, and trains the 3D full-resolution ResEnc M configuration using five-fold cross-validation. Each fold is optimized for 500 epochs, and the predictions from all five folds are ensembled during inference. This branch therefore benefits from both the higher-capacity residual encoder preset and the robustness of cross-validation ensembling, while remaining within the official nnUNetv2 design space.

During inference, predictions are exported as NIfTI files following the required Task 2 format. The ResEnc M five-fold ensemble achieves a DSC of 0.8543, an HD of 10.1049, and an ASD of 0.5485 on the validation set, demonstrating that scaling the official nnUNetv2 residual encoder preset and aggregating complementary fold predictions substantially improves volumetric echocardiographic segmentation.

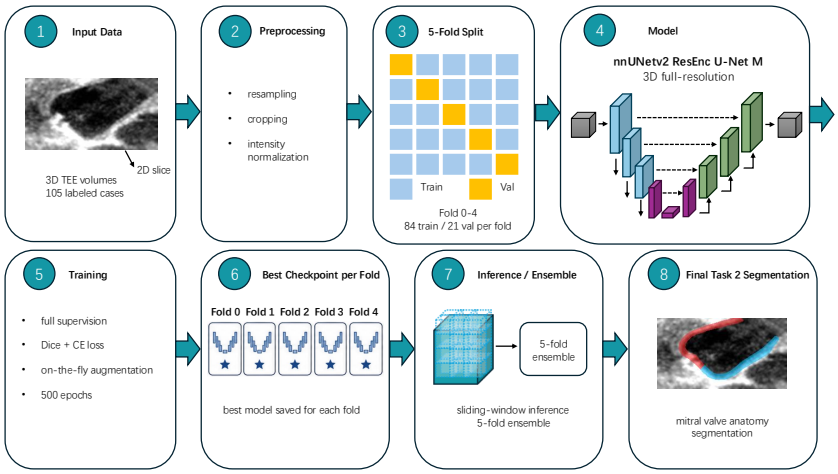

Fig. 2. Task 2 pipeline for 3D TEE segmentation. The workflow preprocesses 105 labeled 3D TEE volumes, uses a five-fold split with 84 training and 21 validation cases per fold, trains the 3D full-resolution nnUNetv2 ResEnc U-Net M for 500 epochs, saves the best checkpoint from each fold, and performs sliding-window inference with five-fold ensembling. The final panels show the saved best checkpoints, the sliding-window inference/ensemble step, and the resulting mitral-valve mask overlay on a representative 2D slice as output.

2.3 Task 3: UNet++ Video Segmentation

Video segmentation requires generating frame-wise binary masks from surgical video frames. Our initial attempt employed a DINOv3-based feature extraction pipeline [21], motivated by the transferability of self-supervised vision representations [2, 18]. However, validation results revealed limited effectiveness for this task. We therefore adopt a UNet++ segmentation model [27] with an ImageNet-pretrained [7] EfficientNet-B5 encoder [23], which provides nested skip connections for multi-scale feature fusion. The complete pipeline is illustrated in Fig. 3, encompassing base model training, pseudo-label self-training, final fine-tuning, inference, and post-processing.

The base model is trained within a Mean Teacher semi-supervised framework [24], where the student model is optimized by supervised Dice–Focal loss and the teacher model is updated via exponential moving average. We use AdamW [15] with cosine learning-rate scheduling [14] for optimization. Pseudo labels are filtered by confidence thresholds and small-region removal, following the principle that unlabeled supervision should be introduced gradually and conservatively [12, 22]. The base checkpoint is subsequently refined through strict pseudo-label self-training and HD-aware dice-recovery before the final fine-tuning stage.

