Events Meet Dynamic Mode Decomposition: Capturing the Spatiotemporal Dynamics of Moving Objects

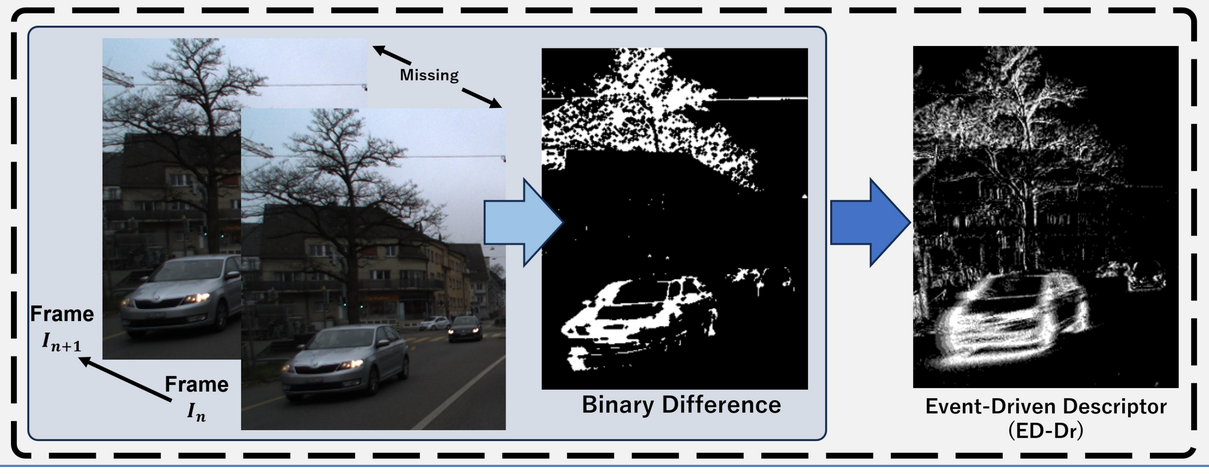
Zhouning Du, Israr Ulhaq, Thanh Thi Huyen Phan, Yuichiro Yoshimura, Jigyasa Chand, Truong Vinh Truong Duy
Tokyo Research Center, AISIN CORPORATION
contact: zhouning.du@aisin.co.jp

Motivation

• Traditional RGB methods miss critical motion details between frames, leading to inaccurate segmentation in moving scenes.

• Event cameras provide high-resolution, asynchronous data, capturing motion changes missed by conventional techniques.

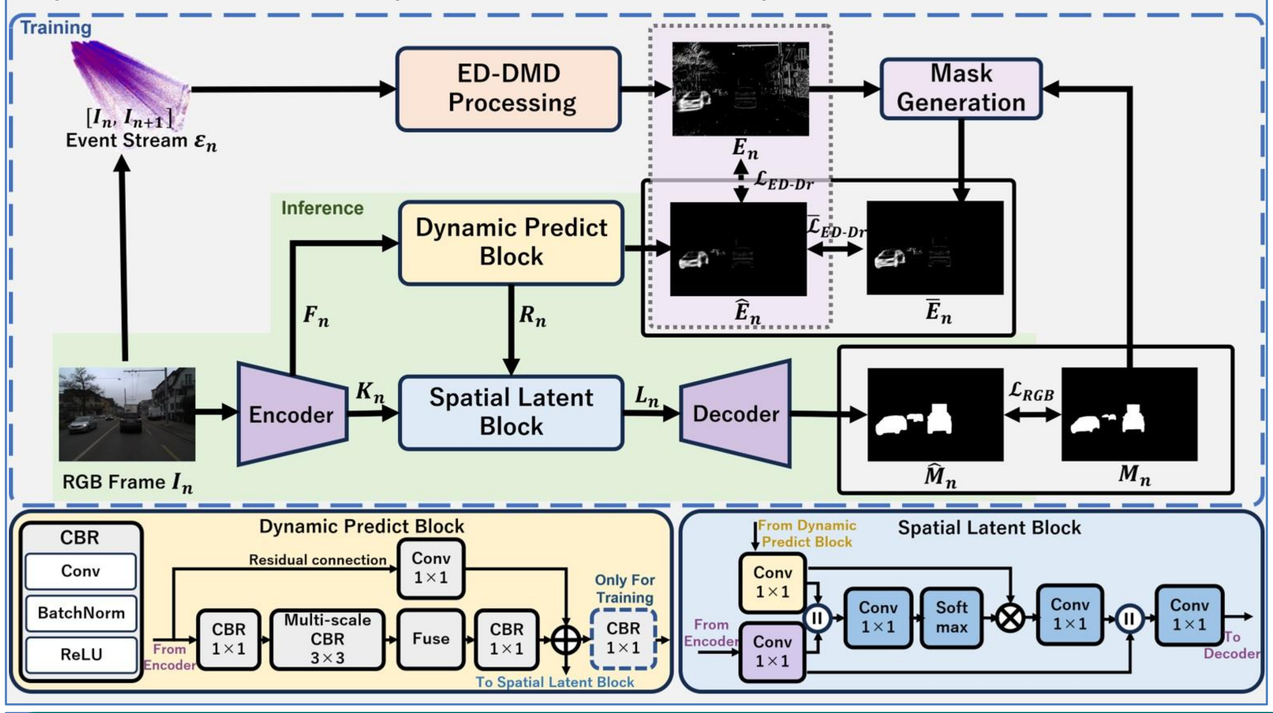
• Our goal is to **enhance moving object segmentation** by **combining event data with RGB frames** using **ED-DMD** and **MAS-Net**.



