清華大學
Tsinghua University

Event-Enhanced Synthetic Aperture Imaging

Siqi Li, Shaoyi Du, Junhai Yong, Yue Gao

Submitted to Science China: Information Science

Introduction

- **Synthetic Aperture Imaging (SAI)**
  - **See through dense foreground occlusions and reconstruct scene appearance.**

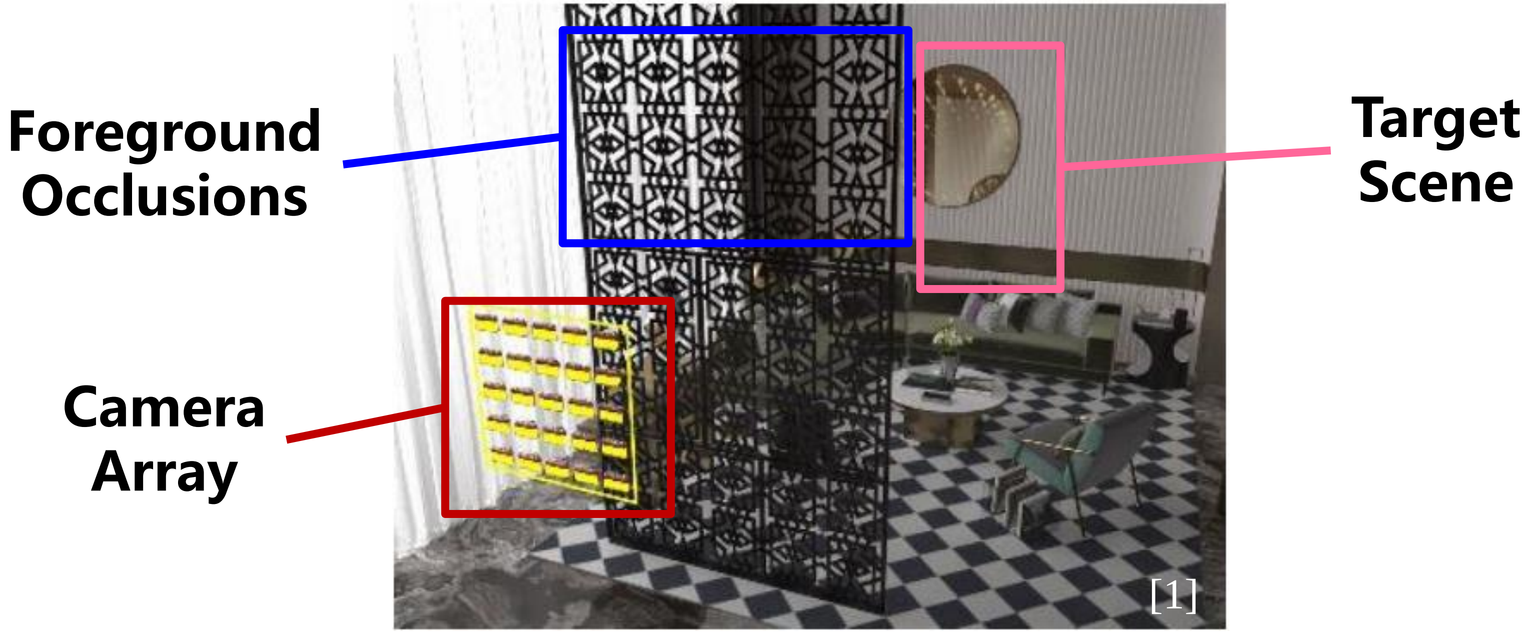
Input Occluded Visual Data

Unobstructed Scene Image

Introduction

Frame-based SAI Method

- **Using images captured by a camera array as input.**



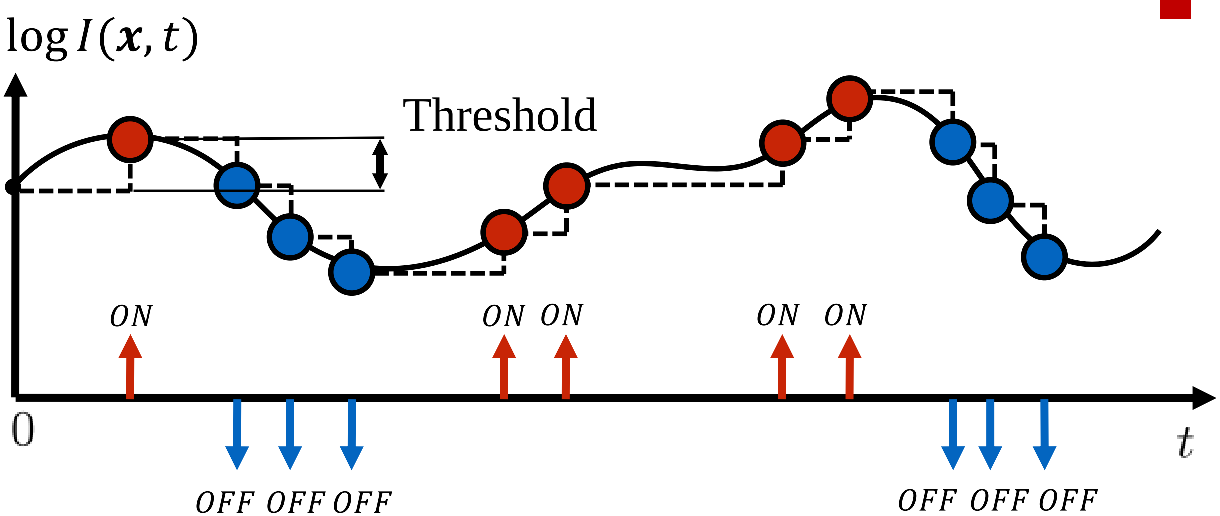
[1] Wang Y., Wu T., Yang J., *et al.* DeOccNet: Learning to See Through Foreground Occlusions in Light Fields. In: Proceedings of the IEEE WACV, 2020, 118-127.

