EGMNET: ERROR-FREE AND GLOBAL MEMORY EVOLUTION FOR BREAST ULTRASOUND VIDEO SEGMENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Breast ultrasound video segmentation remains challenging due to severe appearance variations and temporal discontinuity, where lesions may temporarily disappear, change their appearance, or be interfered by newly emerging targets, resulting in inaccurate lesion re-identification and slow segmentation recovery after reappearance. Existing memory-based methods alleviate this issue by storing historical representations, but recursive memory evolution may introduce accumulated update deviations and weaken previously learned lesion information during long-term propagation. Moreover, current-frame-driven memory updating lacks sufficient global temporal awareness, while redundant information in memory state limits representation efficiency. To address these challenges, we propose EGMNet, an efficient memory optimization framework for breast ultrasound video segmentation, consisting of three complementary components. First, we introduce EME, which employs continuous-time dynamics to derive an analytical memory transition process, enabling accurate spatial memory updating with reduced approximation deviation. Second, we develop GMA based on recursive least squares, which formulates memory updating as a projection optimization problem and incorporates global temporal information for adaptive memory evolution. Third, we propose CMC to exploit the intrinsic low-dimensional structure of memory updates, reducing redundant information storage while preserving informative representations. Extensive experiments on the BUS video segmentation dataset demonstrate that EGMNet achieves significant improvements over existing state-of-the-art methods. Besides, The proposed EGMNet provides stronger robustness under lesion disappearance and appearance variation scenarios, enabling rapid target re-capture and more reliable long-term ultrasound video segmentation.

1 INTRODUCTION

Breast ultrasound (BUS) imaging is widely used for breast lesion diagnosis due to its non-invasive, real-time, and low-cost characteristics. Compared with static images, ultrasound videos provide temporal information describing lesion deformation and appearance variations, which is crucial for consistent lesion segmentation. However, speckle noise, low contrast, and large temporal appearance changes make BUS video segmentation highly challenging Noble & Boukerroui (2006). Prior works have demonstrated that temporal modeling plays a vital role in maintaining lesion consistency across frames Li et al. (2022); Su et al. (2024); Kim et al. (2025). While early approaches based on 3D convolution or motion estimation Çiçek et al. (2016); Dosovitskiy et al. (2015) struggled with long-range temporal dependency modeling, memory-based video segmentation methods have alleviated this issue by explicitly storing and retrieving historical representations. For instance, STM introduced the key-value memory paradigm Oh et al. (2019), and subsequent approaches like AOT and XMem further improved long-term association through transformer matching and hierarchical memory mechanisms Yang et al. (2021); Cheng & Schwing (2022). Recently, medical methods such as MemSAM have adapted these memory-guided mechanisms for ultrasound video segmentation Deng et al. (2024). However, existing approaches predominantly focus on memory retrieval and

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

feature aggregation, leaving the mechanism for effective long-term memory evolution insufficiently explored.

To address this, state-space-based approaches have provided efficient solutions for long-sequence modeling. Vision-LSTM and Video-Mamba, further explored recurrent memory state evolution for visual understanding Alkin et al. (2025); Li et al. (2024), while Vivim demonstrated their potential in ultrasound video segmentation Yang et al. (2025). Meanwhile, LiVOS and GDKVM investigated efficient memory updating strategies for video segmentation through gated matching and key-value memory evolution Liu et al. (2025); Wang et al. (2025). Although the Delta-rule effectively improves the adaptability of memory evolution, its recursive updating process still performs iterative memory state refinement rather than directly obtaining the closed-form optimal mapping. Therefore, the updated memory state may inevitably deviate from the ideal solution during long-term temporal propagation. As a result, when ultrasound lesions temporarily disappear, undergo appearance variations, or new targets appear, the model requires additional frames to recover the original lesion representation after its reappearance. Motivated by this observation, we aim to further improve memory evolution by preserving historical information more accurately while maintaining adaptive temporal updating.

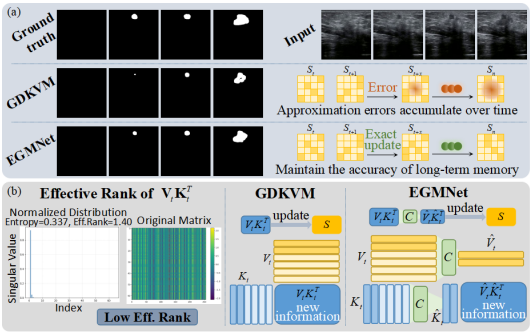


Figure 1: Motivation of EGMNet. (a) Existing methods like GDKVM rely on recursive discrete updates, where approximation errors accumulate over time. In contrast, EGMNet employs an analytical exact update to maintain long-term memory fidelity. (b) Existing methods like GDKVM update memory state using the full key-value interaction matrix $\mathbf{K}_t\mathbf{V}_t^T$. However, this matrix exhibits a low effective rank with high spatial redundancy. EGMNet exploits this property via Compact Memory Compression (CMC) to project information into a compact space, reducing redundancy while preserving essential features.

Furthermore, we observe that the newly introduced information in GDKVM, represented by the key-value interaction matrix $\mathbf{KV}^T$, often exhibits a low effective rank due to the high spatial redundancy and repetitive structures in ultrasound videos. Therefore, the actual information required for memory adaptation is much smaller than the complete key-value interaction space. This observation motivates us to identify informative components from key-value interactions and perform selective memory updating, thereby reducing redundant computation and improving memory representation efficiency.

To address these limitations, we propose EGMNet, a memory optimization framework for breast ultrasound video segmentation. Specifically, we introduce EME, which reformulates memory transition as a continuous dynamical system and analytically solves the memory state evolution under the zero-order hold assumption. This design avoids accumulated recursive deviation and preserves historical lesion representations for more reliable recovery. Furthermore, we propose GMA, which models the key-value memory state as a projection relationship and introduces recursive least squares optimization to capture global temporal correlations without repeatedly processing historical frames. Finally, we develop CMC, which exploits the low effective rank property of key-value interactions and projects redundant tokens into a compact representative space, improving memory efficiency and information density.

The main contributions of this work are summarized as follows:

- We propose a memory optimization framework for breast ultrasound video segmentation that improves long-term temporal consistency under lesion disappearance and appearance variation.
- We introduce Error-Free Memory Evolution (EME), which replaces recursive discrete memory transition with analytical continuous dynamics, enabling accurate and stable historical feature preservation.

- We propose Global Memory Adaptation (GMA) based on recursive least squares optimization to establish globally adaptive key-value memory updating.
- We design Compact Memory Compression (CMC) to exploit the low-rank property of key-value interactions and reduce redundant memory storage.

