GDKVM: Echocardiography Video Segmentation via Spatiotemporal Key-Value Memory with Gated Delta Rule

Supplementary Material

A. Ejection Fraction Estimation

The Left Ventricular Ejection Fraction (LV_{EF}) is a critical measure of cardiac pump function. Its calculation requires determining the End-Diastolic Volume (V_{ED}), the volume of the left ventricle when it is maximally filled, and the End-Systolic Volume (V_{ES}), the volume when it is maximally contracted.

The general formula to compute the LV_{EF} is:

$$LV_{EF} = \frac{V_{ED} - V_{ES}}{V_{ED}} \times 100\%. \tag{7}$$

Biplane Simpson's method This method is the gold standard recommended by the American Society of Echocardiography [17] because it involves fewer geometric assumptions and is more accurate, especially for abnormally shaped ventricles. The method uses two orthogonal views: the apical four-chamber (a4c) and the apical two-chamber (a2c) view. The left ventricle is divided into a series of n disks of equal height along its long axis, L. Each disk is assumed to be an elliptical cylinder. The total volume V is the sum of the volumes of all disks. The volume is calculated using the diameters of the i-th disk measured from the four-chamber (D_i^{4c}) and two-chamber (D_i^{2c}) views. The volume formula is given by:

$$V = \frac{\pi}{4} \sum_{i=1}^{n} \left(D_i^{4c} \cdot D_i^{2c} \cdot \frac{L}{n} \right). \tag{8}$$

Single-plane Simpson's method This is a simplified version of the method, applied when only one view (typically the apical four-chamber view) is of sufficient quality for analysis. This method uses only a single view and assumes the left ventricle is a solid of revolution. Therefore, each disk is treated as a perfect circular cylinder. The volume calculation relies on the diameter of each disk (D_i^{4c}) as measured from the single available plane. The volume formula is:

$$V = \frac{\pi}{4} \sum_{i=1}^{n} \left((D_i^{4c})^2 \cdot \frac{L}{n} \right). \tag{9}$$

For both methods, V_{ED} and V_{ES} are calculated separately using the appropriate formula. These values are then used in the general LV_{EF} Eq. (7) to determine the ejection fraction.

B. Experiments Continued

B.1. The Clinical Metric

We also performed linear regression and Bland-Altman plots for the clinical metric LV_{EF} on the EchoNet-Dynamic dataset. As shown in the Fig. 11 and Tab. 5, our method demonstrated excellent performance and favorable clinical metrics.

Method	EchoNet-Dynamic	
	corr	bias ± std (%)
XMem++ [3]	0.692	5.77 ± 10.4
Cutie [7]	0.695	4.30 ± 11.7
VideoMamba [19]	0.764	-2.92 ± 9.38
Vision LSTM [2]	0.768	-2.42±9.10
PKEchoNet [40]	0.852	1.29±9.62
DSA [22]	0.868	1.42±9.18
MemSAM [10]	0.859	1.09±9.44
SimLVSeg [26]	0.794	0.91±9.56
GDKVM	**0.872**	**-0.70±9.15**

Table 5. Clinical metrics comparison against different state-of-the-art methods on EchoNet-Dynamic.

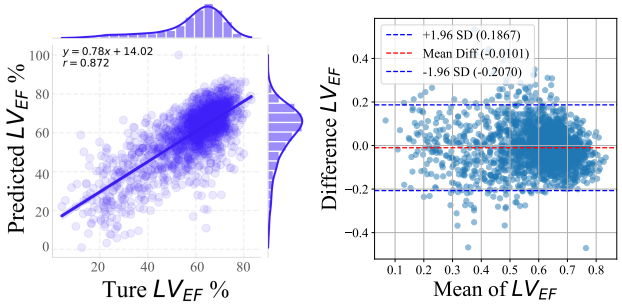


Figure 11. Linear regression and Bland-Altman plots for clinical metric LV_{EF} on EchoNet-Dynamic.

B.2. Visual Comparison with SOTA

We also conducted experiments on EchoNet-Dynamic, and the visualization is shown in Fig. 12.

B.3. Visualization of the Weights

We also conducted experiments on the weights of parameters $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ over training steps on the EchoNet-Dynamic dataset, as shown in Fig. 13.