Valid visual information is insufficient when facing dense occlusions due to the frame rate limitation

Introduction

Event Camera: Bio-Inspired Vision Sensor

- Response to pixel-wise brightness intensity changes
- Output in the form of event stream.



- **Advantages:**
  - High temporal resolution (about 1 μs)
  - High dynamic range (about 140 dB)
  - Low power consumption
  - Free from motion blur

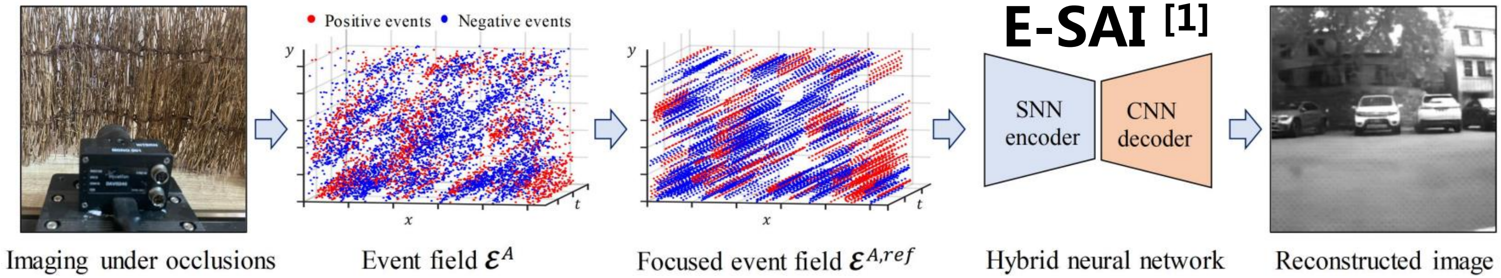
A moving event camera could record sufficient valid visual data of the occluded scene due to its high temporal resolution

Introduction

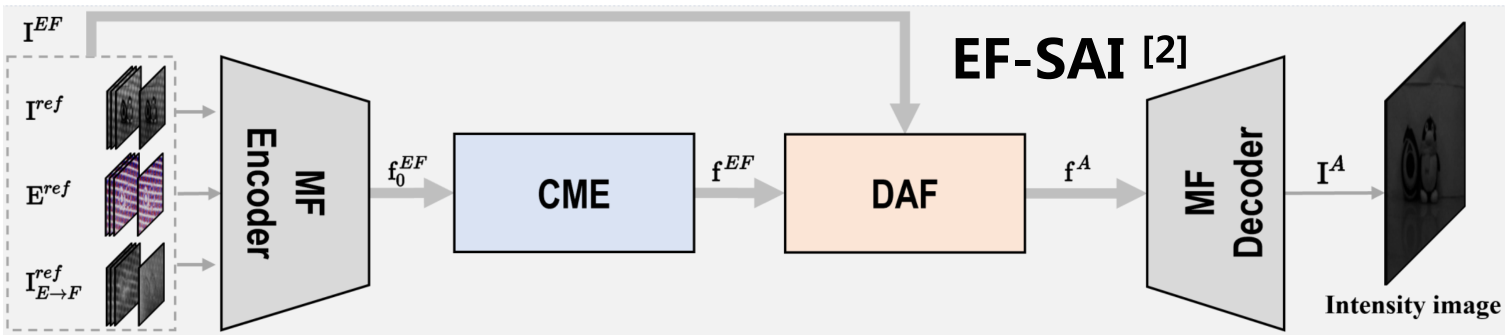
Event-based SAI Method

- **Using only events or events and frames as input.**

E-SAI [1]



