Paper ID #370

RGB-Event Fusion for Robust Lane Detection

Jingtao Dong, Hao Zhuang, Hao Yang, Liyuan Pan†

Beijing Institute of Technology

Introduction

Our task

Exploiting image and event data for robust lane detection in challenging scenarios.

Key words

Event Camera, Multimodal, Lane Detection, Autonomous Dataset.

Lack of synchronized image-event lane detection dataset

Dataset	Year	Modality	E-I Pair	MI	MS	ML	MM	Real	Annotations	Scale	Resolution
Caltech Lanes [2]	2008	RGB	✗	✗	✗	✗	✗	✓	Key point coordinates	1,225	640×480
Tusimple [49]	2017	RGB	✗	✗	✗	✗	✗	✓	Key point coordinates	6,498	1280×720
CULane [39]	2017	RGB	✗	✗	✗	✗	✓	✓	Key point coordinates	133,235	1640×590
VPGNet [22]	2017	RGB	✗	✗	✗	✓	✓	✓	Key point coordinates	21,097	346×260
LLAMAS [4]	2019	RGB	✗	✗	✗	✓	✓	✓	Pixel level	100,442	128×128
CurveLanes [61]	2020	RGB	✗	✓	✓	✓	✓	✓	Key point coordinates	150,000	1280×800
DET [9]	2020	Event	✗	✓	✓	✓	✗	✓	Pixel level	5,424	1280×800
DSEC-Lane	2025	RGB+Event	✓	✓	✓	✓	✓	✓	Key point coordinates	11,862	640×480

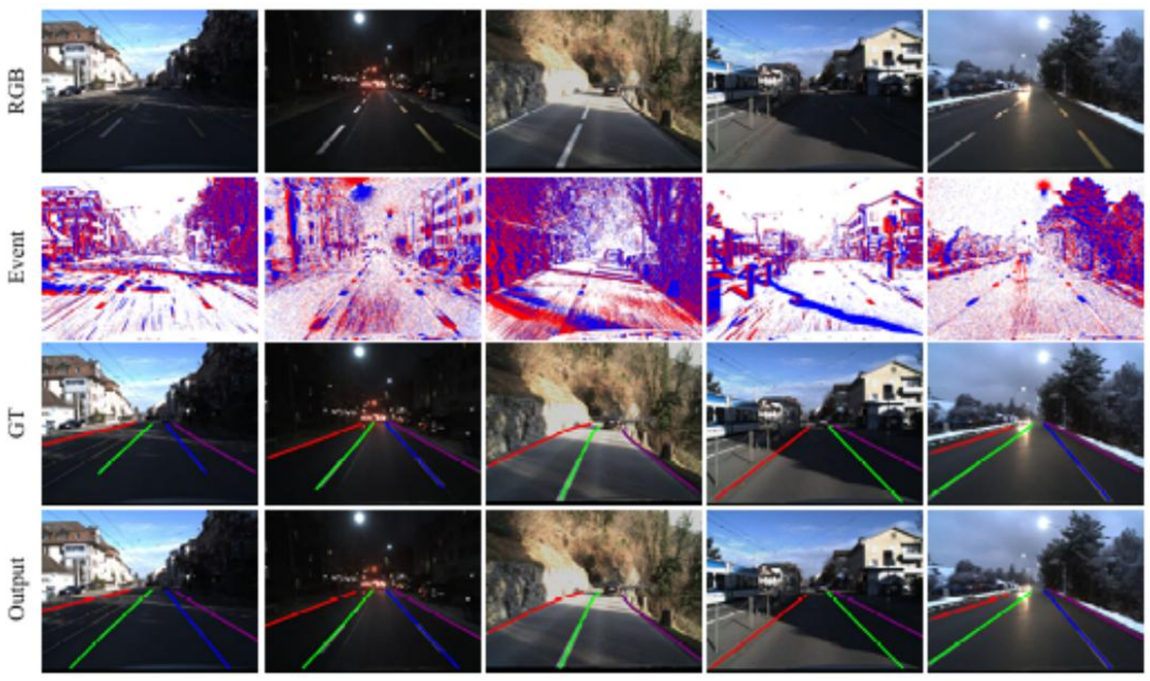
Definitions: E-I Pair (Event-Image Pairing), MI (Multi-Illumination), MS (Multi-Scenes), ML (Multi-Line Type), MM(Multiple Moments).