2 Related Works

Video segmentation requires maintaining consistent object representations under appearance variations and long-term temporal changes. Memory-based methods address this challenge by explicitly storing historical features and retrieving relevant representations during temporal propagation. STM introduced the key-value memory paradigm, which retrieves previous representations through attention matching Oh et al. (2019). XMem further incorporated hierarchical short-term and long-term memory interaction for robust long-video segmentation Cheng et al. (2024), followed by XMem++ and LiVOS, which explored more efficient memory management and lightweight matching Bekuzarov et al. (2023); Liu et al. (2025). More recently, GDKVM introduced a spatiotemporal key-value memory with a Delta-rule-based update mechanism for echocardiography video segmentation, demonstrating the effectiveness of adaptive memory evolution in medical videos Wang et al. (2025). Despite their effectiveness in improving temporal consistency, these approaches primarily rely on discrete memory updates, which may accumulate update deviations during long-term propagation and limit the effective utilization of historical information. Such limitations can become particularly pronounced when targets temporarily disappear, undergo substantial appearance changes, or reappear after occlusion, where reliable recovery depends on the fidelity of the stored representations.

Efficient memory evolution has also been actively investigated in long-sequence modeling. Linear attention and state-space models formulate sequence modeling as recurrent memory-state evolution, enabling efficient temporal representation with reduced computational complexity Gu et al. (2021); Gu & Dao (2023); Katharopoulos et al. (2020). Vision-LSTM and VideoMamba further explored recurrent and selective memory-state evolution for visual understanding Alkin et al. (2025); Li et al. (2024), while Vivim demonstrated the applicability of state-space modeling to ultrasound video segmentation Yang et al. (2025). Beyond these approaches, EFLA introduced a continuous-time formulation for linear attention and derived an analytical solution for memory-state transitions, thereby avoiding additional approximation errors caused by iterative finite-step updates in long-sequence modeling Lei et al. (2025). Although EFLA was not specifically developed for video segmentation, its continuous-time perspective provides a promising direction for improving the accuracy and stability of memory evolution in ultrasound video analysis.

In addition to accurate memory evolution, efficient memory representation is important for reducing redundant updates in long-term video segmentation. Existing key-value memory mechanisms typically maintain dense feature interactions, potentially retaining redundant information arising from repetitive spatial structures. Studies on Robust Principal Component Analysis (RPCA) Candès et al. (2011) and effective rank estimation Roy & Vetterli (2007) suggest that high-dimensional visual observations often exhibit low-dimensional structure, while similar low-rank characteristics have also been observed in transformer attention interaction matrices Bhojanapalli et al. (2020). In breast ultrasound video analysis, substantial spatial and temporal redundancy is likewise present, motivating efforts to emphasize informative temporal regions Lin et al. (2023). These observations suggest that the key-value memory state may contain redundant interactions that can be compressed without substantially sacrificing informative temporal representations. Therefore, jointly improving the fidelity of memory-state evolution and reducing redundant memory updates provides a natural direction toward more accurate and efficient temporal memory modeling for ultrasound video segmentation.

3 Methods

The key encoder uses a ResNet-50 He et al. (2016) backbone to extract Keys, while the value encoder encodes the input frame and previous prediction mask into Values. Keys and Values are integrated into a recurrent memory state. During inference, only previously predicted masks are used without accessing ground-truth annotations.

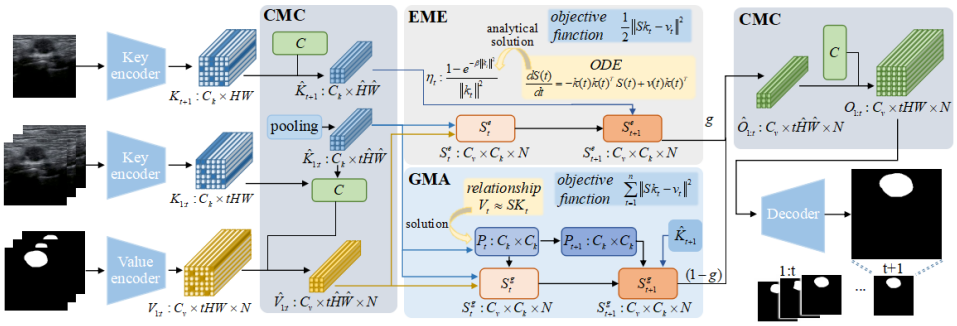


Figure 2: Overview of EGMNet. Input sequences are encoded and compressed into a low-rank memory representation by the CMC. Within the memory bank, the EME models continuous temporal dynamics via an ODE-based analytical solution, while the GMA captures long-range dependencies through global relationship modeling. The aggregated memory is subsequently projected back to the original feature space via CMC.

3.1 Error-Free Memory Evolution

The key-value memory mechanism in GDKVM updates the memory state through Delta-rule, which enables efficient temporal information propagation by dynamically integrating current frame features into historical representations. Specifically, the memory state is recursively updated as:

$$\mathbf{S}_{t+1} = \mathbf{S}_t(\alpha_t(I - \beta_t \mathbf{k}_t \mathbf{k}_t^T)) + \beta_t \mathbf{v_t} \mathbf{k}_t^T, \tag{1}$$

where $\mathbf{S}_t$ denotes the memory state at frame t, $\mathbf{k}_t$ and $\mathbf{v}_t$ represent the corresponding key and value features, α_t is a trainable attenuation matrix for historical information. The first term in the update equation preserves previous memory information, while the second term performs error correction using the current observation. Although the Delta-rule offers an efficient recursive formulation, since each video frame is independently sampled and the memory state transition is approximated by a finite update step β_t that is solely projected from the previous memory state, its discrete approximation of continuous memory evolution accumulates errors over time, which can severely distort historical target features and impair segmentation recovery when the target temporarily disappears.

To address this limitation, we formulate the memory update as an optimization problem that minimizes the reconstruction error of the current observation. The objective function for the memory state $\mathbf{S}$ at time t is defined as:

$$\mathcal{L}_t^e(\mathbf{S}) = \frac{1}{2} \|\mathbf{S}\mathbf{k}_t - \mathbf{v}_t\|^2, \tag{2}$$

where $\mathbf{k}_t$ and $\mathbf{v}_t$ denote the key and value features of the current frame.

