北京大学计算机学院
School of Computer Science

A Computational Photography Laboratory at Peking University
http://camera.pku.edu.cn

Event-based Vision Workshop @ CVPR'23

NeuCAP

Neuromorphic Camera Aided Photography

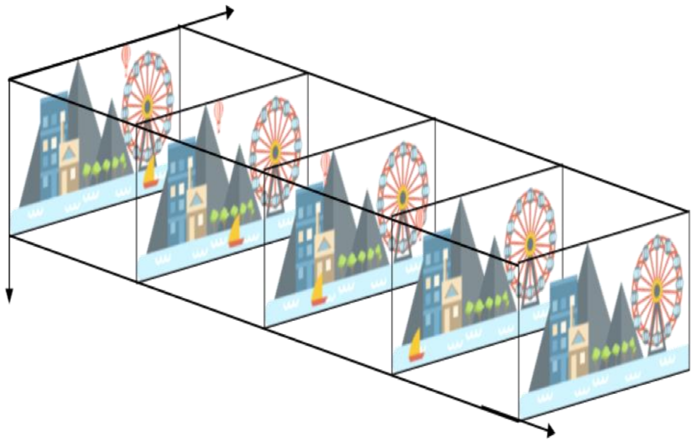
Boxin Shi

shiboxin@pku.edu.cn

* Courtesy of APSIPA Distinguished Lecturer Program

Frame-based Cameras are Problematic

The past 60 years of research have been devoted to frame-based cameras … but they are not good enough!

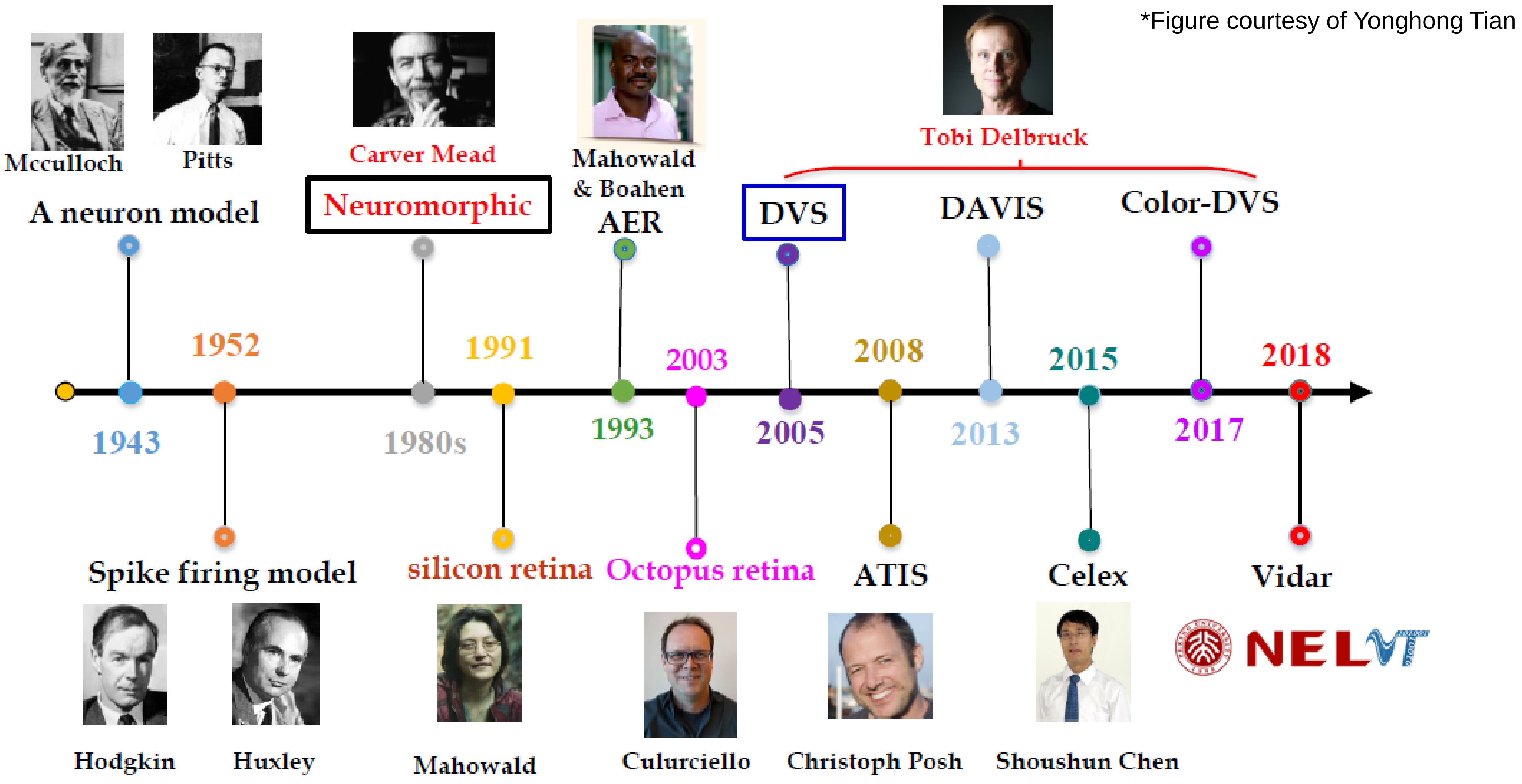


Redundant

High Latency

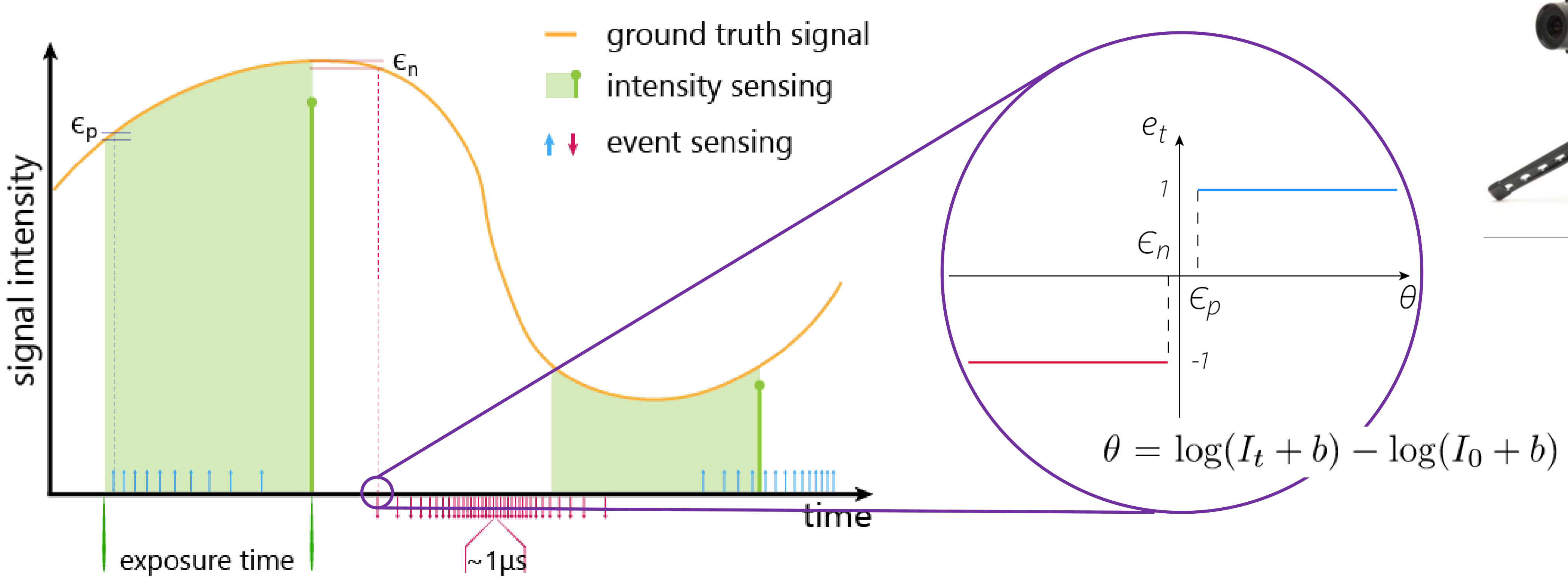
Low Dynamic Range

Neuromorphic Vision



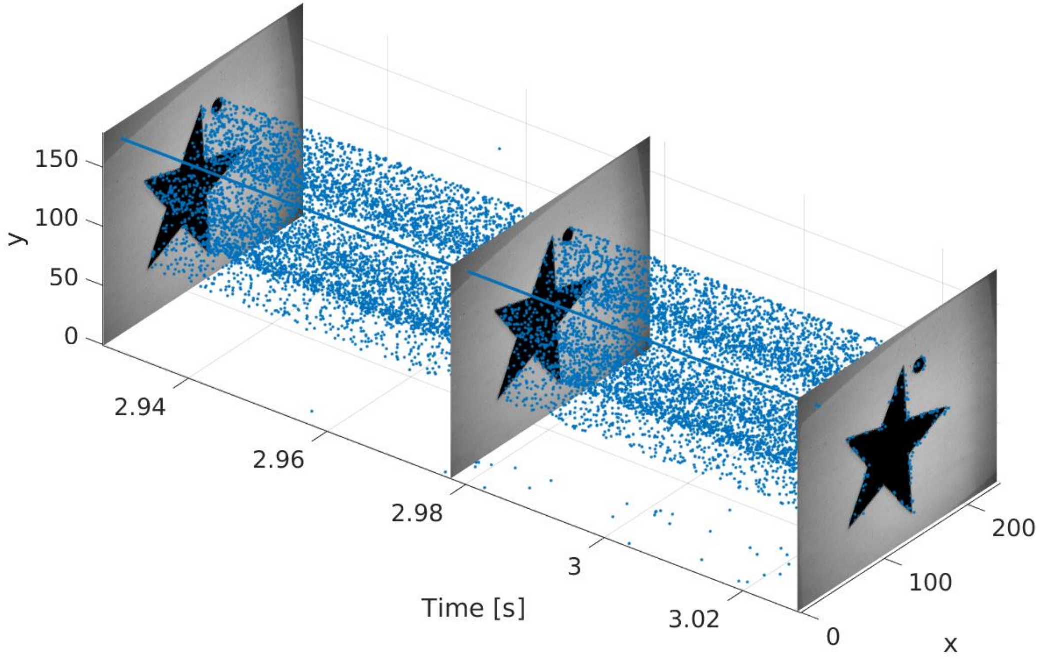
Neuromorphic Camera Example 1: Event Camera

- Events are generated any time a single pixel sees a change in brightness larger than the threshold.



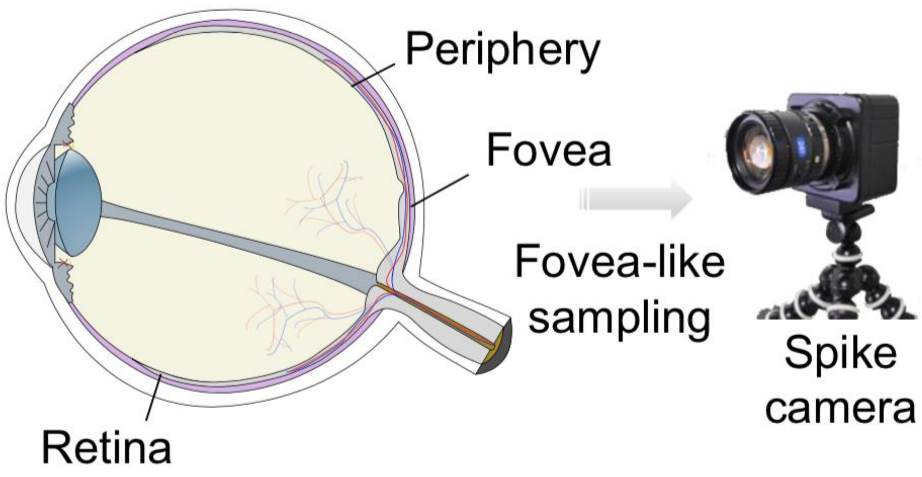
Neuromorphic Camera Example 1: Event Camera

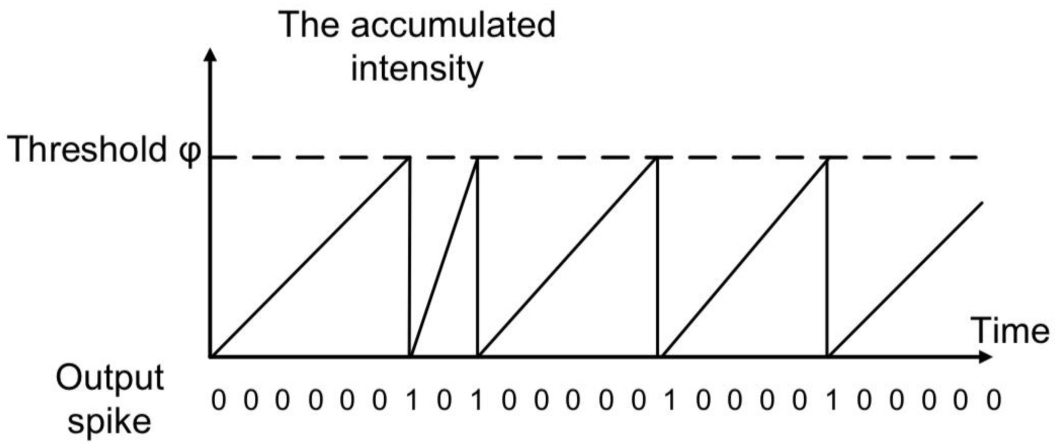
- Event camera has novel sensor that measures only motion in the scene. First commercialized in 2008 under the name of Dynamic Vision Sensor (DVS).



*Video courtesy of Elias Mueggler

Neuromorphic Camera Example 2: Spike Camera





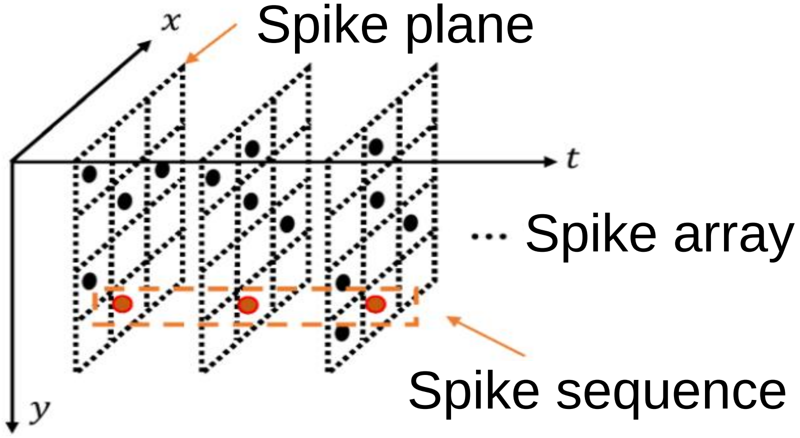
$$\int I dt \geq \phi$$

[Zhu et al., CVPR'20]

- For a pixel, the light intensity is accumulated, if the accumulated intensity reaches the dispatch threshold φ, a spike is fired and the accumulator is reset.

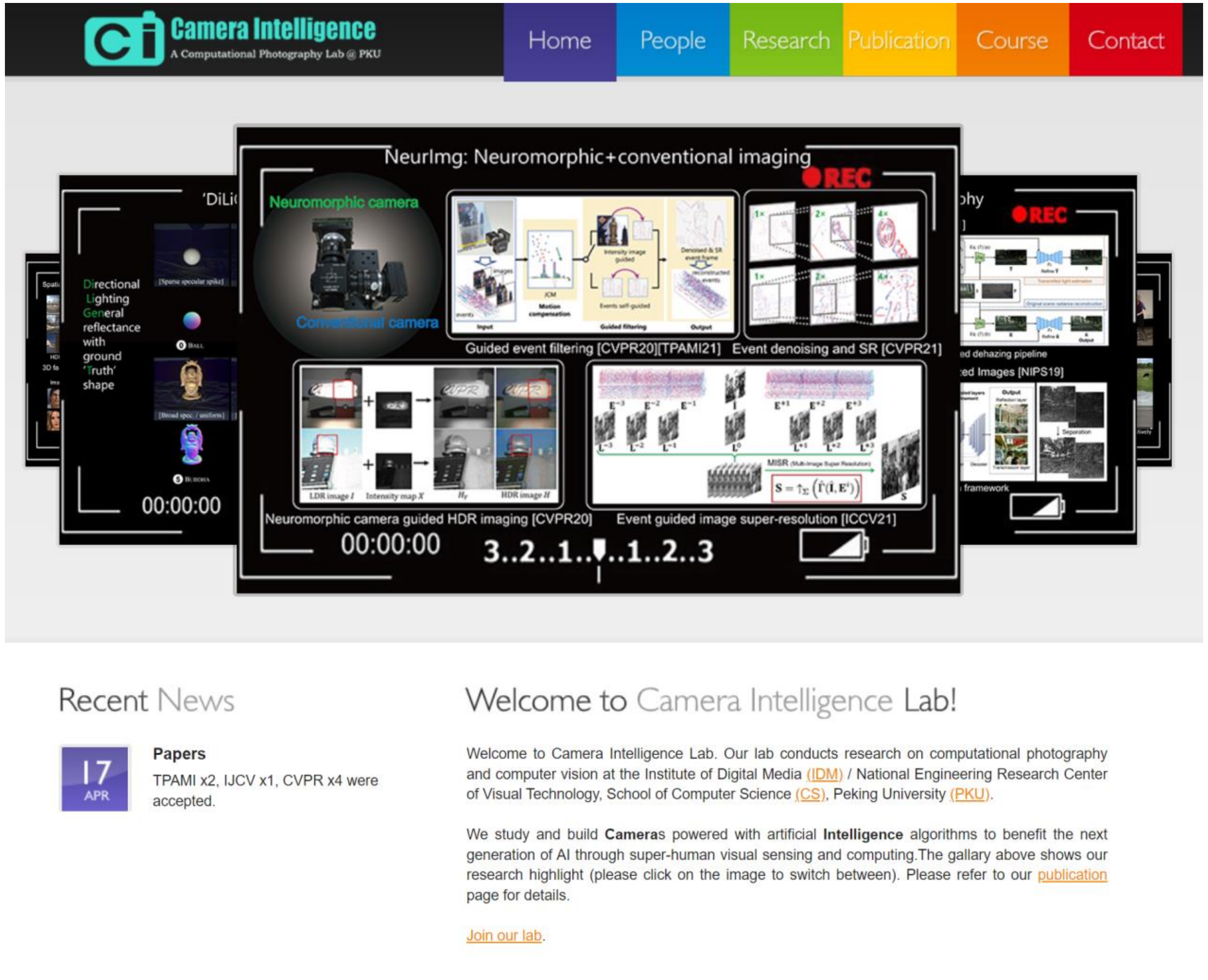
Neuromorphic Camera Example 2: Spike Camera

