Task-Specific Segmentation Frameworks for Multi-Modal Mitral Valve Tasks in MVAA2026

Sihan Chen[1] [0009-0003-2083-7940] and Jianning Chi[1] [0000-0002-9748-5619]

School of Robotics Science and Engineering, Northeastern University, Shenyang, China
{chensh11@mails.neu.edu.cn, chijianning@mail.neu.edu.cn}

Abstract. Accurate mitral valve segmentation is important for structural heart disease assessment, treatment planning, and image-guided intervention. Accordingly, the MVAA2026 competition includes three tasks covering cardiac CT, 3D transesophageal echocardiography (TEE), and surgical video frames, corresponding to preoperative anatomical assessment, valve morphology evaluation, and intraoperative scene understanding. Task 1 focuses on semi-supervised 3D mitral valve segmentation from cardiac CT volumes, which provides volumetric anatomical information for preoperative evaluation. It remains challenging because the mitral valve occupies a small region in 3D cardiac volumes and available annotations are limited. To solve this problem, we develop a semi-supervised 3D CT pipeline that combines volumetric masked anatomy modeling, tri-view pseudo-label fusion, refined pseudo-label learning, and 3D CubeMix consistency training. Task 2 targets fully supervised mitral valve leaflet segmentation from 3D TEE, which is valuable for valve morphology assessment and intraoperative evaluation. It is difficult due to ultrasound artifacts and thin leaflet structures. To solve this problem, we design a 3D TEE segmentation method using cross-resolution global-local modeling, boundary-aware shape guidance, and a five-fold SegResNet ensemble. Task 3 aims at semi-supervised mitral valve region segmentation from 2D surgical video frames, which can support surgical scene understanding and image-guided procedure analysis. It is challenged by sparse annotations and confusing chamber-like backgrounds. To solve this problem, we propose a 2D surgical frame segmentation pipeline based on ImageNet-initialized Mean Teacher learning, SAM-guided pseudo-label refinement, and prototype-guided pseudo-label denoising. On the official validation platform, our submissions achieved Dice scores of 0.8418, 0.8047, and 0.7983 for Tasks 1–3, respectively. The corresponding HD and ASD were 5.6920 and 0.3122 for Task 1, 16.9773 and 0.9108 for Task 2, and 175.1968 and 20.3188 for Task 3. These results indicate that task-specific segmentation strategies can support multi-modal clinical mitral valve segmentation across CT, ultrasound, and surgical imaging scenarios.

Keywords: Mitral valve segmentation · Cardiac CT · Transesophageal echocardiography · Surgical video segmentation · Semi-supervised learning

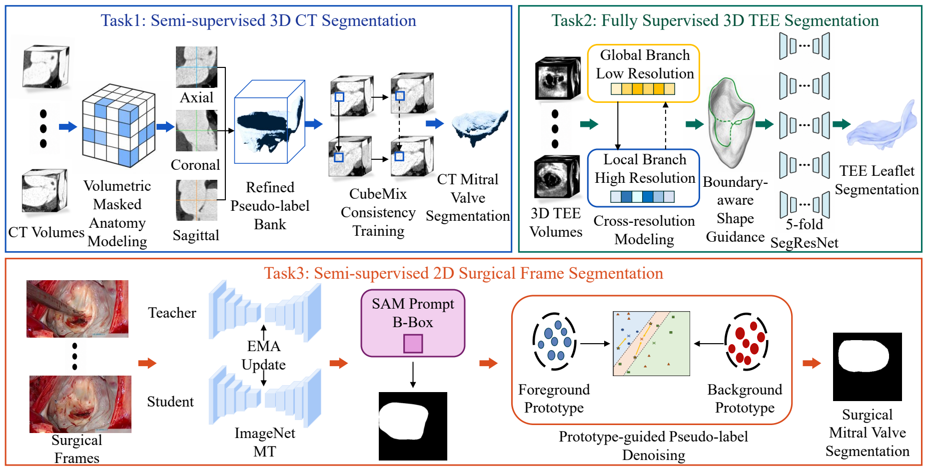

Fig. 1. Overview of the proposed task-specific segmentation pipelines for the three MVAA2026 tasks.

1 Overview

The mitral valve plays a critical role in cardiac function, and accurate segmentation of its anatomical structures is essential for disease assessment, treatment planning, and image-guided intervention. Cardiac CT, 3D transesophageal echocardiography (TEE), and surgical videos provide complementary information across preoperative assessment, valve morphology evaluation, and intra-operative analysis. Recent cardiac imaging benchmarks have further promoted automated anatomical analysis across modalities [2, 4, 5].

