Efficient Meshflow and Optical Flow Estimation from Event Cameras

Xinglong Luo[1,4*], Ao Luo[2,4*], Zhengning Wang[1†] , Chunyu Lin[3], Bing Zeng[1], and Shuaicheng Liu[1,4†]

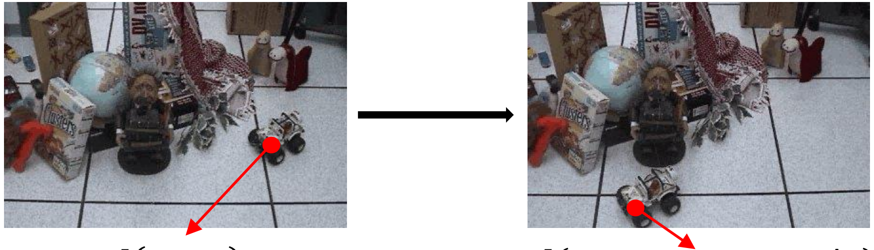
[1] University of Electronic Science and Technology of China,

[2] Southwest Jiaotong University, [3] Beijing Jiaotong University , [4] Megvii Technology

[*] Equal contribution. [†] Corresponding Author.

Introduction-Meshflow

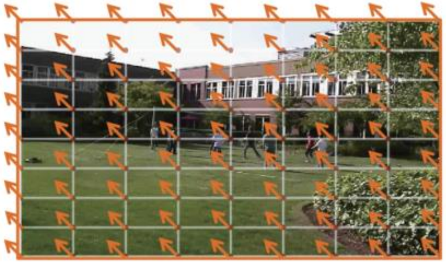
- **Optical Flow** is the movements of **all pixels** on the picture plane over time.



$I(x, y, t)$ $\quad\quad$ $I(x + \boldsymbol{u}, y + \boldsymbol{v}, t + \Delta t)$

- **Meshflow**[1] is a spatially smooth sparse motion field that only records motion vectors at **mesh vertices**.

Neighboring pixel movement are very comparable.



The global motions can be obtained by capturing movement of mesh vertices.

Meshflow is **widely** applied in **various vision applications**.

- Image Alignment

- Video Stabilization

- HDR Imaging

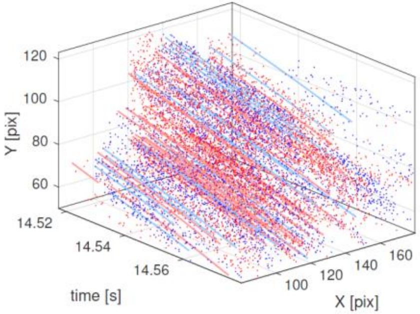
......

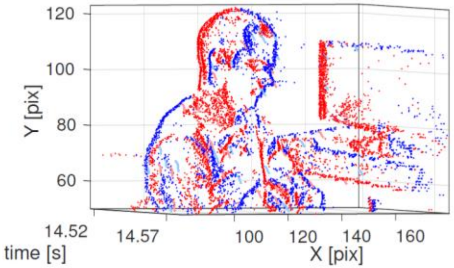
[1] Liu S, et al. Meshflow: Minimum latency online video stabilization. Proc. of ECCV, 2016.

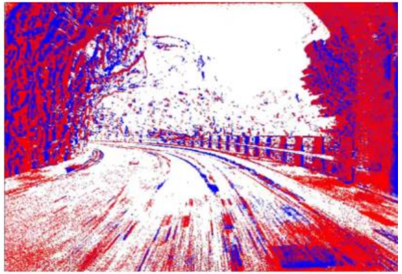
Introduction-Event Camera

- RGB images often loss fine **texture details** and suffer **motion blurs** under extreme scenes.

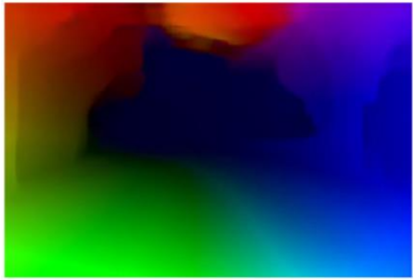

- **Event Cameras**[1] have several advantages over traditional cameras: **high temporal resolution**, **high dynamic range**, and low power consumption.