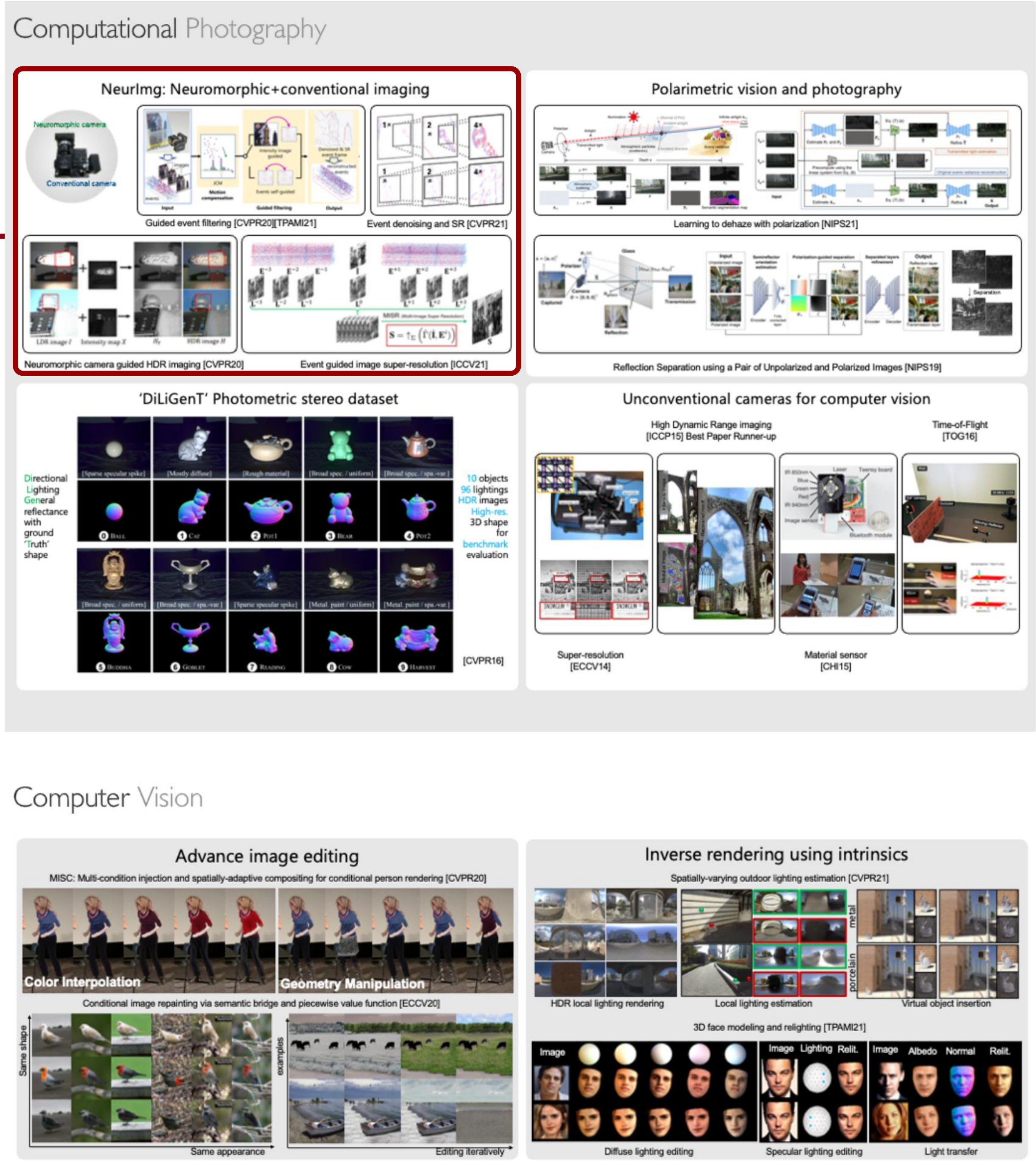
- The first spike camera is designed by **Peking university in 2018**, with spatial resolution of 400x250 and temporal resolution of 40KHz.



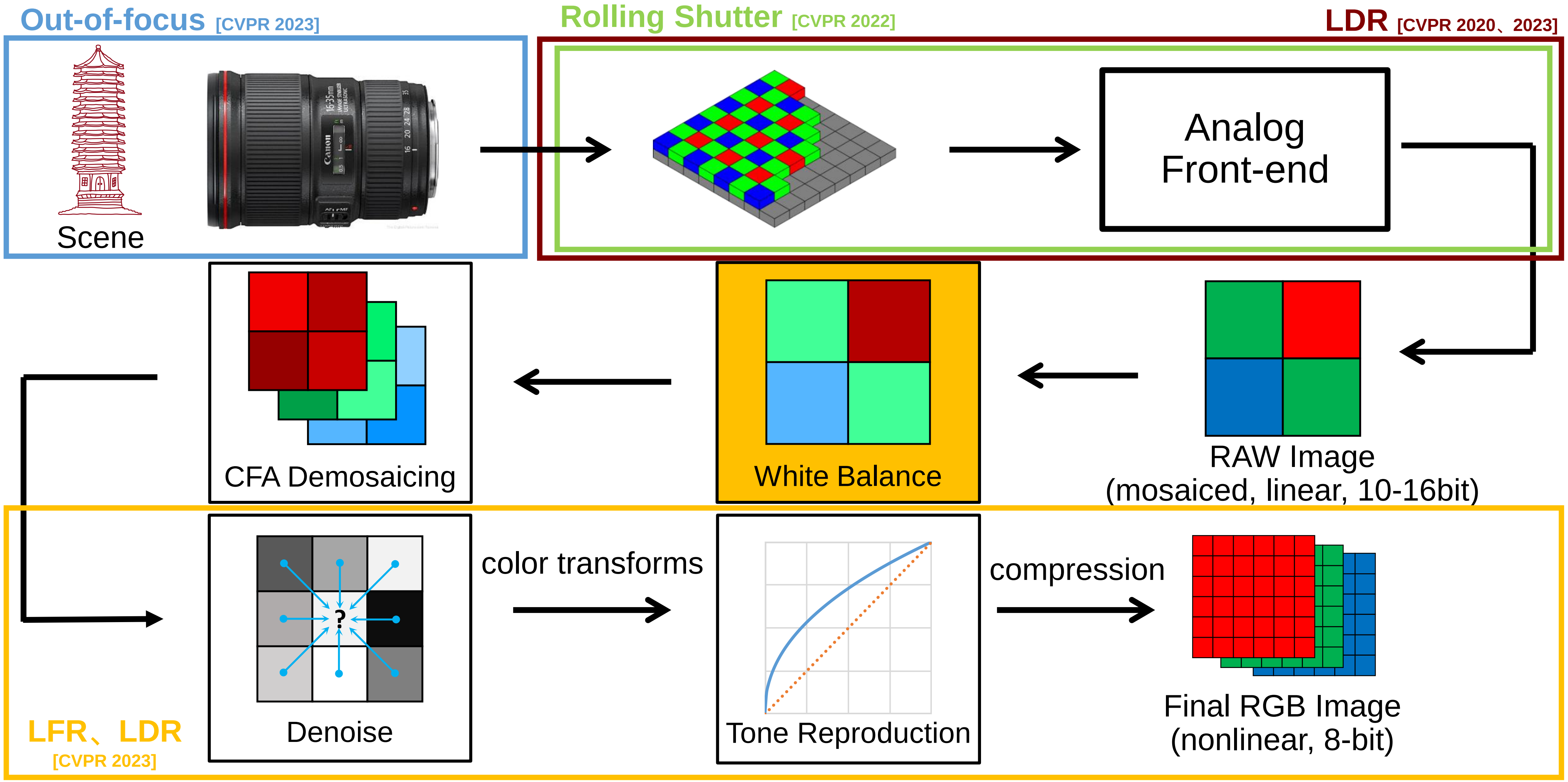
http://camera.pku.edu.cn

Camera Intelligence @ PKU



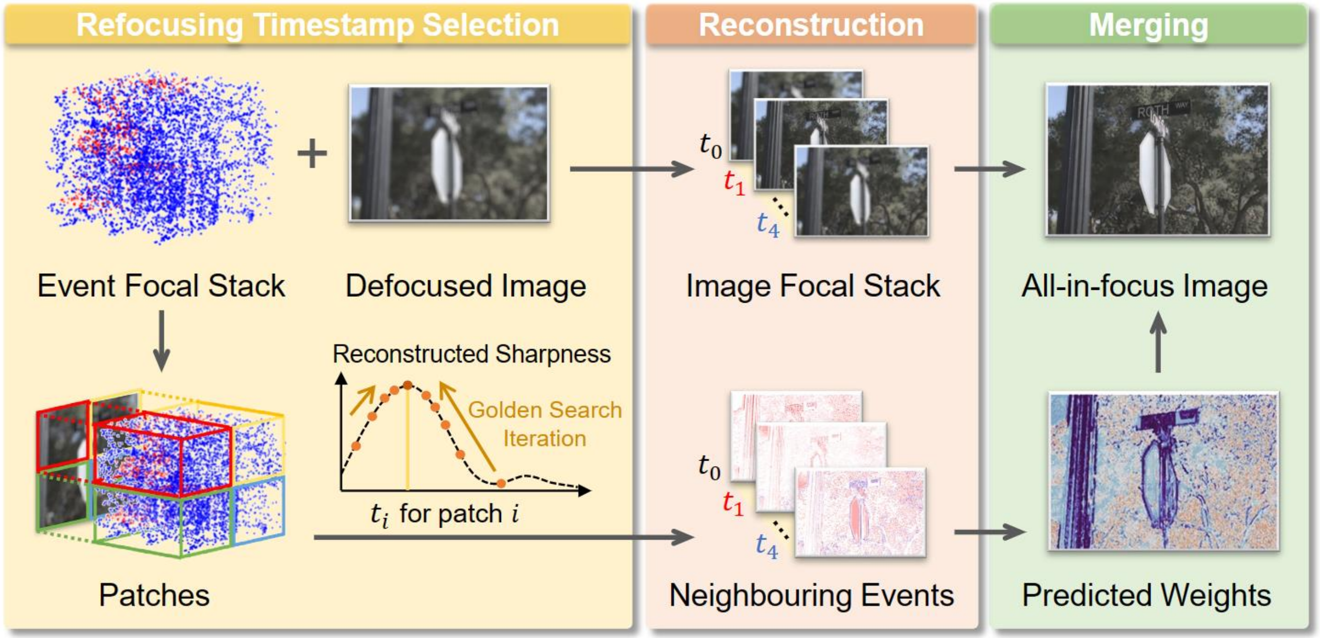


Camera Pipeline



Camera Pipeline

Out-of-focus [CVPR 2023]



Event Focal Stack

All-in-focus Imaging

CVPR 2023

All-in-focus Imaging from Event Focal Stack

Hanyue Lou[1,2 #] Minggui Teng[1,2 #] Yixin Yang[1,2] Boxin Shi[1,2]

[1] National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University

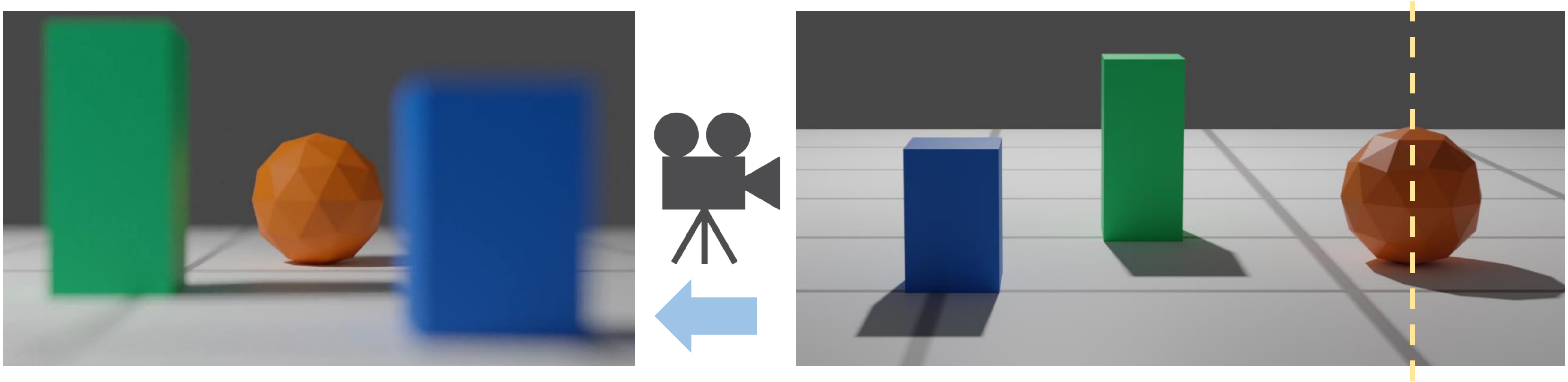
[2] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

