EINSTEIN CENTER Digital Future

Technische Universität Berlin

Unsupervised Joint Learning of Optical Flow and Intensity with Event Cameras

Shuang Guo, Friedhelm Hamann and Guillermo Gallego

Project page

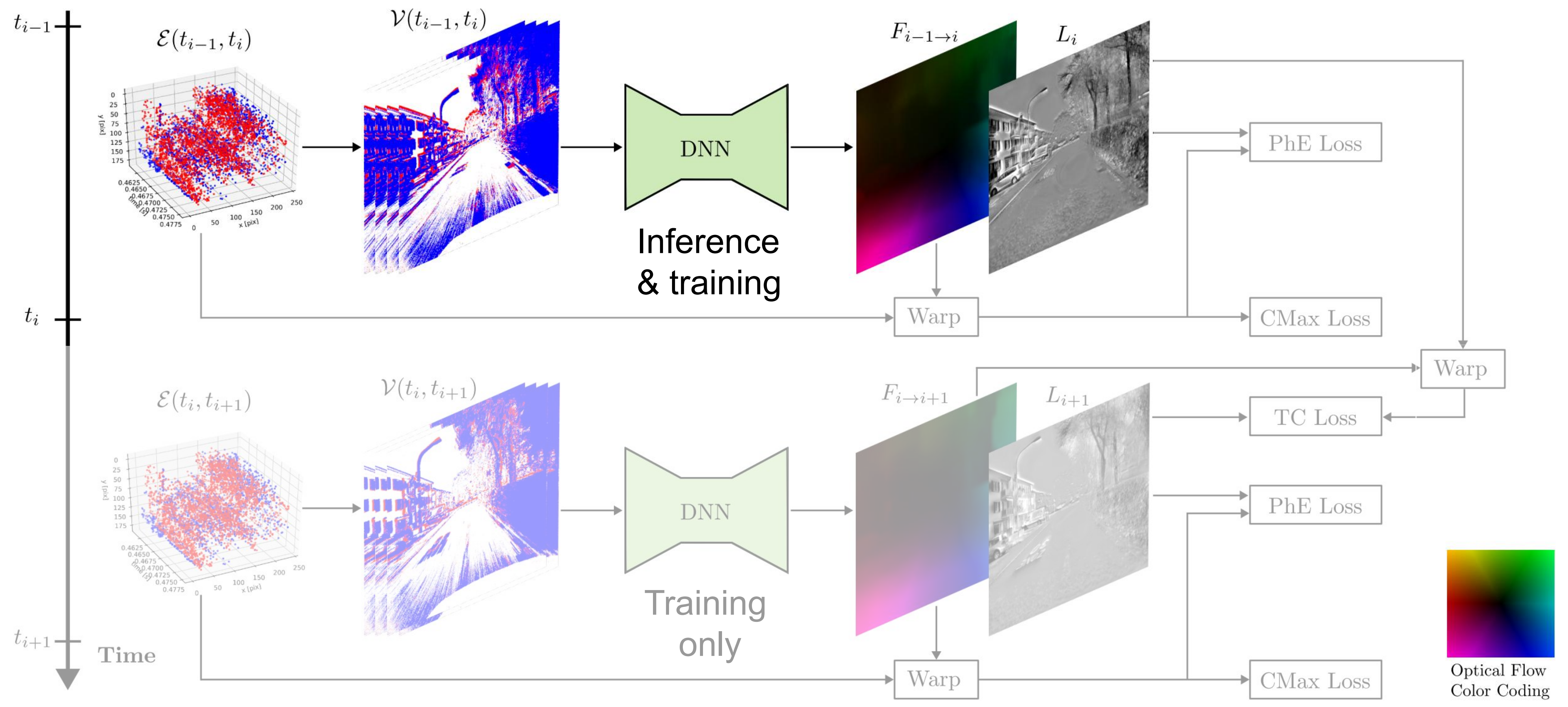
ICCV OCT 19-23, 2025 HONOLULU HAWAII

science of intelligence

Robotics Institute Germany

Summary of E2FAI: Events to Flow And Intensity

- **Appearance** and **motion** are **inherently linked in event cameras**: **either both** are present and recorded in the event data, **or neither** is captured.
- **Therefore, we do not** treat the recovery of these two visual quantities as **separate** tasks.
- We propose the **1st unsupervised learning framework** that **jointly** estimates optical flow (motion) and image intensity (appearance) using **a single network**.
- We derive **event-based photometric error**, and combine it with **contrast maximization**, yielding a **comprehensive and well-behaved loss function**.