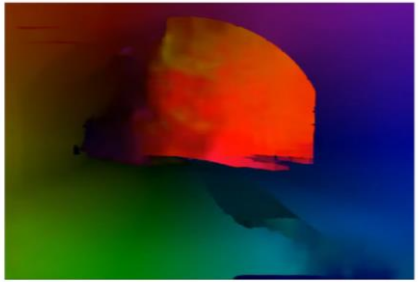





Events


RGB Image


ERAFT

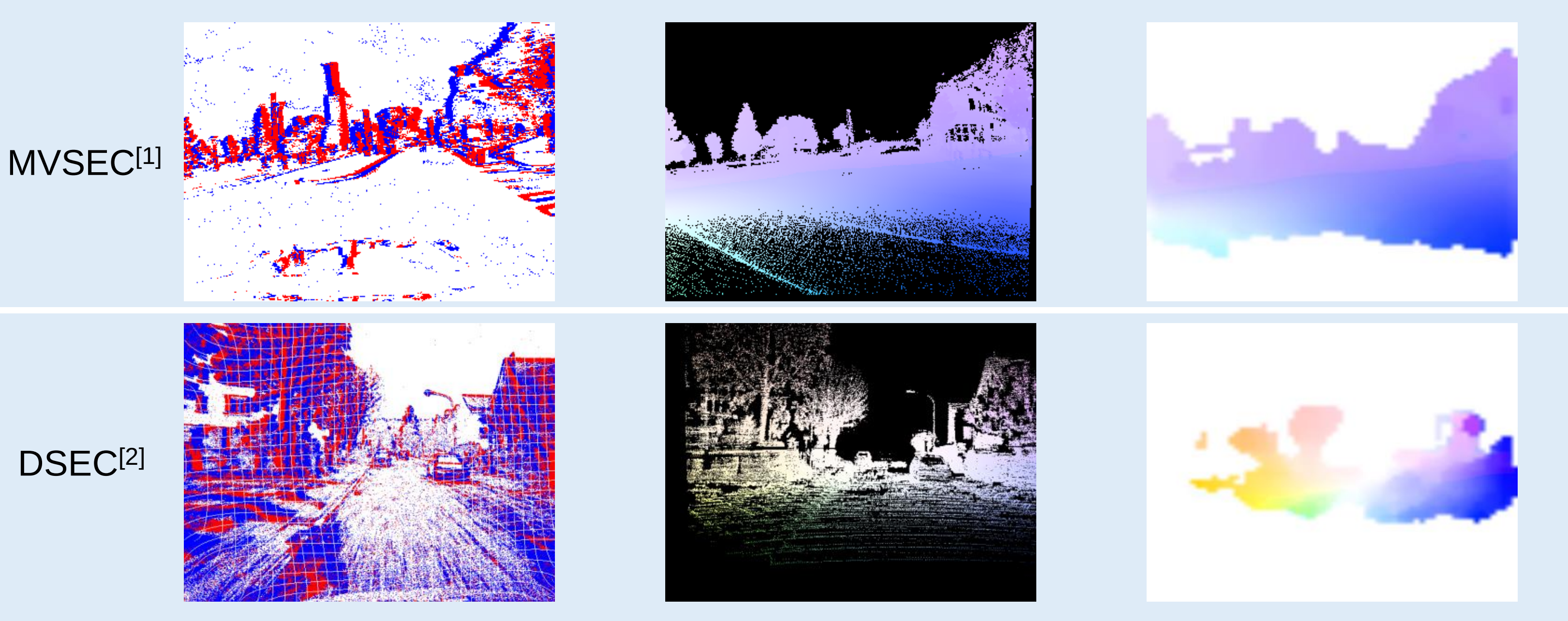

RAFT

Event cameras are **well-suited** for motion estimation under **extreme scenes**.[2]

[1] Gallego G, et al. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. Proc. of CVPR, 2018.
[2] Gehrig M, et al. E-raft: Dense optical flow from event cameras. International Conference on 3D Vision (3DV), IEEE, 2021.

Method 1 - HREM Dataset

- None of the event-based datasets involve the estimation of meshflow.
- Existing event-flow datasets can’t provide **high-quality** meshflow labels extracting form optical flow.

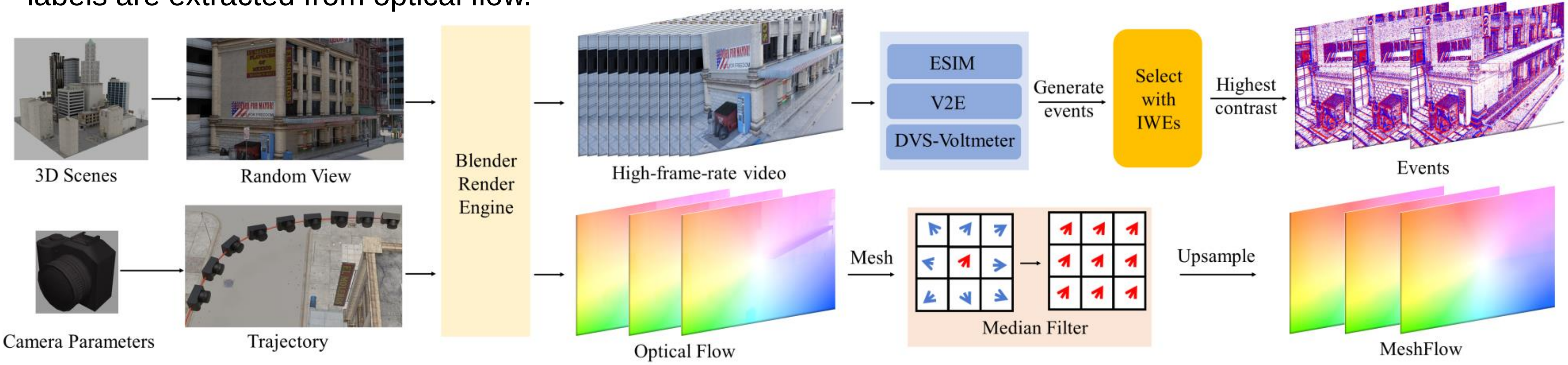


Events Optical Flow Meshflow

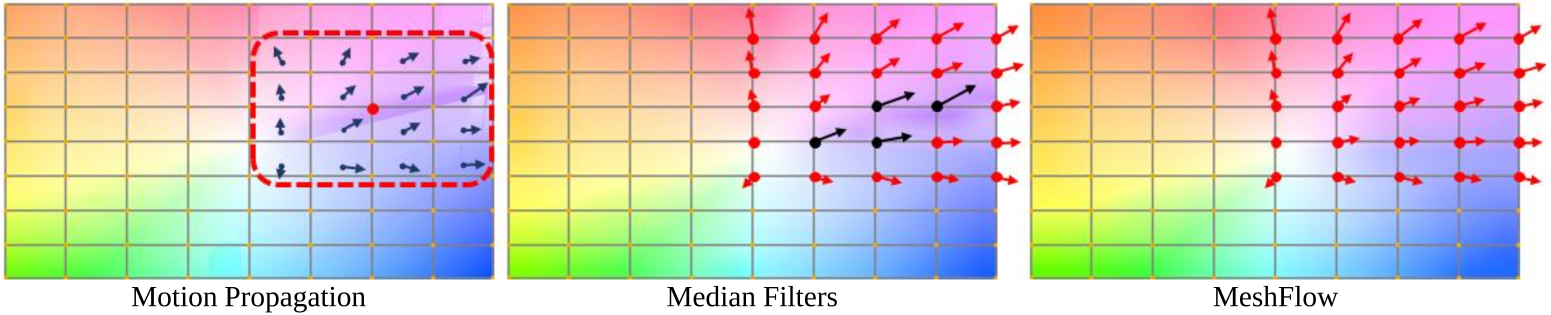
[1] Alex Z, et al. The multi-vehicle stereo event camera dataset: An event camera dataset for 3d perception. IEEE Robotics and Automation Letters, 2018.
[2] Mathias G, et al. Dsec: A stereo event camera dataset for driving scenarios. IEEE Robotics and Automation Letters, 2021.