The MVAA2026 competition [1] contains three mitral-valve segmentation tasks covering 3D cardiac CT, 3D TEE, and 2D surgical frames. The proposed methods for the three tasks are shown in Fig. 1.

2 Method

Across the three tasks, our designs follow a shared principle: constructing reliable anatomical supervision according to modality characteristics. CT segmentation uses volumetric representation learning and multi-planar pseudo-label consistency, TEE segmentation uses global-local anatomical modeling and boundary-aware refinement, and surgical video segmentation uses task-specific teacher learning with foundation-model priors. Each strategy is adapted to modality-specific challenges.

2.1 Task 1: Semi-supervised 3D CT Mitral Valve Segmentation

Task 1 aims to segment the mitral valve region from 3D cardiac CT volumes. The small foreground region and limited voxel-level annotations make this task challenging. We design a semi-supervised pipeline consisting of volumetric masked anatomy modeling, tri-view pseudo-label refinement, and 3D Mutual CubeMix consistency training.

Volumetric Masked Anatomy Modeling. Before pseudo-label generation, we exploit unlabeled CT volumes through volumetric masked anatomy modeling (VMAM) [14, 15]. Given a 128^3 CT crop X, we randomly sample 3–7 cuboid regions with side lengths of 12%–45% of the corresponding spatial dimensions. The masked input is generated as

$$\tilde{X} = X \odot (1 - M), \tag{1}$$

where M denotes the binary cuboid mask. The masked volume is processed by a 3D residual encoder-decoder network [6, 7, 12] with three reconstruction heads predicting CT intensity, low-frequency anatomy, and 3D edge-gradient maps. The low-frequency target is obtained by $7 \times 7 \times 7$ average pooling, and the edge target is computed using finite differences.

The VMAM objective is

$$\mathcal{L}_{\text{VMAM}} = \|M \odot (\hat{X} - X)\|_1 + \alpha \|M \odot (\hat{X}_{low} - X_{low})\|_1 + \beta \|M \odot (\hat{X}_{edge} - X_{edge})\|_1. \tag{2}$$

where $\hat{X}$, $\hat{X}_{low}$, and $\hat{X}_{edge}$ denote the reconstructed targets. The three reconstruction losses use weights of 1.0, 0.5, and 0.5, respectively. The pretrained encoder is transferred to downstream segmentation training.

Tri-view Pseudo-label Refinement. After VMAM pretraining, a teacher model generates pseudo-labels for unlabeled CT volumes. To reduce view-specific errors, we fuse predictions from multiple anatomical orientations [28, 27].

Three 2.5D specialists process axial, coronal, and sagittal views using five adjacent slices, while an additional 3D specialist provides volumetric predictions. All probability maps are resampled to the same voxel space and fused in the log-odds domain:

$$P_{\text{fuse}} = \sigma \left(w_{3D} \text{logit}(P_{3D}) + \sum_{v \in \{a,c,s\}} w_v \text{logit}(P_v) \right). \tag{3}$$

The 3D branch weight is $w_{3D} = 0.45$, and the three planar-view weights jointly sum to 0.55. The fused probability map is refined using thresholding, morphological operations, and component filtering, then stored in the pseudo-label bank.

3D Mutual CubeMix. The final stage optimizes the segmentation model using labeled CT volumes and refined pseudo-labels. We apply 3D Mutual CubeMix [10] by exchanging randomly sampled cuboids between two unlabeled CT crops and applying the same mask operation to their pseudo-labels.

The cuboid side length ratios are randomly sampled from 0.25 to 0.55 of the spatial dimensions. The training objective is

$$\mathcal{L}_{\mathrm{T1}} = \mathcal{L}_{\mathrm{sup}} + \lambda_u \mathcal{L}_{\mathrm{pseudo}} + \lambda_m \mathcal{L}_{\mathrm{CubeMix}}. \tag{4}$$

where $\mathcal{L}_{\mathrm{sup}}$, $\mathcal{L}_{\mathrm{pseudo}}$, and $\mathcal{L}_{\mathrm{CubeMix}}$ denote the supervised segmentation loss, pseudo-label supervision loss, and CubeMix consistency loss, respectively. The maximum weights are set to 1.0, 0.4, and 0.25, respectively, with a 10-epoch ramp-up schedule to reduce early confirmation bias.