[#] Contributed equally to this work as first authors

Motivation

Traditional Camera

Scene



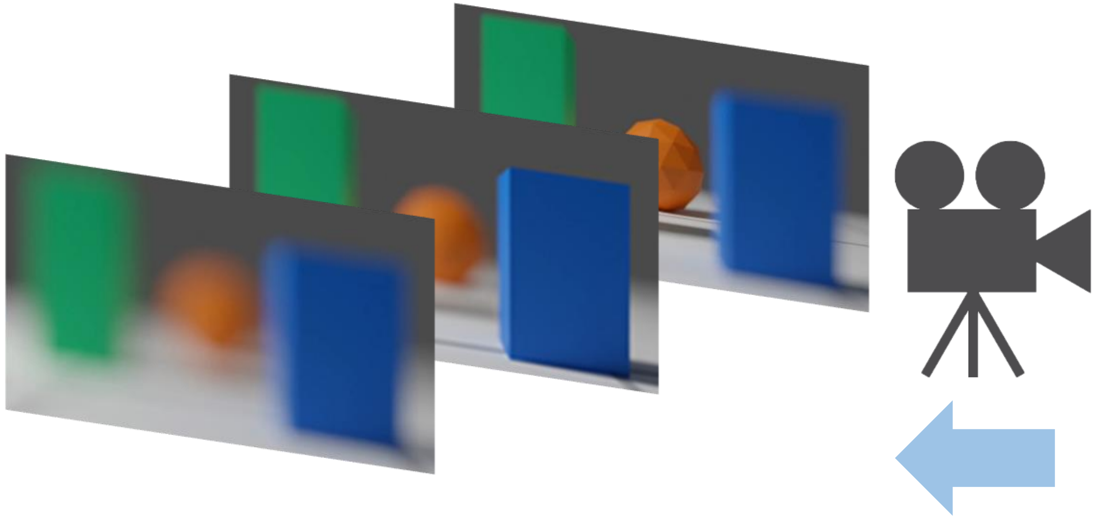
One depth in focus

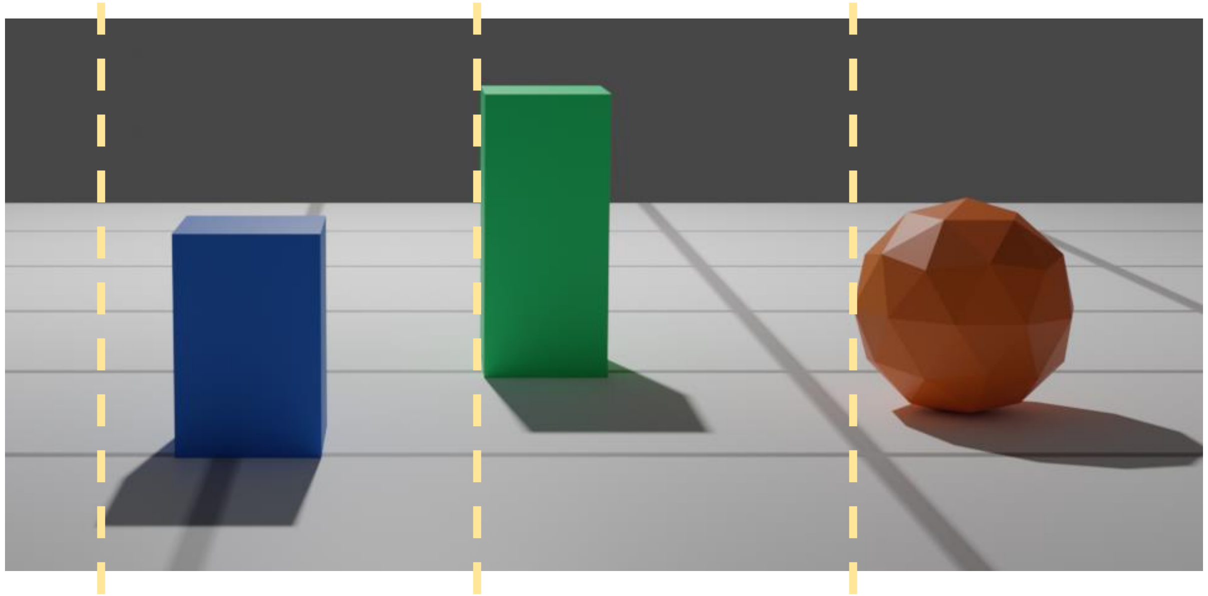
Focal Depth

Motivation

Frame-based Focal Stack

Scene



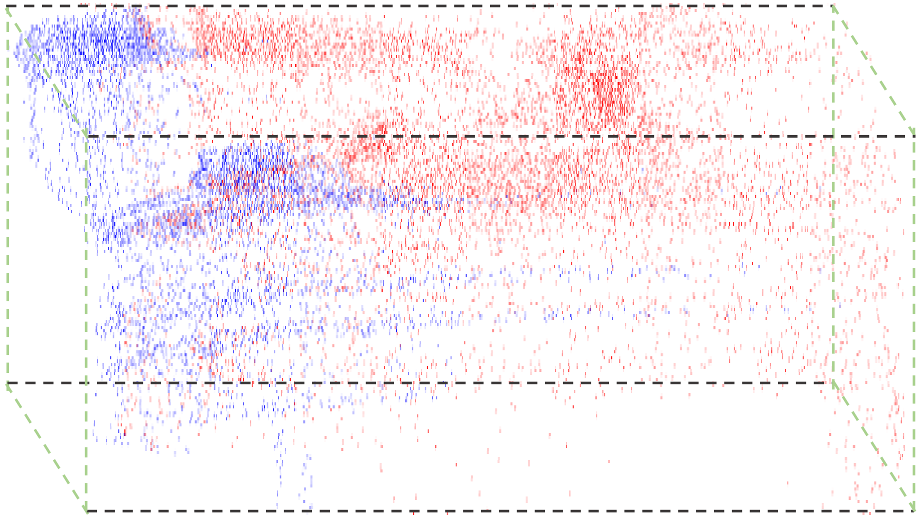


Multiple depths in focus

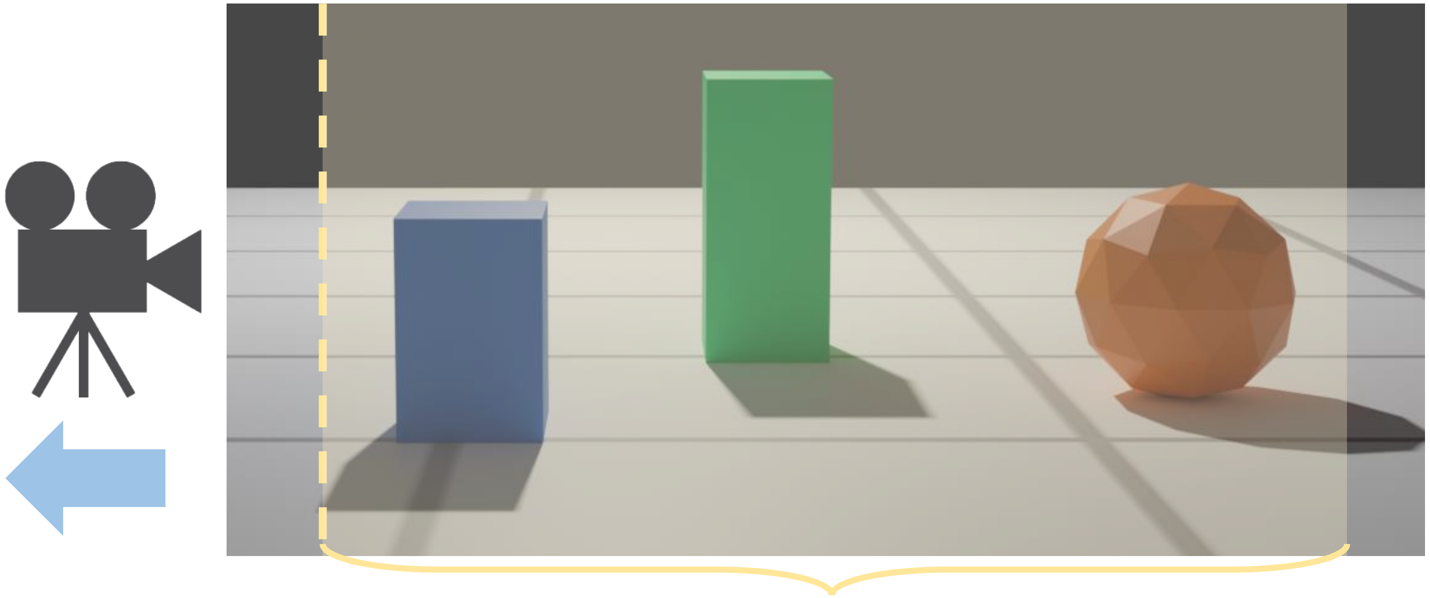
Focal Depths

Motivation

Event Focal Stack



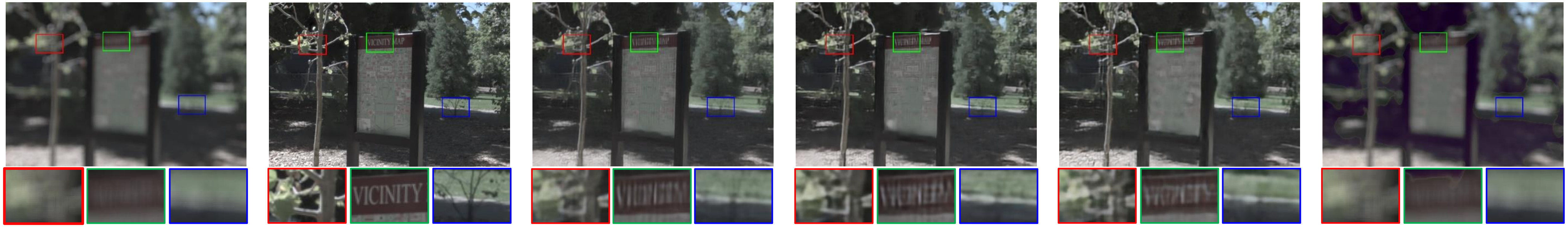
Scene



All depths in focus

Focal Depths

Previous Works



Defocused Image | Ground Truth | DRBNet | IFAN | KPAC | APL

Image-based methods: High frequency information cannot be recovered due to the ill-posed nature of single image defocus deblur.

Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Previous Works

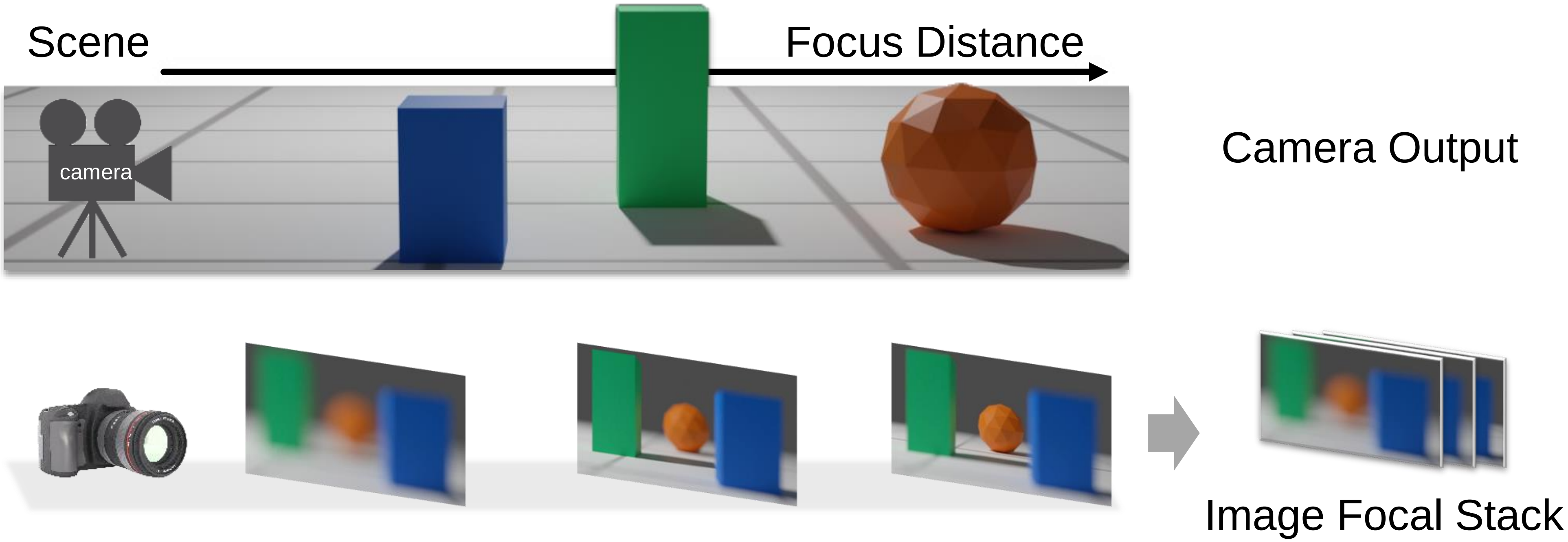
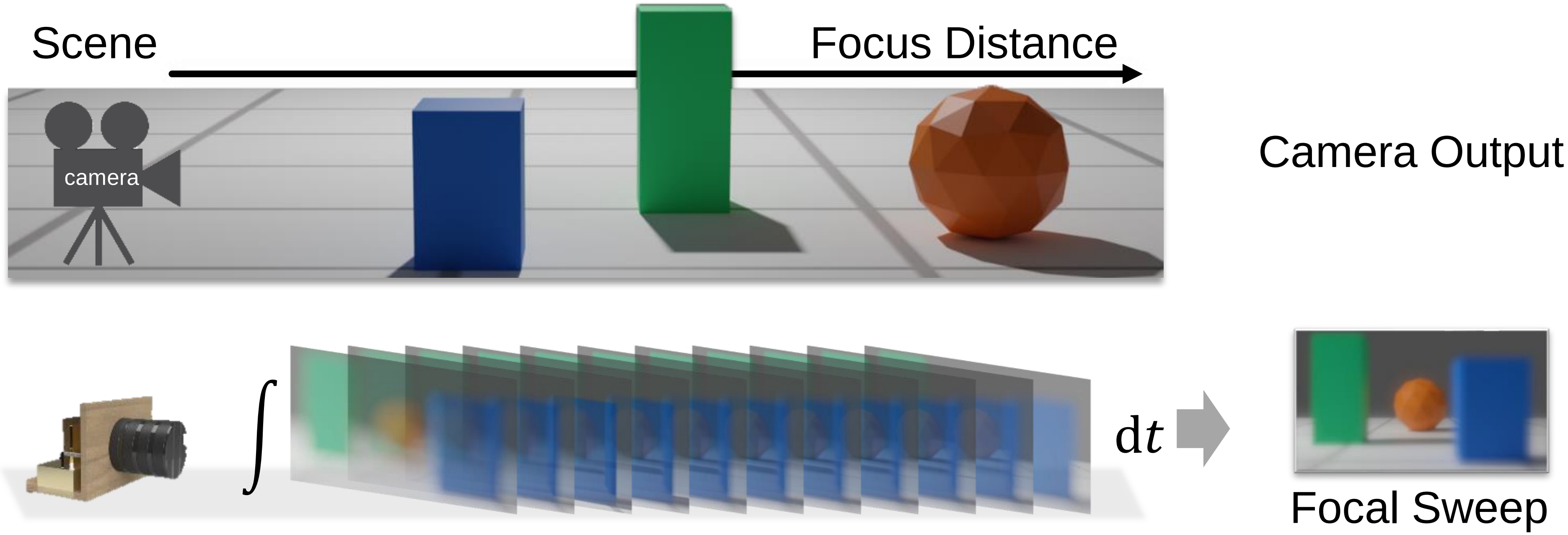


Image focal stack:

Requires careful operation & consumes long time;

Information is only collected at a discrete number of focal distances.

Previous Works



Focal sweep:
By sweeping the focal point in a single exposure, information is collected from all depths, but the depth information is lost.

Problem Formulation

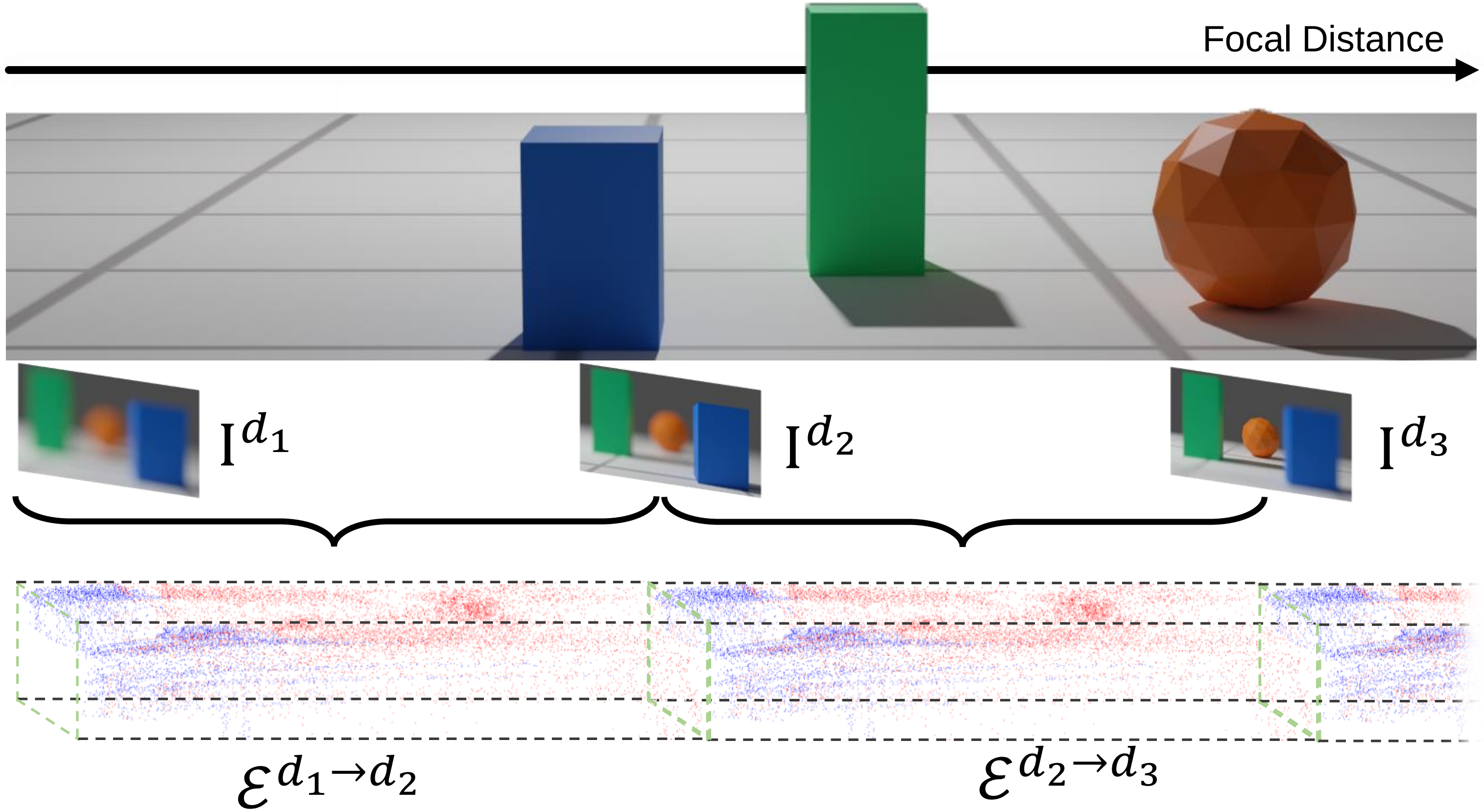
$$\log(\mathrm{I}^{d_j}) = \log(\mathrm{I}^{d_i}) + \sum \mathcal{E}^{d_i \to d_j}$$

The latent intensity frame changes as the focal point moves.
Events encode the logarithm change of latent intensity.

Scene

Latent Intensity Frame

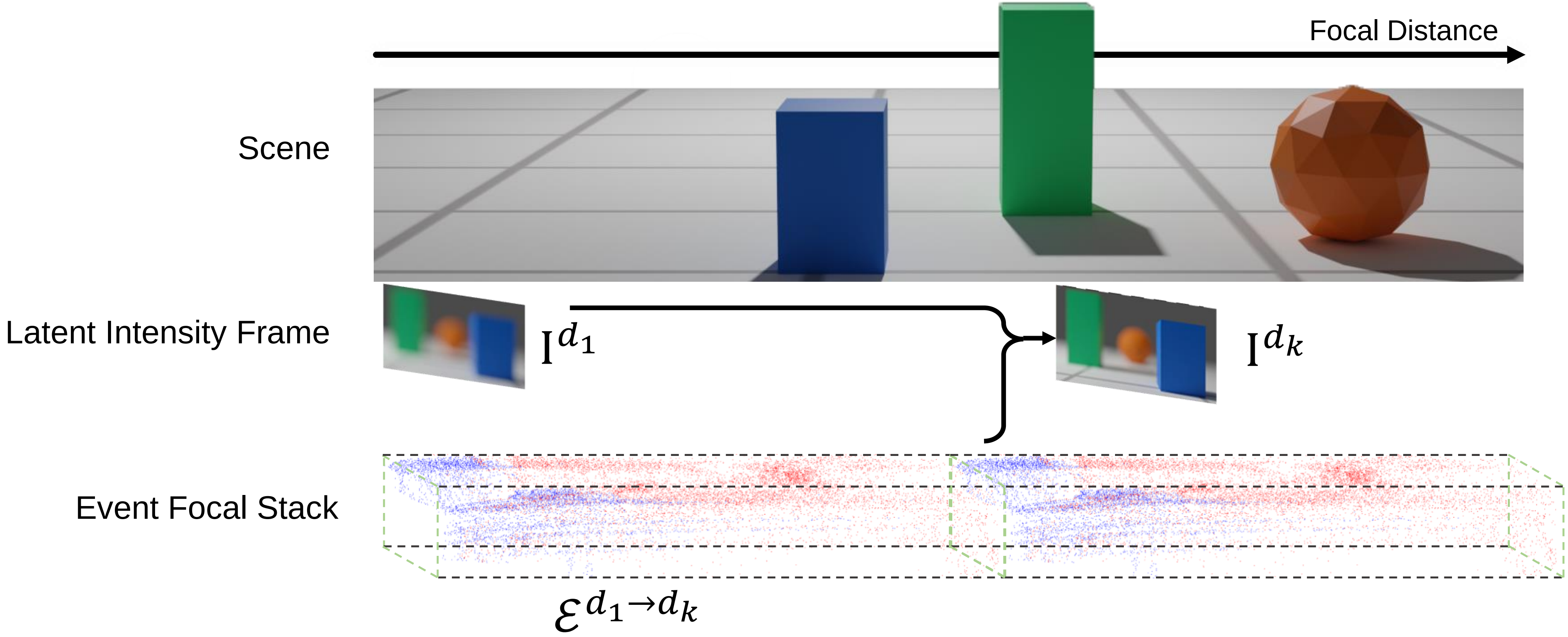
Event Focal Stack



Problem Formulation

$$\log(\mathrm{I}^{d_j}) = \log(\mathrm{I}^{d_i}) + \sum \mathcal{E}^{d_i \to d_j}$$

Given a defocused image and the EFS, we can reconstruct the latent image focused at any distance.