EF-SAI [2]



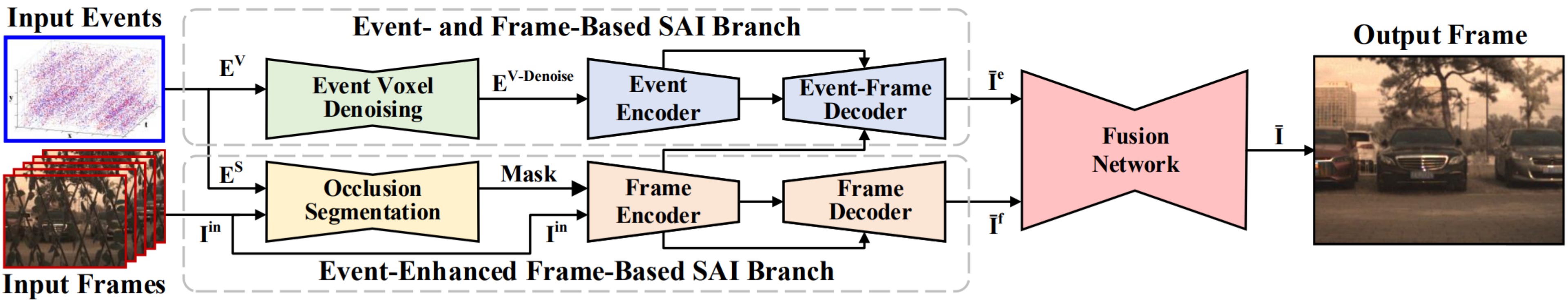
[1] Yu L., Zhang X., Liao W., *et al*. Learning to See Through with Events. IEEE T-PAMI, 2023, 45(7): 8660-8678.
[2] Liao W., Zhang X., Yu L ., *et al*. Synthetic Aperture Imaging With Events and Frames. In: Proceedings of the IEEE CVPR, 2021, 17735-17744.

Directly use occluded inputs leading to artifacts
Not robust to the noise caused by event camera

Method

Two-Branch Event-Enhanced SAI Method

- Takes occluded images and event stream as input



- The event-enhanced frame-based SAI branch aims to reconstruct scene image under sparse occlusions.
- The event- and frame-based SAI branch mainly focuses on dense occluded scenarios.
- The full model could achieve satisfactory performance under various occluded scenarios.

Method

Event- and Frame-Based SAI Branch

- More rely on visual information in occlusion frames, mainly targeting sparse occlusion environments
- Using event surface as guidance to predict occlusion masks of input occluded frames
- Event surface could be formulated as:

$$E_i^s = \sum_{i \in \{i | T_{i-1} < t_i < T_{i+1}\}} \delta(x - x_i, y - y_i),$$

- The input occluded frames are extracted using partial convolutional layers with the guidance of occlusion masks. Thus, invalid occlusions could be filtered out.

$$x_{out} = \begin{cases} \mathbf{W}^{\mathrm{T}}(\mathbf{F} \odot \mathbf{M}) \frac{\mathrm{sum}(\mathbf{1})}{\mathrm{sum}(\mathbf{M})} + b & , \mathrm{sum}(\mathbf{M}) > 0 \\ 0 & , \mathrm{otherwise} \end{cases},$$

Method

Event-Enhanced Frame-Based SAI Branch

- Uses event stream to provide extra visual information, mainly focus on dense occlusion scenarios.
- Uses an event representation method to convert event stream into grid-based event voxel.

$$t_i' = \lfloor (N_b - 1)\frac{t_i - T_{m-k}}{T_{m+k} - T_{m-k}} \rfloor$$

$$E^V(x, y, t) = \sum_i p_i \delta(x - x_i)\delta(y - y_i)\delta(t - t_i')$$

- Uses an event stream denoising module to generate denoised event voxel.
- Combine event feature and frame feature to obtain unobstructed scene image.