2.2 Task 2: Fully Supervised 3D TEE Leaflet Segmentation

Task 2 focuses on mitral valve leaflet segmentation from 3D TEE volumes. Ultrasound artifacts and thin leaflet structures make boundary delineation challenging. We design a global-local 3D TEE segmentation pipeline combining cross-resolution context modeling, boundary-aware shape regularization, and multi-fold ensemble inference.

Global-local Modeling. To balance global anatomical context and local leaflet details, we adopt a cross-resolution dual-branch design. The local branch receives a high-resolution crop $X_l \in \mathbb{R}^{1\times128\times128\times128}$ for leaflet segmentation, while the global branch processes a downsampled whole volume $X_g \in \mathbb{R}^{1\times64\times64\times64}$ to provide anatomical context.

Let F_l and F_g denote local and global features. The global features are projected and fused into the local decoder:

$$F_{\mathrm{fuse}} = \phi_l(F_l) + \mathrm{Up}(\phi_g(F_g)) \tag{5}$$

where ϕ_l and ϕ_g are feature projections. The global branch is optimized as an auxiliary task:

$$\mathcal{L}_{\mathrm{CR}} = \mathcal{L}_{\mathrm{local}} + \lambda_g \mathcal{L}_{\mathrm{global}} \tag{6}$$

During inference, the global feature is extracted from the whole-volume input and reused for local sliding-window predictions.

Shape Regularization. To improve leaflet boundary accuracy under ultrasound artifacts, we introduce boundary-aware shape regularization. A boundary target and signed distance field (SDF) are derived from the ground-truth mask to provide surface and geometric supervision:

$$\mathcal{L}_{\mathrm{surf}} = \lambda_b \mathcal{L}_{\mathrm{BCE}}(\hat{B}, B(Y)) + \lambda_d |\hat{D} - D(Y)|_1 \tag{7}$$

where $\hat{B}$ and $\hat{D}$ represent the predicted boundary and distance maps. In addition, class-specific feature prototypes are introduced to reduce confusion between adjacent leaflet classes. The shape-aware objective is formulated as:

$$\mathcal{L}_{\text{shape}} = \mathcal{L}_{\text{seg}} + \mathcal{L}_{\text{surf}} + \lambda_p \mathcal{L}_{\text{proto}} \tag{8}$$

The prototype-based regularization is designed to reduce feature ambiguity between adjacent leaflet structures. This design can help reduce uncertainty in anatomically close regions, such as the posterior mitral leaflet, where weak boundaries and limited spatial separation may affect segmentation accuracy.

Five-fold Ensemble. To improve robustness under limited fully supervised data, we train a five-fold SegResNet ensemble [11]. Each fold selects the checkpoint with the best validation performance and is optimized with:

$$\mathcal{L}_{\text{T2}} = \mathcal{L}_{\text{DiceCE}} + \lambda_f \mathcal{L}_{\text{Focal}} + \lambda_b \mathcal{L}_{\text{Boundary}} + \lambda_o \mathcal{L}_{\text{Overlap}} \tag{9}$$

During inference, each fold generates a probability map using sliding-window inference with test-time augmentation. The final prediction is obtained by averaging all fold outputs:

$$P_{\text{ens}} = \frac{1}{K} \sum_{k=1}^{K} P_k, \quad K = 5 \tag{10}$$

The final segmentation is generated by voxel-wise $\arg\max$.

2.3 Task 3: Semi-supervised 2D Surgical Frame Segmentation

Task 3 is a frame-wise binary segmentation task on surgical video frames [3]. Limited labeled frames and chamber-like backgrounds make this task challenging. We develop a semi-supervised pipeline consisting of ImageNet-initialized Mean Teacher learning [9], SAM-guided pseudo-label refinement [16, 18], and prototype-guided pseudo-label denoising [28, 27].

Mean Teacher Learning. We initialize an encoder-decoder segmentation network [6] with ImageNet-pretrained weights. Labeled frames are optimized using Dice-Focal loss [8], while unlabeled frames are incorporated through Mean Teacher learning. The teacher is updated by exponential moving average of the student and generates pseudo-labels from weakly augmented images, while the student learns from strongly augmented views.

