ICCV
OCT 19-23, 2025
HONOLULU
HAWAII

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

Rui Wang[1] Yimu Sun[1] Jingxing Guo[1] Huisi Wu[1*] Jing Qin[2]
[1]College of Computer Science and Software Engineering, Shenzhen University
[2]Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University

Contents

01 | Problem Setting

02 | Method

03 | Experiments

04 | Discussion

Problem Setting

Medical

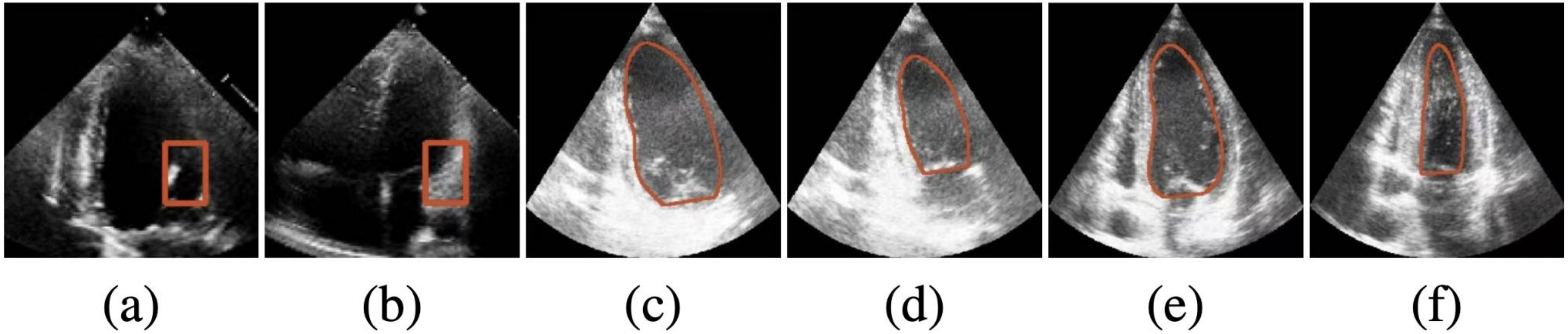
- (a) Speckle Noise, (b) Blurred Contours, and (c-f) Pronounced Variations in the Target’s Morphology Across the Cardiac Cycle

Video

- (a) Extended Temporal Contexts, (b) Efficiency–Accuracy Trade-off in Recall, (c) Computational Burden