Method 1 - HREM Dataset

Dataset Generation: Camera trajectories and dynamic objects are randomly placed in 3D virtual scenes to render **high-frame-rate videos** and **optical flow** labels, then **events** are generated from videos and **meshflow** labels are extracted from optical flow.



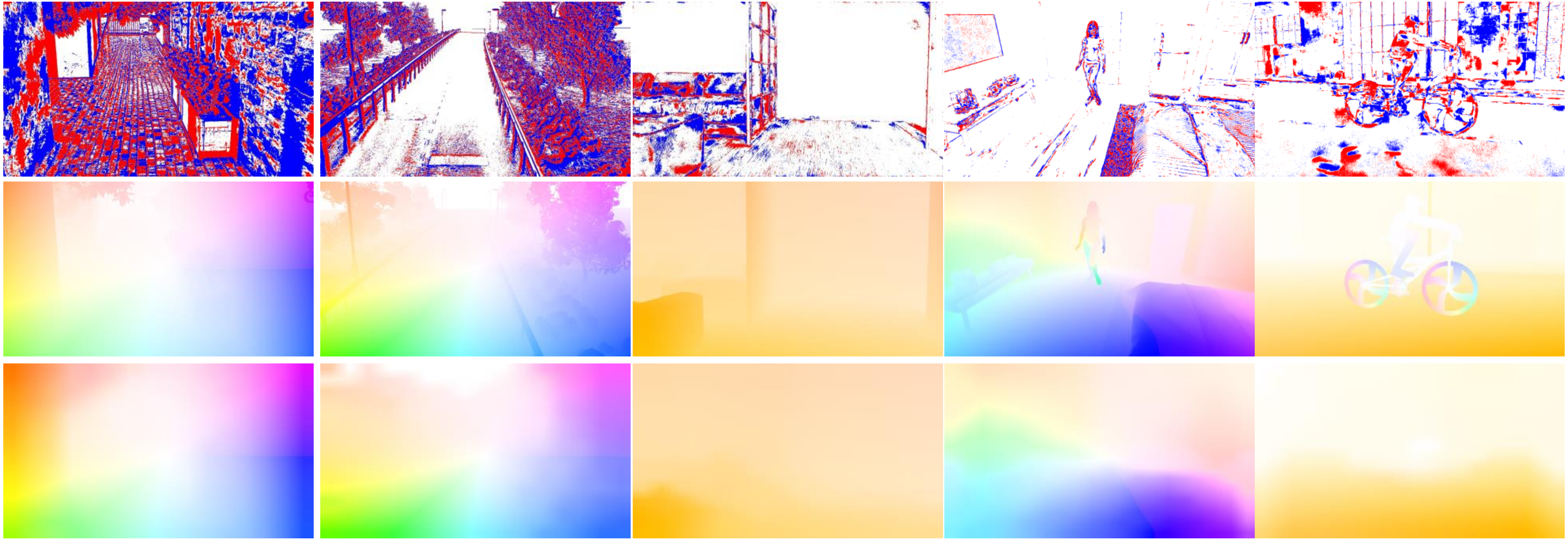
Meshflow: Propagate the median movements to the **mesh vertices**, and filter outliers using median filters.



Method 1 - HREM Dataset

High-**R**esolution **E**vent **M**eshflow (**HREM**) Dataset

Scene	Resolution	Motion Pattern	Dynamic Objects	Extreme Conditions	Dense Optical Flow	Meshflow
Indoor, outdoor	1280x720	Random	✓	✓	✓	✓