To derive a continuous-time formulation from the discrete key-value observations, we employ a Zero-Order Hold (ZOH) Iserles (1996) representation, where each pair of encoded features (k_t, v_t) is treated as a piecewise-constant input over its corresponding memory transition interval. Specifically, for $t \leq \tau < t+1$, the discrete observations are converted into continuous-time inputs as $k(\tau) = k_t, v(\tau) = v_t$. This assumption does not imply that the encoder outputs are intrinsically constant over time; rather, it provides a continuous-time surrogate of the discrete observations, allowing the memory dynamics to be modeled and analytically solved within each transition interval. Under this continuous-time formulation, the evolution of the memory state can be described by an ordinary differential equation (ODE), which characterizes how the memory continuously evolves under the held key-value inputs:

$$\frac{d\mathbf{S}(t)}{dt} = -\mathbf{S}(t)\mathbf{k}(t)\mathbf{k}(t)^T + \mathbf{v}(t)\mathbf{k}(t)^T. \tag{3}$$

Under the ZOH assumption, we regard the discrete memory objective as defining an idealized continuous-time surrogate dynamics. The first term in Eq. 3 describes the forgetting process induced by the current key representation, while the second term represents the incorporation of new

observations into the memory state. By analytically solving Eq. 3, we obtain the exact state transition under the assumed continuous-time dynamics, rather than relying on a finite-step recursive approximation.

The continuous memory dynamics can be analytically solved within each update interval:

$$\mathbf{S}_{t+1} = \mathbf{S}_t e^{-\beta_t \mathbf{k}_t \mathbf{k}_t^T} + \int_0^{\beta_t} \mathbf{v}_t \mathbf{k}_t^T e^{-(\beta_t - \tau)\mathbf{k}_t \mathbf{k}_t^T} d\tau. \tag{4}$$

Since $\mathbf{k}_t \mathbf{k}_t^T$ is a rank-one matrix, the matrix exponential term can be efficiently calculated:

$$e^{-\beta_t \mathbf{k}_t \mathbf{k}_t^T} = \mathbf{I} - \eta_t \mathbf{k}_t \mathbf{k}_t^T, \tag{5}$$

where the adaptive coefficient is obtained directly from the analytical solution:

$$\eta_t = \frac{1 - e^{-\beta_t ||\mathbf{k}_t||^2}}{||\mathbf{k}_t||^2}. \tag{6}$$

Therefore, the final memory update can be expressed as:

$$\mathbf{S}_{t+1}^e = \mathbf{S}_t^e (\mathbf{I} - \eta_t \mathbf{k}_t \mathbf{k}_t^T) + \eta_t \mathbf{v}_t \mathbf{k}_t^T. \tag{7}$$

Under the ZOH-induced continuous dynamics, the analytical solution eliminates the additional numerical discretization error introduced by finite-step approximation. Therefore, the resulting update coefficient η_t is directly determined by this exact solution, enabling the memory state to evolve according to the underlying continuous dynamics without introducing approximation error from numerical discretization.

By integrating this error-free memory evolution (EME), the model can preserve historical tumor representations with higher fidelity. Consequently, when the target disappears temporarily and subsequently reappears, the stored memory provides more reliable temporal cues for segmentation recovery.

3.2 Global Memory Adaptation

The EME introduced above provides a stable memory transition by analytically solving the continuous dynamics. However, the update magnitude is still determined by the current key representation, which mainly reflects local frame information and lacks the ability to adaptively capture global temporal statistics.

We treat the spatial memory state as a projection matrix that establishes the mapping relationship between key and value spaces, formulated as$\mathbf{V} \approx \mathbf{SK}$, where $\mathbf{S}$ represents the latent projection from key space to value space. Therefore, instead of only adjusting the memory according to the current observation, we aim to optimize $\mathbf{S}$ by considering all historical key-value correspondences. To achieve efficient online updating without repeatedly processing the complete video history, we introduce Recursive Least Squares (RLS) to obtain a globally adaptive memory evolution strategy.

Given historical key-value pairs $\{(\mathbf{k}_t, \mathbf{v}_t)\}_{t=1}^n$, we formulate the memory construction as an unconstrained linear least squares problem:

$$\mathcal{L}_n^g(\mathbf{S}) = \sum_{t=1}^{n} \|\mathbf{S}\mathbf{k}_t - \mathbf{v}_t\|^2. \tag{8}$$

By minimizing the accumulated projection error and solving the corresponding normal equation, the optimal memory projection matrix is obtained as:

$$\mathbf{S} = \left(\sum_{tf=1}^{n} \mathbf{v}_t \mathbf{k}_t^T \right) \left(\sum_{t=1}^{n} \mathbf{k}_t \mathbf{k}_t^T \right)^{-1}. \tag{9}$$

This solution indicates that the memory state is determined by the global correlation between key and value spaces rather than simple feature accumulation.

However, directly recomputing Eq. 9 requires storing all historical tokens. Moreover, when the number of accumulated tokens becomes large, the contribution of each individual token is gradually diluted. Therefore, we adopt Recursive Least Squares Benesty et al. (2011) to update the projection matrix online. The inverse correlation matrix is defined as:

$$\mathbf{P}_t = \left(\sum_{t=1}^{n} \mathbf{k}_t \mathbf{k}_t^T \right)^{-1}. \tag{10}$$

When a new key-value pair arrives, the covariance matrix is updated using the Sherman-Morrison formula:

$$\mathbf{P}_{t+1} = \mathbf{P}_t - \frac{\mathbf{P}_t \mathbf{k}_{t+1} \mathbf{k}_{t+1}^T \mathbf{P}_t}{1 + \mathbf{k}_{t+1}^T \mathbf{P}_t \mathbf{k}_{t+1}}. \tag{11}$$

Based on the updated covariance estimation, the memory state can be recursively optimized:

$$\mathbf{S}_{t+1}^g = \mathbf{S}_t^g + \beta_t^g (\mathbf{v}_{t+1} - \mathbf{S}_t^g \mathbf{k}_{t+1}) \mathbf{k}_{t+1}^T \mathbf{P}_{t+1}, \tag{12}$$

where β_t^g controls the update intensity.

To fully exploit parallel computation for practical video segmentation, we further rewrite the recursive optimization into a matrix-based memory update. Given the current key and value matrices $\mathbf{K} \in \mathbb{R}^{d_K \times n}$ and $\mathbf{V} \in \mathbb{R}^{d_V \times n}$, the GMA updates the memory state as follows:

$$\mathbf{S}_{t+1}^g = \mathbf{S}_t^g + \beta_t^g (\mathbf{V}_{t+1} - \mathbf{S}_t^g \mathbf{K}_{t+1}) \mathbf{K}_{t+1}^T \mathbf{P}_{t+1}, \tag{13}$$

$$\mathbf{P}_{t+1} = \left(\mathbf{P}_t^{-1} + \mathbf{K}_{t+1} \mathbf{K}_{t+1}^T \right)^{-1}. \tag{14}$$

We set the initial state $\mathbf{P}_0 = \epsilon \mathbf{I}$, with $\mathbf{I}$ being the identity matrix and ϵ a small positive scalar, to ensure that $\mathbf{P}_t$ remains non-singular and invertible in subsequent steps.