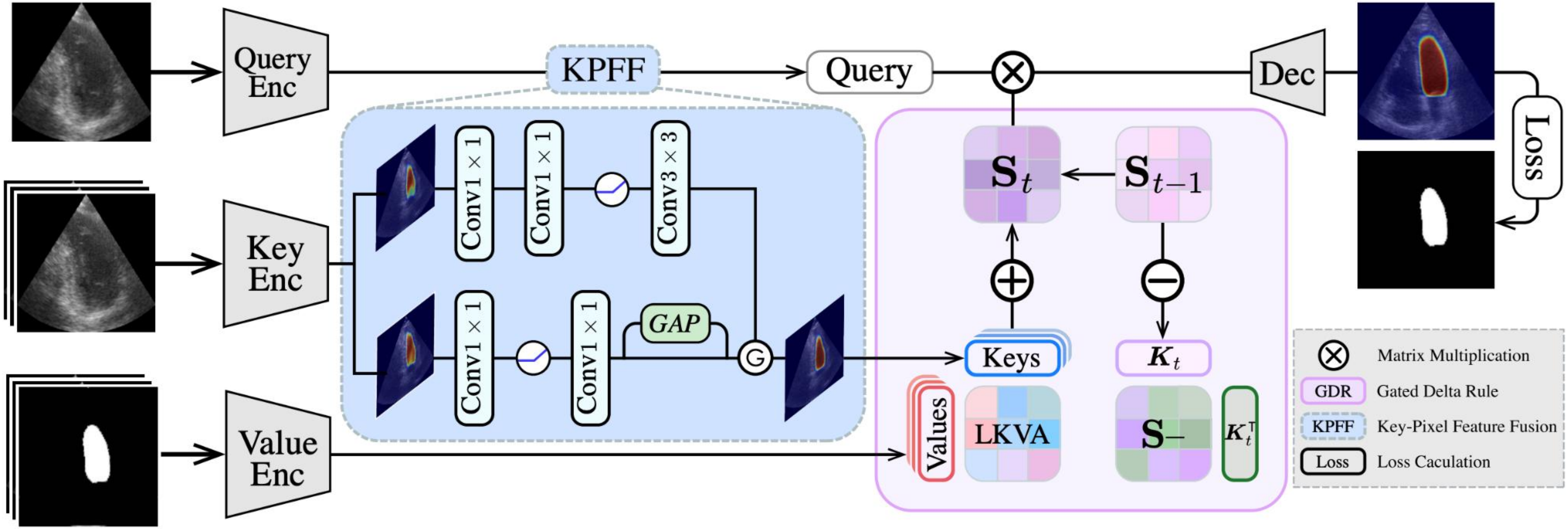
Task

- Clinical Simulation Setting — Absence of Ground Truth at Inference; Training via Boundary-Frame Prediction and Loss Computation



Method

- **Overview**



Method

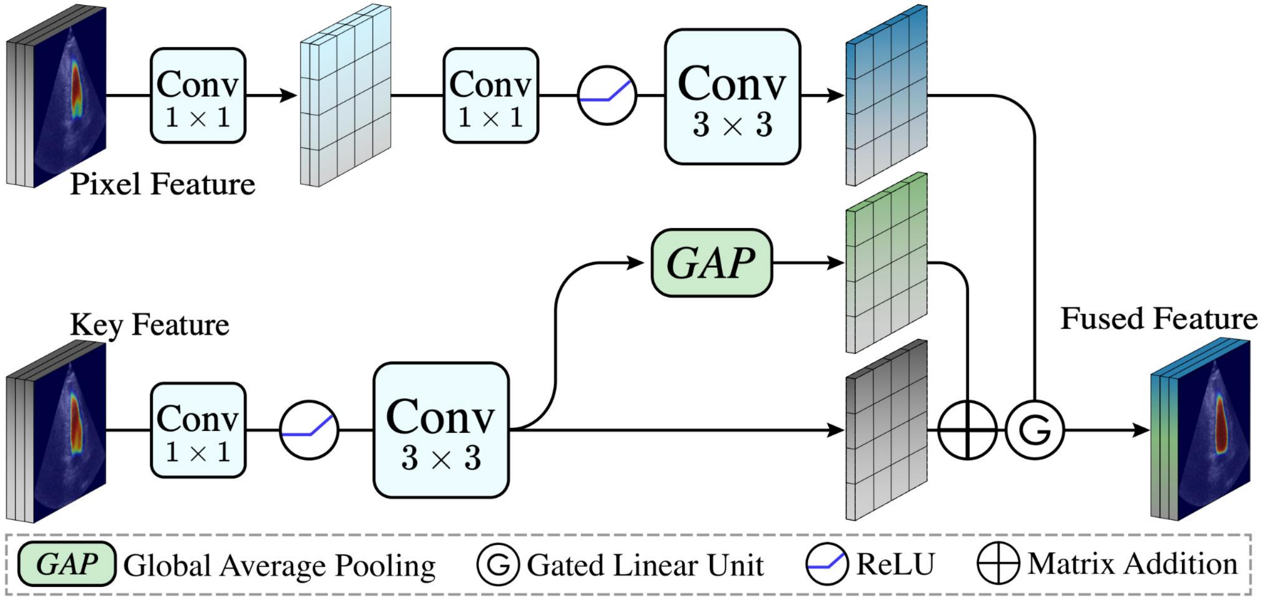
- Linear Key-Value Association

$$\boldsymbol{O}_t = \sum_{i=1}^{t} \frac{\exp(\boldsymbol{K}_i^\intercal \boldsymbol{Q}_t)}{\sum_{j=1}^{t} \exp(\boldsymbol{K}_j^\intercal \boldsymbol{Q}_t)} \boldsymbol{V}_i, \tag{1}$$