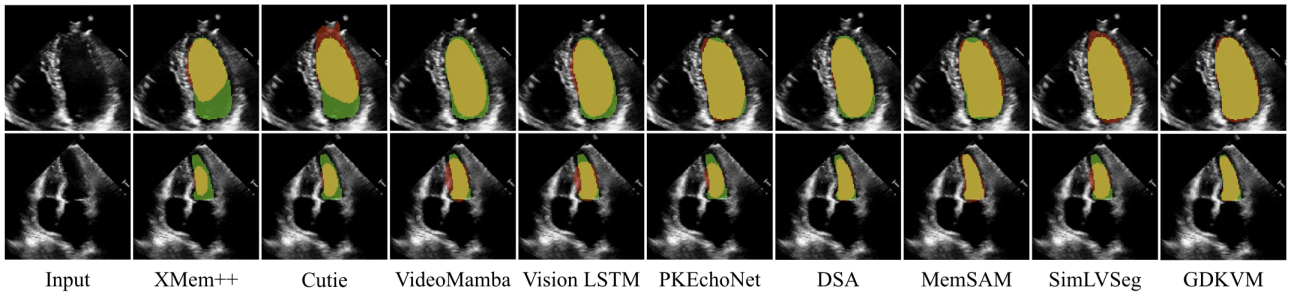
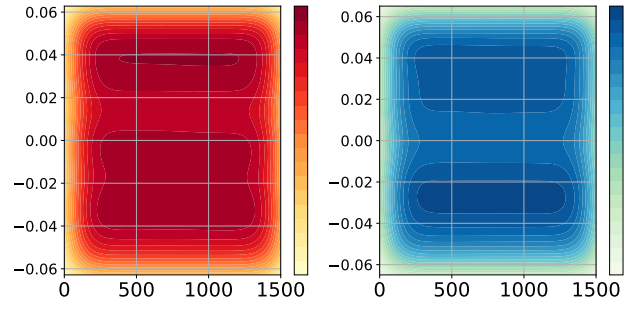
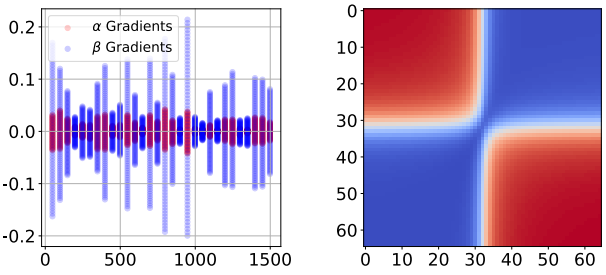


Figure 12. Visual comparison with state-of-the-art methods on the EchoNet-Dynamic test set. Green, red, and yellow regions represent the ground truth, prediction, and overlapping regions, respectively.



(a) Distribution of parameter $\boldsymbol{\alpha}_t$. Darker regions indicate higher density.

(b) Distribution of parameter $\boldsymbol{\beta}_t$. Darker regions indicate higher density.



(c) Distribution of gradient $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$.

(d) Pearson Correlation of gradient $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$.

Figure 13. Weights of parameters $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ over training steps on EchoNet-Dynamic.

These figures show how $\boldsymbol{\alpha}_t$ (controlling memory decay) and $\boldsymbol{\beta}_t$ (balancing old and new information) evolve over training. Their values fluctuate roughly between -0.06 and $+0.06$, and they often become more concentrated or exhibit recurring "hot spots" as training progresses. Such distributions suggest that the model continually adjusts these gating factors to align with changing task demands, eventually reaching relatively stable (or task-optimal) regions.

Although the gradients for these parameters frequently display positive or negative spikes, most gradient values remain within a stable range. This indicates that the network periodically makes significant adjustments to "forget" old information or to "retain" it when needed. Large gradient magnitudes can pose training challenges, highlighting the importance of proper learning rate settings, gradient clipping, or regularization.

Across different value ranges, the gradients of α and β can correlate positively or negatively, forming visually distinct blocks or diagonal patterns. In certain ranges, an increase in α (faster memory decay) may coincide with an increase in β (boosting new information), whereas in other ranges α and β change in opposite directions. These patterns reflect how the network coordinates the interplay between forgetting past data and incorporating new inputs.

First, α and β are learnable rather than fixed, undergoing notable shifts throughout training to accommodate evolving tasks. Second, their correlated gradients reveal how the model dynamically manages memory decay and integration to address varying environments or data patterns. Third, occasional extreme gradient values suggest training instability when balancing "forget–retain" operations, emphasizing the need for methods that mitigate abrupt parameter changes. If further experiments show that final values of α and β strongly influence performance—for instance, by adopting larger α in scenarios requiring rapid forgetting—it would confirm that this learnable gating mechanism indeed supports the flexible discarding of outdated information.