Finally, the local EME branch and the global GMA memory branch are fused through a learnable gate:

$$\mathbf{O}_t = g \mathbf{S}_{t+1}^e \mathbf{K}_{t+1} + (1 - g) \mathbf{S}_{t+1}^g \mathbf{K}_{t+1}, \tag{15}$$

where $\mathbf{S}_{t+1}^e$ and $\mathbf{S}_{t+1}^g$ represent the memory state updated by EME and GMA respectively, and $g \in [0, 1]$ balances the contribution between EME and GMA. The EME branch preserves precise frame-level correspondence, while the GMA branch captures global temporal evolution. Their complementary interaction enables robust tumor segmentation under long-term appearance variation.

Overall, the proposed GMA complements the EME introduced previously. The EME branch focuses on preserving accurate local temporal transitions, while the GMA branch captures global temporal correlations beyond individual frames. Their combination enables the memory state to simultaneously maintain local precision and long-term stability, which is essential for robust tumor segmentation under complex ultrasound video variations.

3.3 Compact Memory Compression

Although the proposed EME and GMA improve the temporal stability and representation capability of the memory state, the storage efficiency is still limited by redundant key-value interactions. Specifically, the memory state relies on the key-value interaction matrix $\mathbf{M} = \mathbf{V}\mathbf{K}^T$ stored in the memory to update the memory representations. However, due to the spatial redundancy and repetitive appearance patterns in ultrasound videos, the effective rank of $\mathbf{M}$ is usually much smaller than its original dimensionality. This indicates that only a small number of representative feature directions contain the dominant information, while a considerable amount of token-level interactions introduce redundant information into the memory state. Therefore, directly maintaining all key-value interactions not only increases computational costs but also reduces the information density of stored representations. To address this issue, we propose compact memory compression (CMC), construct a compact memory representation by projecting the original token space into a representative subspace, thereby preserving the dominant key-value correlations while removing redundant feature interactions.

According to the low-rank property of feature correlation matrices, the newly introduced key-value interaction can be represented in a compact token space. Specifically, we introduce a contraction matrix $\mathbf{C} \in \mathbb{R}^{m \times n}$, where n denotes the original number of spatial tokens and $m \ll n$ represents the number of representative tokens. The matrix $\mathbf{C}$ aggregates highly correlated tokens and projects the original key $\mathbf{K}$ and value $\mathbf{V}$ features into a compact space:

$$\hat{\mathbf{K}} = \mathbf{C}\mathbf{K}, \quad \hat{\mathbf{V}} = \mathbf{C}\mathbf{V}. \tag{16}$$

Compared with the original key $\mathbf{K}$ and value $\mathbf{V}$ features, the compact representation preserves the dominant correlation structure through representative tokens while suppressing redundant token-level interactions and irrelevant noise components. Therefore, the compressed key-value pairs $\left(\hat{\mathbf{K}}, \hat{\mathbf{V}}\right)$ are directly fed into the memory evolution branches, where EME and GMA update the memory state in the compact token space. Finally, the updated memory representations are jointly aggregated and projected back to the original token space through the expansion transformation:

$$\mathbf{O}_t = \mathbf{C}^T \left(g\mathbf{S}^e_{t+1}\hat{\mathbf{K}}_{t+1} + (1-g)\mathbf{S}^g_{t+1}\hat{\mathbf{K}}_{t+1} \right) = \mathbf{C}^T \hat{\mathbf{O}}_t, \tag{17}$$

By introducing this compact representation, the proposed module reduces the number of tokens involved in memory evolution, thereby decreasing computational cost and storage overhead. Meanwhile, removing redundant and noisy information improves the feature density of memory state, allowing subsequent memory updates to focus on more informative temporal representations.

In practice, the contraction matrix $\mathbf{C}$ can be constructed either as a learnable parameter matrix or in a data-dependent manner. The former is jointly optimized through backpropagation, while the latter is obtained by downsampling the key features $\mathbf{K}$ into representative tokens $\mathbf{K} * down \in \mathbb{R}^{d \times m}$ and computing $\mathbf{C} = \mathbf{K} * down^T \mathbf{K}$. In our experiments, we use $m = 2 \times 2$, retaining only four representative vectors. Both strategies achieve comparable performance; we therefore adopt the downsampling-based construction to provide content-aware compression without introducing additional learnable parameters.

By operating in a compact token space, CMC reduces redundant memory representations and computational cost while preserving dominant temporal patterns in the key-value correlation space, thereby improving memory efficiency and supporting more reliable tumor representation recovery under appearance variations and temporary target disappearance.

4 Experiments

4.1 Datasets

We evaluate the effectiveness of EGMNet on two breast ultrasound video segmentation datasets: a large-scale private dataset, BUV-1K, and the public BUV2022 benchmark Li et al. (2022). BUV-1K comprises 1,085 videos containing 50,375 frames at a resolution of 256×256 with expert pixel-level annotations, and it is randomly split into training, validation, and testing sets with ratios of 60%, 20%, and 20%, respectively, with the splitting performed at the patient level to prevent data from the same patient from appearing across different subsets. The BUV2022 dataset contains 63 video sequences with 4,619 frames and pixel-level expert annotations, and we follow its official split to ensure fair comparison with previous methods.

4.2 Evaluation Metrics

To comprehensively evaluate the segmentation performance, we adopt five widely used metrics for medical image and video segmentation, including Region Similarity ($\mathcal{J}$), Contour Accuracy ($\mathcal{F}$), Dice Similarity Coefficient (Dice), 95th percentile Hausdorff Distance (HD95), and Average Surface Distance (ASD). Specifically, $\mathcal{J}$ and Dice measure the overlap between the predicted segmentation and ground truth mask, reflecting the region-level segmentation accuracy. In contrast, $\mathcal{F}$, HD95, and ASD evaluate the boundary quality and geometric consistency between prediction and annotation, which are particularly important for breast lesions with irregular shapes and ambiguous boundaries.