The objective is

$$\mathcal{L}_{\text{MT}} = \mathcal{L}_{\text{sup}} + \lambda_u \mathcal{L}_{\text{unsup}}, \tag{11}$$

where $\mathcal{L}_{\text{sup}}$ and $\mathcal{L}_{\text{unsup}}$ denote supervised segmentation loss and consistency loss, respectively.

SAM-guided Pseudo-label Refinement. Mean Teacher predictions may contain inaccurate boundaries and false-positive regions. We therefore introduce SAM as a structural refinement prior rather than the final predictor [16, 18].

For each unlabeled frame, the teacher prediction is thresholded to obtain a coarse foreground mask. Its bounding box is expanded by 12 pixels as the SAM box prompt, and the median coordinate of foreground pixels is used as a positive point prompt. SAM generates candidate masks, which are filtered according to prediction agreement, foreground constraints, geometric plausibility, and confidence score. A SAM mask replaces the teacher refinement only when it achieves a higher quality score.

The refined pseudo-label bank contains 1000 accepted samples from 1379 unlabeled frames (72.5% acceptance rate). Among them, only 4 samples are replaced by SAM-generated masks, while the others retain teacher-morphology refined masks.

The student is optimized using labeled masks and refined pseudo-labels:

$$\mathcal{L}_{\mathrm{SAM}} = \mathcal{L}_{\mathrm{sup}} + \lambda_p \mathcal{L}_{\mathrm{pseudo}} + \lambda_b \mathcal{L}_{\mathrm{boundary}}, \tag{12}$$

where $\mathcal{L}_{\mathrm{boundary}}$ improves boundary localization.

Prototype-guided Denoising. To further reduce unreliable pseudo-label regions around chamber-like backgrounds, we introduce prototype-guided denoising. A foreground prototype and a background/context prototype are maintained from encoder features. Pixel features are compared with these prototypes to assign reliability weights for pseudo-label supervision.

The denoising objective is

$$\mathcal{L}_{\mathrm{denoise}} = \mathcal{L}_{\mathrm{sup}} + \lambda_p \sum_i w_i \mathcal{L}_{\mathrm{pseudo}}(p_i, \hat{y}_i) + \lambda_b \mathcal{L}_{\mathrm{boundary}}, \tag{13}$$

where $\hat{y}_i$ denotes the refined pseudo-label, p_i is the predicted foreground probability, and w_i represents the prototype-based reliability weight.

3 Results

We evaluated the proposed methods on the official validation set released by the MVAA challenge. The reported metrics include Dice similarity coefficient (DSC), Hausdorff distance (HD), and average surface distance (ASD). Higher DSC indicates better region overlap, while lower HD and ASD indicate more accurate boundary localization.

Table 1. Official validation results of the submissions for the three MVAA tasks.

Task	DSC ↑	HD ↓	ASD ↓
Task 1: 3D cardiac CT segmentation	0.8418	5.6920	0.3122
Task 2: 3D TEE segmentation	0.8047	16.9773	0.9108
Task 3: 2D surgical frame segmentation	0.7983	175.1968	20.3188

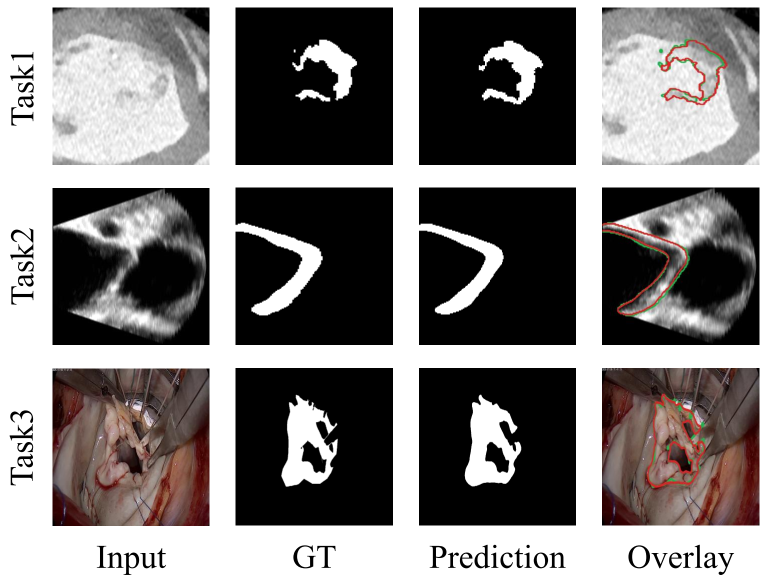

Fig. 2. Qualitative segmentation results on the MVAA validation set. Rows correspond to the three tasks, and columns show input, ground truth, prediction, and overlay. Green and red contours denote ground truth and prediction, respectively.