$$\begin{aligned}\boldsymbol{O}_t &= \sum_{i=1}^{t} \frac{\phi(\boldsymbol{K}_i)^\intercal \phi(\boldsymbol{Q}_t)}{\sum_{j=1}^{t} \phi(\boldsymbol{K}_j)^\intercal \phi(\boldsymbol{Q}_t)} \boldsymbol{V}_i \\ &= \frac{(\sum_{i=1}^{t} \boldsymbol{V}_i \phi(\boldsymbol{K}_i)^\intercal)\phi(\boldsymbol{Q}_t)}{(\sum_{j=1}^{t} \phi(\boldsymbol{K}_j)^\intercal)\phi(\boldsymbol{Q}_t)} \\ &= \frac{\mathbf{S}_t \phi(\boldsymbol{Q}_t)}{\boldsymbol{Z}_t^\intercal \phi(\boldsymbol{Q}_t)},\end{aligned} \tag{2}$$

$$\begin{aligned}\mathbf{S}_t &= \mathbf{S}_{t-1} + \boldsymbol{V}_t \boldsymbol{K}_t^\intercal \in \mathbb{R}^{C_v \times C_k}, \\ \boldsymbol{O}_t &= \mathbf{S}_t \boldsymbol{K}_t \in \mathbb{R}^{HW \times C_v}.\end{aligned} \tag{3}$$

