Semi-supervised Echocardiography Video Segmentation via Anchor Semantic Awareness and Continuous Pseudo-label Reforging

Supplementary Material

1. More Implementation Details

As shown in Table 1, we provide the key hyperparameters used during training and their tuning ranges. E_0 and E_1 control the burn-in phase of pseudo-label introduction and the end of weight ramp-up, respectively. A small E_0 prematurely introduces pseudo labels of varying quality, leading to unstable model convergence. E_0 = 10 and E_1 = 20 achieve the best performance on the validation set and are therefore selected as the default configuration.

We compared the maximum value $\lambda_{\max}$ of the pseudo-label weight $\lambda(e)$ between 0.1, 0.5, and 1.0, and found that $\lambda_{\max}$ = 0.5 fully utilized pseudo-label information during training while preventing the model from over-trusting low-quality pseudo-label. Linearly decreasing the confidence threshold τ from 0.9 to 0.6 can filter out excessive noise samples early on while ensuring high-quality pseudo labels later on. The rest of the common model hyperparameters were cross-validated within a common range to ensure training stability and convergence speed [2].

Table 1. Details of hyperparameters used during training, we conducted experiments within the scope.

Hyperparameters	Symbol	Scope	Value
Burn-in	E_0	{0, 5, 10, 15}	10
Ramp-up	E_1	{10, 20, 30}	20
Pseudo-label weight	$\lambda_{\max}$	{0.1, 0.5, 1.0}	0.5
Confidence threshold	τ_0	{0.7, 0.8, 0.9}	0.9
	τ_1	{0.4, 0.5, 0.6}	0.6
Learning rate	-	{1e-3, 1e-4}	1e-4
Batch Size	-	{2, 4, 8}	4

2. Detailed Introduction to LVEF

Left ventricular ejection fraction (LVEF) calculation strictly adhered to the American Society of Echocardiography guidelines [3], employing Simpson's biplane method for volumetric analysis. This standardized approach requires simultaneous processing of apical two-chamber (a2c) and four-chamber (a4c) view videos, implementing a biplane disk segmentation protocol that discretizes the ventricular structure into multi-level cylindrical units. The total volume (V) is derived through cumulative summation of individually calculated disk volumes, formally expressed as:

$$V = \frac{\pi}{4} \sum_{1}^{n} D_i^{2c} \times D_i^{4c} \times \frac{L}{n} \tag{1}$$

where D_i^{2c} and D_i^{4c} denote the chamber diameters across the two-chamber and four-chamber apical views respectively, and L indicates the length of the long axis, n is the number of disks. Finally, the calculation of left ventricular ejection fraction LVEF is as follows:

$$LV_{EF} = \frac{V_{ED} - V_{ES}}{V_{ED}} \times 100\% \tag{2}$$

where V_{ED} and V_{ES} denote the volumes at the end-diastole (ED) and end-systole (ES) respectively.

3. More Experiments

As shown in Table 2, we conduct ablation experiments on different components on the EchoNet-Dynamic [4] test set to further demonstrate the effectiveness of the module.

Table 2. Ablation study of various designs on the EchoNet-Dynamic test set.

Method	ASA		CPR	mDice	mHD	ASD
	TSF	ARC				
I				87.46	4.97	1.90
II	✓			91.88	3.53	1.34
III	✓	✓		92.56	3.25	1.19
IV	✓	✓	✓	**93.63**	**3.05**	**1.02**

We conducted ablation experiments on clustering methods in ASA, as shown in Table 3, the results show that the KNN achieves the best performance and efficiency.

Table 3. Ablation study of different clustering methods on the CAMUS-Semi test set.

Method	mDice	mHD	corr	FPS
K-medoids [5]	93.54	3.40	87.59	42
GMM [6]	92.51	3.88	84.75	16
K-means [1]	94.09	3.35	90.02	39
KNN [7]	**94.89**	**3.12**	**91.25**	**46**

4. More Visual Comparison Results

We present more visual comparisons with various SOTA methods on full echocardiography sequences from the CAMUS and EchoNet-Dynamic test sets, as illustrated in Figure 1 and Figure 2. Notably, even when only the ED and ES frames are labeled, our segmented left ventricle maintains smooth across the entire video, which is essential for identifying ED and ES frames in clinical practice.