Method	BUV-1K Dataset					BUV2022 Dataset					Efficiency	
	Dice↑	$\mathcal{J}\uparrow$	$\mathcal{F}\uparrow$	HD95↓	ASD↓	Dice↑	$\mathcal{J}\uparrow$	$\mathcal{F}\uparrow$	HD95↓	ASD↓	Mem.↓	FPS↑
DPSTT	73.41	66.25	58.17	20.99	8.06	59.35	48.48	33.91	50.16	18.66	626	61.05
FLA-Net	70.68	64.06	57.01	25.48	10.62	68.82	59.64	44.62	32.90	13.00	1596	102.76
Bas	69.31	56.73	31.20	44.37	15.88	69.77	57.16	30.22	66.22	24.30	484	65.36
Vision LSTM	61.62	54.56	48.83	33.89	16.18	66.85	54.35	31.62	87.59	30.32	1580	5.45
Cutie	69.51	61.96	54.13	21.50	8.42	73.70	62.65	48.49	35.45	11.74	853	39.91
LiVOS	75.52	68.69	62.26	17.52	6.51	76.13	66.10	51.02	33.95	12.71	1122	50.74
GDKVM	75.35	68.55	62.00	16.96	6.62	72.67	61.84	47.97	44.15	17.16	1118	53.53
EGMNet	**77.89**	**71.24**	**64.70**	**15.83**	**5.59**	**77.40**	**67.71**	**51.44**	**30.66**	**11.53**	980	54.44

Table 1: Quantitative comparison on the BUV-1K and BUV2022 datasets. FPS and Mem. represent inference speed and GPU memory usage (MB), respectively. The segmentation performance is highlighted in bold.

4.3 Experimental Results Comparison

Our evaluation benchmarks EGMNet against state-of-the-art methods, ranging from medical-specific models like DPSTT Li et al. (2022), FLA-Net Lin et al. (2023), Bas Pang et al. (2025) and GDKVM Wang et al. (2025) to advanced general video segmentation approaches such as Vision LSTM Alkin et al. (2025), Cutie Cheng et al. (2024), and LiVOS Liu et al. (2025).

Table 1 presents the quantitative comparison on the BUV-1K and BUV2022 datasets. EGMNet achieves the best performance on all five metrics on BUV-1K, with HD95 and ASD reduced by 1.13 and 0.92, respectively, compared with the second-best results. On BUV2022, EGMNet also outperforms previous memory-based methods, improving Dice by 4.73% percentage points and reducing HD95 by 13.49 compared with GDKVM. Despite the additional overhead of the dual-branch memory design, EGMNet achieves higher inference speed and lower GPU memory consumption than GDKVM, benefiting from the compact memory representation provided by CMC.

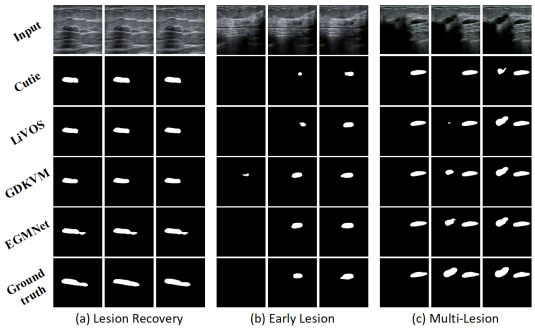


Figure 3: Visual comparison with state-of-the-art methods on the BUV-1K test set.

A visual comparison of EGMNet with other methods is presented in Figure 3. From left to right, (a) illustrates lesion recovery under temporary target disappearance or appearance variation, (b) shows the segmentation performance during the early stage of lesion emergence, and (c) demonstrates the capability of handling multiple lesions simultaneously. Compared with other methods, EGMNet provides more consistent and accurate lesion boundaries. Benefiting from error-free memory evolution and global temporal adaptation, our method can preserve historical lesion representations more effectively, enabling rapid recovery when previously observed lesions reappear.

4.4 Memory Stability Analysis

Figs. 4 further evaluate robustness against Gaussian and speckle noise. Unlike the steep degradation in GDKVM and LiVOS, EGMNet exhibits a much gentler performance drop while maintaining a substantial lead over the baselines. This stability confirms that CMC improves feature density by removing irrelevant token interactions, while EME's analytical solution and GMA's global projection prevent error amplification under corrupted inputs, ensuring reliability in clinical scenarios.As shown in Figure 5, all models maintain comparable Dice scores during the clean initial segment. While the models exhibit comparable performance in the initial clean segment (A1), GDKVM and LiVOS show noticeable performance degradation and fluctuations after composite noise is introduced in the middle segment. In contrast, EGMNet maintains a relatively stable Dice trajectory with a smaller performance decline. After the composite noise subsides, EGMNet gradually recovers its

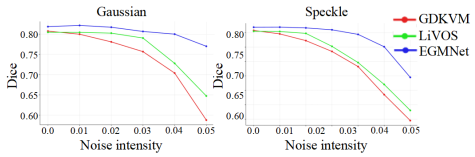


Figure 4: Robustness evaluation under varying intensities of additive Gaussian and multiplicative speckle noise.

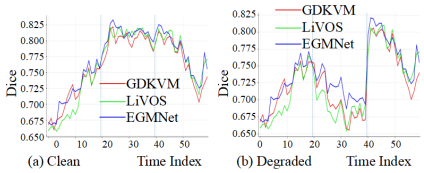


Figure 5: Memory stability analysis on the BUV-1K dataset. (a) Temporal Dice curves on clean sequences. (b) Temporal Dice curves when the middle segment is perturbed with BUS-related degradations.

Variant	Dice↑	$\mathcal{J}\uparrow$	HD95↓	ASD↓
EME-only	76.52	69.75	16.65	5.94
GMA-only	76.25	69.52	16.55	6.19
w/o EME	76.80	70.21	15.56	5.60
w/o GMA	76.58	69.93	16.52	5.99
w/o CMC	77.22	70.71	16.43	5.90
Full	77.89	71.24	15.83	5.59

Table 2: Ablation study of different configurations on the BUV-1K dataset.

Setting	Dice↑	$\mathcal{J}\uparrow$	$\mathcal{F}\uparrow$	HD95↓	ASD↓
2×2	77.89	71.24	64.70	15.83	5.59
4×4	77.57	70.92	64.52	16.44	5.76
6×6	77.80	70.96	64.14	16.26	5.68

Table 3: Ablation study of different compression resolutions in CMC. Each setting represents the spatial resolution of the feature map after applying CMC compression.

segmentation performance in the subsequent clean segment (A2), with recovery occurring slightly earlier than the competing methods. This demonstrates that EGMNet can maintain target recognition under composite noise and recover the target representation shortly after the interference is removed.