-**Collecting a** large-scale, high-quality real-world event-image lane detection dataset represents a challenge.

Lanes	One	Two	Three	Four	Total
Count	566	4033	4232	3031	11862
Percent (%)	4.77	34.00	35.68	25.55	100

Distribution of lane annotation counts per image in our dataset.

-The dataset contains manually annotated lane line instance point sets, with points sampled at consistent height intervals.



Examples of our DSEC-Lane dataset

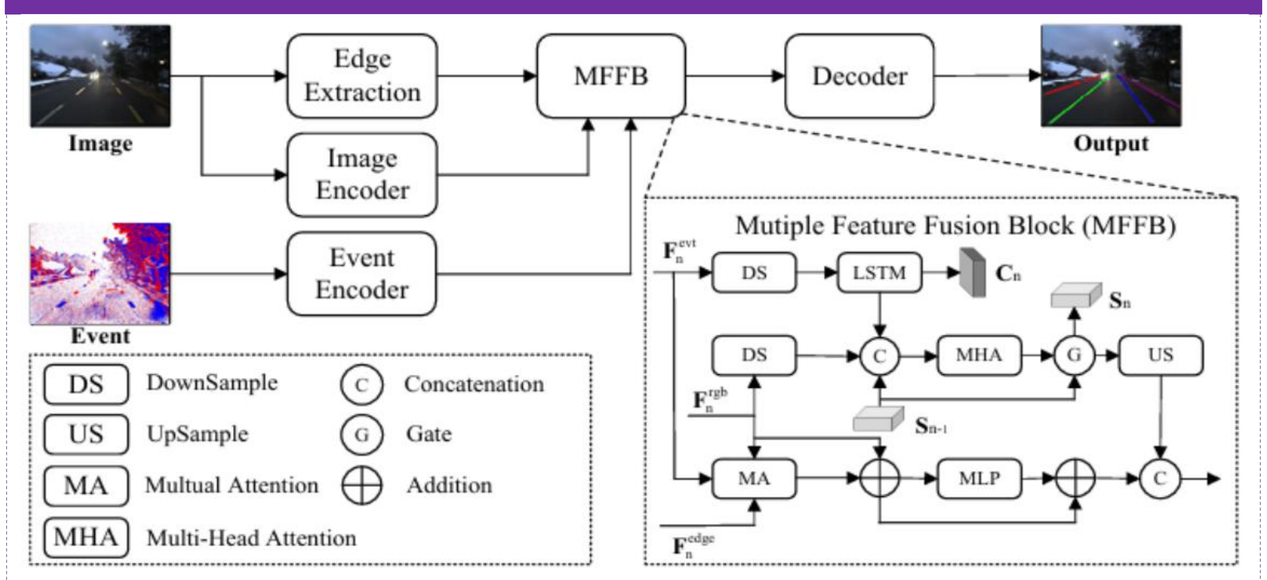
Ensure spatial and temporal sync of image and event data

$$P_{event,rgb} = K_{event} * R_{rgb} * R_{event,rgb} * R_{event}^{T} * K_{rgb}^{-1}$$

-We apply a homography transformation to map the scene from the RGB frame to the event camera frame, followed by annotating the lane lines on the transformed RGB image.

-We compress the raw event stream along the time dimension with Δt=50ms,which corresponds to the frame rate of an standard camera,allowing for better alignment.

Method



Contributions

- We introduce DSEC-Lane, a dataset with 11,862 samples for lane detection.
- We propose MFFB, fusing RGB frames and event streams using time-aware self-attention for robust lane detection.
- We establish the first RGB-Event fusion benchmark for lane detection, achieving state-of-the-art results on DSEC-Lane.

