This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Supplementary Material 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

1 Additional Experimental Results

This supplementary material provides additional experiments and analyses corresponding to the reviewer comments. We further evaluate the detailed effectiveness of the tri-view fusion strategy in Task 1, the robustness of Task 2 under different data splits, and the pseudo-label refinement statistics of Task 3.

1.1 Tri-view Pseudo-label Fusion Analysis for Task 1

To further analyze the contribution of multi-planar anatomical information, we evaluate different fusion strategies during the pseudo-label refinement stage of Task 1. The compared settings include individual volumetric or planar predictions, dual-view combinations, tri-view fusion, and the combination with the 3D branch. The results are summarized in Table 1.

Table 1. Detailed comparison of different fusion strategies for Task 1. Higher DSC and lower HD/ASD indicate better segmentation performance.

Fusion Setting	DSC ↑	HD ↓	ASD ↓
3D only	0.8141	14.0702	0.9335
Axial only	0.8189	13.6626	0.9148
Coronal only	0.7913	26.4451	1.3283
Sagittal only	0.8080	24.8057	1.1029
Axial + Coronal	0.8269	10.0879	0.7793
Axial + Sagittal	0.8241	13.4889	0.8003
Coronal + Sagittal	0.8290	11.0453	0.8096
Tri-view	**0.8370**	**9.1097**	**0.6875**
3D + Tri-view	0.8366	11.2493	0.7865

The results show that multi-planar fusion consistently improves segmentation performance compared with single-view predictions. Tri-view fusion achieves the

best performance among the evaluated fusion strategies, with a DSC of 0.8370 and lower surface errors. Although the 3D branch provides additional volumetric context, its combination with tri-view fusion does not further improve the performance, indicating that high-resolution multi-planar information is particularly beneficial for the small mitral valve structure.

1.2 Internal Five-fold Validation for Task 2

To evaluate robustness under different training splits, we perform internal five-fold cross-validation on Task 2. The average performance across five folds is reported in Table 2.

Table 2. Internal five-fold cross-validation results for Task 2. Values are reported as mean±standard deviation.

Evaluation	DSC ↑	HD ↓	ASD ↓
Internal 5-fold result	0.8047±0.0217	16.9773±1.59	0.9108±0.09

The small variation across folds indicates that the proposed global-local TEE segmentation framework maintains stable performance under different training partitions.

1.3 SAM-guided Pseudo-label Refinement Statistics for Task 3

We further summarize the pseudo-label refinement process of Task 3. The statistics of accepted and rejected pseudo-label candidates are shown in Table 3.

Table 3. Statistics of SAM-guided pseudo-label refinement in Task 3.

Pseudo-label statistic	Value
Unlabeled frames	1379
Accepted pseudo labels	1000
Rejected candidates	379
Acceptance rate	72.5%
SAM-selected pseudo labels	4
Teacher-morphology retained	996

Among 1379 unlabeled frames, 1000 samples satisfy the quality criteria and are retained in the pseudo-label bank. Only four samples are replaced by SAM-generated masks, while the remaining samples retain teacher predictions with morphology refinement. These statistics demonstrate that SAM mainly serves as a selective structural refinement tool rather than replacing the task-specific teacher model.