Table 1 summarizes the official validation performance of our submissions, and Fig. 2 provides representative qualitative examples for the three tasks.

For Task 1, the 3D CT segmentation model achieved a DSC of 0.8418, with HD of 5.6920 and ASD of 0.3122. These quantitative results indicate accurate volumetric segmentation and low surface-distance error for the mitral valve region in cardiac CT. The qualitative example in Fig. 2 shows that the predicted mask captures the main CT target region and remains close to the ground-truth contour, which is consistent with the low surface-distance errors reported in Table 1.

For Task 2, the 3D TEE leaflet segmentation model obtained a DSC of 0.8047, HD of 16.9773, and ASD of 0.9108. Compared with CT, TEE segmentation is more affected by ultrasound artifacts and thin leaflet geometry, which makes boundary localization more challenging. The qualitative result shows that the prediction preserves the major leaflet structure and follows the ground-truth contour with reasonable spatial consistency. These observations indicate that the proposed 3D TEE pipeline provides stable leaflet segmentation under noisy ultrasound appearance.

For Task 3, the 2D surgical frame segmentation model achieved a DSC of 0.7983, HD of 175.1968, and ASD of 20.3188. The DSC shows that the main mitral valve region can be segmented in the surgical scene, while the higher HD and ASD reflect remaining boundary deviations and distant false-positive

Table 2. Ablation study of the proposed Task 1 pipeline. VMAM, Tri-view, and CubeMix are progressively added to the baseline model.

Method	DSC ↑	HD ↓	ASD ↓
Baseline	0.8076	18.9557	0.9584
+ VMAM	0.8241	18.4732	0.6956
+ Tri-view	0.8370	9.1097	0.6875
+ CubeMix (Full)	**0.8418**	**5.6920**	**0.3122**

regions. This observation is consistent with the qualitative example in Fig. 2, where the prediction captures the main foreground region but still shows local contour differences in visually complex surgical backgrounds. The relatively high surface-distance errors are mainly caused by challenging surgical imaging conditions, including specular reflections, occlusions, and visually similar chamber-like regions, where small false-positive components can substantially increase HD and ASD. These results indicate that surgical-frame segmentation remains the most challenging among the three tasks, especially for suppressing false positives around chamber-like structures.

Ablation Study on Task 1. To evaluate the contribution of each component in the proposed CT segmentation pipeline, we conduct an ablation study by progressively adding VMAM pretraining, tri-view pseudo-label refinement, and 3D Mutual CubeMix. The results are summarized in Table 2.

The baseline model achieves a DSC of 0.8076, while adding VMAM improves DSC to 0.8241 and reduces ASD from 0.9584 to 0.6956, demonstrating the benefit of CT-specific representation pretraining. Introducing tri-view pseudo-label refinement further improves DSC to 0.8370 and significantly reduces HD, indicating that multi-planar anatomical information helps improve spatial localization. Finally, 3D Mutual CubeMix achieves the best performance with a DSC of 0.8418, HD of 5.6920, and ASD of 0.3122, showing the effectiveness of consistency learning with refined pseudo-labels. Additional analyses, including detailed tri-view fusion comparisons, Task 2 cross-validation results, and Task 3 pseudo-label refinement statistics, are provided in the supplementary material.

4 Conclusion

We presented task-specific segmentation pipelines for the three MVAA2026 mitral valve segmentation tasks across cardiac CT, 3D TEE and surgical video frames. The validation results show that the proposed methods can address multi-modal mitral valve segmentation under limited annotations and modality-specific imaging challenges. By providing segmentation results across preoperative CT assessment, 3D ultrasound-based valve evaluation, and surgical scene analysis, these pipelines may support computer-assisted clinical mitral valve diagnosis and intervention.