Method

Single image + EFS

↓ Golden search

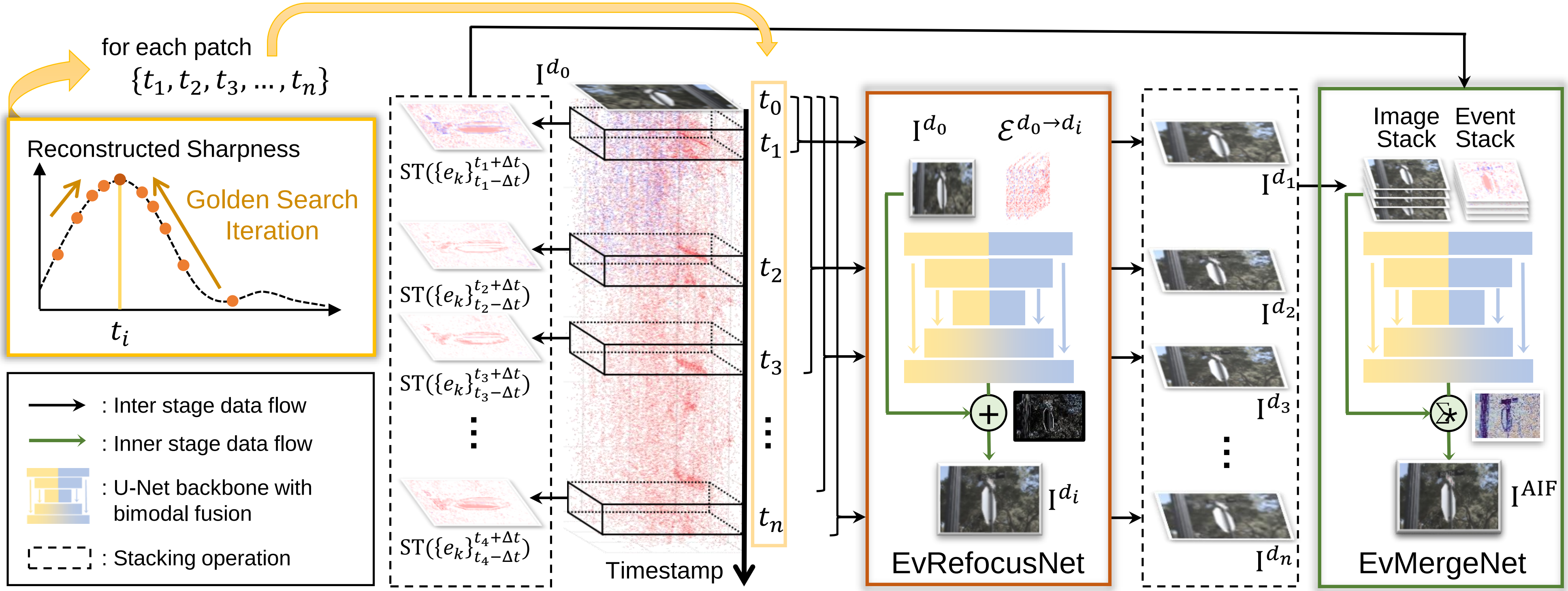
Refocus timestamps

↓ Reconstruction

Image focal stack

↓ Merge

All-in-focus Image

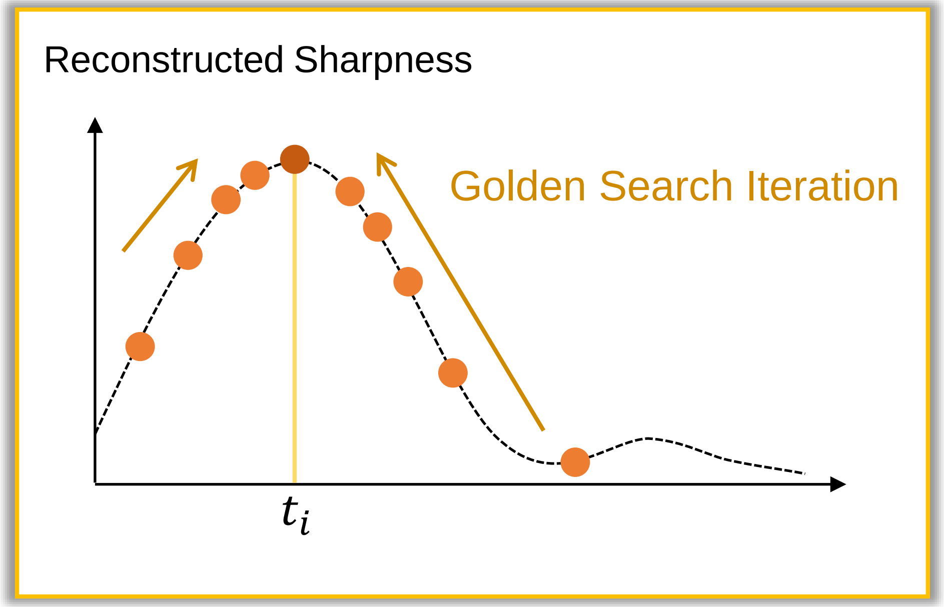


Golden Search for Refocus Timestamps

Algorithm 1 Refocusing time selection with EFS

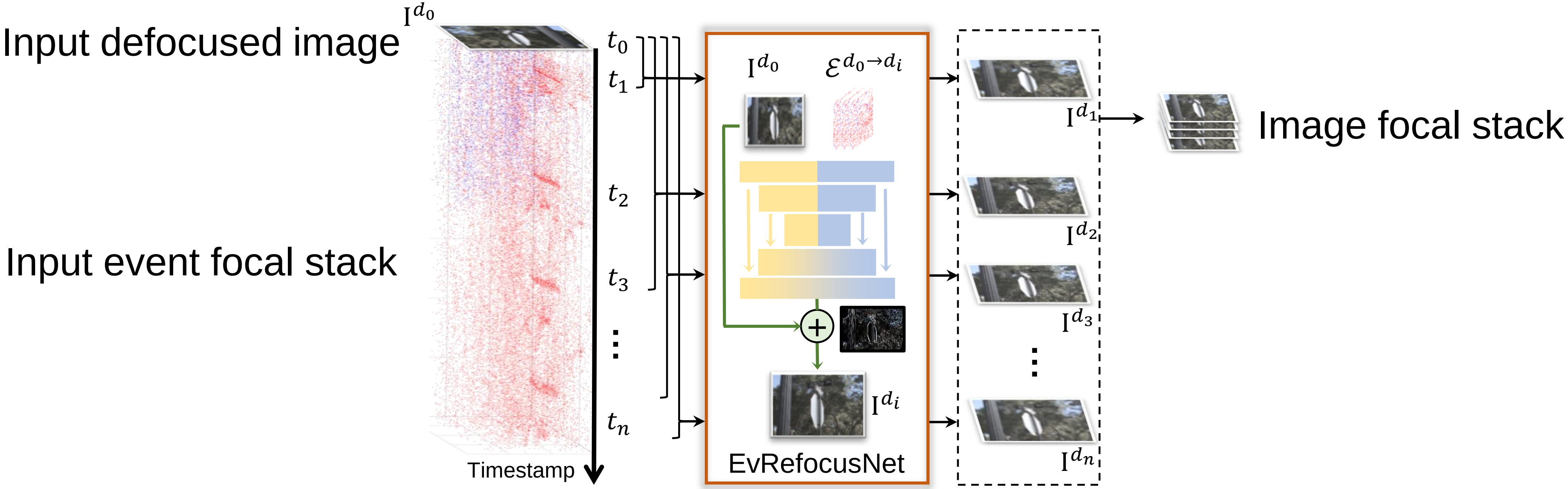
Data: threshold μ, golden ratio φ = 1.618
Input: EFS $\mathcal{E}$ and an RGB image $\mathbf{I}^d$
Result: Refocusing timestamp t_r

$L \leftarrow 0, R \leftarrow N_e$
while $R - L > \mu$ **do**
  $t_1, t_2 \leftarrow R - (R - L)/\varphi, L + (R - L)/\varphi$
  Reconstruct $\mathbf{I}^{d_1}, \mathbf{I}^{d_2}$ with Equation (5)
  if $\mathbb{D}(\mathbf{I}^{d_1}) > \mathbb{D}(\mathbf{I}^{d_2})$ **then** $R \leftarrow t_2$
  else $L \leftarrow t_1$
  end if
  $t_r \leftarrow (L + R)/2$
end while



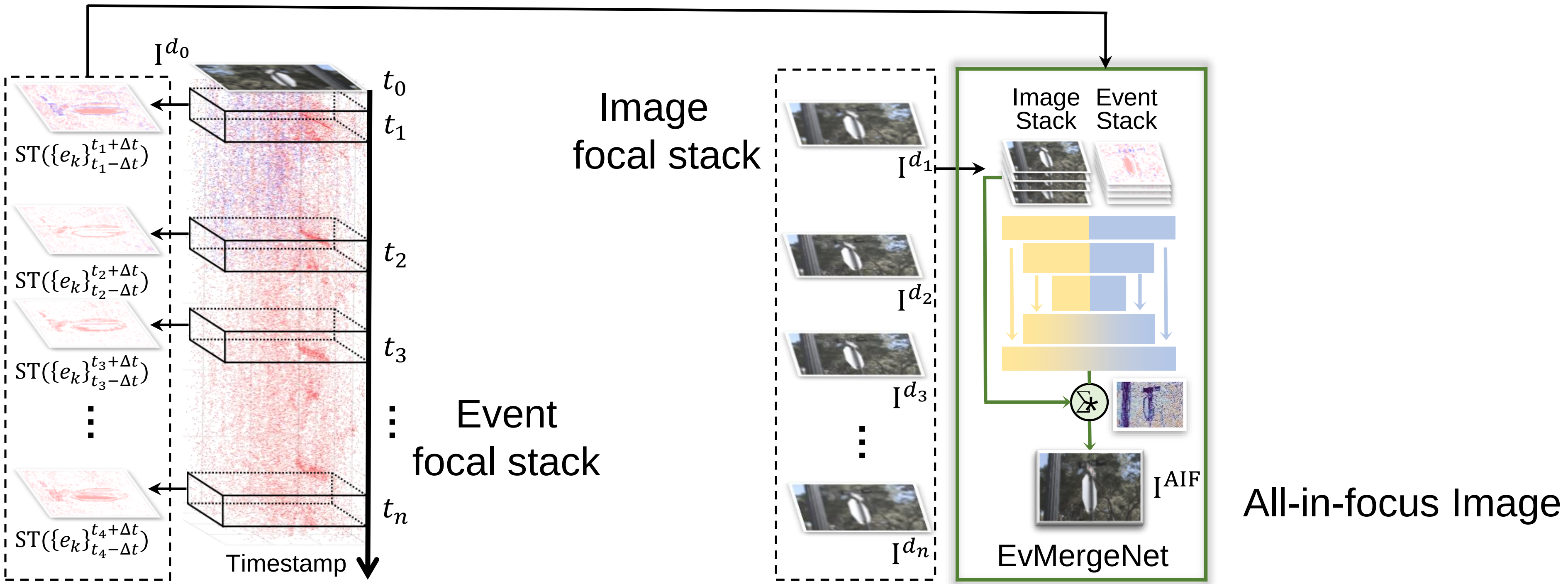
Step 1： For each patch of the image, we use the Golden Rate Search Algorithm to find the moment when it was in focus, getting N×N refocus timestamps (one for each patch).

Image Refocusing



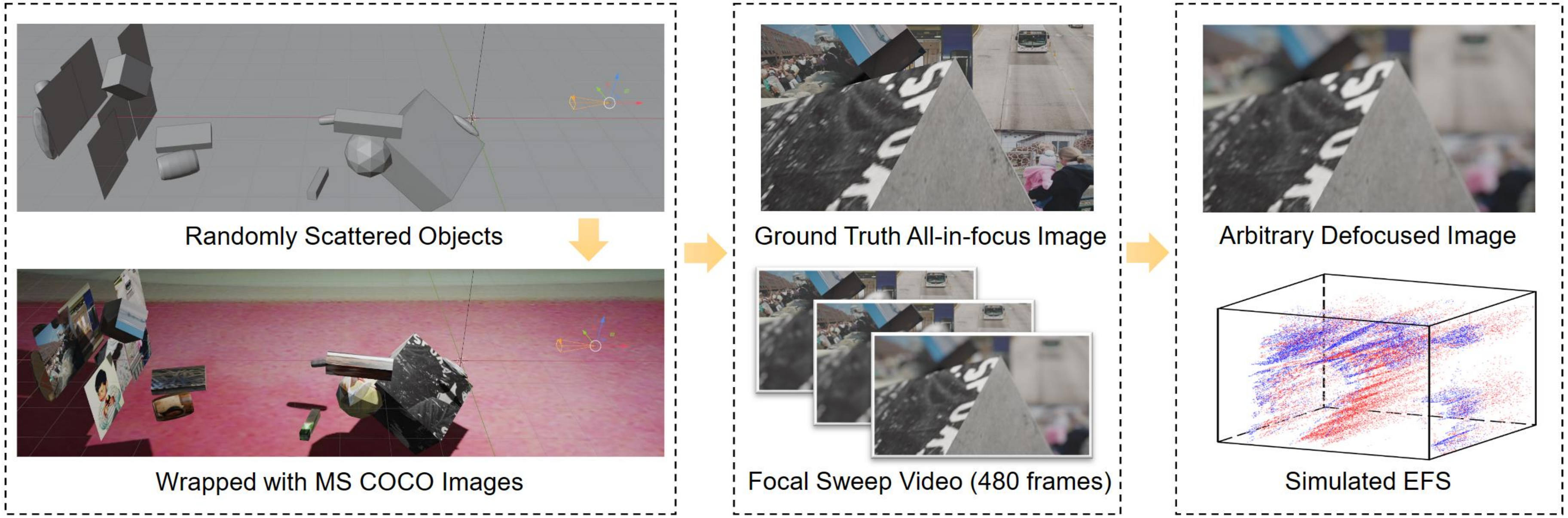
Step 2： For each refocus timestamp, we use EvRefocusNet to reconstruct a refocused image from the input image + event focal stack, forming an image focal stack.

Merging Focal Stack



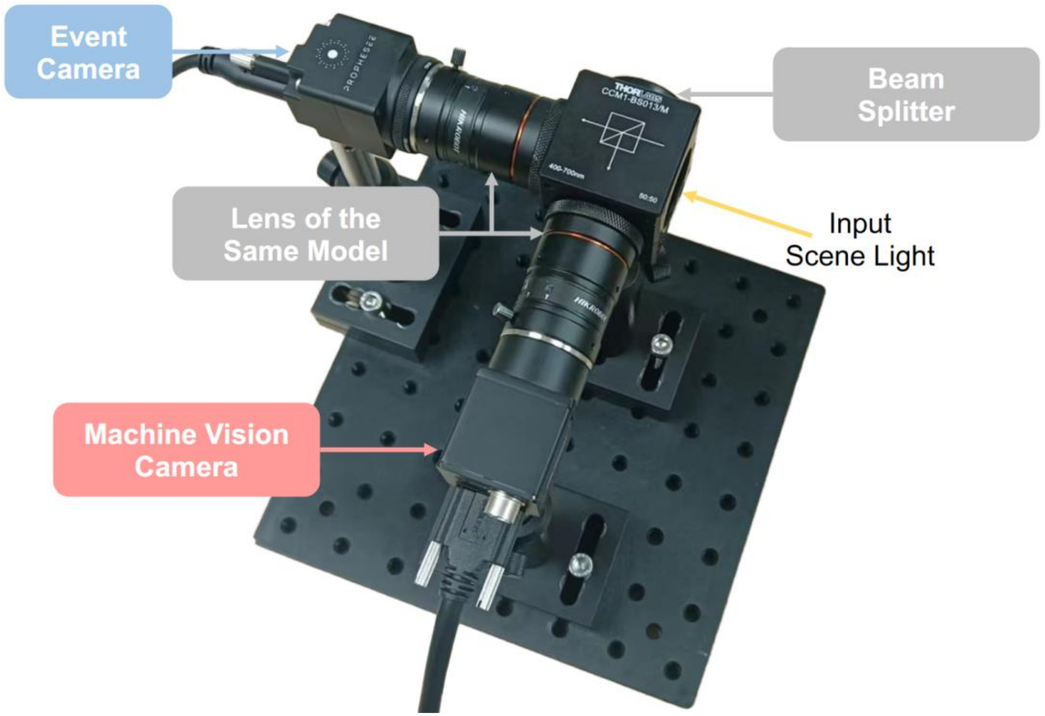
Step 3： We predict merging weights with EvMergeNet, using image focal stack and events as input. Then, we merge the focal stack with weights to get all-in-focus result.

Training Dataset



- We use Blender to render high-frame-rate focal sweep videos & ground truth all-in-focus images.
- We use MS COCO dataset images to enrich object textures.
- We use DVS-Voltmeter [Lin *et al.*, ECCV'22] to simulate events from focal sweep videos. To prevent overfitting, we use random simulator parameters in each video.

Test Dataset



- We use a hybrid camera system to capture real test data.