Total Loss:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{PhE}} + \lambda_2 \mathcal{L}_{\text{CMax}} + \lambda_3 \mathcal{L}_{\text{FTV}} + \lambda_4 \mathcal{L}_{\text{ITV}} + \lambda_5 \mathcal{L}_{\text{TC}}$$

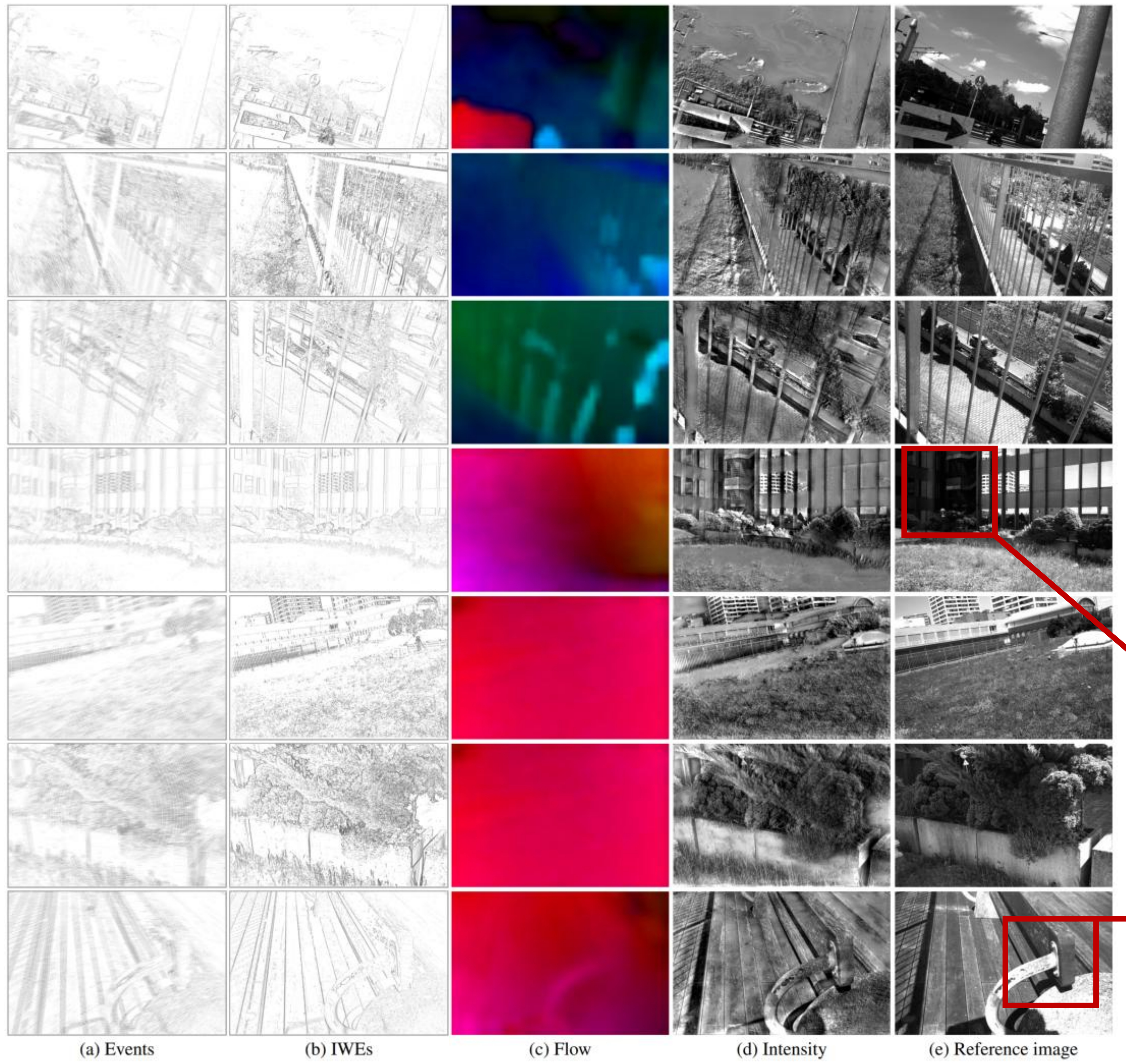
● **Event-based Photometric Error (PhE):**