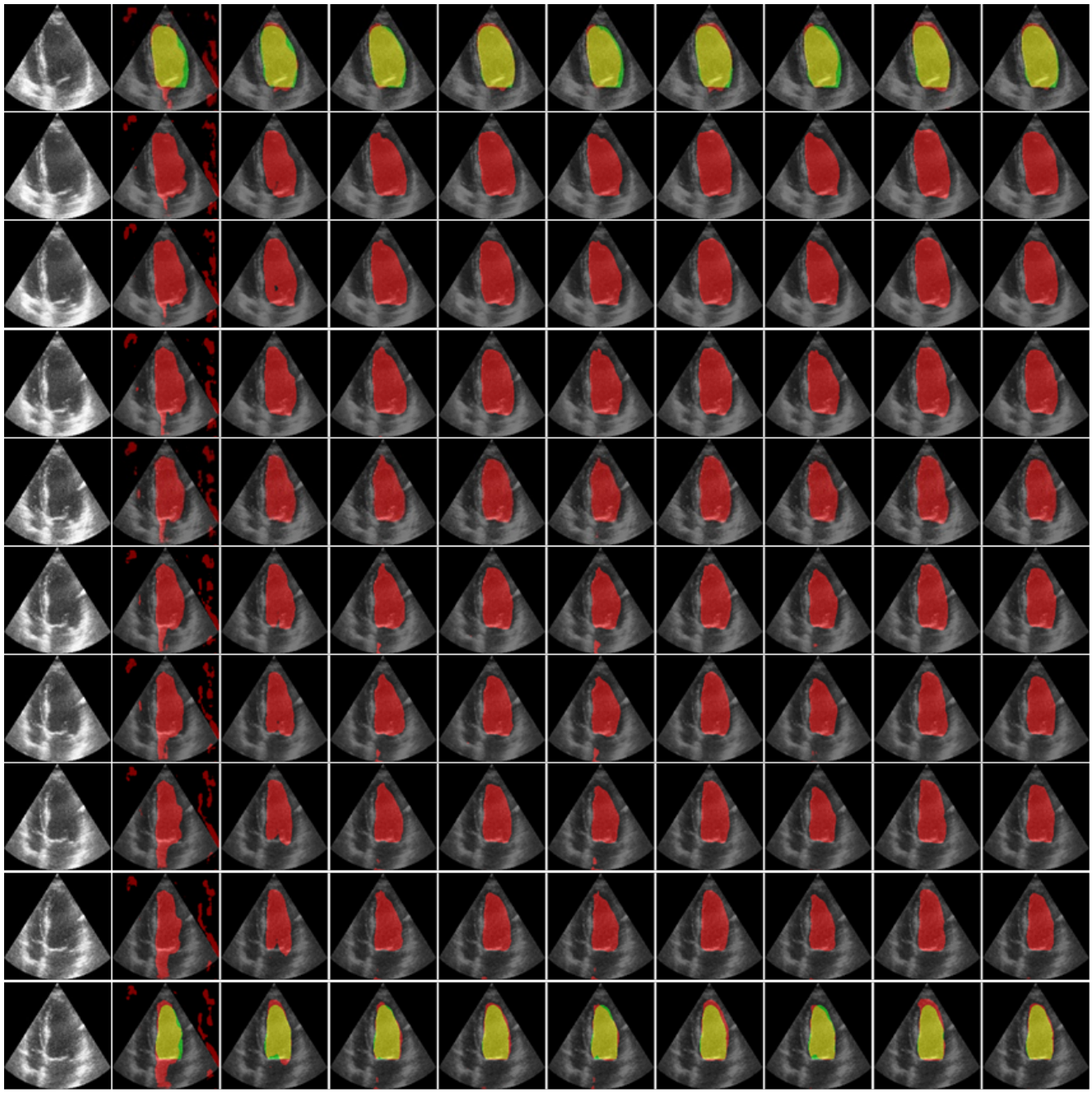

Figure 1. More visual comparison results of our method with other SOTA methods for the whole echocardiography sequences on the CAMUS-Semi test set. Green, red, and yellow regions represent the ground truth, prediction, and overlapping regions, respectively.