Method

Fusion Network

- **A UNet-like network to fuse the results from both branches and obtain final output.**

Loss Functions

- **Pixel-level Manhattan distance to maintain low-level features such as color and texture**

$$\mathcal{L}_{pix} = \frac{1}{H \cdot W \cdot C} \sum_{i,j} \| I_{i,j}^{out} - I_{i,j}^{gt} \|_1$$

- **Perceptual loss to maintain high-level vision features**

$$\mathcal{L}_{per} = \sum_{j} \frac{w_j}{H_j \cdot W_j \cdot C_j} \| \mathcal{F}_j(I^{gt}) - \mathcal{F}_j(I^{out}) \|_2^2$$

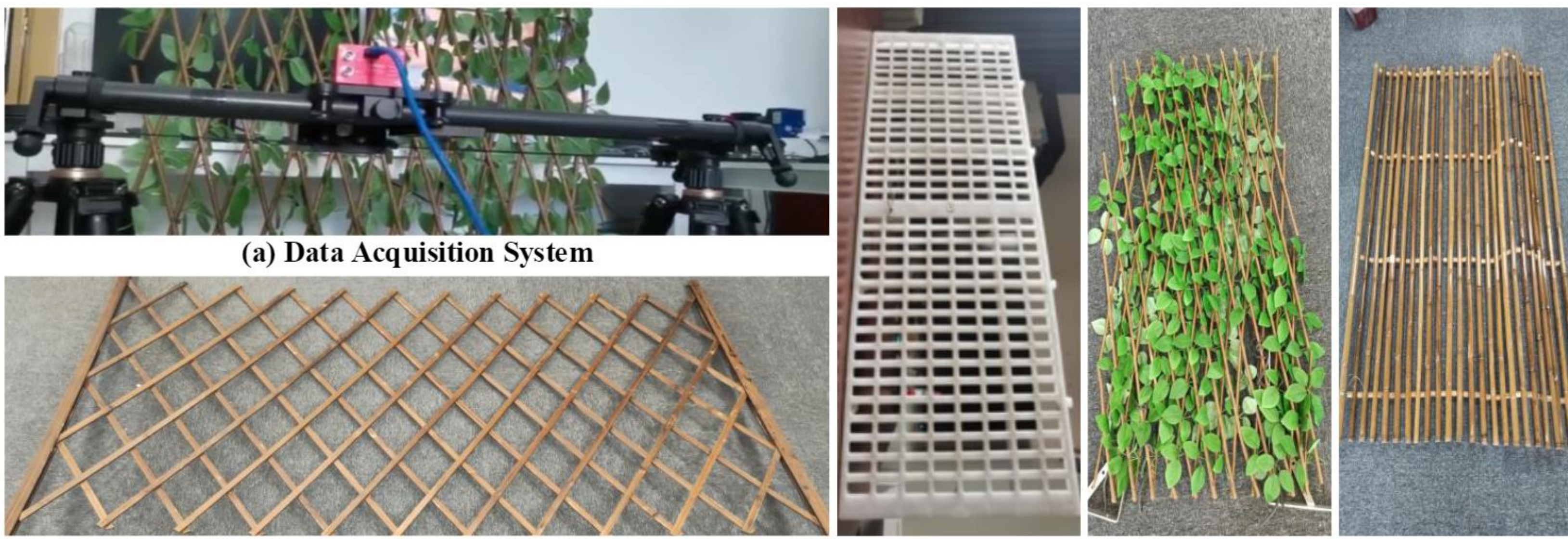
- **Total loss function**

$$\mathcal{L} = \mathcal{L}_{pix} + \lambda \mathcal{L}_{per}$$

$THU^{ERGB\text{-}SAI}$ Dataset

$THU^{ERGB\text{-}SAI}$ Dataset

- **We collect a real-world event-enhanced SAI dataset with larger scale and RGB information**