Results on Synthetic Data

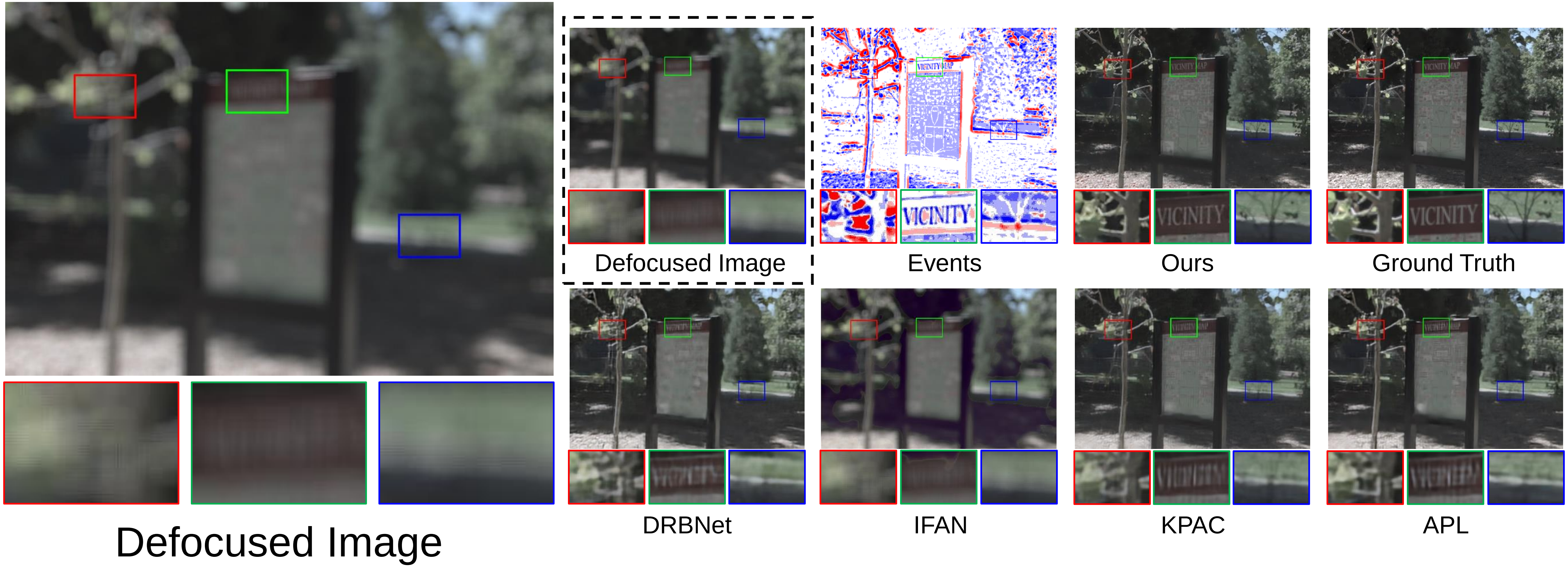


Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Results on Synthetic Data

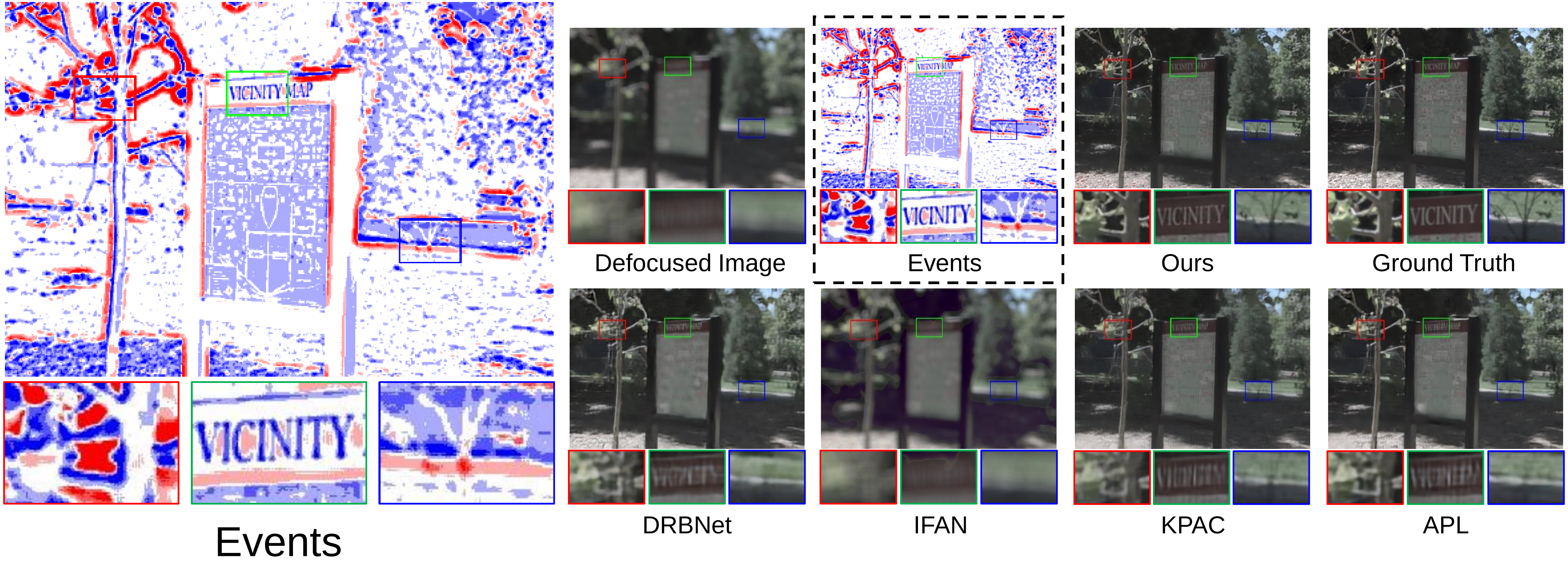


Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Results on Synthetic Data

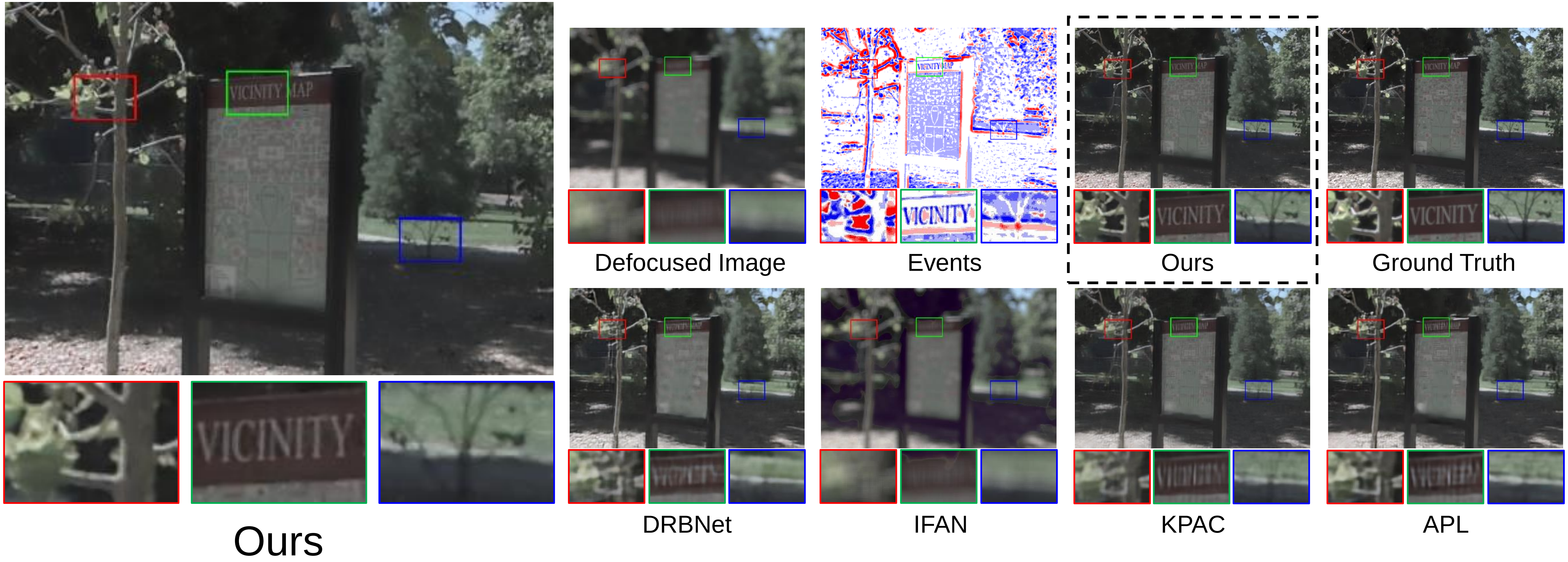


Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Results on Synthetic Data

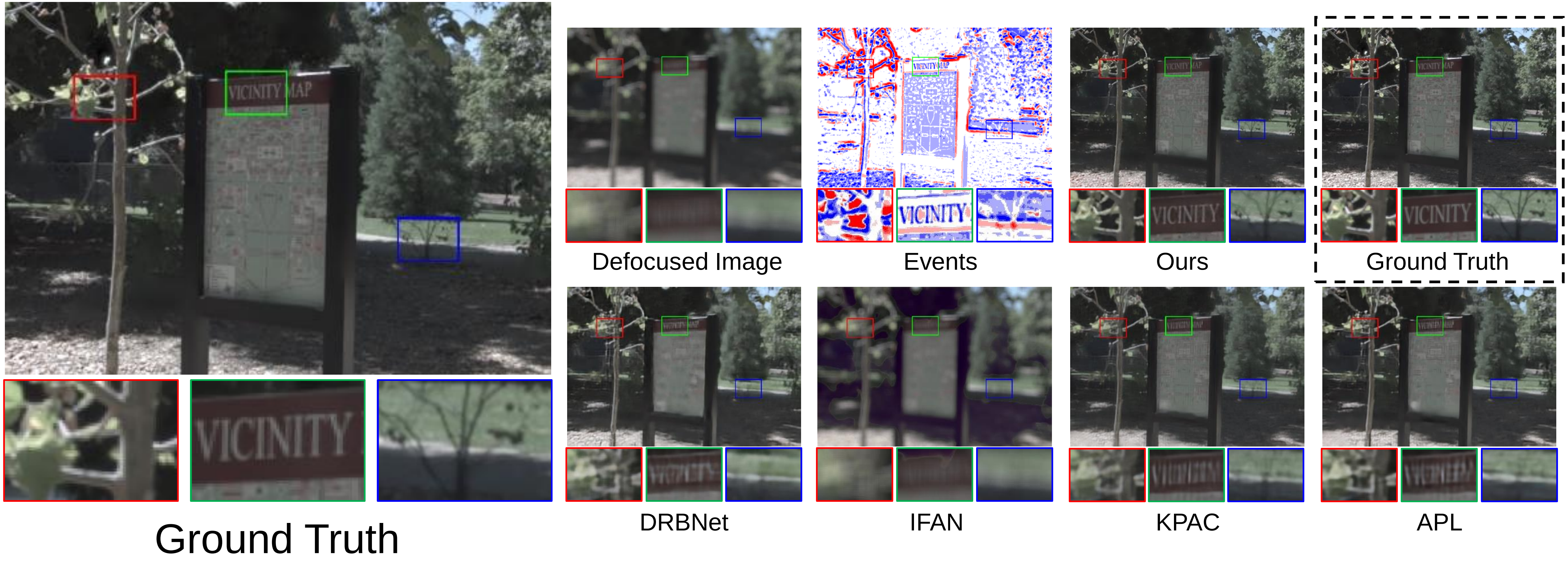


Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Results on Synthetic Data

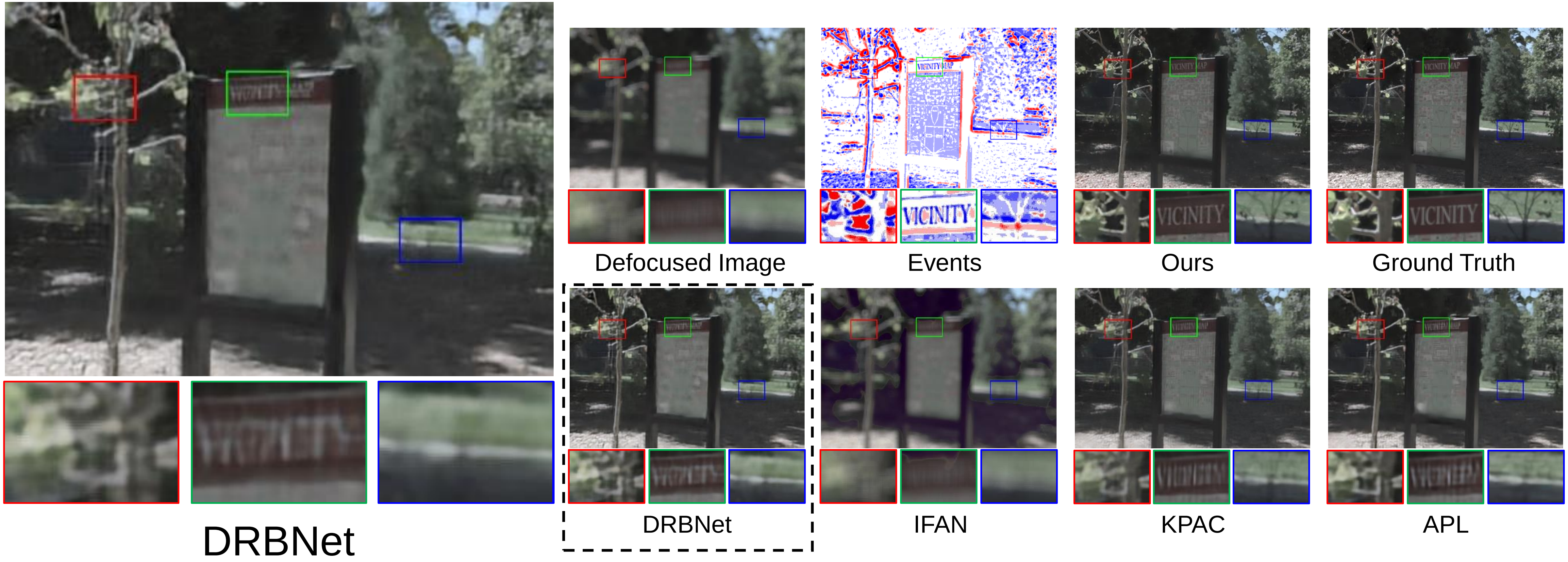


Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Results on Synthetic Data

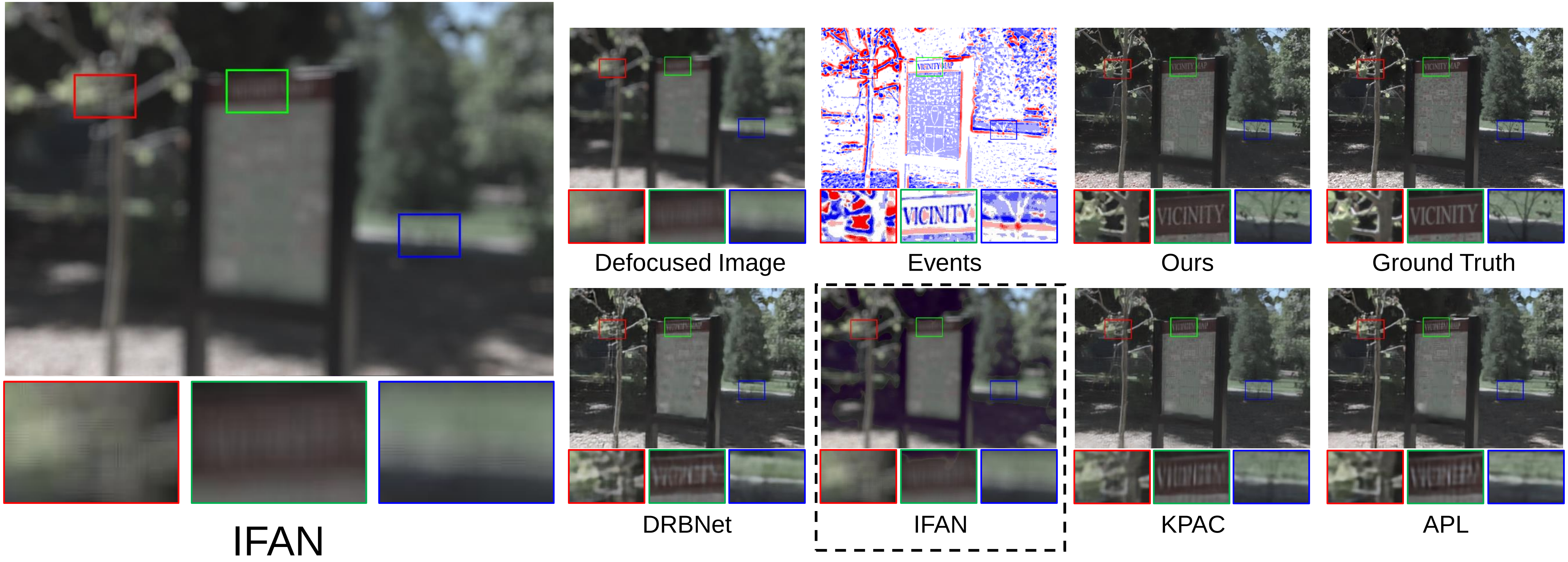


Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Results on Synthetic Data