- Key-Pixel Feature Fusion



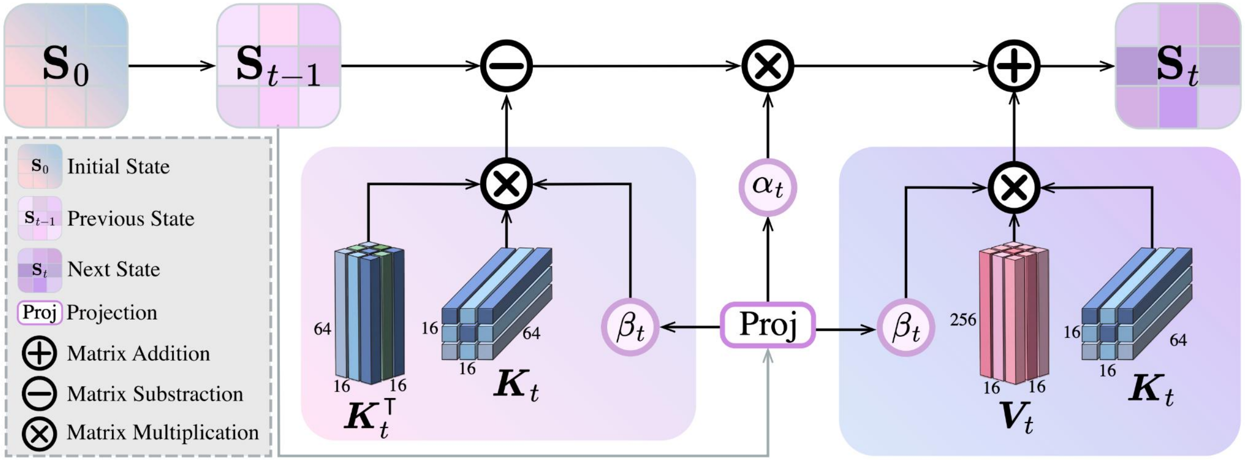
Method

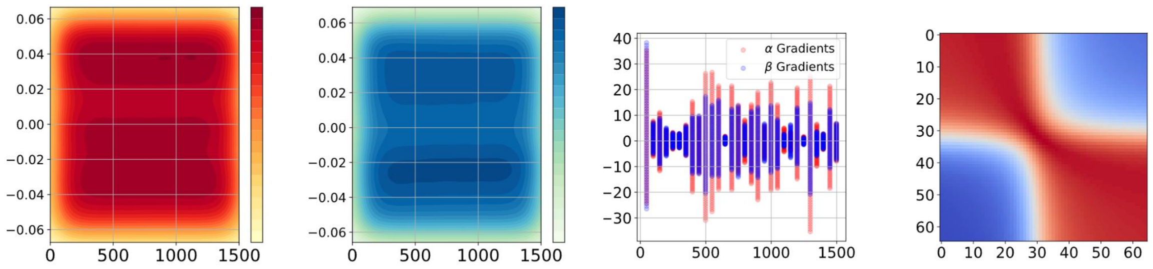
- Gated Delta Rule

$$\begin{aligned}\mathbf{S}_t &= \mathbf{S}_{t-1} \\ &\quad - \underbrace{(\mathbf{S}_{t-1}\boldsymbol{K}_t)}_{\boldsymbol{V}_t^{\text{old}}}\boldsymbol{K}_t^\intercal + \underbrace{(\boldsymbol{\beta}_t\boldsymbol{V}_t + (\mathbf{I}-\boldsymbol{\beta}_t)\,\mathbf{S}_{t-1}\boldsymbol{K}_t)}_{\boldsymbol{V}_t^{\text{new}}}\boldsymbol{K}_t^\intercal \\ &= \mathbf{S}_{t-1}\,(\mathbf{I}-\boldsymbol{\beta}_t\boldsymbol{K}_t\boldsymbol{K}_t^\intercal) + \boldsymbol{\beta}_t\boldsymbol{V}_t\boldsymbol{K}_t^\intercal,\end{aligned} \tag{4}$$

$$\mathbf{S}_t = \mathbf{S}_{t-1}\,(\boldsymbol{\alpha}_t\,(\mathbf{I}-\boldsymbol{\beta}_t\boldsymbol{K}_t\boldsymbol{K}_t^\intercal)) + \boldsymbol{\beta}_t\boldsymbol{V}_t\boldsymbol{K}_t^\intercal. \tag{5}$$

Strategy	Recurrence Equation	mDice	Inf. Time
Baseline	$\mathbf{S}_t = \mathbf{S}_{t-1} + \boldsymbol{V}_t\boldsymbol{K}_t^\intercal$	93.30	151.61 ms
Sanity Check	$\mathbf{S}_t = \mathbf{S}_{t-1} - (\mathbf{S}_{t-1}\boldsymbol{K}_t)\boldsymbol{K}_t^\intercal + \boldsymbol{V}_t\boldsymbol{K}_t^\intercal$	74.68	155.09 ms
w/o α_t	$\mathbf{S}_t = \mathbf{S}_{t-1}\,(\mathbf{I}-\beta_t\boldsymbol{K}_t\boldsymbol{K}_t^\intercal) + \beta_t\boldsymbol{V}_t\boldsymbol{K}_t^\intercal$	94.57	158.77 ms
w/o β_t	$\mathbf{S}_t = \alpha_t\mathbf{S}_{t-1} + \boldsymbol{V}_t\boldsymbol{K}_t^\intercal$	94.26	156.90 ms
GDR	$\mathbf{S}_t = \mathbf{S}_{t-1}\,(\alpha_t\,(\mathbf{I}-\beta_t\boldsymbol{K}_t\boldsymbol{K}_t^\intercal)) + \beta_t\boldsymbol{V}_t\boldsymbol{K}_t^\intercal$	95.11	160.62 ms





(a) Distribution of parameter $\boldsymbol{\alpha}_t$.

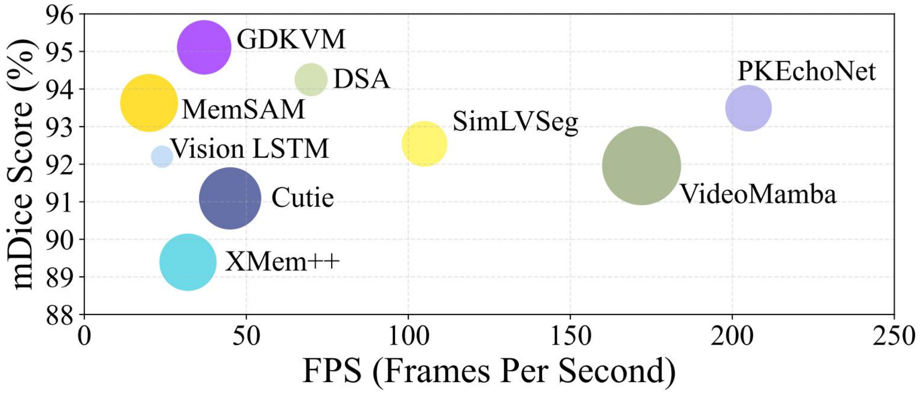
(b) Distribution of parameter $\boldsymbol{\beta}_t$.