(a) Data Acquisition System

(b) Sparse Occlusion

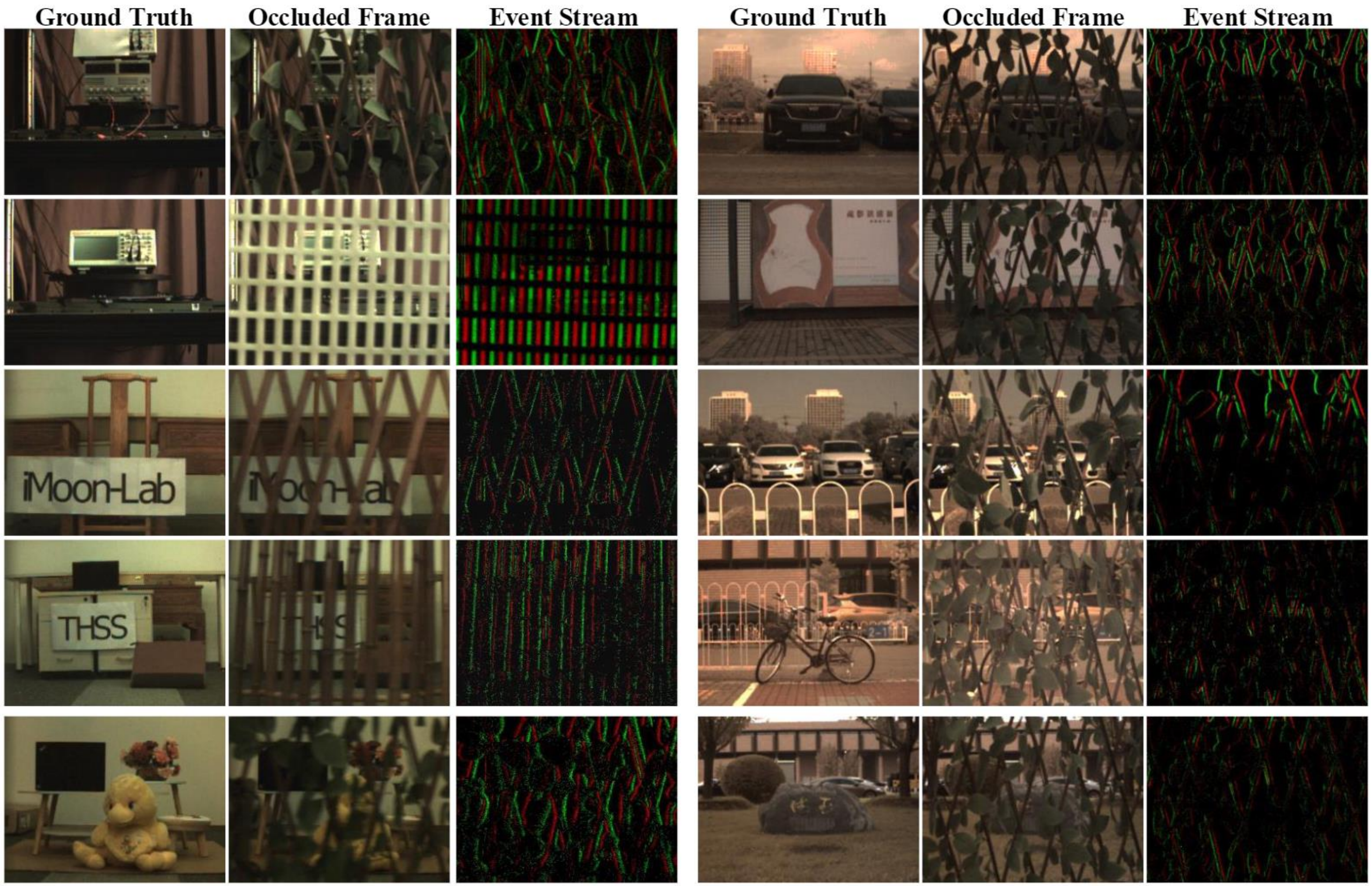
(c) Dense Occlusions

Dataset	Events	Frames	RGB	Type of Occlusions	Dataset Scale
E-SAI [26]	✓	✗	✗	1	588
EF-SAI [28]	✓	✓	✗	3	988
$THU^{ERGB\text{-}SAI}$ (Ours)	✓	✓	✓	4	2560