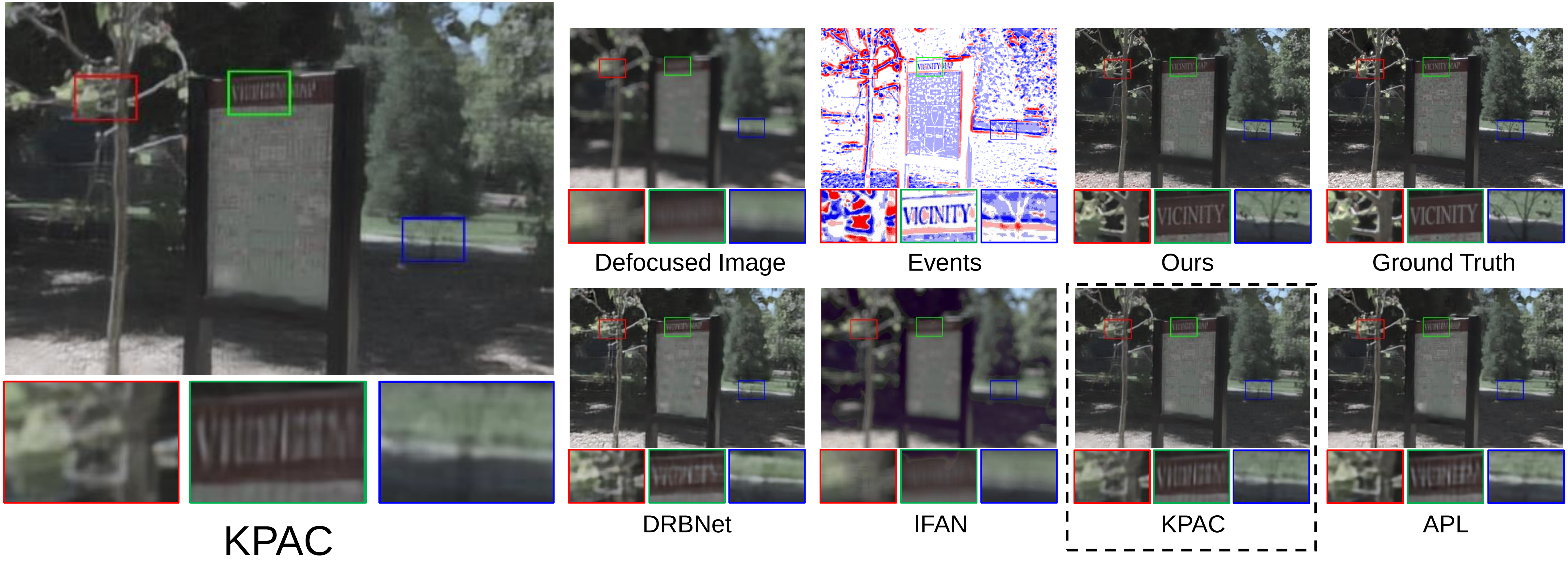


Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Results on Synthetic Data

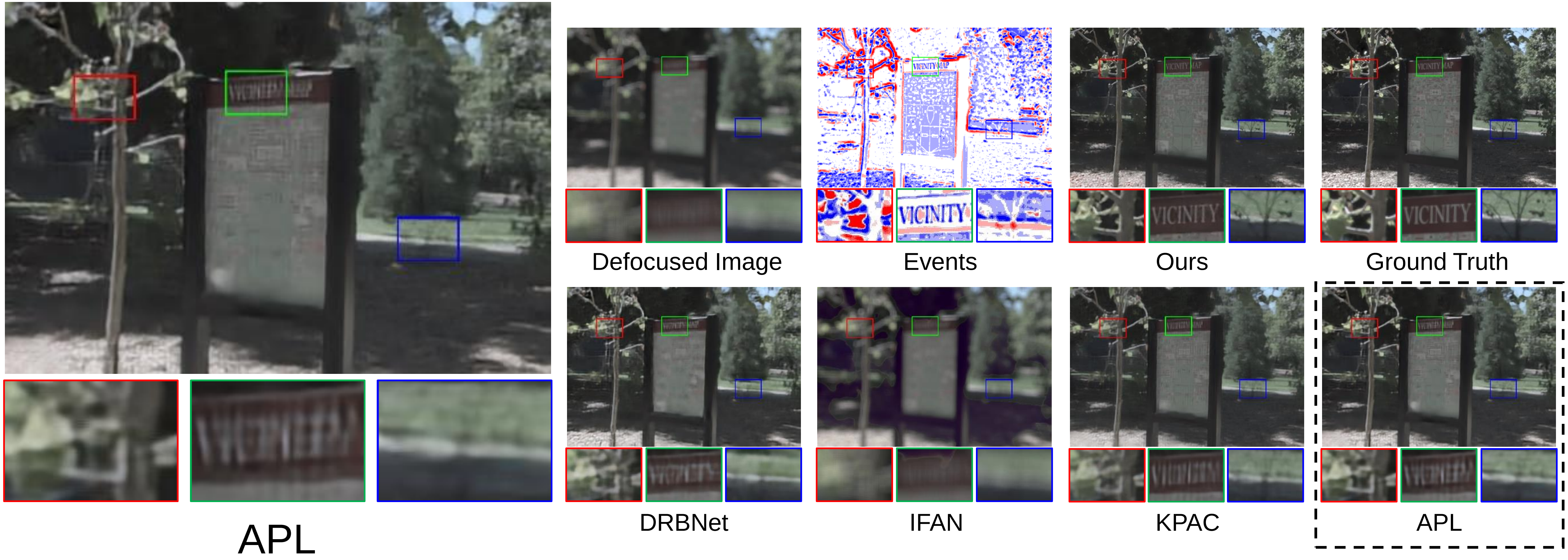
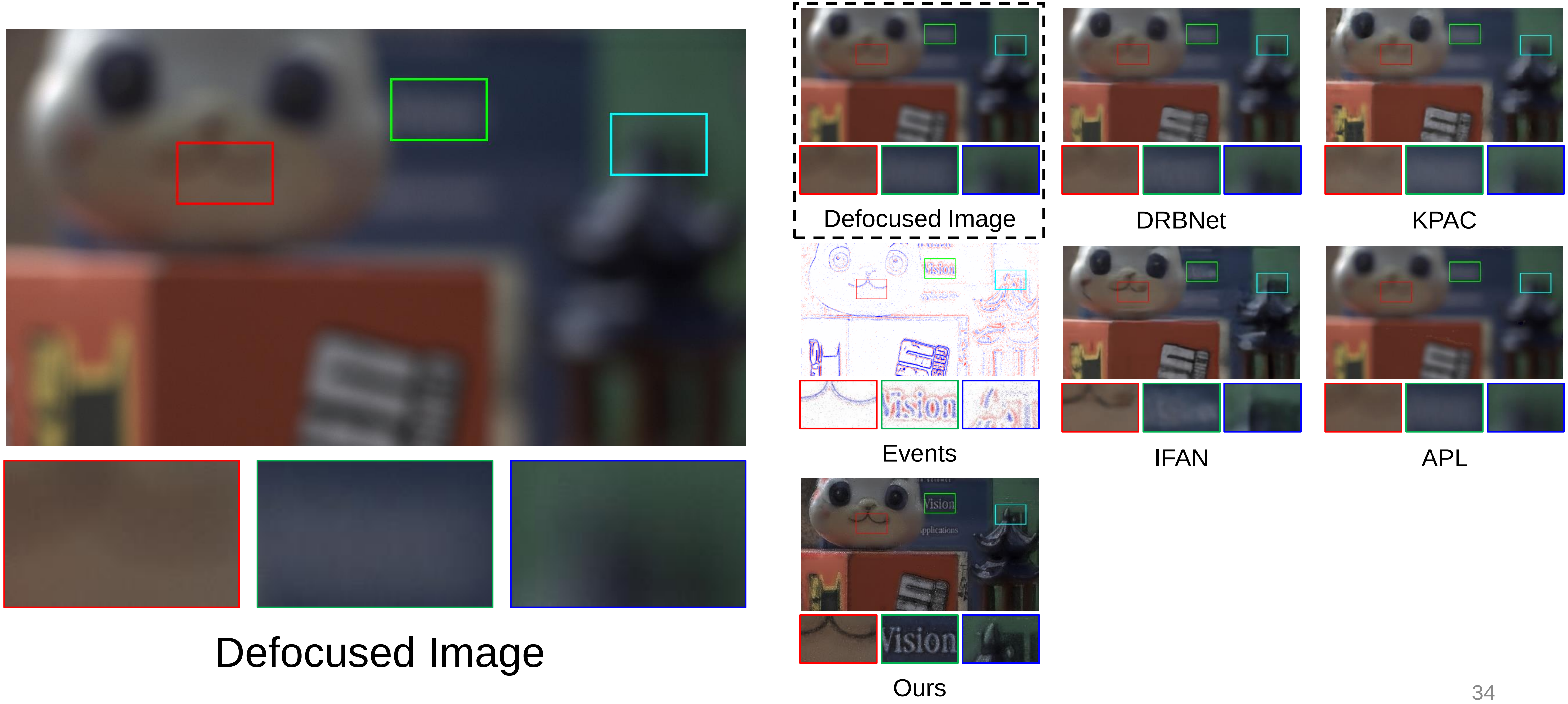


Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Results on Real Data



Defocused Image

Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Results on Real Data

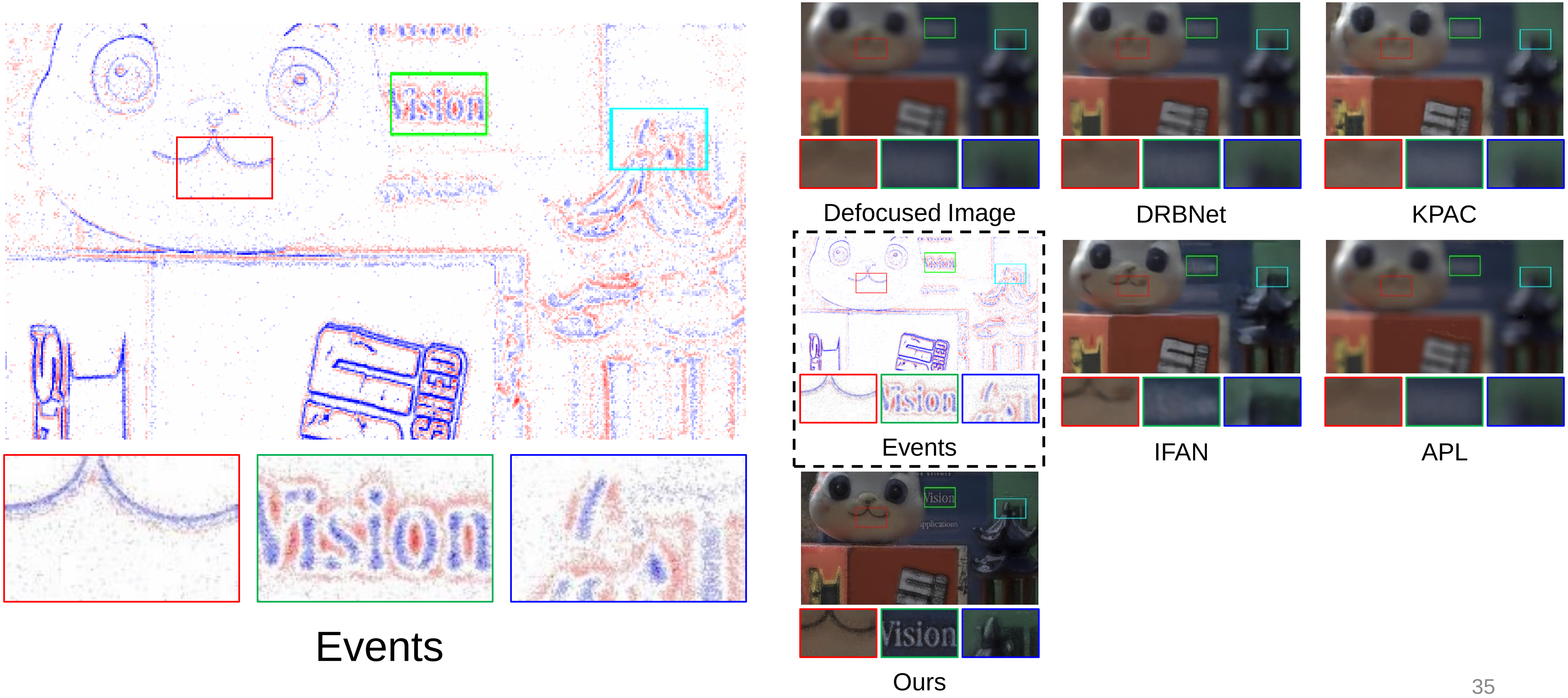


Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Results on Real Data

DRBNet

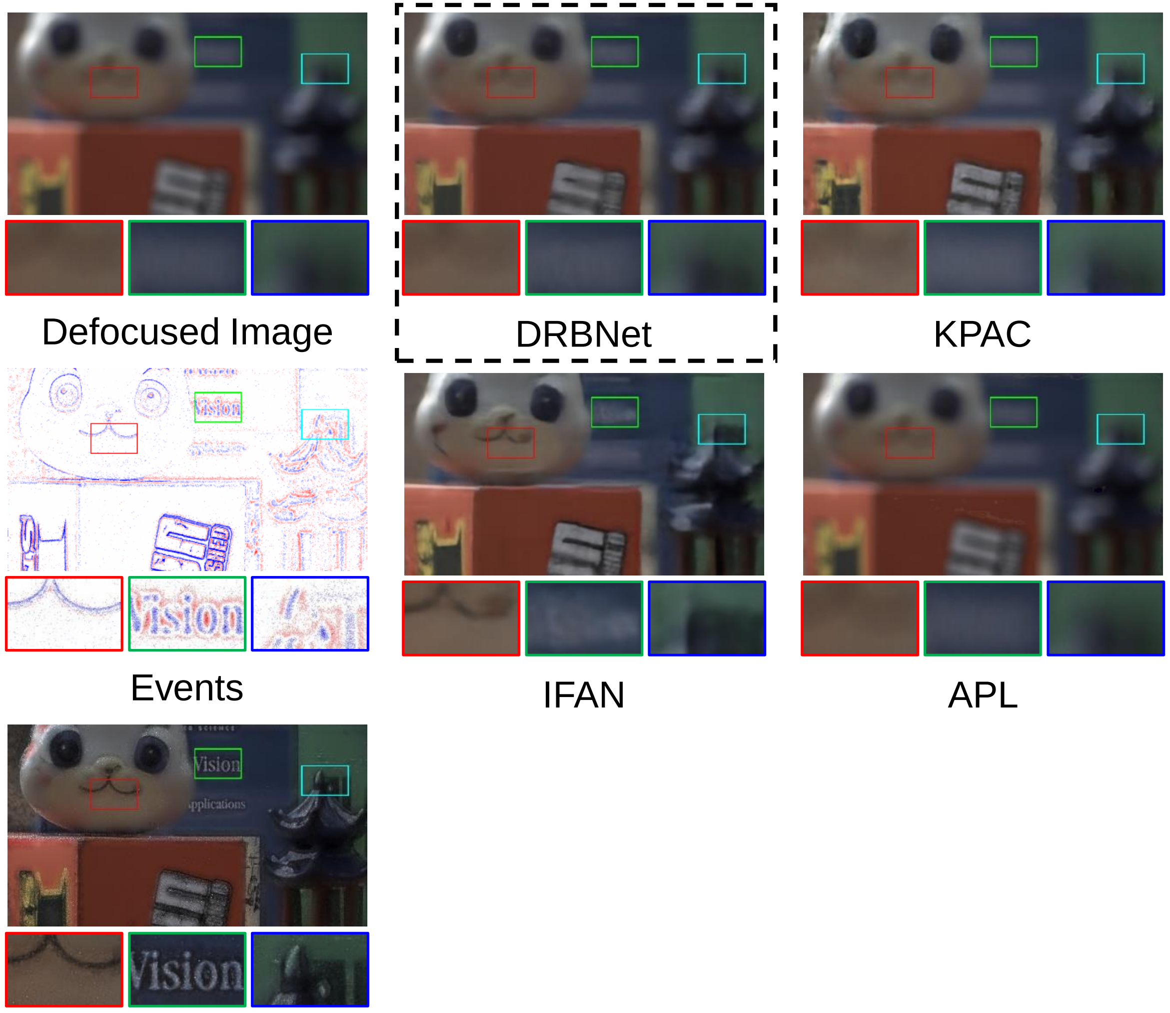


Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Results on Real Data

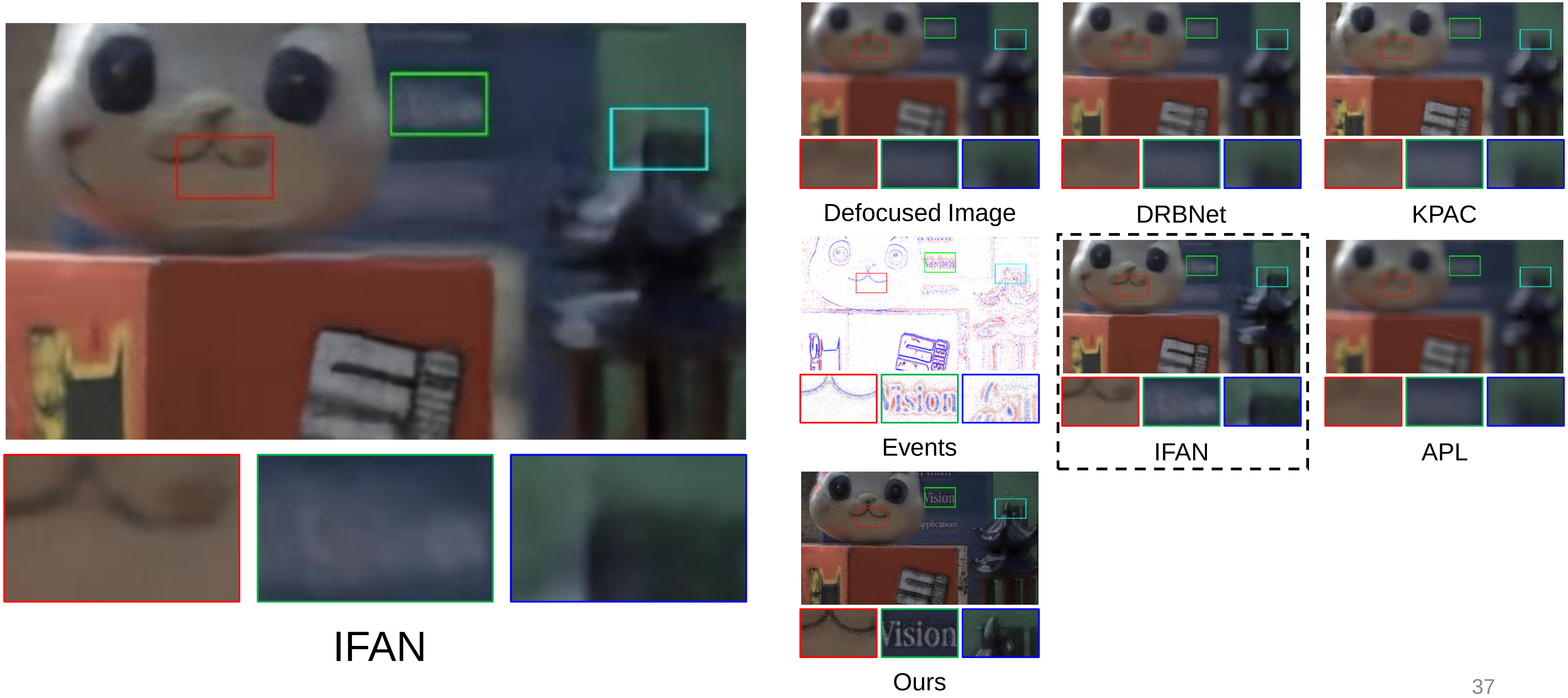
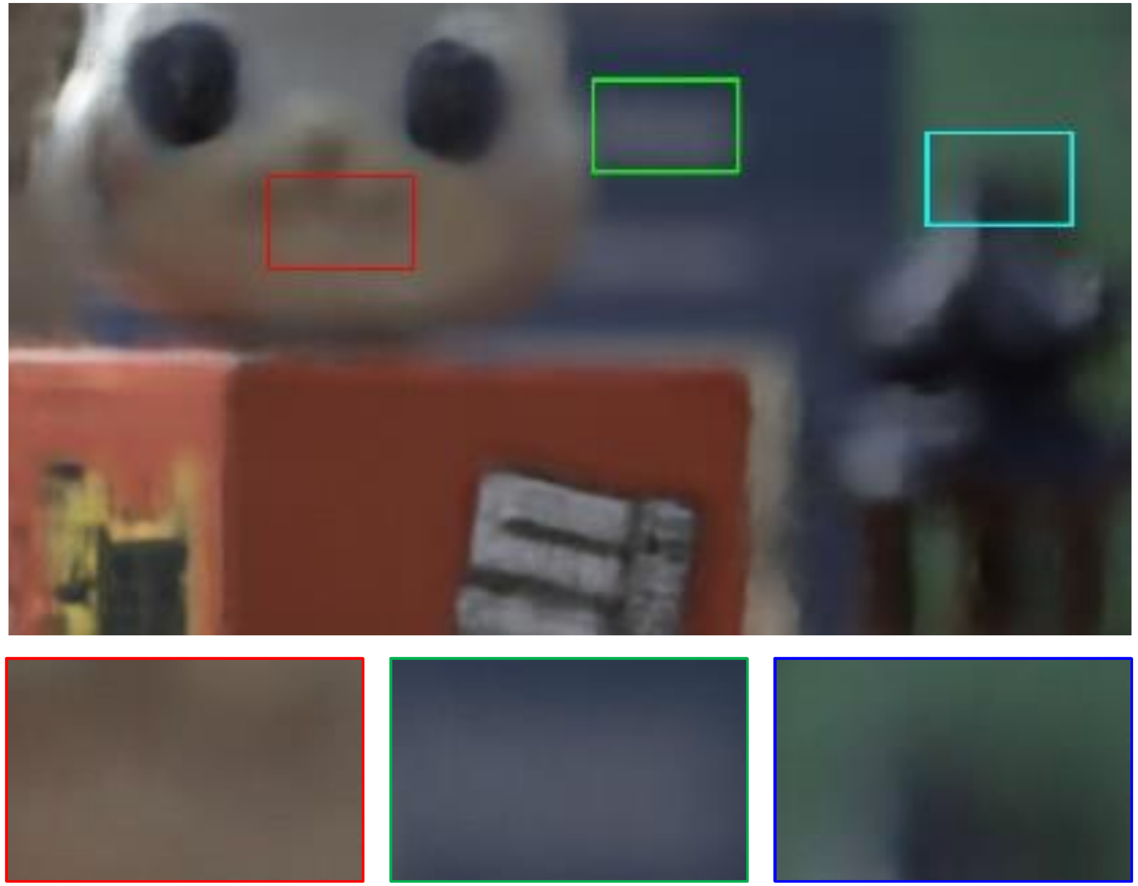


Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Results on Real Data



KPAC

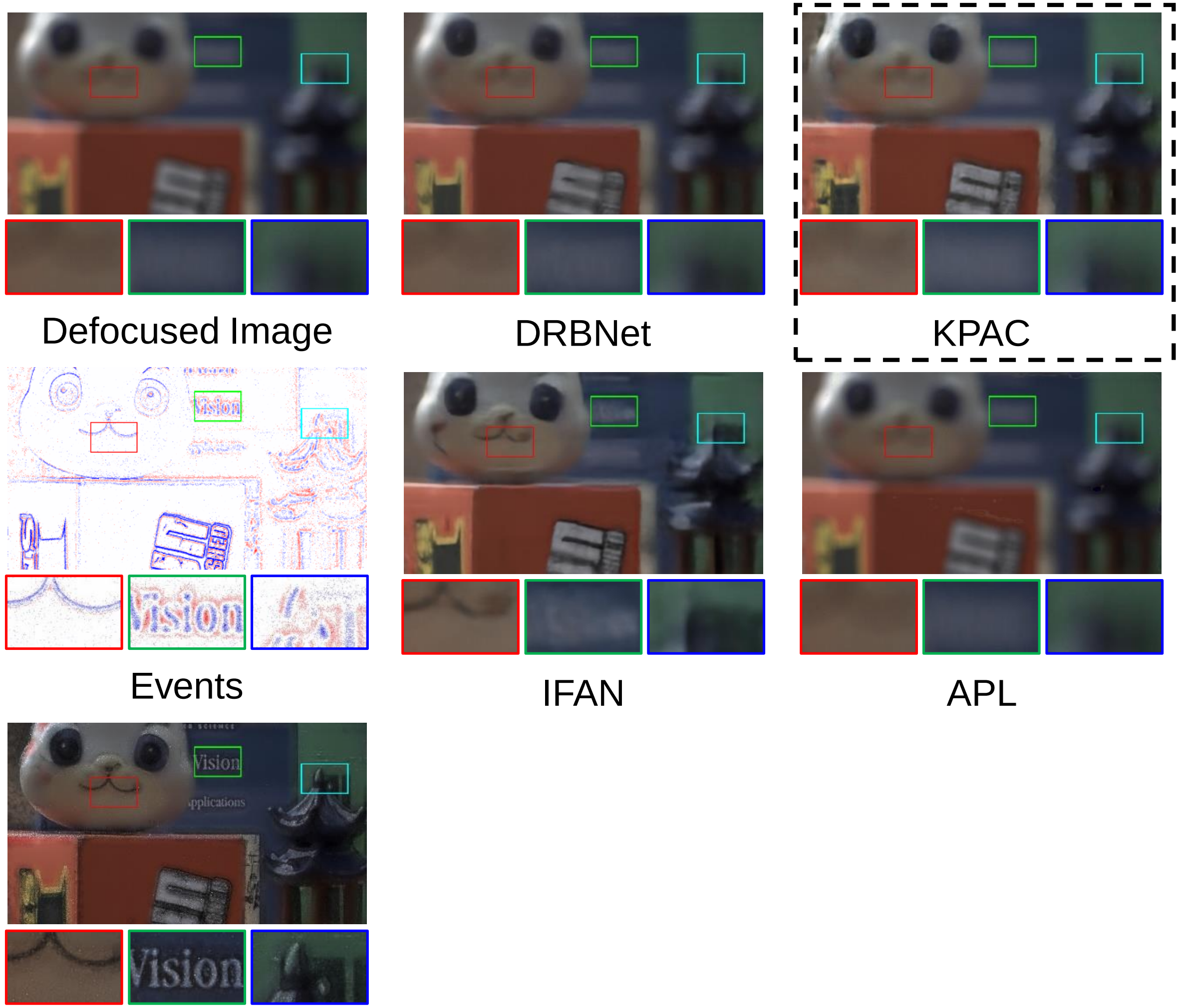


Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Results on Real Data

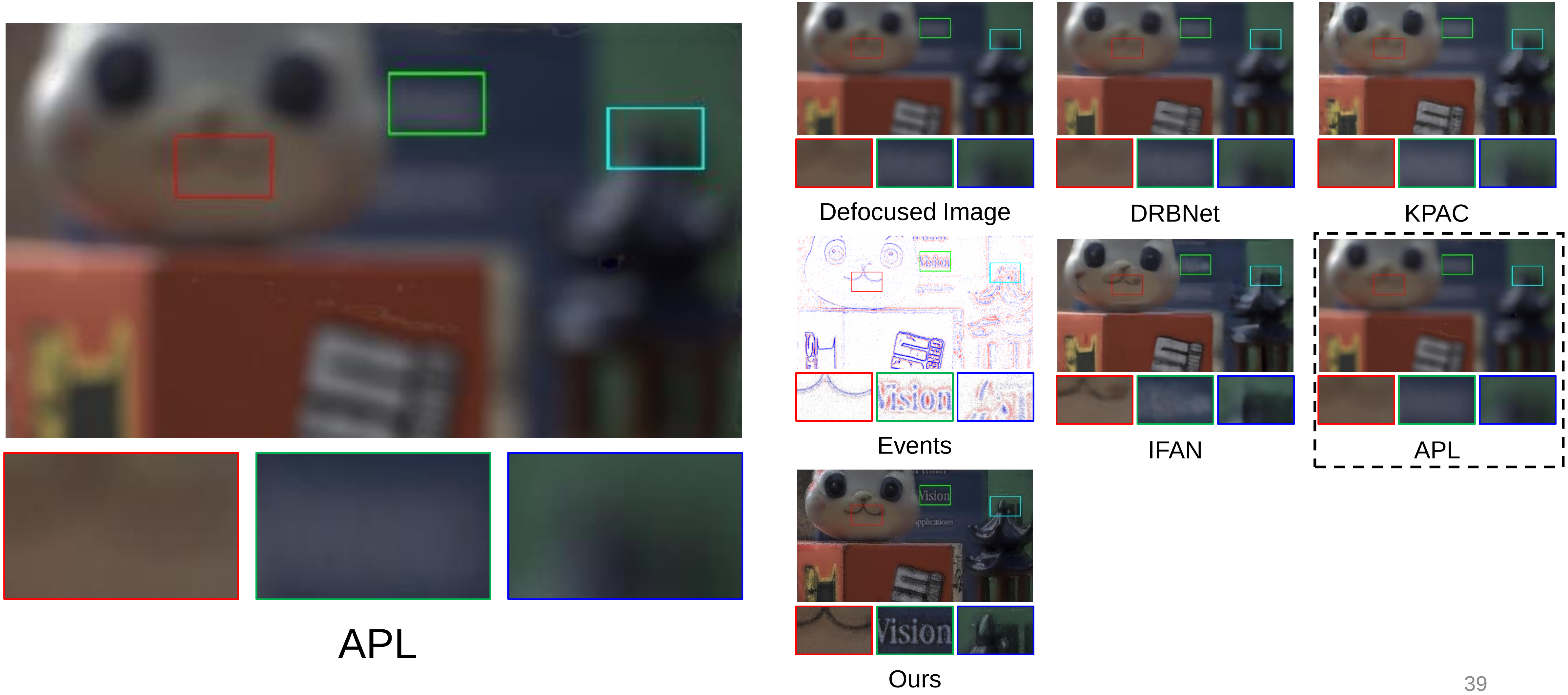


Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

Results on Real Data

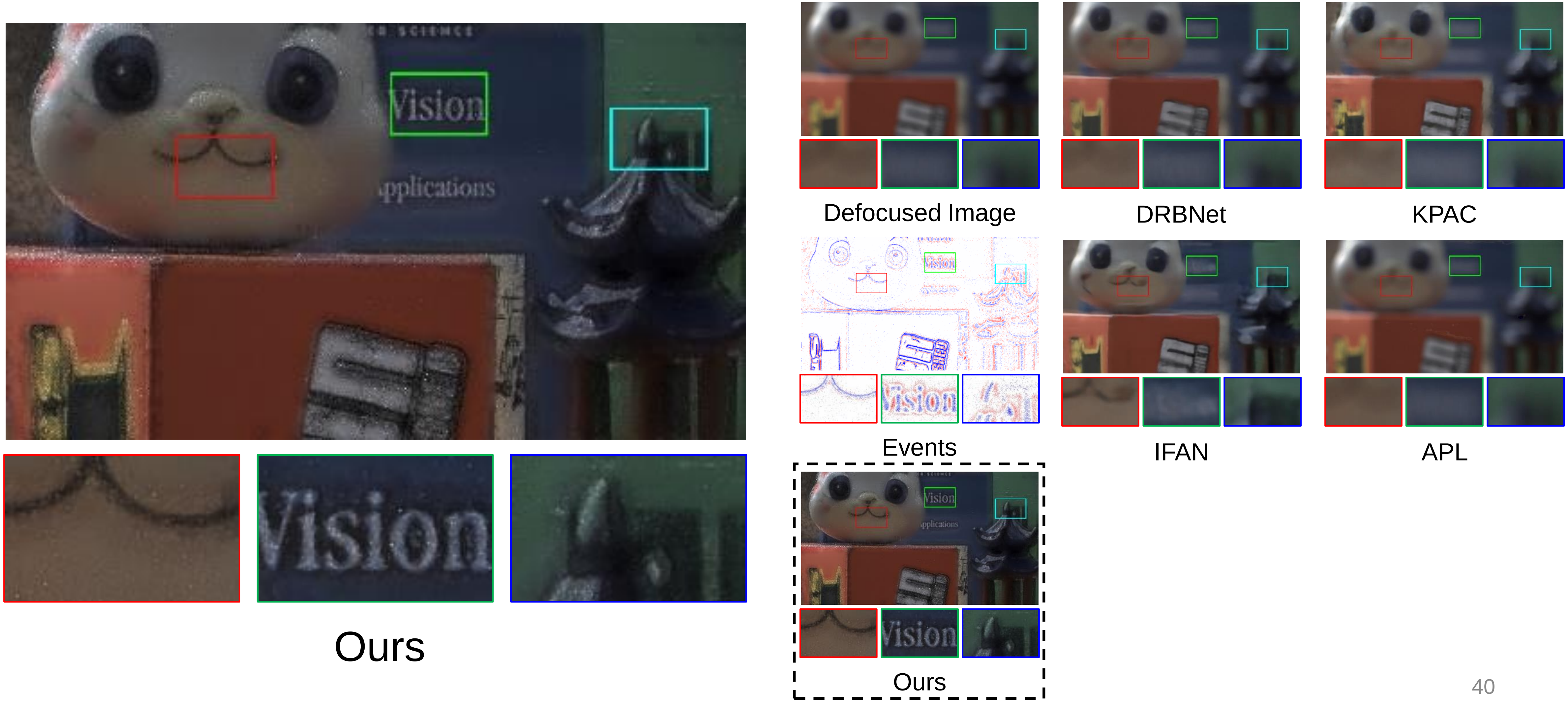
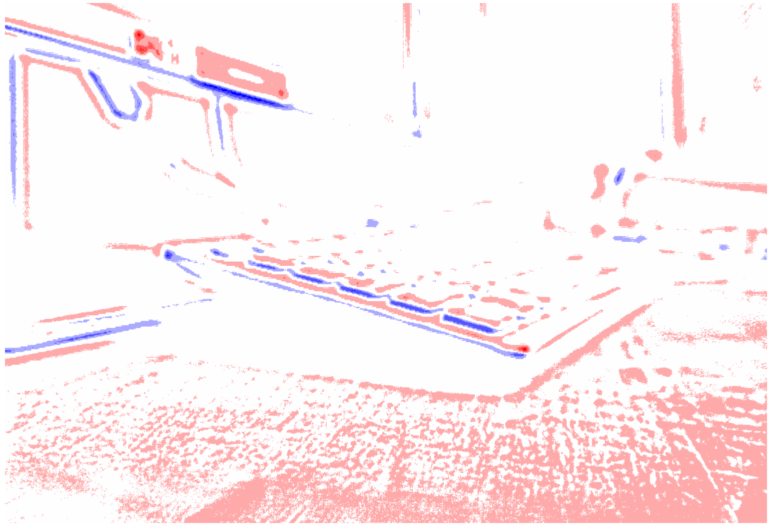


Image-based defocus: [DRBNet, CVPR 2022], [IFAN, CVPR 2021], [KPAC, ICCV 2021], [APL, ECCV 2022]