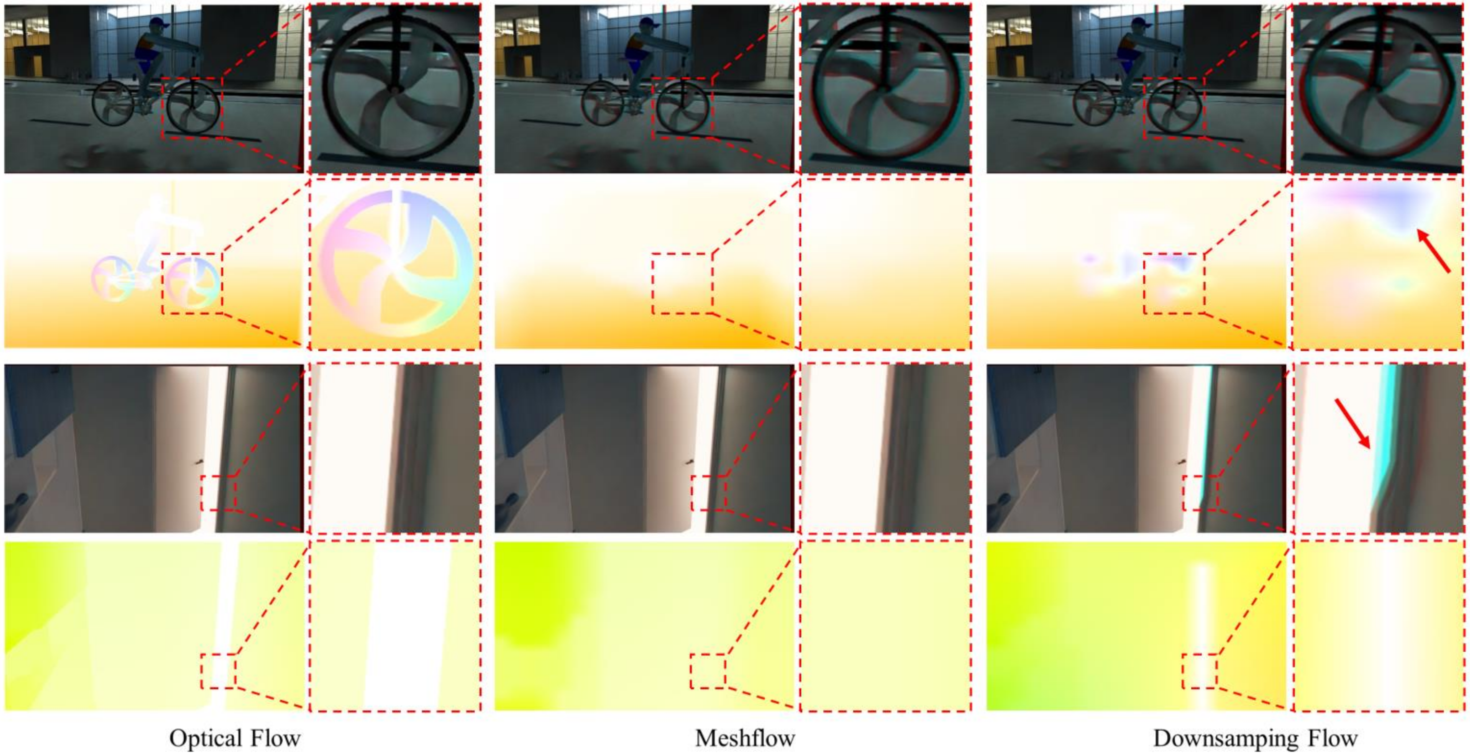


The first row is Events, the second is Optical Flow, and the third is Meshflow.

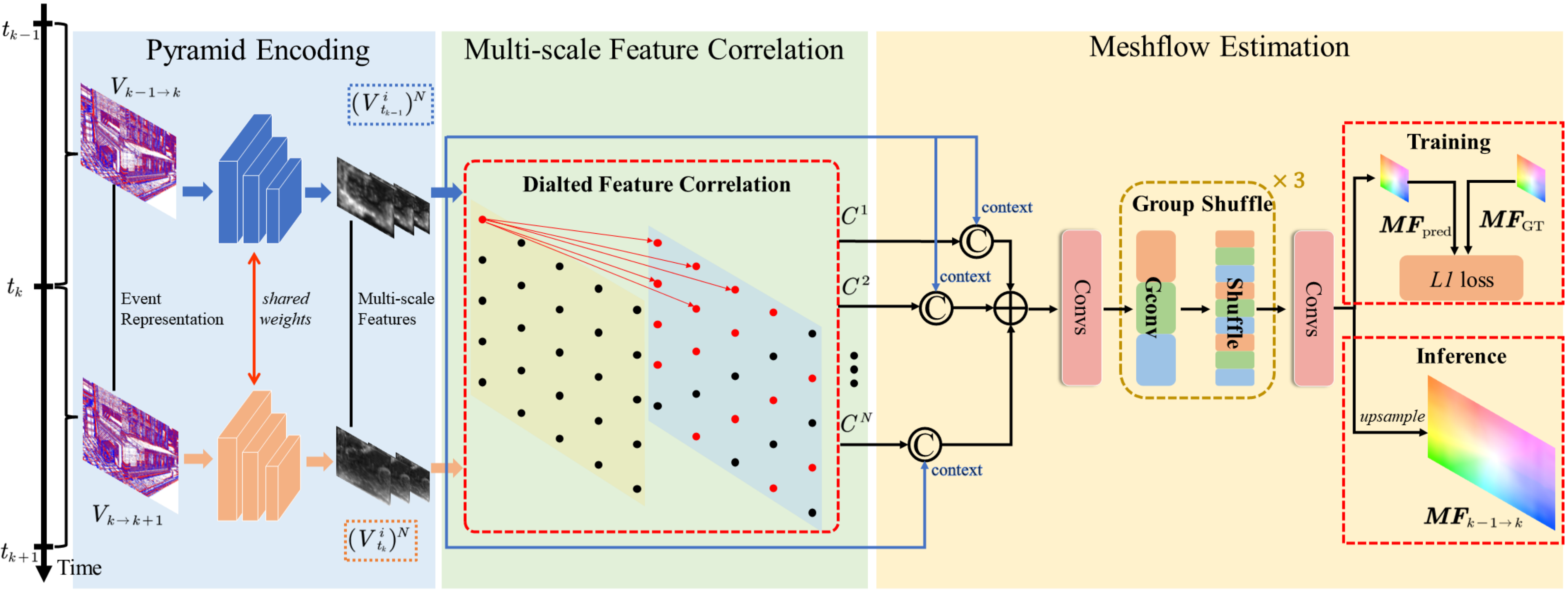
Method 1 - HREM Dataset

Ablation experiment of HREM dataset:

- **Meshflow** achieves **image alignment** comparable to optical flow.
- Downsampling flow via bilinear interpolation **twists the shape** of objects compared to meshflow.



Method 2 - EEMFlow



EEMFlow (Event-based Meshflow Estimation):

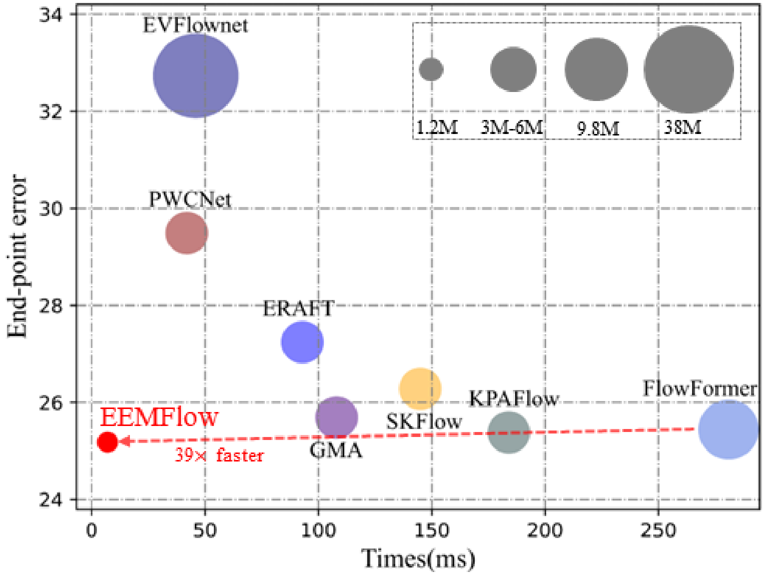
- The light-weight encoder
- Building cost volume with **dilated feature correlation**
- Using **group shuffle convolutions** during decoding