$$\mathcal{L}_{\text{PhE}}(L, F) \doteq \frac{1}{N_e} \sum_{k=1}^{N_e} \left| \underbrace{\left(L(\mathbf{x}'_k(F)) - L(\mathbf{x}'_{k-1}(F))\right)}_{\text{EGM Predicted } \hat{\Delta L}} - \underbrace{p_k C}_{\text{Measured } \Delta L} \right|$$

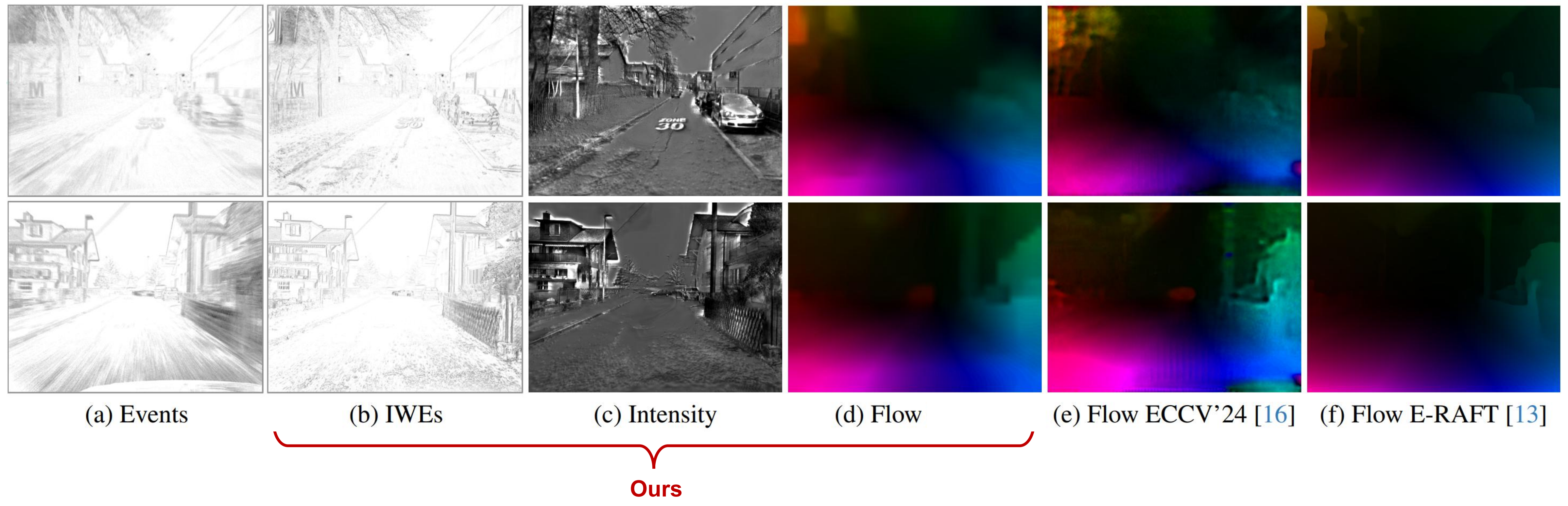
is **a function of intensity and flow**, which **enables the joint estimation of both quantities**.

● We also have **Contrast Maximization** (CMax), **Total Variation** (TV) regularizers and **Temporal Consistency** (TC) terms.