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

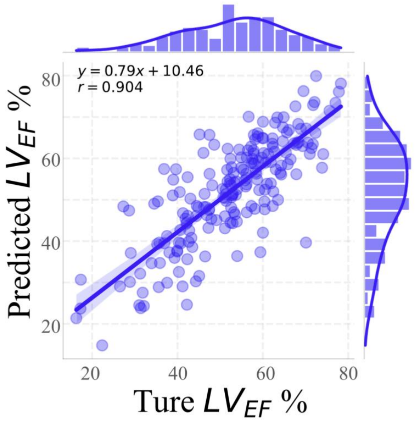
(d) Pearson Correlation of gradient $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$.

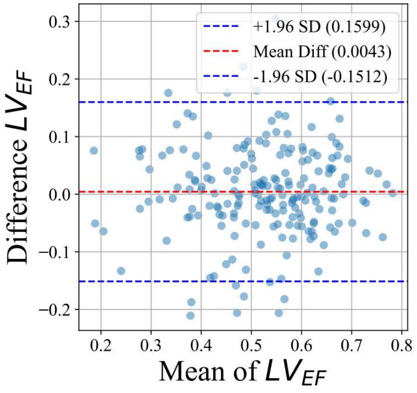
Experiments

Method	Venue & Year	CAMUS				EchoNet-Dynamic			
		mDice	mIoU	HD	ASD	mDice	mIoU	HD	ASD
XMem++ [3]	ICCV'23	89.38	85.81	4.03	4.87	87.51	83.57	3.14	2.69
Cutie [7]	CVPR'24	91.09	87.97	3.89	3.74	88.96	85.63	2.89	2.24
VideoMamba [19]	ECCV'24	91.96	89.04	3.48	3.31	90.22	87.03	2.79	2.05
Vision LSTM [2]	ICLR'25	92.14	89.11	3.79	3.39	90.24	89.14	2.65	1.69
PKEchoNet [40]	AAAI'23	93.49	90.95	3.42	2.93	92.60	89.89	2.53	1.48
DSA [22]	TMI'24	94.25	91.80	3.27	2.37	92.91	90.26	2.46	1.44
MemSAM [10]	CVPR'24	93.63	90.97	3.47	2.60	92.71	89.90	2.56	1.51
SimLVSeg [26]	UMB'24	92.54	89.71	3.65	3.12	91.91	89.08	2.65	1.65
GDKVM	-	**95.11**	**92.97**	**3.05**	**1.98**	**93.46**	**90.86**	**2.38**	**1.36**

