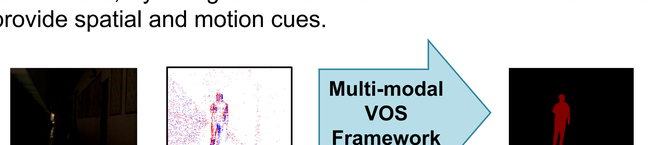
EVOLVE: Event-Guided Deformable Feature Transfer and Dual-Memory Refinement for Low-Light Video Object Segmentation

Jong-Hyeon Baek, Jiwon Oh and Yeong Jun Koh
Chungnam National University

ICCV OCT 19-23, 2025 HONOLULU HAWAII

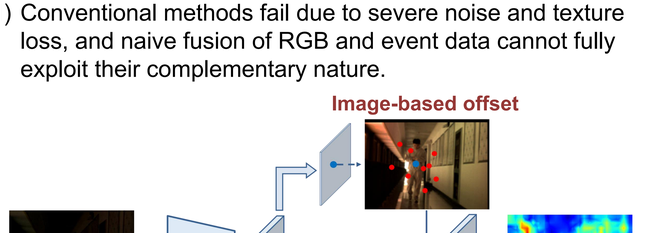
Introduction:

Low-Light Video Object Segmentation (VOS) aims to segment objects consistently across frames under severe noise and texture loss, by using both RGB frames and event streams to provide spatial and motion cues.

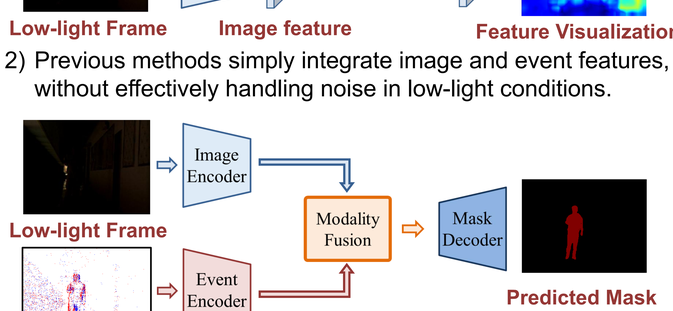


Motivation:

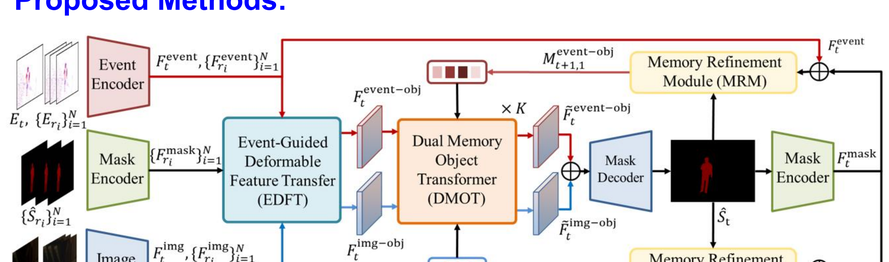
1) Conventional methods fail due to severe noise and texture loss, and naive fusion of RGB and event data cannot fully exploit their complementary nature.



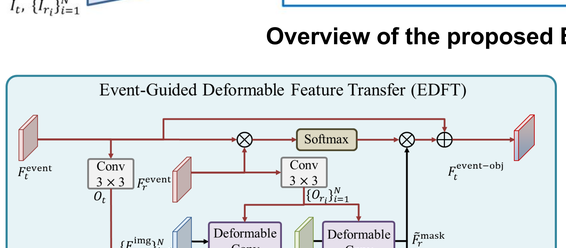
2) Previous methods simply integrate image and event features, without effectively handling noise in low-light conditions.



Proposed Methods:



Overview of the proposed EVOLVE



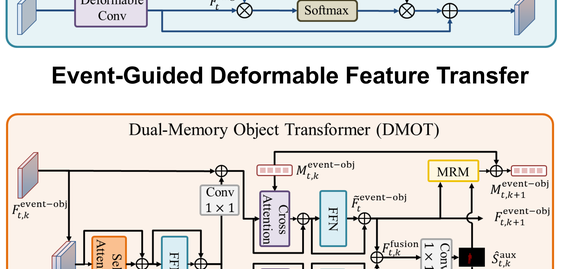
Event-Guided Deformable Feature Transfer

$$\tilde{F}_t^{\text{img}} = \text{DeformConv}\left(F_t^{\text{img}}, O_t\right)$$
$$\tilde{F}_{r_i}^{\text{img}} = \text{DeformConv}\left(F_{r_i}^{\text{img}}, O_{r_i}\right)\ i = 1, \ldots, N.$$
$$\tilde{F}_{r_i}^{\text{mask}} = \text{DeformConv}\left(F_{r_i}^{\text{mask}}, O_{r_i}\right)\ i = 1, \ldots, N.$$