4.5 Ablation Study

To evaluate the effectiveness of each component in EGMNet, we conduct comprehensive ablation studies, as shown in Table 4 and Table 3. Removing EME decreases segmentation performance, indicating that error-free memory evolution effectively reduces accumulated updating deviations and preserves reliable lesion representations over time. Similarly, the degradation observed without GMA highlights the importance of global memory adaptation, where the GMA enhances long-range temporal correlation modeling and improves lesion recovery under appearance variations. Notably, removing the CMC primarily degrades region overlap performance, demonstrating its unique capability to suppress redundant spatial information and ultrasound noise while retaining discriminative lesion features. Further analysis of CMC variants with different compression resolutions confirms that removing redundant background responses consistently improves memory representation quality, providing an trade-off between compact memory storage and segmentation accuracy. Overall, the full EGMNet achieves consistent improvements across both overlap and boundary metrics, verifying that accurate memory evolution, global temporal adaptation, and compact representation jointly enhance segmentation robustness.

5 Conclusion

This paper proposes EGMNet, an efficient memory optimization framework for breast ultrasound video segmentation, addressing the limitations of existing methods including accumulated discretization errors, insufficient global temporal awareness, and redundant storage via three key components: EME, deriving an analytical ODE solution to eliminate recursive update errors; GMA, employing recursive least squares for globally adaptive memory evolution; and CMC, exploiting low-rank property to compress redundant key-value interactions. Experiments demonstrate that EGMNet achieves superior robustness, rapidly adapting to sudden lesion appearance changes and maintaining high segmentation accuracy under various noise corruptions, highlighting its potential for reliable clinical ultrasound analysis.

6 ETHICS STATEMENT

The publicly available dataset "BUV2022Dataset" used in this work was obtained from open-access sources, and its usage strictly follows the terms and conditions set by the dataset provider, adhering to established academic and research community standards. For the private dataset "BUV-1KDataset," all patient privacy information was removed during preprocessing, and data splits were performed at the patient level to prevent data leakage. The collection and usage of this dataset comply with relevant ethical guidelines and institutional review board requirements. No personally identifiable information or sensitive patient data is involved in any dataset used in this work.

7 REPRODUCIBILITY STATEMENT

We have taken steps to ensure our results are reproducible. All model and algorithmic details, training procedures, hyperparameters, evaluation protocols, and metrics are specified in the main text. The appendix provides complete implementation notes, ablation studies, and additional qualitative results. An anonymized GitHub repository containing the source code and configuration files will be made publicly available upon publication to facilitate reproduction of our results.

REFERENCES

Benedikt Alkin, Maximilian Beck, Korbinian Pöppel, Sepp Hochreiter, and Johannes Brandstetter. Vision-lstm: xlstm as generic vision backbone. In *International Conference on Learning Representations*, volume 2025, pp. 23804–23826, 2025.

Maksym Bekuzarov, Ariana Bermudez, Joon-Young Lee, and Hao Li. Xmem++: Production-level video segmentation from few annotated frames. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 635–644, 2023.

Jacob Benesty, Constantin Paleologu, Tomas Gänsler, and Silviu Ciochină. Recursive least-squares algorithms. In *A perspective on stereophonic acoustic echo cancellation*, pp. 63–69. Springer, 2011.

Srinadh Bhojanapalli, Chulhee Yun, Ankit Singh Rawat, Sashank Reddi, and Sanjiv Kumar. Low-rank bottleneck in multi-head attention models. In *International conference on machine learning*, pp. 864–873. PMLR, 2020.

Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37, 2011.

Ho Kei Cheng and Alexander G Schwing. Xmem: Long-term video object segmentation with an atkinson-shiffrin memory model. In *European conference on computer vision*, pp. 640–658. Springer, 2022.

Ho Kei Cheng, Seoung Wug Oh, Brian Price, Joon-Young Lee, and Alexander Schwing. Putting the object back into video object segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3151–3161, 2024.

Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pp. 424–432. Springer, 2016.

Xiaolong Deng, Huisi Wu, Runhao Zeng, and Jing Qin. Memsam: Taming segment anything model for echocardiography video segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9622–9631, June 2024.

Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pp. 2758–2766, 2015.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Arieh Iserles. *A first course in the numerical analysis of differential equations*. Cambridge University Press, 1996.

Angelos Katharopoulos, Apoorv Vyas, Nikolaos Pappas, and François Fleuret. Transformers are rnns: Fast autoregressive transformers with linear attention. In *International conference on machine learning*, pp. 5156–5165. PMLR, 2020.

Sekeun Kim, Pengfei Jin, Cheng Chen, Kyungsang Kim, Zhiliang Lyu, Hui Ren, Sunghwan Kim, Zhengliang Liu, Aoxiao Zhong, Tianming Liu, et al. Medivista: Medical video segmentation via temporal fusion sam adaptation for echocardiography. *IEEE Journal of Biomedical and Health Informatics*, 2025.

Jingdi Lei, Di Zhang, and Soujanya Poria. Error-free linear attention is a free lunch: Exact solution from continuous-time dynamics. *arXiv preprint arXiv:2512.12602*, 2025.

Jialu Li, Qingqing Zheng, Mingshuang Li, Ping Liu, Qiong Wang, Litao Sun, and Lei Zhu. Rethinking breast lesion segmentation in ultrasound: A new video dataset and a baseline network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 391–400. Springer, 2022.

Kunchang Li, Xinhao Li, Yi Wang, Yinan He, Yali Wang, Limin Wang, and Yu Qiao. Videomamba: State space model for efficient video understanding. In *European conference on computer vision*, pp. 237–255. Springer, 2024.

Junhao Lin, Qian Dai, Lei Zhu, Huazhu Fu, Qiong Wang, Weibin Li, Wenhao Rao, Xiaoyang Huang, and Liansheng Wang. Shifting more attention to breast lesion segmentation in ultrasound videos. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 497–507. Springer, 2023.

Qin Liu, Jianfeng Wang, Zhengyuan Yang, Linjie Li, Kevin Lin, Marc Niethammer, and Lijuan Wang. Livos: Light video object segmentation with gated linear matching. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 8668–8678, 2025.

J Alison Noble and Djamal Boukerroui. Ultrasound image segmentation: a survey. *IEEE Transactions on medical imaging*, 25(8):987–1010, 2006.

Seoung Wug Oh, Joon-Young Lee, Ning Xu, and Seon Joo Kim. Video object segmentation using space-time memory networks. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9226–9235, 2019.

Yan Pang, Yunhao Li, Teng Huang, Jiaming Liang, Ziyu Ding, Hao Chen, Baoliang Zhao, Ying Hu, Zheng Zhang, and Qiong Wang. Efficient breast lesion segmentation from ultrasound videos across multiple source-limited platforms. *IEEE Journal of Biomedical and Health Informatics*, 2025.