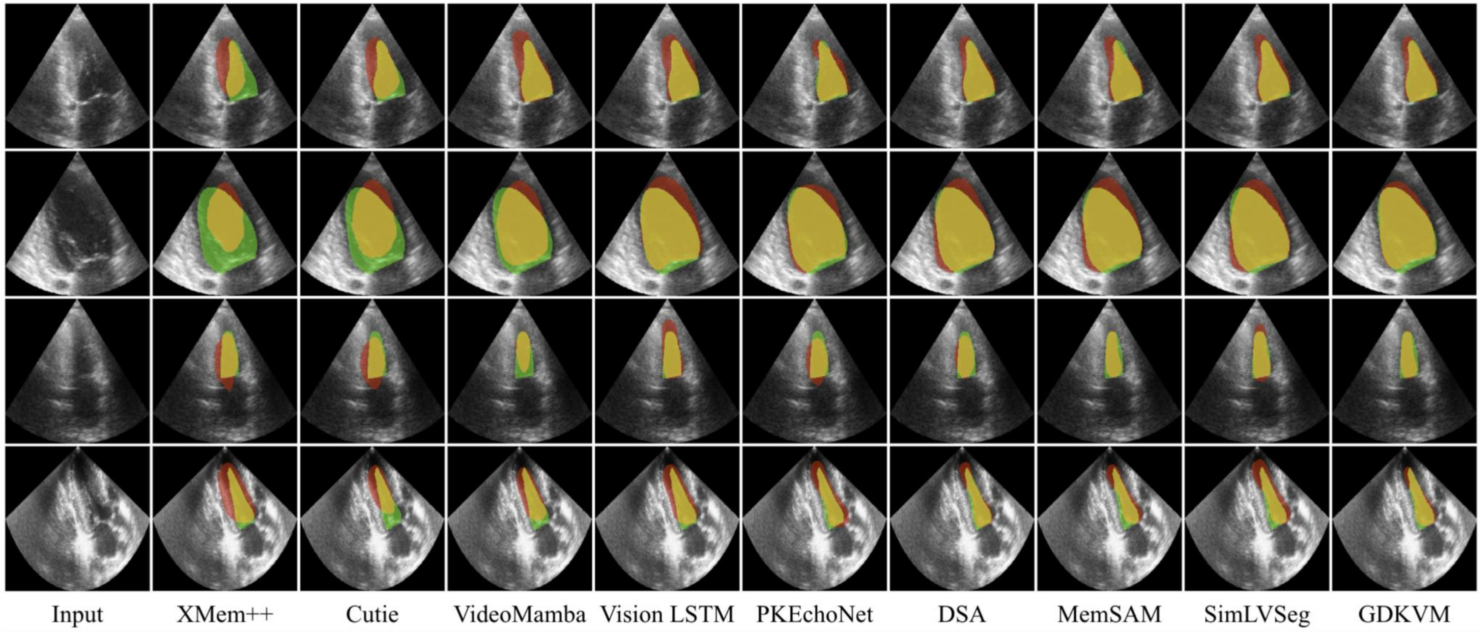


Method	CAMUS	
	corr	bias ± std (%)
XMem++ [3]	0.746	1.70 ± 21.9
Cutie [7]	0.787	1.67 ± 21.7
VideoMamba [19]	0.780	-4.49 ± 19.4
Vision LSTM [2]	0.806	-0.31±18.8
PKEchoNet [40]	0.862	-1.53±16.4
DSA [22]	0.891	0.86±13.4
MemSAM [10]	0.878	-0.89±12.3
SimLVSeg [26]	0.895	1.83±13.8
GDKVM	**0.904**	**-0.19±11.3**

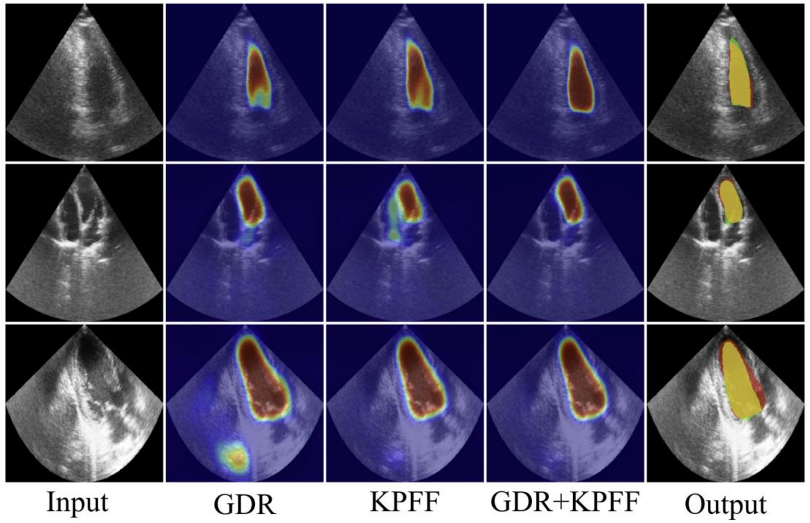




Experiments



LKVA	GDR	KPFF	mDice	mIoU	HD	ASD
✓			93.10	90.46	3.65	2.85
✓	✓		94.49	92.11	3.21	2.19
✓		✓	93.30	90.78	3.55	2.74
✓	✓	✓	**95.11**	**92.97**	**3.05**	**1.98**



Discussion

- **Generality: Extend to broader ultrasound datasets.**
- **Harder video tasks: Tackle longer sequences, complex rhythms, and difficult cases.**
- **Hardware-aware: Optimize the matrix state for parallel acceleration.**

Thanks!