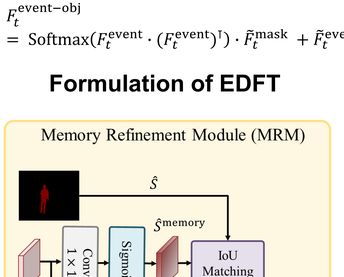
$$F_t^{\text{img-obj}} = \text{Softmax}\left(\tilde{F}_t^{\text{img}} \cdot \left(\tilde{F}_t^{\text{img}}\right)^{\intercal}\right) \cdot \tilde{F}_t^{\text{mask}} + \tilde{F}_t^{\text{img}}$$

$$F_t^{\text{event-obj}} = \text{Softmax}\left(F_t^{\text{event}} \cdot \left(F_t^{\text{event}}\right)^{\intercal}\right) \cdot \tilde{F}_t^{\text{mask}} + \tilde{F}_t^{\text{event}}$$

Formulation of EDFT



Dual-Memory Object Transformer



Memory Refinement Module

Results:

Setting	Methods	LLE-VOS Indoor Scenes			LLE-VOS Outdoor Scenes			LLE-DAVIS		
		$\mathcal{J}$	$\mathcal{F}$	$\mathcal{J}\&\mathcal{F}$	$\mathcal{J}$	$\mathcal{F}$	$\mathcal{J}\&\mathcal{F}$	$\mathcal{J}$	$\mathcal{F}$	$\mathcal{J}\&\mathcal{F}$
Image	STCN [7]	0.486	0.321	0.403	0.400	0.309	0.354	0.424	0.453	0.438
	XMem [6]	0.664	0.528	0.596	0.507	0.456	0.481	0.465	0.477	0.471
	AOT [41]	0.699	0.618	0.659	0.592	0.571	0.581	0.540	0.578	0.559
	DeAOT [40]	0.716	0.643	0.680	0.580	0.580	0.580	0.541	0.571	0.556
	Cutie [9]	0.706	0.708	0.707	0.534	0.484	0.511	0.560	0.531	0.545
	OneVOS [14]	0.695	0.582	0.639	0.542	0.513	0.527	0.526	0.553	0.539
Image + Event	STCN [7]	0.522	0.360	0.441	0.460	0.354	0.407	0.450	0.498	0.474
	XMem [6]	0.732	0.616	0.674	0.525	0.465	0.495	0.507	0.534	0.521
	AOT [41]	0.745	0.674	0.709	0.590	0.574	0.582	0.555	0.614	0.584
	DeAOT [40]	0.746	0.678	0.712	0.596	0.604	0.600	0.566	0.608	0.587
	Cutie [9]	0.784	0.682	0.733	0.626	0.599	0.612	0.654	0.592	0.623
	OneVOS [14]	0.785	0.663	0.724	0.634	0.610	0.622	0.603	0.659	0.631
	LLE-VOS [12]	0.789	0.710	0.749	0.604	0.588	0.596	0.602	0.654	0.628
	EVOLVE	**0.847**	**0.817**	**0.832**	**0.703**	**0.725**	**0.714**	**0.677**	**0.721**	**0.699**

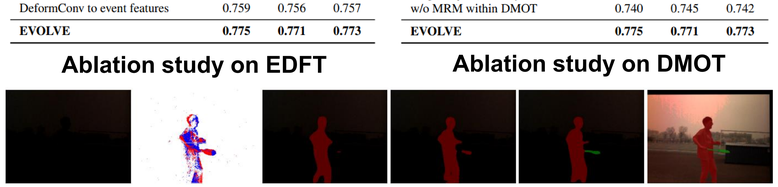
Quantitative results on LLE-VOS and LLE-DAVIS

Methods	$\mathcal{J}$	$\mathcal{F}$	$\mathcal{J}\&\mathcal{F}$
w/o DeformConv	0.733	0.727	0.739
Offsets from image features	0.719	0.728	0.724
DeformConv to event features	0.759	0.756	0.757
EVOLVE	**0.775**	**0.771**	**0.773**

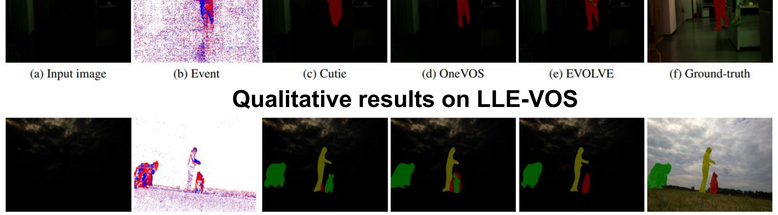
Ablation study on EDFT

Methods	$\mathcal{J}$	$\mathcal{F}$	$\mathcal{J}\&\mathcal{F}$
w/o cross-attention with memories	0.721	0.744	0.733
Single object memory	0.740	0.741	0.740
w/o MRM within DMOT	0.740	0.745	0.742
EVOLVE	**0.775**	**0.771**	**0.773**

Ablation study on DMOT



Qualitative results on LLE-VOS



Qualitative results on LLE-DAVIS