Olivier Roy and Martin Vetterli. The effective rank: A measure of effective dimensionality. In *2007 15th European signal processing conference*, pp. 606–610. IEEE, 2007.

Chenkai Su, Yuxiang Zhou, Jinlian Ma, Haoyu Chi, Xin Jing, Junyan Jiao, and Qiqi Yan. Janet: A joint attention network for balancing accuracy and speed in left ventricular ultrasound video segmentation. *Computers in Biology and Medicine*, 169:107856, 2024.

Rui Wang, Yimu Sun, Jingxing Guo, Huisi Wu, and Jing Qin. Gdkvm: Echocardiography video segmentation via spatiotemporal key-value memory with gated delta rule. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12191–12200, 2025.

Yijun Yang, Zhaohu Xing, Lequan Yu, Huazhu Fu, Chunwang Huang, and Lei Zhu. Vivim: a video vision mamba for ultrasound video segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.

Zongxin Yang, Yunchao Wei, and Yi Yang. Associating objects with transformers for video object segmentation. *Advances in neural information processing systems*, 34:2491–2502, 2021.

8 Supplementary Material

8.1 Abstract

This supplementary material provides additional implementation details, theoretical analysis, and experimental evaluations of EGMNet for ultrasound video segmentation. We first describe the training configurations and data augmentation strategies used in all experiments. Then, we present an efficient implementation of Global Memory Adaptation (GMA) by reformulating the covariance update with the Woodbury identity and Cholesky decomposition, which reduces computational overhead and improves numerical stability during long-term memory evolution. Furthermore, we provide detailed noise generation procedures, additional ablation studies, and visualization results to further analyze the robustness and effectiveness of EGMNet under challenging ultrasound conditions.

8.2 Implementation Details

All experiments are conducted on an NVIDIA RTX 4090 GPU. The model is optimized using the AdamW optimizer with a learning rate of 1×10^{-4}. We train the model using 4,000 training items with a batch size of 5 and apply gradient norm clipping with a maximum value of 3 to improve training stability. During training, standard data augmentation strategies, including random rotation, scaling, contrast adjustment, and gamma augmentation, are randomly applied with a probability of 0.5.

8.3 Implementation Details of GMA

In the original formulation of GMA, the covariance matrix is updated as:

$$\mathbf{P}_{t+1} = \left(\mathbf{P}_t^{-1} + \mathbf{K}_{t+1}\mathbf{K}_{t+1}^T\right)^{-1}, \tag{18}$$

where $\mathbf{P}_{t+1}$ models the global statistics of historical key distributions. However, directly implementing Eq. 18 requires the inversion of a $d_K \times d_K$ matrix, resulting in a computational complexity of $\mathcal{O}(d_K^3)$. Since the proposed CMC module compresses the key representation into a smaller number of informative tokens, where the number of updated tokens n is usually much smaller than the key dimension ($n \ll d_K$), we derive an equivalent token-space formulation to exploit this compact property.

Using the Woodbury matrix identity, Eq. 18 can be reformulated as:

$$\mathbf{P}_{t+1} = \mathbf{P}_t - \mathbf{G}_{t+1}(\mathbf{I} + \mathbf{K}_{t+1}^T\mathbf{G}_{t+1})^{-1}\mathbf{G}_{t+1}^T, \tag{19}$$

$$\mathbf{G}_{t+1} = \mathbf{P}_t\mathbf{K}_{t+1}. \tag{20}$$

The inverse operation is therefore transferred from the original $d_K \times d_K$ covariance space to the compact token space $\mathbf{I} + \mathbf{K}_{t+1}^T\mathbf{G}_{t+1} \in \mathbb{R}^{n\times n}$. To further improve numerical stability, we solve this system through Cholesky decomposition:

$$\mathbf{I} + \mathbf{K}_{t+1}^T \mathbf{G}_{t+1} = \mathbf{L}_{t+1} \mathbf{L}_{t+1}^T, \tag{21}$$

followed by the triangular solution:

$$\mathbf{L}_{t+1} \mathbf{Y}_{t+1} = \mathbf{G}_{t+1}^T. \tag{22}$$

The covariance update is finally computed as:

$$\mathbf{P}_{t+1} = \mathbf{P}_t - \mathbf{Y}_{t+1}^T \mathbf{Y}_{t+1}. \tag{23}$$

This equivalent formulation preserves the original optimization objective while avoiding explicit high-dimensional matrix inversion. Specifically, the direct implementation has a complexity of $\mathcal{O}(d_K^3)$, whereas the Cholesky-based formulation reduces the dominant cost to approximately $\mathcal{O}(d_K n^2 + n^3)$ by operating on the compressed token space. Therefore, when the number of updated tokens is significantly smaller than the key dimension, the proposed formulation provides lower computational overhead. Conversely, for cases with large token numbers, the original covariance update can still be preferable.

Beyond computational efficiency, the proposed Cholesky-based solver also improves numerical stability during long-term memory evolution. By decomposing $\mathbf{I} + \mathbf{K}_{t+1}^T \mathbf{G}_{t+1}$ into $\mathbf{L}_{t+1} \mathbf{L}_{t+1}^T$, the correction term $\mathbf{Y}_{t+1}^T \mathbf{Y}_{t+1}$ is guaranteed to be symmetric positive semi-definite, which preserves the positive definiteness of the covariance matrix and ensures a valid inverse covariance representation throughout online updates. Moreover, solving the triangular system $\mathbf{L}_{t+1} \mathbf{Y}_{t+1} = \mathbf{G}_{t+1}^T$ avoids explicit matrix inversion and reduces the accumulation of floating-point errors, which is particularly beneficial for long ultrasound video sequences.

Both implementations maintain the same theoretical GMA update, and the choice can be determined according to the relationship between n and d_K. In EGMNet, we adopt the Woodbury-based Cholesky solver for covariance update. Benefiting from CMC, the number of updated tokens is effectively reduced, making the condition $n \ll d_K$ easier to hold. Consequently, the proposed solver reduces computational overhead while providing a more efficient, numerically stable, and reliable covariance update.

8.4 Noise Generation Details

To provide more details about the noise robustness experiments presented in the main manuscript, we describe the noise generation process used in our evaluation. Specifically, two types of noise commonly observed in ultrasound imaging, including Gaussian noise and speckle noise, are considered. For each noise intensity, the corrupted frames are generated from the original ultrasound frames, and the segmentation performance is evaluated using the Dice coefficient. The horizontal axis in Figure 6 represents the noise intensity, where each value corresponds to the predefined "noise_levels" used in our implementation. The vertical axis denotes the Dice score obtained under the corresponding noise condition.