Given an input frame I, the UNet++ branch predicts a binary logit map z and the probability map $P = \sigma(z)$. The final fine-tuning stage leverages all 180 labeled frames and 593 strictly selected pseudo-labeled masks. The video

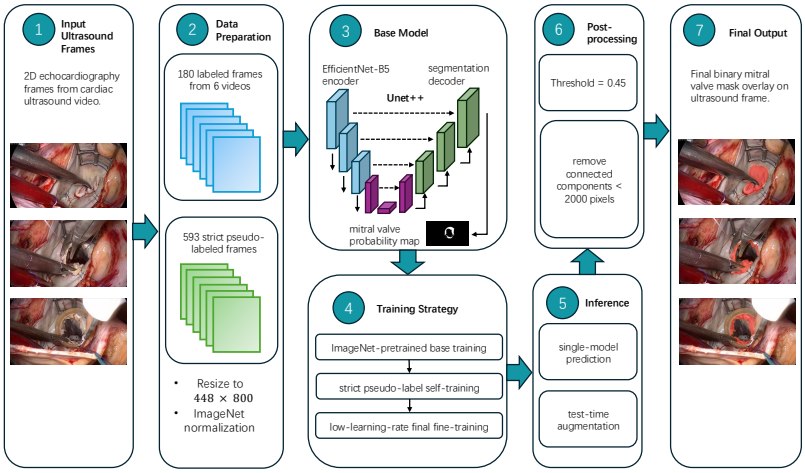

Fig. 3. Task 3 pipeline for video segmentation. Frames are resized to 448 × 800 and normalized with ImageNet statistics. UNet++ with an EfficientNet-B5 encoder is trained using 180 labeled frames and 593 strict pseudo-labeled frames, followed by pseudo-label self-training and low-learning-rate fine-tuning. Inference uses single-model prediction with test-time augmentation, threshold 0.45, and removal of connected components smaller than 2000 pixels.

segmentation objective is formulated as:

$$\mathcal{L}_{video} = 0.7\,\mathcal{L}_{Dice}(z, Y) + 0.3\,\mathcal{L}_{Focal}(z, Y), \tag{5}$$

where Dice loss captures region-level overlap [17] and Focal loss ($\gamma = 2.0$, $\alpha = 0.75$) emphasizes difficult foreground pixels [13]. During inference, the probability map is thresholded at 0.45 with test-time augmentation enabled, and connected components smaller than 2000 pixels are removed. Compared with the DINOv3-based pipeline, the UNet++ branch improves DSC from 0.7168 to 0.8105 and reduces ASD from 21.4949 to 14.7360, confirming that task-specific dense prediction architectures can outperform general-purpose foundation features for frame-wise surgical video segmentation.

3 Experiments

3.1 Experimental Setup

Datasets. We evaluate MVAA-Seg on the MICCAI 2026 Mitral Valve Anatomy Analysis (MVAA) Challenge validation set, which encompasses three tasks with distinct imaging modalities. Task 1 targets CT segmentation with 27 labeled volumetric scans and a substantially larger unlabeled pool for semi-supervised learning. Task 2 addresses 3D TEE segmentation, where volumetric echocardiographic data exhibit severe speckle noise and weak tissue boundaries. Task 3 focuses on surgical video segmentation, requiring frame-wise binary mask prediction under varying illumination, viewpoint changes, and instrument occlusion.

Table 1. Ablation study on Task 1 CT segmentation.

Configuration	DSC ↑	HD ↓	ASD ↓
Baseline	0.6184	9.6213	0.9594
Fully supervised plain nnUNetv2 5-fold	0.8583	4.6146	0.2769
Semi-supervised ResEnc M 5-fold + pseudo labels	**0.8585**	**4.5231**	**0.2724**

All experiments are conducted on an NVIDIA RTX A6000 GPU (48GB). For Task 1, we first train a fully supervised plain nnUNetv2 five-fold model on the 27 labeled CT scans to generate probability maps for unlabeled cases, select the top 627 pseudo labels according to confidence and anatomical plausibility, and then train a five-fold nnUNetv2 ResEnc M model on the combined real and pseudo-labeled set. For Task 2, we use the official 3D full-resolution nnUNetv2 ResEnc M configuration [8, 9], train five folds for 500 epochs per fold, and ensemble all fold predictions during inference. For Task 3, we use UNet++ [27] with an ImageNet-pretrained EfficientNet-B5 encoder [23], optimized by Dice–Focal loss [17, 13] with AdamW [15] and cosine scheduling. Following the official protocol, performance is measured by Dice similarity coefficient (DSC, ↑), Hausdorff distance (HD, ↓), and average surface distance (ASD, ↓).