Method 2 - EEMFlow

Results for Event-based Meshflow Estimation：

- EEMFlow performs well across scenes and speeds, achieving **SOTA** on Average EPE metrics.
- EEMFlow‘s **inference** is fast, about **39 times faster** compared to the well-performing FlowFormer[1].

Method $dt=1$	Parameters (M)	Time (ms)	Outdoor Slow	Outdoor Fast	Indoor Slow	Indoor Fast	Avg
EVFlownet [52]	38.2	46	3.55	16.16	2.93	11.65	8.57
PWCNet [40]	3.36	42	3.91	14.49	2.86	11.89	8.29
ERAFT [11]	5.27	93	4.15	13.32	2.91	10.34	7.68
SKFlow [41]	6.28	145	3.76	11.78	7.24	8.81	7.24
GMA [17]	5.89	108	2.18	12.07	2.02	9.34	6.40
KPAFlow [32]	6.00	184	2.03	12.25	1.95	9.02	6.31
FlowFormer [15]	9.87	281	2.06	11.71	1.88	8.66	6.08
EEMFlow(Ours)	1.24	7	2.42	9.09	2.00	8.46	5.50
Method $dt=4$	**ΔP (M)**	**ΔT (ms)**	**Outdoor Slow**	**Outdoor Fast**	**Indoor Slow**	**Indoor Fast**	**Avg**
EVFlownet [52]	+624%	+51%	18.25	49.32	16.16	47.19	32.73
PWCNet [40]	-36%	-55%	16.40	46.17	14.49	40.90	29.49
ERAFT [11]	0%	0%	15.21	40.83	13.32	39.61	27.24
SKFlow [41]	+19%	+56%	14.93	39.24	11.71	39.22	26.28
GMA [17]	+11%	+16%	14.13	38.89	12.07	37.68	25.69
KPAFlow [32]	+14%	+99%	14.04	38.03	12.25	37.20	25.38
FlowFormer [15]	+88%	+202%	13.89	38.55	10.77	38.53	25.44
EEMFlow(Ours)	-76%	-92%	13.97	37.33	12.09	34.39	24.45

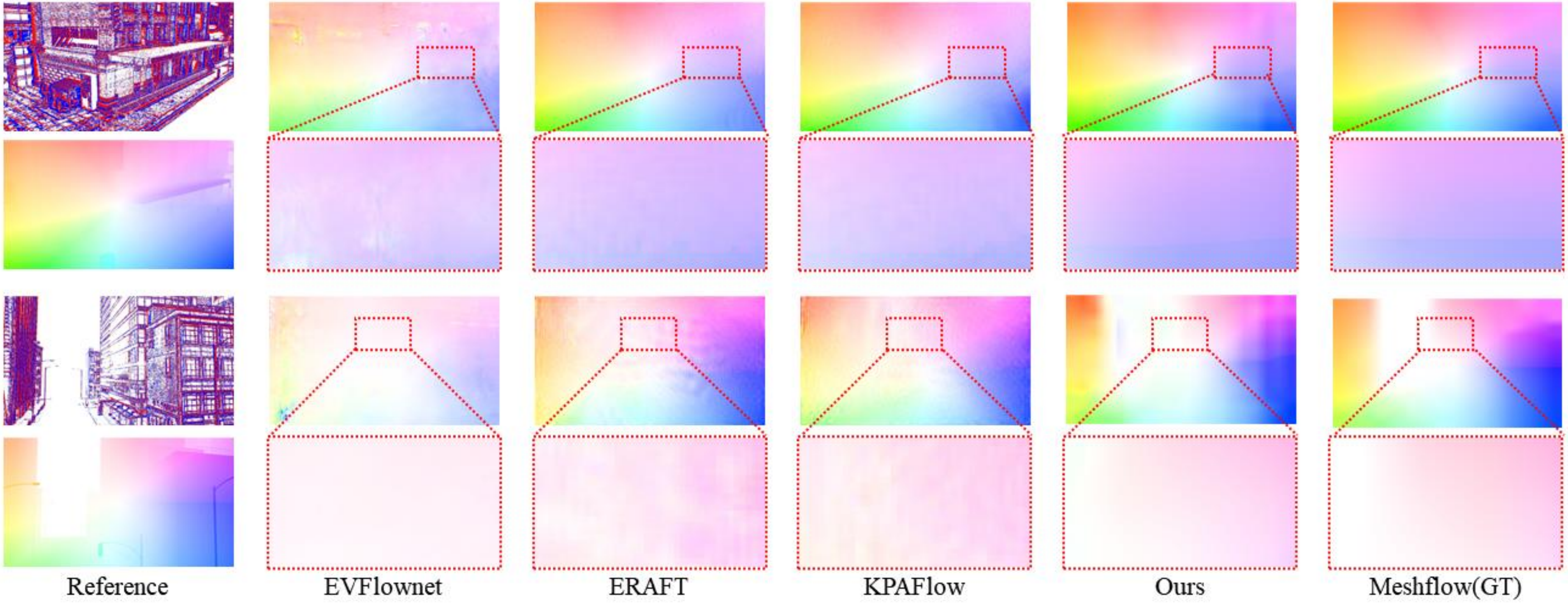


[1] Zhaoyang H, et al. FlowFormer: A transformer architecture for optical flow. Proc. of ECCV, 2022.