Results

Method	Overall		Normal		Overexposure		Underexposure	
	Acc.(%)↑	F1(%)↑	Acc.(%)↑	F1(%)↑	Acc.(%)↑	F1(%)↑	Acc.(%)↑	F1(%)↑
LaneATT [46]	84.80	73.42	85.48	70.04	83.16	73.11	85.05	74.21
CLRNet [66]	86.48	79.97	86.32	79.95	87.97	78.51	85.82	81.28
ADNet [60]	86.38	79.41	86.02	79.42	87.73	77.36	86.75	81.16
SRLane [6]	86.39	79.74	86.52	80.51	87.94	77.20	84.42	78.61
CLRKDNet [41]	86.62	79.32	86.48	79.70	87.86	77.07	86.10	79.69
Ours-S	86.92	81.41	86.67	81.48	87.99	**79.07**	86.10	**82.80**
Ours-B	**87.32**	**81.57**	**87.35**	**82.14**	**88.05**	78.03	**86.73**	82.21

Lane-detection results on the DSEC-Lane dataset. **Ours-S** and **Ours-B** represent the small and base versions of our proposed method.

Method	Acc. (%)↑	F1 (%)↑
Add	86.30	80.44
Concat	86.81	80.34
CBAM [57]	86.58	80.38
ECANet [54]	86.24	80.79
CMT [55]	86.54	80.70
Ours	**87.32**	**81.57**

Guidance	Operator	Acc. (%)↑	F1 (%)↑
w/o edge	-	86.00	80.49
w/ edge	Canny	86.45	80.91
	Sobel	**87.32**	**81.57**

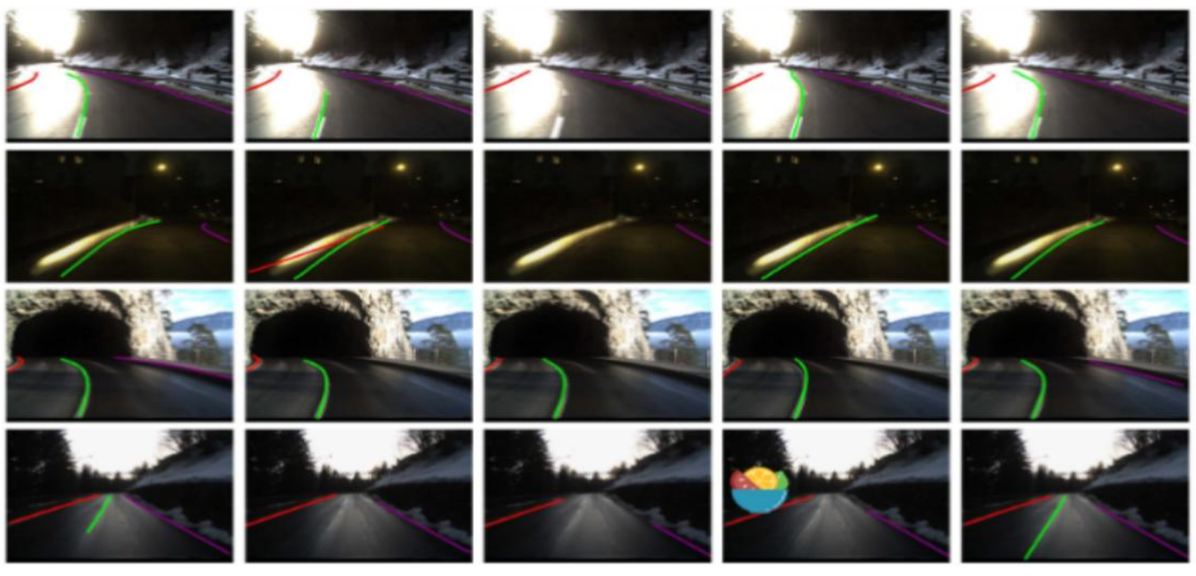
Ablation study on edge guidance and different Event and Image fusion methods.

Representation	Acc. (%)↑	F1 (%)↑
Voxel(bin=3)	86.01	80.38
Voxel(bin=6)	86.31	80.60
Ours(Timestamp)	**87.32**	**81.57**

Ablation study on different event representations.

Baseline	Time	Edge	Acc.(%)↑	F1(%)↑
✓	✗	✗	86.81	80.34
✓	✓	✗	86.70	81.27
✓	✗	✓	87.22	81.00
✓	✓	✓	**87.32**	**81.57**

Ablation study on model components.



GT LaneATT CLRNet CLRKDNet Ours