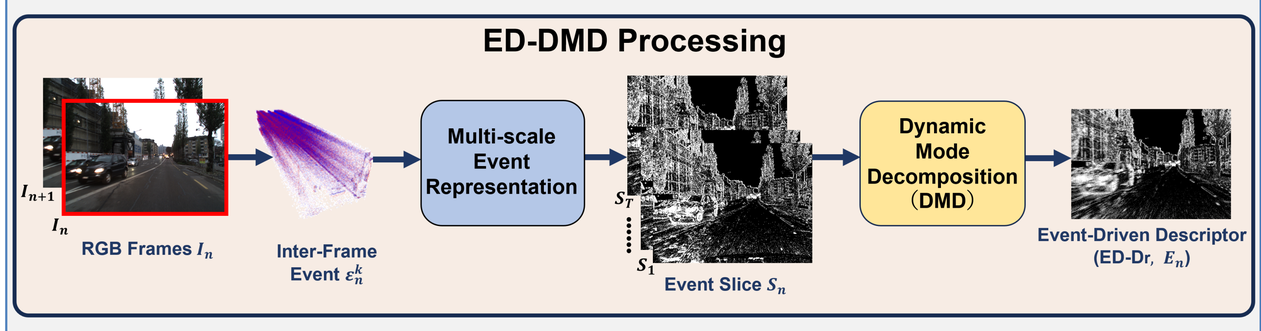
Motion-aware segmentation network (MAS-Net)

• **Training Phase:** **Blue dashed line.** Network learns motion dynamics and static features from ED-Dr and RGB using DPB and SLB.

• **Inference Phase:** **Green background.** Uses only RGB inputs, DPB predicts dynamics from learned priors, no event data required.



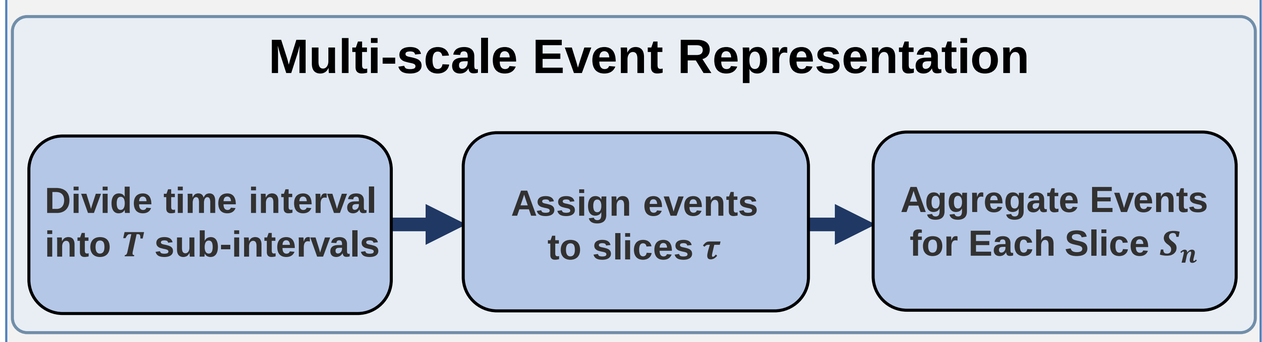
Event-driven dynamic mode decomposition(ED-DMD)



Multi-scale event representation

• **Objective:**
Capture missing dynamics between RGB frames.