Method 2 - EEMFlow

Qualitative comparison for Event-based Meshflow Estimation:

- EEMFlow's subjectsive results are most similar to Meshflow GT.

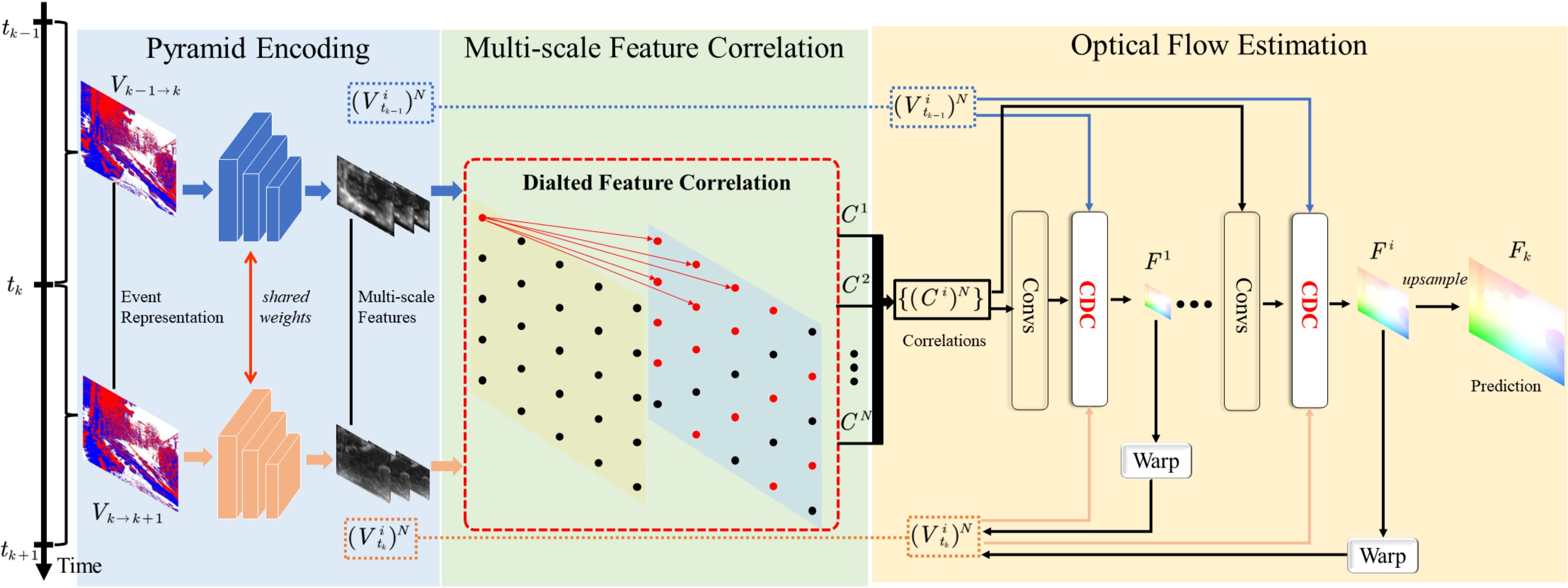


Method 2 - EEMFlow

- In contrast to other methods, the estimation results of EEMFlow for image alignment have minimal shakes

Image Overlaid EVFlownet ERAFT EEMFlow(Ours)

Method 3 – EEMFlow+



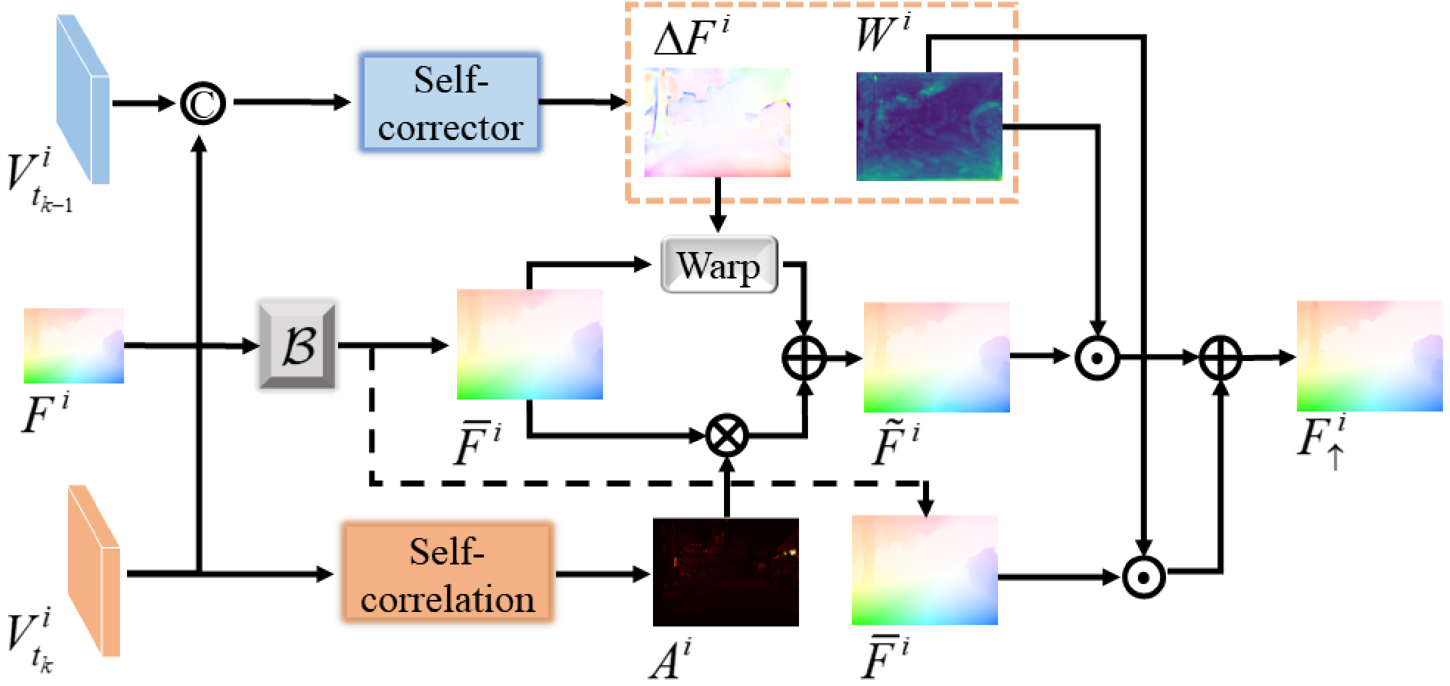
EEMFlow+ (Event-based Optical Flow Estimation):