THU$^{ERGB-SAI}$ Dataset

THU$^{ERGB-SAI}$ Dataset

- Example samples from our dataset



(a) Indoor Scenarios (b) Outdoor Scenarios

Experiments

Quantitative Results

Quantitative results on existing EF-SAI dataset

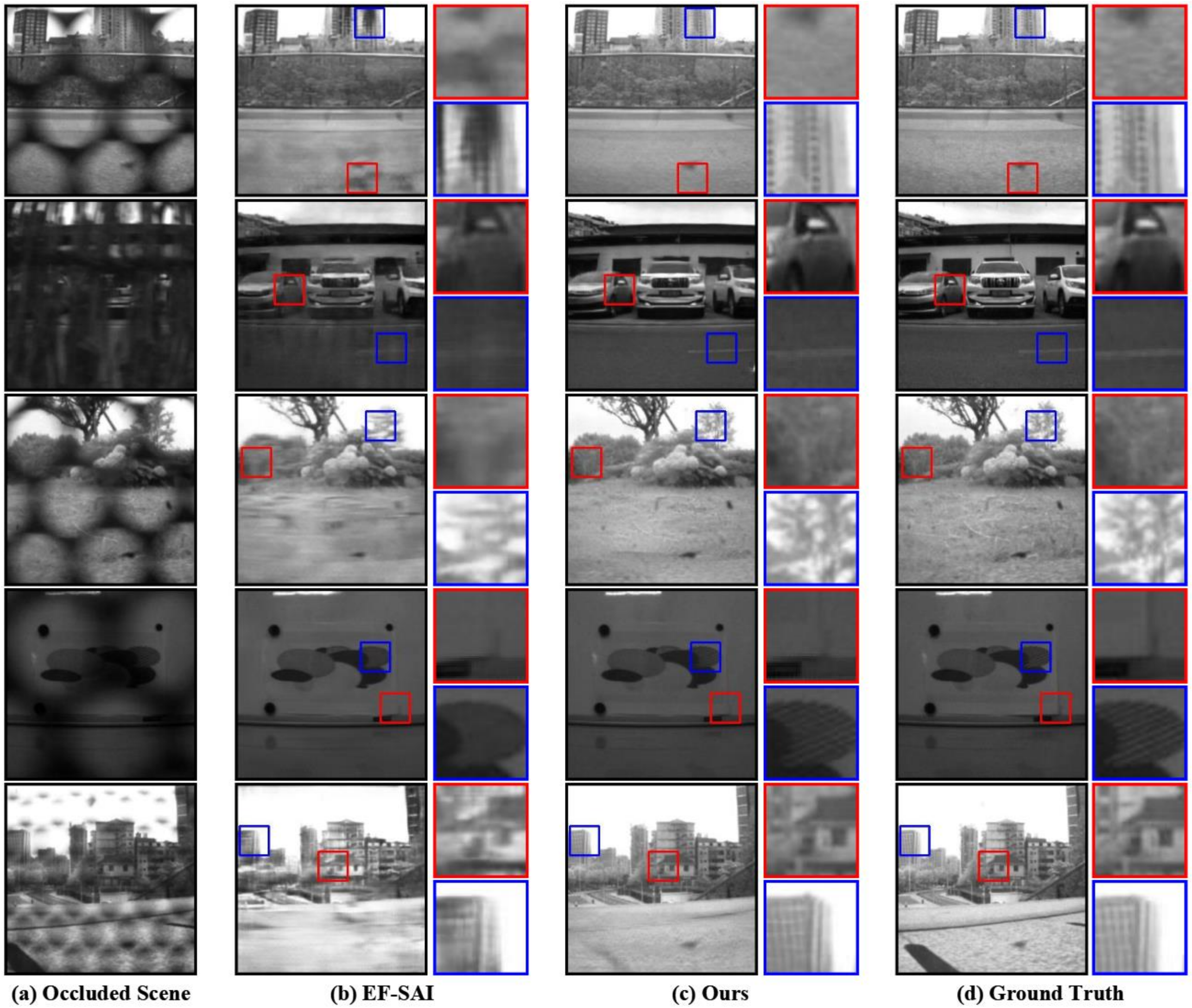
Method	Sparse Occlusions		Dense Occlusions		Total	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
DeOccNet [15]	20.27	0.748	22.84	0.728	21.10	0.741
E-SAI [26]	26.39	0.773	26.72	0.756	26.49	0.767
EF-SAI [28]	31.56	0.873	27.46	0.780	30.24	0.843
Ours	**33.52**	**0.916**	**27.82**	**0.844**	**31.67**	**0.893**

Quantitative results on our THU$^{ERGB\text{-}SAI}$ dataset

Method	Sparse Occlusions		Dense Occlusions		Total	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
DeOccNet [15]	26.37	0.793	25.97	0.776	26.11	0.784
E-SAI [26]	21.16	0.689	21.07	0.681	21.10	0.684
EF-SAI [28]	26.01	0.794	25.58	0.778	25.73	0.783
Ours	**27.39**	**0.824**	**27.18**	**0.817**	**27.25**	**0.819**