3.2 Ablation Study on Task 1 CT Segmentation

Table 1 presents the ablation study on Task 1. The official baseline achieves 0.6184 DSC, 9.6213 HD, and 0.9594 ASD. The fully supervised plain nnUNetv2 five-fold model trained on the 27 labeled CT scans reaches 0.8583 DSC, 4.6146 HD, and 0.2769 ASD. After adding the top 627 confidence-ranked pseudo labels and training the ResEnc M five-fold semi-supervised model, the final Task 1 result further improves to 0.8585 DSC, 4.5231 HD, and 0.2724 ASD. These results show that the selected pseudo labels mainly improve boundary-distance metrics while preserving the strong Dice performance of the fully supervised nnUNetv2 model.

3.3 Comparison of Task 3 Video Segmentation Branches

Table 2 compares the baseline, the previous DINOv3-based branch, and the final UNet++ branch for Task 3. The DINOv3-based branch improves DSC from 0.6440 to 0.7168 and reduces HD from 193.2962 to 96.8771. However, its ASD remains relatively high. By replacing it with UNet++, the DSC further improves to 0.8105 and ASD decreases to 14.7360. Although the DINOv3 branch obtains a slightly lower HD, the UNet++ branch provides the best Dice overlap and average boundary accuracy, and is therefore selected as the final Task 3 model.

3.4 Overall Validation Results

As shown in Table 3, MVAA-Seg achieves substantial improvements over the baseline across all three tasks. The confidence-guided semi-supervised nnUNetv2 with ResEnc M provides the largest relative gain on Task 1, confirming that

Table 2. Comparison of Task 3 video segmentation branches.

Method	DSC ↑	HD ↓	ASD ↓
Baseline	0.6440	193.2962	24.7622
DINOv3-based branch	0.7168	**96.8771**	21.4949
UNet++ branch	**0.8105**	100.1046	**14.7360**

Table 3. Overall validation results on the MVAA Challenge.

Task	Method	DSC ↑	HD ↓	ASD ↓
Task 1 CT	Baseline	0.6184	9.6213	0.9594
	MVAA-Seg	**0.8585**	**4.5231**	**0.2724**
Task 2 TEE	Baseline	0.7451	18.7673	1.6552
	MVAA-Seg	**0.8543**	**10.1049**	**0.5485**
Task 3 Video	Baseline	0.6440	193.2962	24.7622
	MVAA-Seg	**0.8105**	**100.1046**	**14.7360**

confidence-guided pseudo-label selection effectively compensates for limited annotations. The nnUNetv2 ResEnc M branch with five-fold ensembling delivers strong Task 2 performance, demonstrating that systematic pipeline engineering is highly effective for volumetric echocardiographic data. The UNet++ branch yields the most significant DSC improvement on Task 3 (0.6440→0.8105), indicating that task-specific dense prediction architectures outperform general-purpose foundation features for surgical video segmentation. These results validate the design principle of MVAA-Seg: each modality benefits from a task-specific strategy.