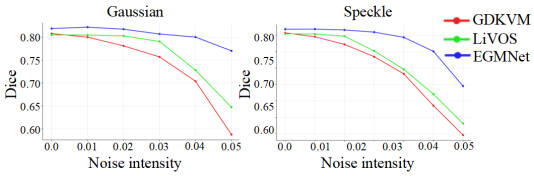


Figure 6: Robustness evaluation under varying intensities of additive Gaussian and multiplicative speckle noise.

Gaussian noise is simulated by adding zero-mean additive perturbations to the original image:

$$I_g = I + \epsilon_g, \tag{24}$$

where I denotes the original ultrasound frame and ϵ_g represents the Gaussian perturbation:

$$\epsilon_g \sim \mathcal{N}(0, \sigma^2). \tag{25}$$

The parameter σ controls the noise intensity and corresponds to the values defined in the "noise_levels" list.

Method	Dice↑	$\mathcal{J}$ ↑	HD95↓	ASD ↓
EME	76.52	69.75	16.65	5.94
GMA	76.25	69.52	16.55	6.19
w/o EME	76.80	70.21	15.56	5.60
w/o GMA	76.58	69.93	16.52	5.99
w/o CMC	77.22	70.71	16.43	5.90
Full	77.89	71.24	15.83	5.59

Table 4: Ablation study of different configurations on the BUV-1K dataset.

Speckle noise is generated by applying multiplicative intensity perturbation, which is consistent with the characteristics of ultrasound image noise:

$$I_s = I + I \odot \epsilon_s, \tag{26}$$

where $\odot$ denotes element-wise multiplication and ϵ_s is the noise perturbation sampled from:

$$\epsilon_s \sim \mathcal{N}(0, \sigma^2). \tag{27}$$

Similar to Gaussian noise, σ is determined by the corresponding noise intensity setting.

In summary, noise intensity in Figure 4 is plotted on the horizontal axis, corresponding to the perturbation magnitude under both noise models. For Gaussian noise, noise intensity determines the standard deviation of the additive noise distribution, while for speckle noise, noise intensity controls the strength of the multiplicative intensity variation applied to the original ultrasound frames. Furthermore, for each noise intensity, the Dice score is averaged over the test set, and the resulting values are reported in the noise robustness curves.

8.5 Additional Ablation Study

To comprehensively evaluate the contribution of each proposed component, we conduct two types of ablation experiments. First, we evaluate the effectiveness of each individual component by constructing single-component variants, where only one method is introduced into the baseline framework while the other two components are removed. For example, the EME method indicates that only EME is activated, whereas GMA and CMC are not included. Second, we perform component removal experiments based on the complete EGMNet. Specifically, "w/o EME", "w/o GMA", and "w/o CMC" denote removing the corresponding module from the full model while keeping the remaining components unchanged.

The single-component variants demonstrate that each proposed module can independently improve the baseline memory segmentation framework. The method using only EME validates the effectiveness of error-free memory evolution by reducing accumulated update errors during temporal propagation. The method using only GMA improves the key-value adaptation ability by exploiting global statistics of historical key distributions.

Compared with the complete EGMNet, removing any component generally leads to a decrease in segmentation performance, demonstrating that the three modules provide complementary improvements. Specifically, removing EME decreases the Dice score from 77.89% to 76.80% and reduces the overall segmentation accuracy, indicating that reliable memory evolution is essential for maintaining historical target representations. Although the w/o EME variant achieves a slightly better HD95 score (15.56 vs. 15.83), the lower region-overlap metrics suggest that it provides less accurate overall segmentation. The removal of GMA leads to performance degradation across most metrics, demonstrating the importance of adaptive key-value mapping under temporal appearance variations. Meanwhile, removing CMC also reduces the segmentation accuracy, showing that compact memory representation helps preserve informative features while suppressing redundant information. Overall, the complete EGMNet achieves the best balance among region-based and boundary-based metrics, validating the complementary effects of the proposed components.

8.6 Experiments Continued

As shown in Figure. 7, EGMNet produces more consistent and lesion-focused activation maps under ultrasound artifacts compared with existing memory-based methods. Although severe artifacts introduce ambiguous responses around the lesion boundary, our method maintains strong activation within the tumor region while suppressing irrelevant background interference. This benefit is attributed to the improved memory evolution mechanism, where accurate state updating and compact representation reduce error accumulation and redundant information during long-term propagation. Consequently, EGMNet can preserve discriminative tumor representations and achieve more reliable localization in challenging ultrasound scenarios.

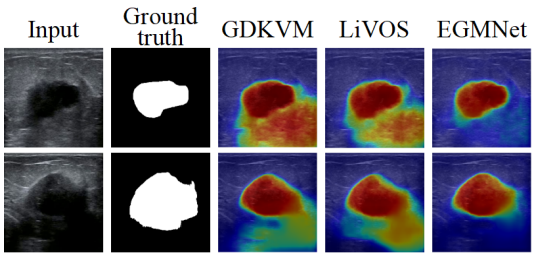


Figure 7: Visualization of activation maps under challenging ultrasound artifacts. Compared with GDKVM and LiVOS, EGMNet generates more concentrated responses within lesion regions, indicating stronger robustness against artifact interference and more reliable tumor representation preservation.

Fig. 8 presents per-frame Dice scores on a sequence formed by splicing two distinct patient videos to simulate lesion disappearance and reappearance during probe movement. While GDKVM and LiVOS exhibit sharp drops and slow recovery due to recursive memory drift, EGMNet adapts within frames, maintaining high performance across transitions. This rapid re-identification is enabled by EME's drift-free memory state transitions for reliable feature preservation, GMA's optimal subspace re-projection via recursive least squares to handle appearance changes, and CMC's suppression of redundant background responses to enhance memory information density during transitions.

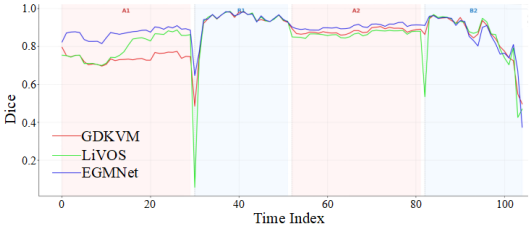


Figure 8: Simulation of lesion disappearance and reappearance. Segments A1 and A2 represent the first and second halves of one ultrasound video, while B1 and B2 correspond to another.