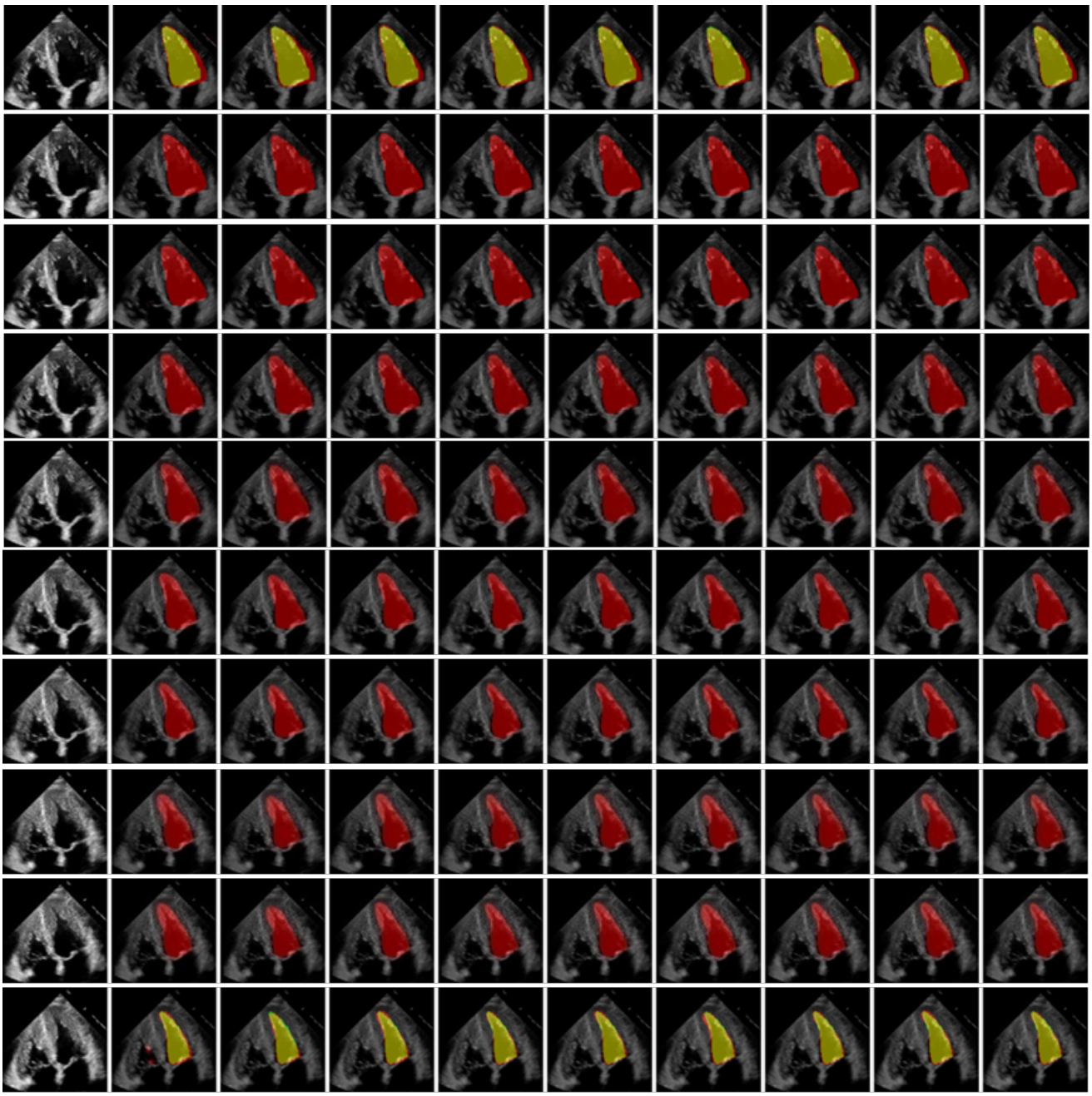

Figure 2. More visual comparison results of our method with other SOTA methods on the EchoNet-Dynamic test set. Each column shows the predictions of one method in chronological order. Green, red, and yellow regions represent the ground truth, prediction, and overlapping regions, respectively.

References

[1] Greg Hamerly and Charles Elkan. Learning the k in k-means. *Advances in neural information processing systems*, 16, 2003. 1

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015. 1

[3] RM Lang, LP Badano, V Mor-Avi, J Afilalo, A Armstrong, L Ernande, et al. an update from the american society of echocardiography and the european association of cardiovascular imaging. *J Am Soc Echocardiogr 2015*, 1:e14, 2015. 1

[4] David Ouyang, Bryan He, Amirata Ghorbani, Neal Yuan, Joseph Ebinger, Curtis P Langlotz, Paul A Heidenreich, Robert A Harrington, David H Liang, Euan A Ashley, et al. Video-based ai for beat-to-beat assessment of cardiac function. *Nature*, 580(7802):252–256, 2020. 1

[5] Hae-Sang Park and Chi-Hyuck Jun. A simple and fast algorithm for k-medoids clustering. *Expert systems with applications*, 36(2):3336–3341, 2009. 1

[6] Douglas A Reynolds et al. Gaussian mixture models. *Encyclopedia of biometrics*, 741(659-663):3, 2009. 1

[7] Michael Steinbach and Pang-Ning Tan. knn: k-nearest neighbors. In *The top ten algorithms in data mining*, pages 165–176. Chapman and Hall/CRC, 2009. 1