- **The coarse to fine residual approach** to progressively refine the flow prediction
- **C**onfidence-induced **D**etail **C**ompletion (**CDC**) module to enhance motion boundary details during flow upsampling.

Method 3 – EEMFlow+

The structure of CDC:

- The self-corrector outputs the correction flow $\Delta \boldsymbol{F}^{i}$ to correct the error region, the confidence map $\boldsymbol{W}^{i}$ to retain corrected regions with high confidence.
- The self-correlation outputs the attention weight $\boldsymbol{A}^{i}$ to focus on the error region, and supplements the flow values of pixels which is warped by the corrected flow $\Delta \boldsymbol{F}^{i}$.

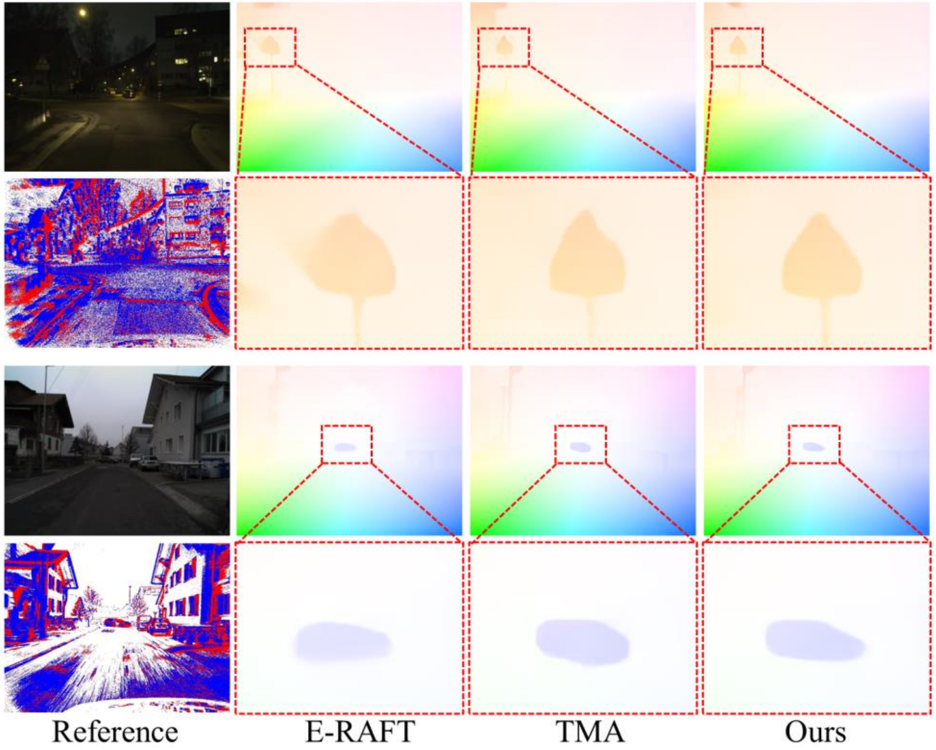


Method 3 – EEMFlow+

Results for Event-based Optical Flow Network：

- EEMFlow+ achieves the SOTA performances on DSEC dataset.
- EEMFlow+ improves inference speed by +4.19x compared to TMA[1] (7.55FPS → 39.2FPS).

Methods	FPS↑	1PE↓	2PE↓	3PE↓	EPE↓	AE↓
MutilCM [39]	-	76.6	48.5	30.9	3.47	14.0
EV-Flownet [52]	22.3	55.4	29.8	18.6	2.32	8.12
OF-EV-SNN [4]	-	53.7	20.2	10.3	1.71	6.34
EVA-Flow [50]	-	15.9	-	3.20	0.88	3.31
ERAFT [11]	11.4	12.7	4.74	2.68	0.79	2.85
ADMFlow [33]	9.88	12.5	4.67	2.65	0.78	2.84
EFlowformer [21]	-	11.2	4.10	2.45	0.76	2.68
TMA [25]	7.55	10.9	3.97	2.30	0.74	2.68
EEMFlow+(Ours)	39.2	11.4	3.93	2.15	0.75	2.67