4 Conclusion

In this work, we present MVAA-Seg, a multi-stage segmentation framework for the MICCAI 2026 MVAA Challenge. The core of our method is a confidence-guided semi-supervised CT segmentation strategy that selects pseudo labels according to foreground confidence, low-percentile confidence, connected-component consistency, and anatomical volume plausibility. We further integrate a five-fold nnUNetv2 ResEnc M branch, trained for 500 epochs per fold, for TEE segmentation and a UNet++ branch for video segmentation. Extensive validation experiments demonstrate that confidence-guided semi-supervised nnUNetv2 improves Task 1 CT segmentation, the five-fold nnUNetv2 ResEnc M ensemble achieves strong performance on Task 2 TEE segmentation, and UNet++ substantially outperforms our previous DINOv3-based video branch on Task 3. Overall, MVAA-Seg provides a practical and robust solution for multi-modal mitral valve anatomy analysis under limited annotation and heterogeneous data conditions.

References

1. Bai, J., Zhou, Z., Ou, Z., Koehler, G., Stock, R., Maier-Hein, K., Elbatel, M., Martí, R., Li, X., Qiu, Y., et al.: Psfhs challenge report: pubic symphysis and fetal head segmentation from intrapartum ultrasound images. Medical Image Analysis **99**, 103353 (2025)
2. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9650–9660 (2021)
3. Chen, G., Bai, J., Ou, Z., Lu, Y., Wang, H.: Psfhs: intrapartum ultrasound image dataset for ai-based segmentation of pubic symphysis and fetal head. Scientific data **11**(1), 436 (2024)
4. Chen, Z., Lu, Y., Long, S., Campello, V.M., Bai, J., Lekadir, K.: Fetal head and pubic symphysis segmentation in intrapartum ultrasound image using a dual-path boundary-guided residual network. IEEE journal of biomedical and health informatics **28**(8), 4648–4659 (2024)
5. Chen, Z., Ou, Z., Lu, Y., Bai, J.: Direction-guided and multi-scale feature screening for fetal head–pubic symphysis segmentation and angle of progression calculation. Expert Systems with Applications **245**, 123096 (2024)
6. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
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. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 488–498. Springer (2024)
10. Jiang, J., Wang, H., Bai, J., Long, S., Chen, S., Campello, V.M., Lekadir, K.: Intrapartum ultrasound image segmentation of pubic symphysis and fetal head using dual student-teacher framework with cnn-vit collaborative learning. In: International conference on medical image computing and computer-assisted intervention. pp. 448–458. Springer (2024)
11. Jiao, Y., Xu, Q., Luo, Y., He, X., Chen, Z., Duan, W.: Tm-unet: Token-memory enhanced sequential modeling for efficient medical image segmentation. In: 2026 IEEE 23rd International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2026)
12. 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)
13. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
14. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)

15. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
16. Luo, Y., Xu, Q., Huang, H., Ouyang, Y., Chen, Z., Duan, W.: Msm-seg: A modality-and-slice memory framework with category-agnostic prompting for multi-modal brain tumor segmentation. arXiv preprint arXiv:2510.10679 (2025)
17. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016)
18. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: Dinov2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193 (2023)
19. Ou, Z., Bai, J., Chen, Z., Lu, Y., Wang, H., Long, S., Chen, G.: Rtseg-net: a lightweight network for real-time segmentation of fetal head and pubic symphysis from intrapartum ultrasound images. Computers in biology and medicine **175**, 108501 (2024)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
21. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: Dinov3. arXiv preprint arXiv:2508.10104 (2025)
22. 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)
23. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)
24. 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)
25. Xu, H., Sun, J., Miao, Q., Hong, Y., Xu, Q., Li, C., Hu, H., Lin, G., Huang, Y.: Synergistic hierarchical and graph attention networks for robust leptomeningeal metastasis detection in brain mri. Applied Soft Computing p. 114861 (2026)
26. Xu, Q., Luo, Y., Duan, W., Chen, Z.: Co-seg++: Mutual prompt-guided collaborative learning for versatile medical segmentation. IEEE Transactions on Medical Imaging (2025)
27. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: International workshop on deep learning in medical image analysis. pp. 3–11. Springer (2018)