References

1. Hammad, M., Bai, J., Zhao, J., Jia, T., Zhang, X., Xu, X., Ma, J., Li, S.: Mitral Valve Anatomy Analysis Using Multimodal Imaging Data. Zenodo, MICCAI 2026. doi:10.5281/zenodo.19726755 (2026)
2. Zeng, A., Wu, C., Lin, G., et al.: ImageCAS: A Large-Scale Dataset and Benchmark for Coronary Artery Segmentation Based on Computed Tomography Angiography Images. Computerized Medical Imaging and Graphics 109, 102287 (2023)
3. Zia, A., Berniker, M., Perreault, C., Nespolo, R., Jarc, A.: Surgical Visual Understanding Challenge (SurgVU). In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. Zenodo (2024). doi:10.5281/zenodo.14054184
4. Zhu, J., Bai, J., Zhou, Z., et al.: RAS Dataset: A 3D Cardiac LGE-MRI Dataset for Segmentation of the Right Atrial Cavity. Scientific Data 11, 401 (2024)
5. Bai, J., et al.: A Benchmark Framework for Right Atrial Cavity Segmentation from LGE-MRIs. IEEE Transactions on Medical Imaging (2025)
6. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015)
7. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings of the 4th International Conference on 3D Vision (3DV), pp. 565–571 (2016)
8. 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 (ICCV), pp. 2980–2988 (2017)
9. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Advances in Neural Information Processing Systems, vol. 30, pp. 1195–1204 (2017)
10. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: CutMix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6023–6032 (2019)
11. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. LNCS, vol. 11384, pp. 311–320. Springer, Cham (2019)
12. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18, 203–211 (2021)
13. Hatamizadeh, A., Yang, D., Roth, H., Xu, D.: UNETR: Transformers for 3D medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 574–584 (2022)
14. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 16000–16009 (2022)
15. Xie, Z., Zhang, Z., Cao, Y., Lin, Y., Bao, J., Yao, Z., Dai, Q., Hu, H.: SimMIM: A simple framework for masked image modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9653–9663 (2022)
16. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4015–4026 (2023)

17. Zhang, K., Liu, D.: Customized Segment Anything Model for medical image segmentation. arXiv preprint arXiv:2304.13785 (2023)
18. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications 15, 654 (2024)
19. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K.V., Carion, N., Wu, C.Y., Girshick, R., Dollár, P., Feichtenhofer, C.: SAM 2: Segment Anything in Images and Videos. arXiv preprint arXiv:2408.00714 (2024)
20. Chen, C., Miao, J., Wu, D., Yan, Z., Kim, S., Hu, J., Zhong, A., Liu, Z., Sun, L., Li, X., Liu, T., Heng, P.A., Li, Q.: MA-SAM: Modality-agnostic SAM adaptation for 3D medical image segmentation. Medical Image Analysis 98, 103310 (2024)
21. Qin, C., Cao, J., Fu, H., Khan, F.S., Anwer, R.M.: DB-SAM: Delving into high quality universal medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. LNCS, vol. 15012. Springer Nature Switzerland (2024)
22. Shen, Y., Li, J., Shao, X., Romillo, B.I., Jindal, A., Dreizin, D., Unberath, M.: FastSAM3D: An efficient Segment Anything Model for 3D volumetric medical images. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. Springer Nature Switzerland (2024)
23. Zheng, X., Zhang, Y., Zhang, H., Liang, H., Bao, X., Jiang, Z., Lao, Q.: Curriculum prompting foundation models for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. LNCS, vol. 15012, pp. 487–497. Springer Nature Switzerland (2024)
24. Xu, Z., Chen, C., Lu, D., Sun, J., Wei, D., Zheng, Y., Li, Q., Tong, R.K.Y.: FM-ABS: Promptable foundation model drives active barely supervised learning for 3D medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. Springer Nature Switzerland (2024)
25. Wang, G., Ye, J., Cheng, J., Li, T., Chen, Z., Cai, J., He, J., Zhuang, B.: SAM-Med3D-MoE: Towards a non-forgetting Segment Anything Model via mixture of experts for 3D medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. Springer Nature Switzerland (2024)
26. Koleilat, T., Asgariandehkordi, H., Rivaz, H., Xiao, Y.: MedCLIP-SAM: Bridging text and image towards universal medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. LNCS, vol. 15012. Springer Nature Switzerland (2024)
27. Zhu, Y., Yang, J., Liu, S., Zhang, R.: Inherent consistent learning for accurate semi-supervised medical image segmentation. In: Proceedings of Machine Learning Research, vol. 227, pp. 1581–1601. PMLR (2024)
28. Su, J., Luo, Z., Lian, S., Lin, D., Li, S.: Mutual learning with reliable pseudo label for semi-supervised medical image segmentation. Medical Image Analysis 95, 103111 (2024)