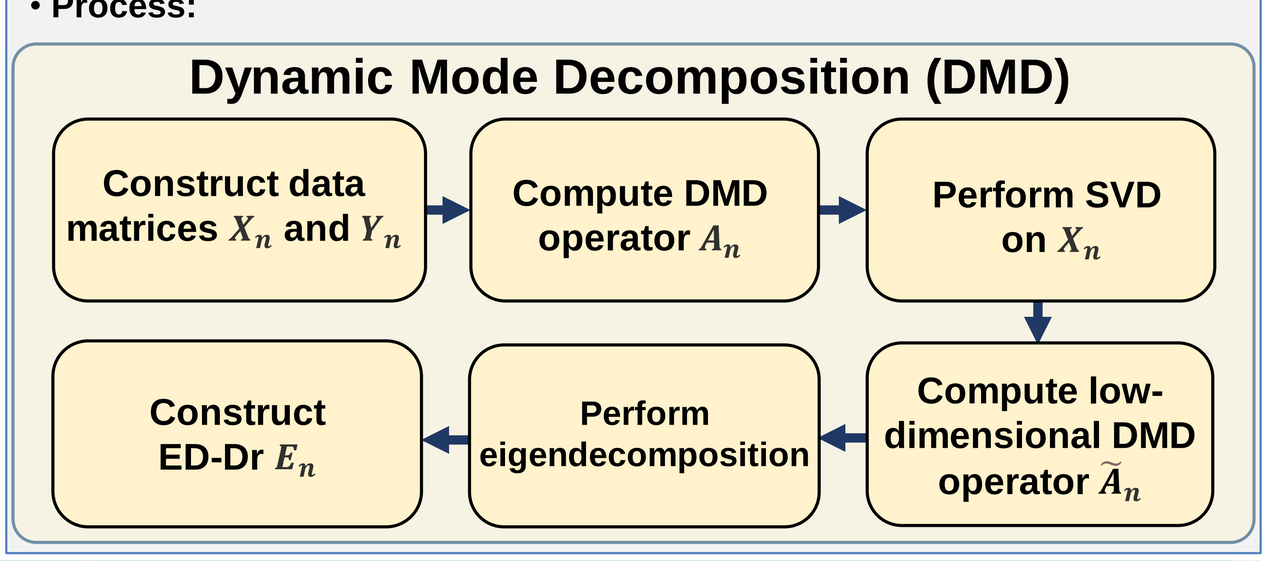
• **Process:**



Event-driven descriptor (ED-Dr)

• **Objective:**
Reconstruct missing dynamics between RGB frames.