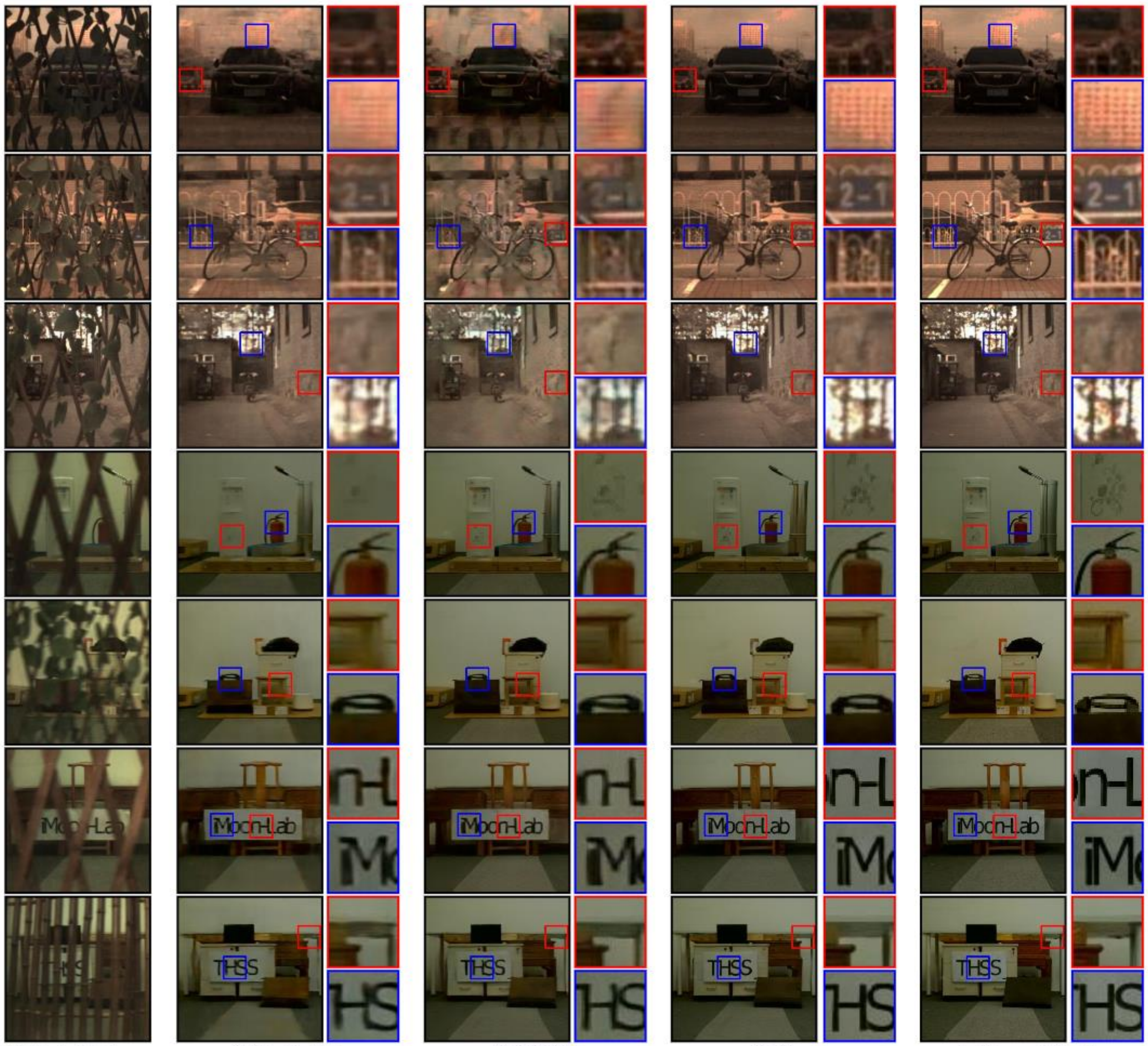
Experiments

- Qualitative results on EF-SAI dataset



Experiments

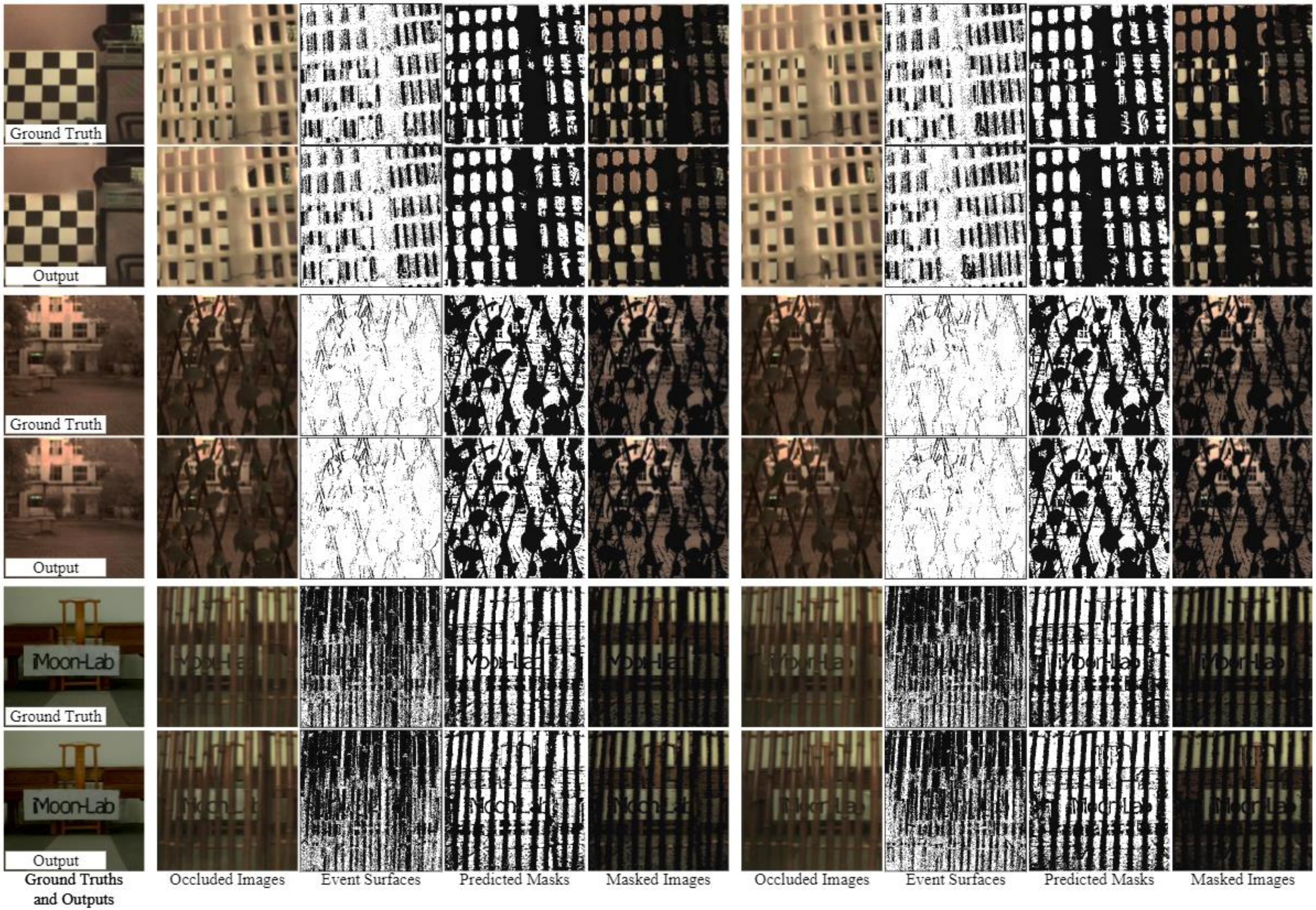
◻ Qualitative results on our THU$^{ERGB\text{-}SAI}$ dataset



(a) Occluded Scene (b) DeOccNet (c) EF-SAI (d) Ours (e) Ground Truth

Experiments

Occlusion segmentation results



Conclusion

Conclusion

- In this paper, we propose a two-branch event-enhanced SAI method for the image de-occlusion task, which contains an event-enhanced frame-based SAI branch, an event- and frame-based SAI branch, and a fusion module.
- The event-enhanced frame-based SAI branch generates SAI results with the guidance of the predicted occlusion mask, which is suitable for sparse occluded scenes. Our proposed method for the first time explicitly tackles the interference foreground occlusion, significantly boost SAI performance.
- The event- and frame-based SAI branch additionally uses event streams to provide extra visual information and could achieve better performance under dense occlusions.
- We collect and construct a large-scale event-enhanced RGB synthetic aperture imaging, which is over 2× larger than the existing largest dataset.