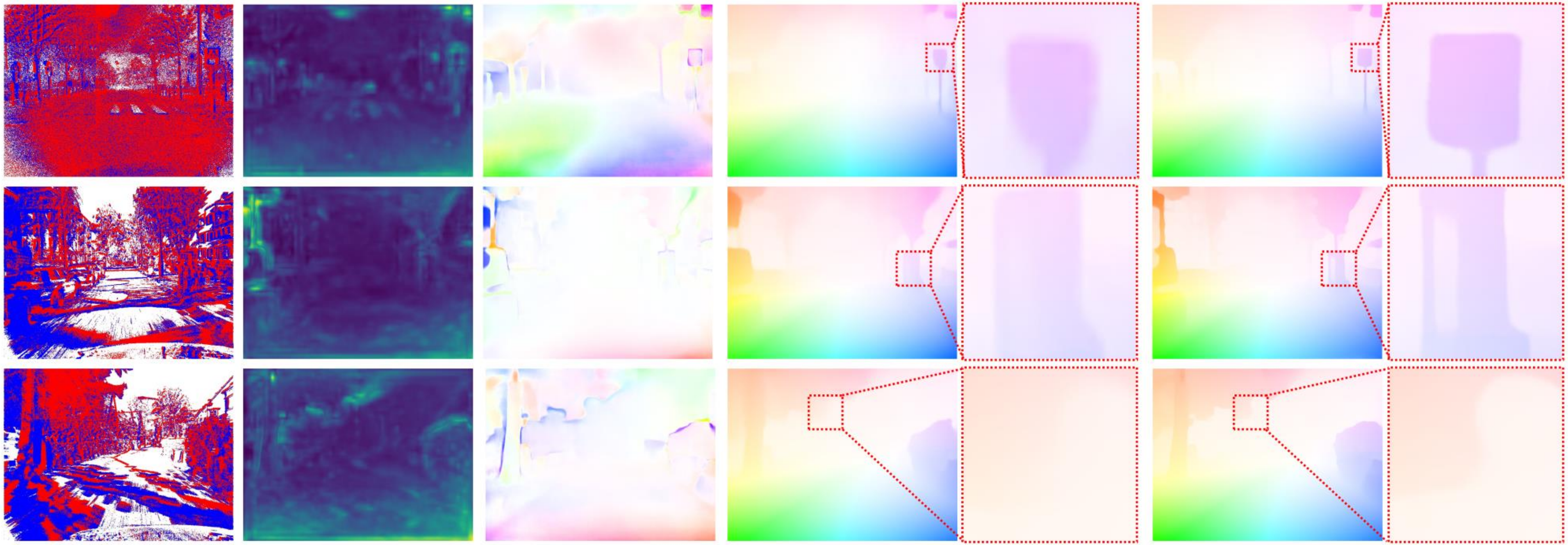


[1] Haotian L, et al. Tma: Temporal motion aggregation for event-based optical flow. Proc. of ICCV, 2023.

Method 3 – EEMFlow+

Ablation experiment for CDC of EEMFlow+:

- The flow prediction corrected by CDC has more object details and better edge contours.



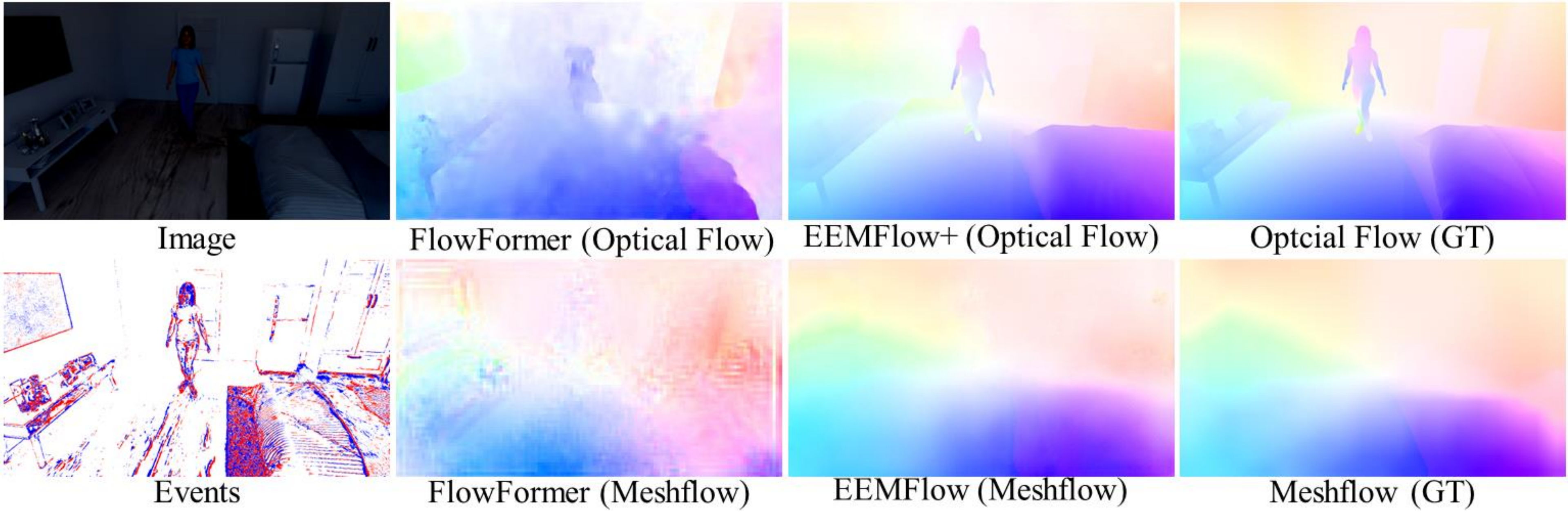
Events | Confidence Map | Corrected Flow | without CDC | with CDC

Method 3 – EEMFlow+

The Advantages of Event-Meshflow Estimation:

- Events over RGB images in extreme scenes.
- Event-meshflow networks are the most accurate and the fastest.

Task	Method	Outdoor		Indoor		Avg
		Slow	Fast	Slow	Fast	
Optical Flow	FlowFormer	6.20	16.06	5.99	15.27	10.88
	EEMFlow+	3.88	11.02	4.03	10.92	7.46
Mesh-flow	FlowFormer	5.99	15.12	5.74	14.95	10.45
	EEMFlow	2.42	9.09	2.00	8.46	5.50



Conclusion

- We are the **first** to study a new problem that estimates **meshflow** from **event camera**.
- We build **the first event-based meshflow dataset**, named as **HREM**, superior in the high resolution, dynamic scenes, complex motion patterns, and physically accurate events and meshflow label.
- We propose an **E**fficient **E**vent-based **M**esh**F**low network (**EEMFlow**), achieving **SOTA** performances and inference speed of **142.9 FPS** (25.5 to 38.7 times faster than compared methods).
- We propose a **C**onfidence-induced **D**etail **C**ompletion module (**CDC**), upgrading EEMFlow to **EEMFlow+** for optical flow estimation, achieving SOTA performances on DSEC dataset at high speed.

Thanks!