• **Process:**



Experimental results

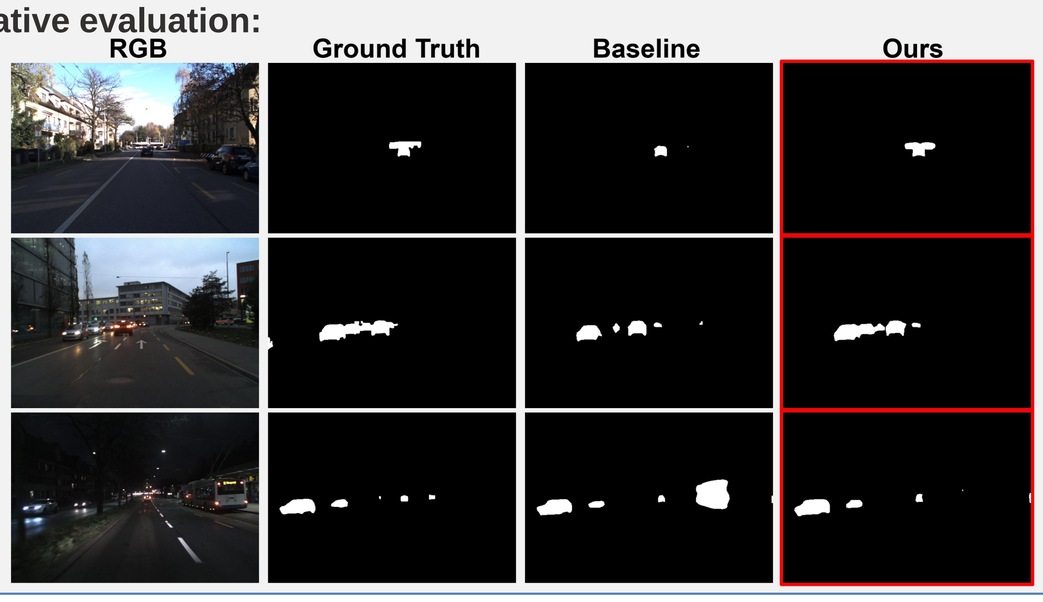
Comparison with SOTA methods

• **Dataset:** DSEC-MOS with 13,314 frames in diverse driving scenarios.

• **Quantitative evaluation:**

Approach	Year	Backbone	Train Data	Inf. Data	MS Inf.	$\mathcal{R}_{mean}$	$\mathcal{F}_{mean}$	$\mathcal{R}\&\mathcal{F}_{(mean)}$
SegFormer-B2[25]	NeurIPS'21	MiT-B2	F	F		0.421	0.595	0.508
SegNeXt-B[11]	NeurIPS'22	MSCAN-B	F	F		0.609	0.749	0.679
Isomer[27]	ICCV'23	Swin-Tiny	OF, F	OF, F		0.559	0.722	0.641
TMO[4]	WACV'23	ResNet-101	OF, F	OF, F		0.562	0.730	0.646
MED-VT[12]	CVPR'23	Swin-B	F	F		0.580	0.810	0.695
EmoFormer[29]	IROS'24	Swin-B	E, F	F		0.665	0.822	0.744
MAS-Net	Ours	Swin-B	E, F	F		**0.680**	**0.835**	**0.758**
HFAN[17]	ECCV'22	MiT-b1,2	OF, F	OF, F	✓	0.626	0.768	0.697
MED-VT[12]	CVPR'23	Swin-B	F	F	✓	0.635	0.815	0.725
EmoFormer[29]	IROS'24	Swin-B	E, F	F	✓	0.689	0.837	0.763
MAS-Net	Ours	Swin-B	E, F	F	✓	**0.701**	**0.842**	**0.772**

• **Qualitative evaluation:**



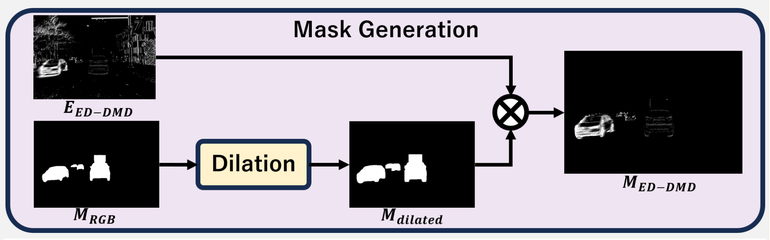
SLB and DPB analysis with ED-Dr

Analyse SLB and DPB with ED-Dr to enhance segmentation by targeting different dynamics.

• M_{RGB} :
RGB frame Mask.

• $E_{\mathrm{ED-Dr}}$:
All motion dynamics.

• $M_{\mathrm{ED-Dr}}$:
Moving object dynamics.



Architecture	Label	$\mathcal{R}_{mean}$	$\mathcal{F}_{mean}$	$\mathcal{R}\&\mathcal{F}_{(mean)}$
SLB	M_{RGB}	0.633	0.804	0.719
SLB+DPB	M_{RGB} + $E_{\mathrm{ED\text{-}Dr}}$	0.651	0.812	0.732
SLB+DPB	M_{RGB} + $M_{\mathrm{ED\text{-}Dr}}$	**0.680**	**0.835**	**0.758**

Backbone analysis

Evaluate backbones for capturing spatial and dynamic features.

Backbone	Param. (M)	Label	$\mathcal{R}_{mean}$	$\mathcal{F}_{mean}$	$\mathcal{R}\&\mathcal{F}_{(mean)}$
MSCAN-T	4	M_{RGB} + $M_{\mathrm{ED\text{-}Dr}}$	0.571	0.717	0.644
MSCAN-S	14		0.605	0.757	0.681
Swin-Tiny	28		0.632	0.793	0.713
MSCAN-B	28		0.643	0.786	0.715
MSCAN-L	49		0.651	0.795	0.723
Swin-S	50		0.651	0.811	0.731
Swin-B	88		**0.680**	**0.835**	**0.758**