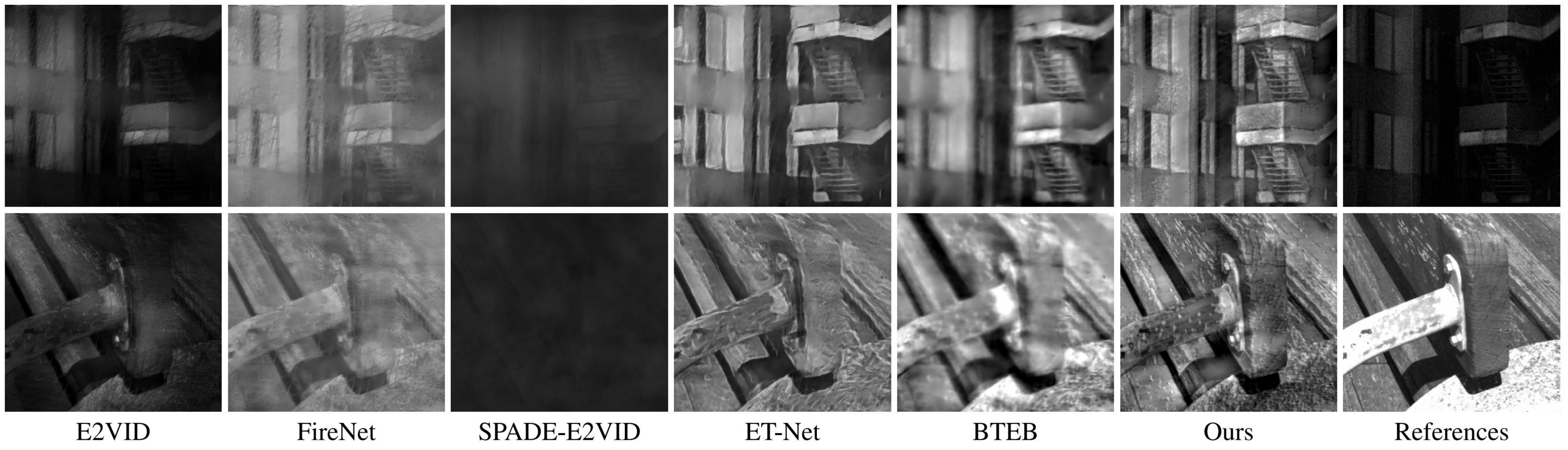
Qualitative Results (BS-ERGB data)



Qualitative Results (DSEC data)



Zoomed-in Image Comparison (BS-ERGB data)



Optical Flow Evaluation (DSEC benchmark)

			All			
Type	Method	t_{inf}[ms]	EPE ↓	AE ↓	%Out ↓	FWL ↑
SL	E-RAFT [13]	46.33	0.79	10.56	2.68	1.29
	IDNet [45]		**0.72**	**2.72**	**2.04**	–
MB/ USL	RTEF [3]		4.88	–	41.95	**2.51**
	MultiCM [37]	$9.9 \cdot 10^3$	3.47	13.98	30.86	1.37
	BTEB [28]		3.86	–	31.45	1.30
	Paredes et al. [29]	40.1	2.33	10.56	17.77	–
	EV-FlowNet [52]		3.86	–	31.45	1.30
	MotionPriorCM [16]	17.86	3.20	8.53	15.21	1.46
	VSA-SM [47]		2.22	8.86	16.83	–
	Ours	**15.12**	**1.78**	**6.44**	**11.24**	1.79

SL: Supervised USL: Unsupervised MB: Model-based

Image Intensity Evaluation (BS-ERGB & HDR)

		BS-ERGB			HDR		
Type	Method	MSE↓	SSIM↑	LPIPS↓	BRISQUE↓	NIQE↓	MANIQA↑
SL	E2VID [31]	0.14	0.33	0.56	**12.63**	4.27	0.30
	FireNet [35]	0.10	0.34	0.53	18.57	3.85	0.30
	SPADE-E2VID [10]	0.09	0.35	0.63	24.51	7.17	0.28
	ET-Net [44]	**0.07**	**0.37**	**0.44**	19.20	**3.45**	**0.32**
USL	BTEB [28]	**0.09**	**0.36**	0.62	51.47	6.24	0.18
	Ours	0.10	0.31	**0.56**	**25.03**	**3.78**	**0.40**

BS-ERGB Dataset: by Tulyakov et al. TimeLens++, CVPR 2022.
HDR data by Rebecq et al., T-PAMI 2021.
DSEC dataset by Gehrig et al., RAL 2021.

Runtime Evaluation [ms]

Resolution	E2VID (2019)	FireNet (2020)	SPADE-E2VID (2021)	ET-Net (2021)	BTEB (2021)	Ours (2024)
640 × 480	10.95	4.94	36.07	173.56	10.59	15.11
1280 × 720	31.04	14.67	105.87	1606.33	29.89	40.78