EchoPOSE: 6D Pose Estimation of Sparse Echocardiograms for Left-Ventricular 3D Shape Reconstruction

Supplementary Material

7. LV Reconstruction Examples

We further report examples of LV shape reconstructions. Figs. 6, 7 and 8 show GHD-based reconstructions using both assumed standard pose described in Fig. 2 and EchoPOSE-predicted 6D image positioning for the MITEA, MITEA+AI, and Routine TTE datasets, respectively.

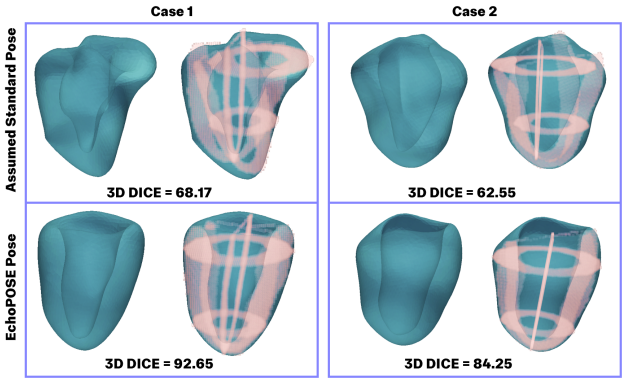

Figure 6. LV shape reconstruction examples from the MITEA dataset using the GHD morphing algorithm, considering assumed standard pose (top) and EchoPOSE slice positioning (bottom).

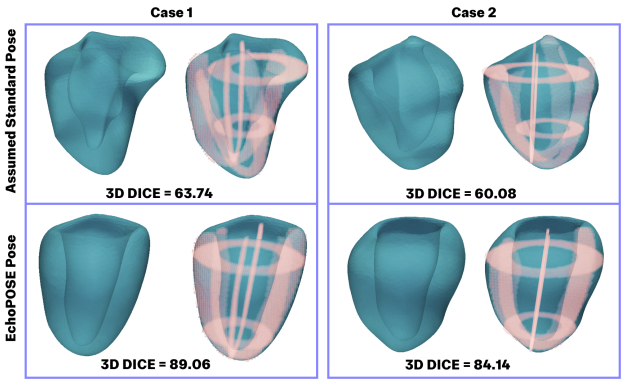

Figure 7. LV shape reconstruction examples from the MITEA+AI dataset using the GHD morphing algorithm, considering assumed standard pose (top) and EchoPOSE slice positioning (bottom).

8. AI Segmentations

We trained a multi-class nnU-Net model for LV myocardium and LV cavity segmentation on 3D echocardiographic slices from the MITEA dataset to produce the MITEA+AI variant, aiming to evaluate the robustness of

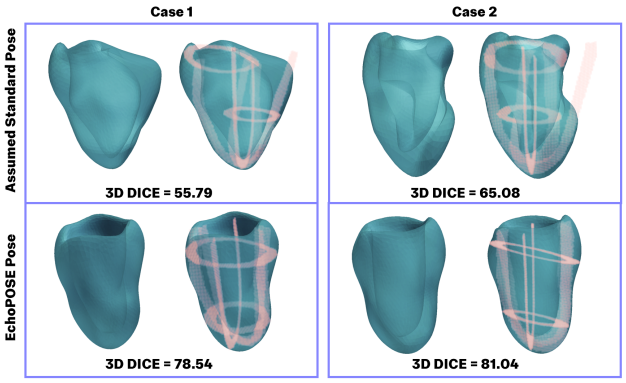

Figure 8. LV shape reconstruction examples from the Routine TTE dataset using the GHD morphing algorithm, considering assumed standard pose (top) and EchoPOSE slice positioning (bottom).

EchoPOSE under non-ideal, automatically generated segmentations. Using the same train/test split described in Sec. 4.2, the nnU-Net achieved a mean Dice (mDice) score of 90%. Fig. 9 shows representative segmentations from the test set.

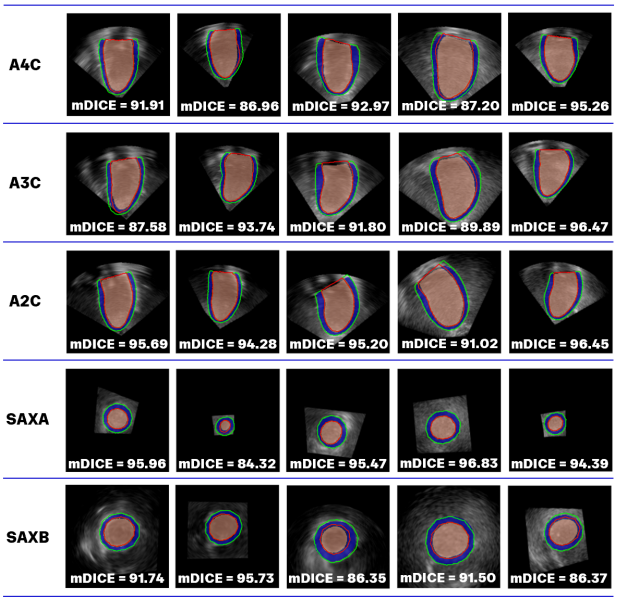

Figure 9. Predicted nnU-Net segmentation of the LV myocardium (blue) and cavity (light red). Ground-truth annotations are displayed as green contours for the myocardium and red contours for the cavity.

9. Architecture Details

EchoPOSE contains 56.1M trainable parameters. The Local Transformer comprises 8 layers, each with 8 heads and a 256-dimensional feature size, totaling 4.47M parameters. The Global Transformer uses 10 layers with 10 heads and a 32-dimensional feature size, contributing 13.78M parameters. Each view is encoded independently using a shared local encoder.

10. Computational Cost

Because each cardiac view is processed independently by the shared local encoder, the computational cost scales approximately linearly with the number of input views, as summarized in Table 4.

Table 4. EchoPose computational cost varying number of views.

Number of Input Views	2	3	4	5
GFLOPs	4.61	7.00	9.44	11.93