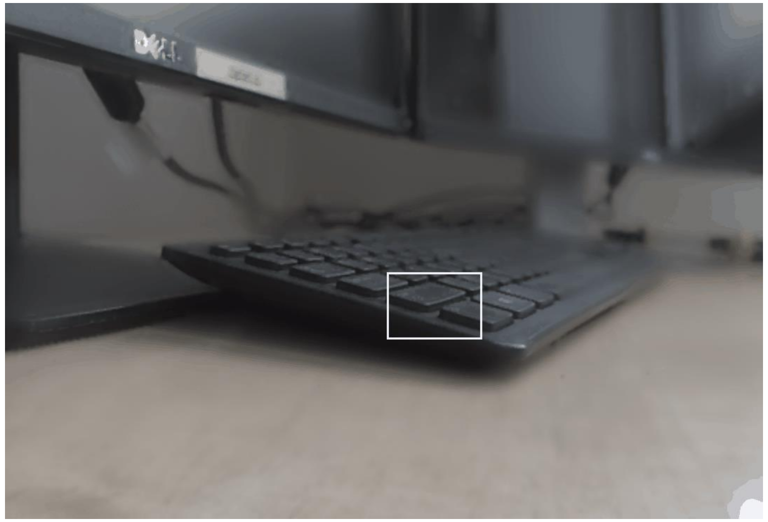
Refocusing on Synthetic Data

Input Events



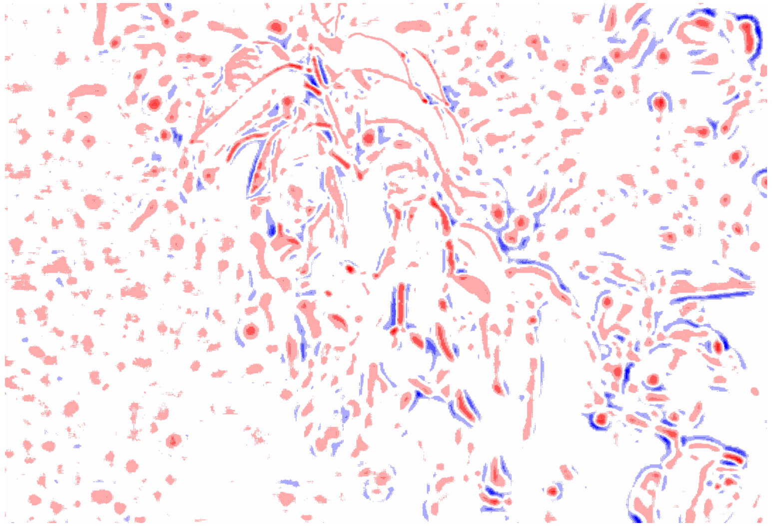
Input Image

Refocus Results

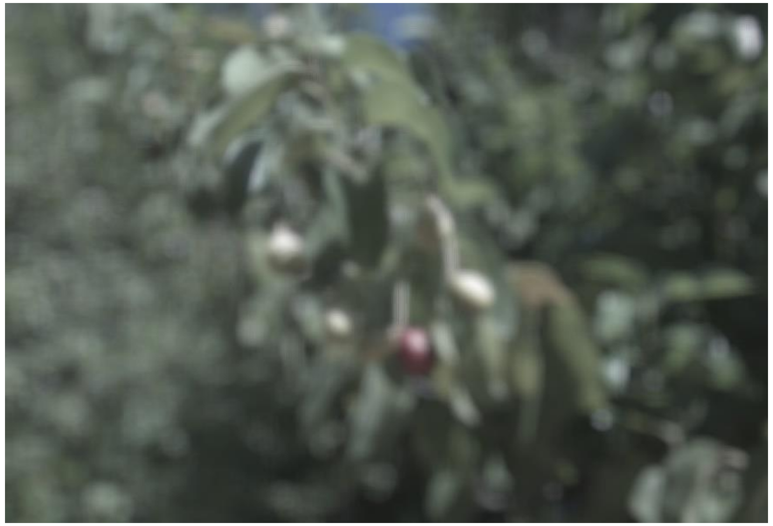


Refocusing on Synthetic Data

Input Events



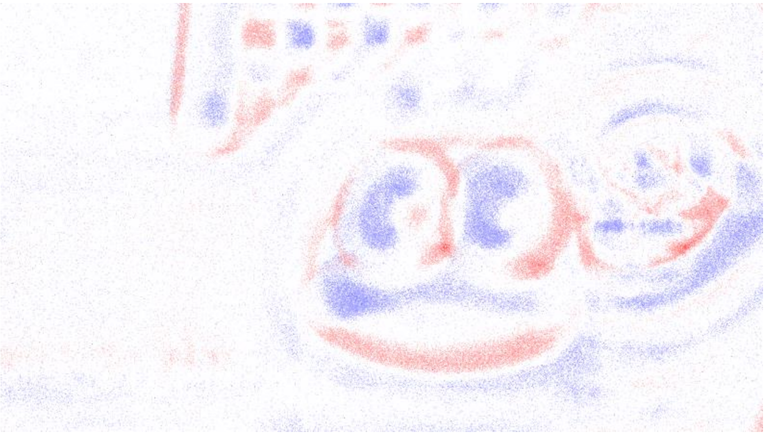
Input Image



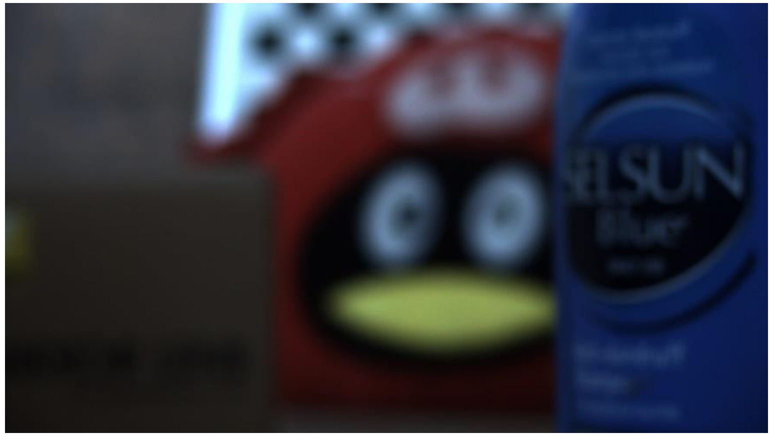
Refocus Results

Refocusing on Real Data

Input Events



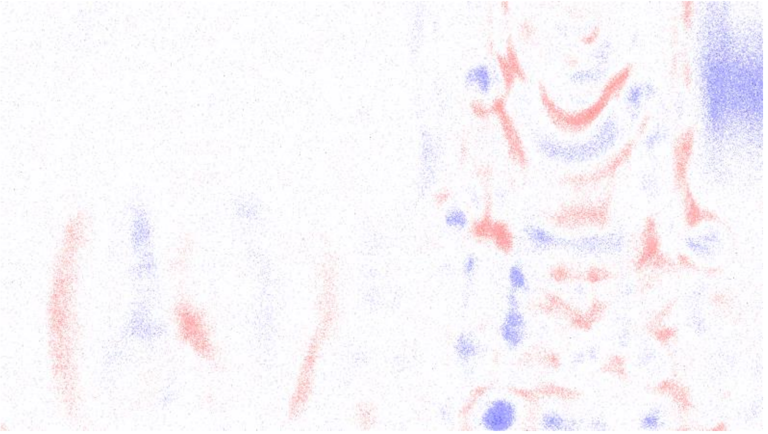
Input Image



Refocus Results

Refocusing on Real Data

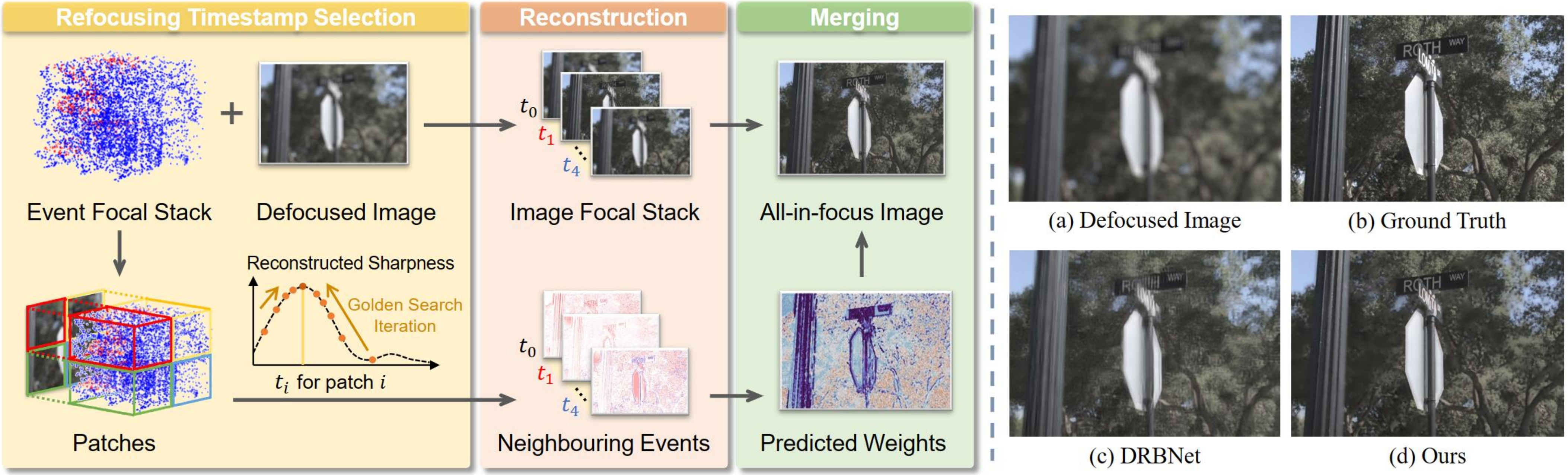
Input Events



Input Image

Refocus Results

Conclusion



- We propose recording Event Focal Stacks (EFS) for high quality all-in-focus imaging.
- We design a three-stage algorithm to exploit the continuous information encoded in event focal stacks.

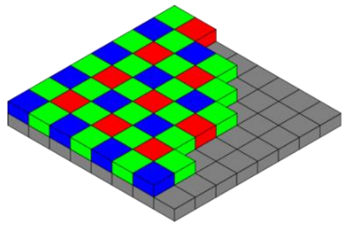
Camera Pipeline

Out-of-focus [CVPR 2023]

Scene

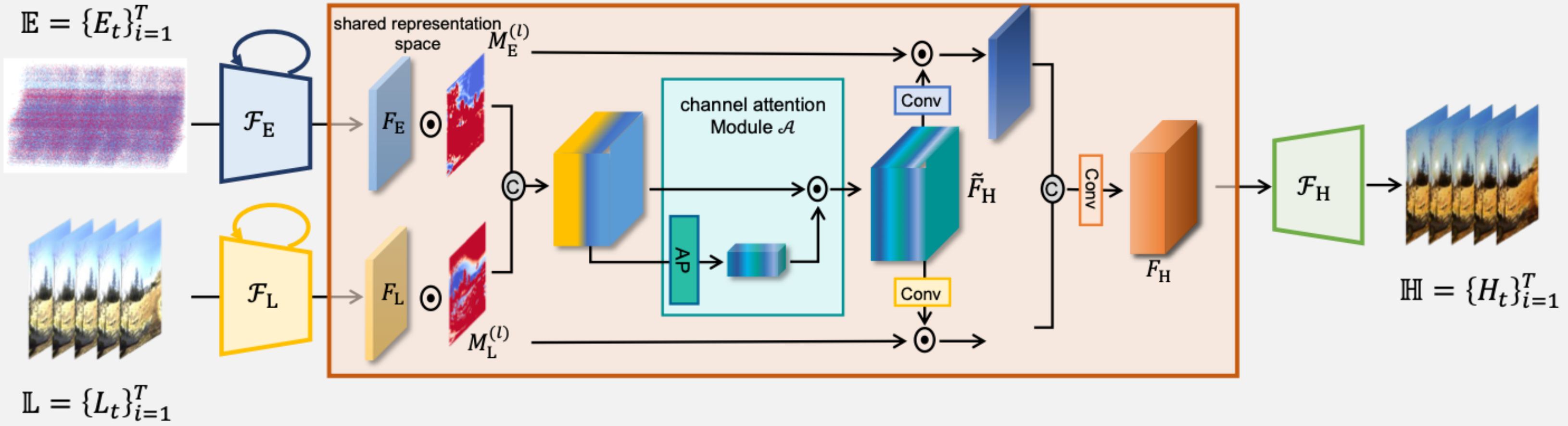
Rolling Shutter [CVPR 2022]

LDR [CVPR 2020、2023]



Analog Front-end

HDR Imaging



Event-RGB HDR

-16bit)

LFR、LDR [CVPR 2023]

Denoise

Tone Reproduction

Final RGB Image (nonlinear, 8-bit)

CVPR 2023

Learning Event Guided High Dynamic Range Video Reconstruction

Yixin Yang[1,2] Jin Han[3,4] Jinxiu Liang[1,2] Imari Sato[3,4] Boxin Shi[1,2]

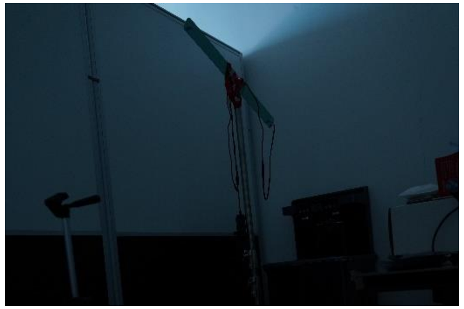
[1] National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University

[2] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

[3] Graduate School of Information Science and Technology, The University of Tokyo

[4] National Institute of Informatics

Dynamic Range in the Real World

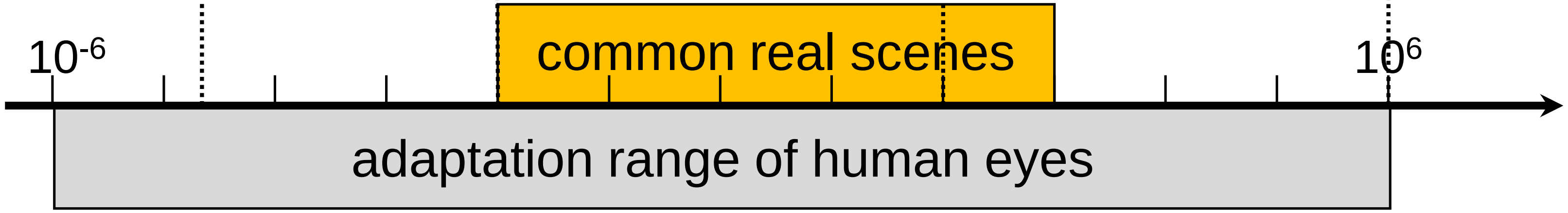


indoor without light

sunlight

moonless sky

outdoor shadow



The Dynamic Range covered by a single image is limited.

Low Exposure Image

High Exposure Image

Related Works

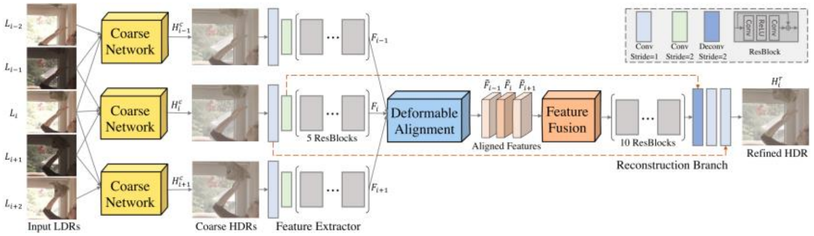
Image-based reconstruction

- Merging multiple LDR images

[Debevec *et al.*, SIGGRAPH'98]

[Li *et al.*, TIP'20]

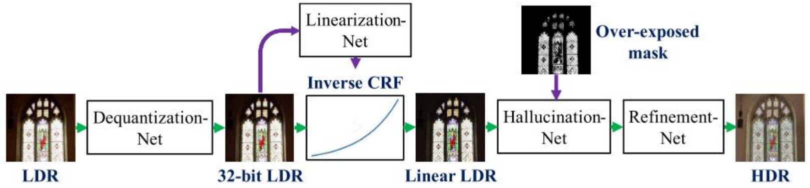
[Chen *et al.*, ICCV'21]



- Inverse tone mapping

[Eilertsen *et al.*, SIGGRAPH Asia'17]

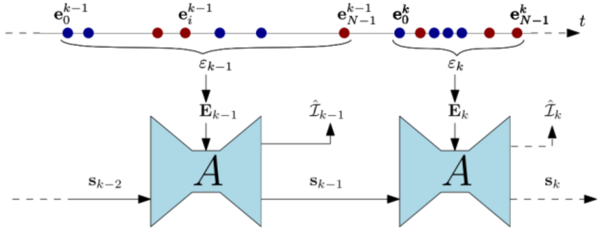
[Liu *et al.*, CVPR'20]



Event-based reconstruction

[Rebecq *et al.*, CVPR'19]

[Mostafavi *et al.*, IJCV'21]



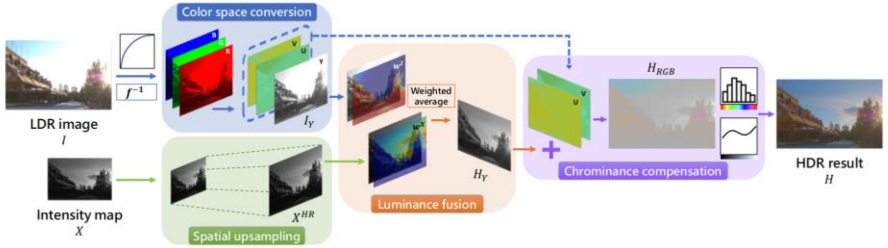
DVS

Event-guided reconstruction

[Han *et al.*, CVPR'20]

[Wang *et al.*, ECCV'20]

[Shaw *et al.*, BMVC'22]



Related Works

Image-based reconstruction

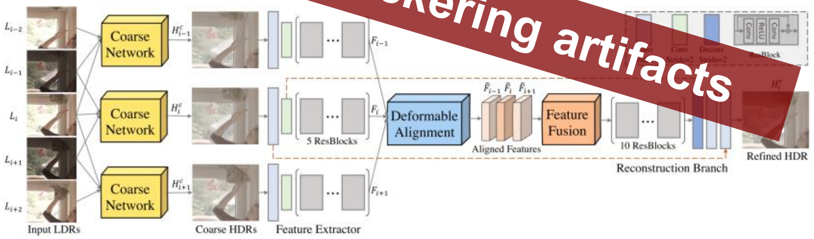
- Merging multiple LDR images

  [Debevec *et al.*, SIGGRAPH'98]

  [Li *et al.*, TIP'20]

  [Chen *et al.*, ICCV'21]

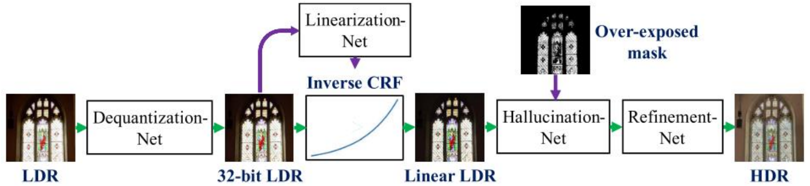
Ghosting and flickering artifacts



- Inverse tone mapping

  [Eilertsen *et al.*, SIGGRAPH Asia'17]

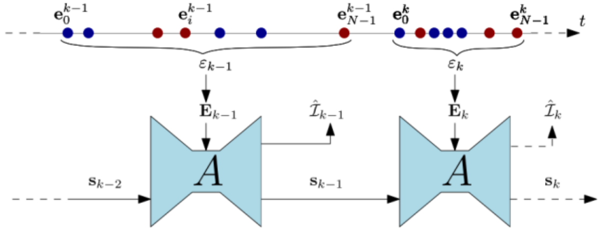
  [Liu *et al.*, CVPR'20]



Event-based reconstruction

[Rebecq *et al.*, CVPR'19]

[Mostafavi *et al.*, IJCV'21]



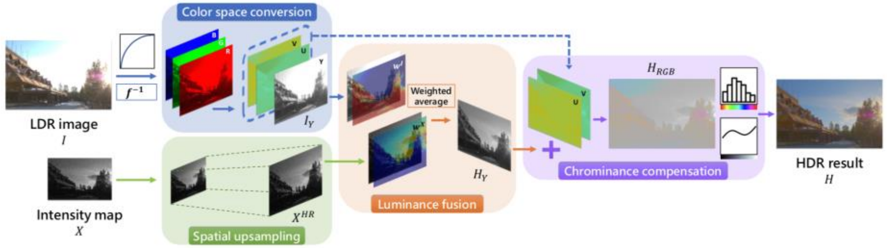
DVS

Event-guided reconstruction

[Han *et al.*, CVPR'20]

[Wang *et al.*, ECCV'20]

[Shaw *et al.*, BMVC'22]



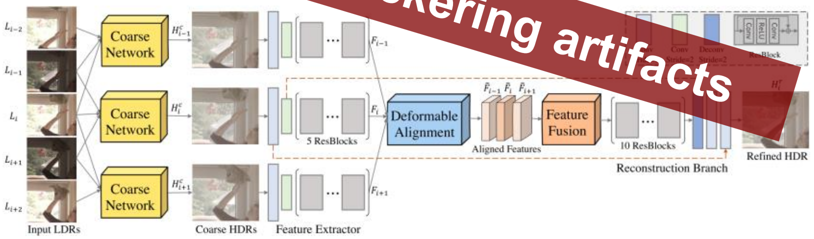
Related Works

Image-based reconstruction

- Merging multiple LDR images

[Debevec *et al.*, SIGGRAPH'98]
[Li *et al.*, TIP'20]
[Chen *et al.*, ICCV'21]

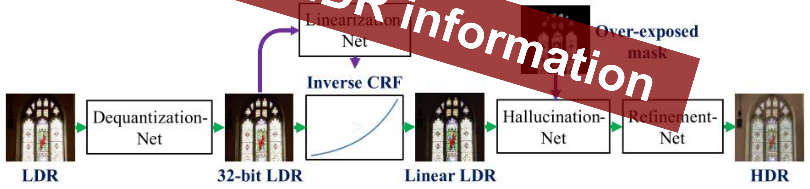
Ghosting and flickering artifacts



- Inverse tone mapping

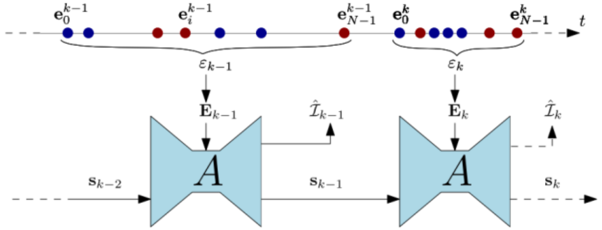
[Eilertsen *et al.*, SIGGRAPH Asia'17]
[Liu *et al.*, CVPR'20]

No reliable HDR information



Event-based reconstruction

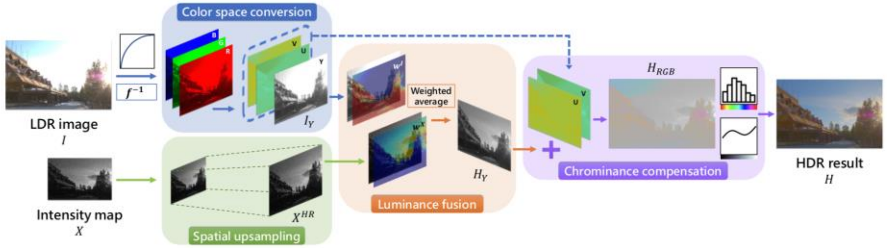
[Rebecq *et al.*, CVPR'19]
[Mostafavi *et al.*, IJCV'21]



DVS

Event-guided reconstruction

[Han *et al.*, CVPR'20]
[Wang *et al.*, ECCV'20]
[Shaw *et al.*, BMVC'22]



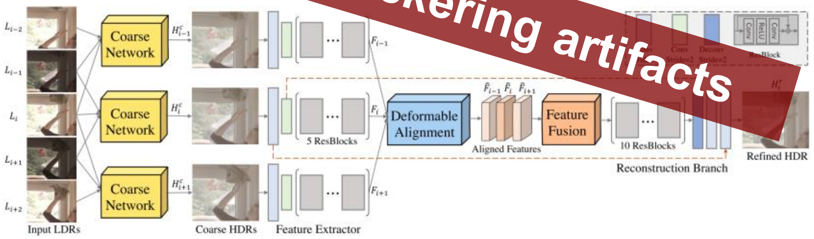
Related Works

Image-based reconstruction

- Merging multiple LDR images

  [Debevec *et al.*, SIGGRAPH'98]
  [Li *et al.*, TIP'20]
  [Chen *et al.*, ICCV'21]

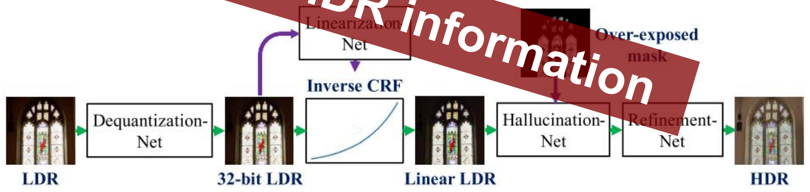
  Ghosting and flickering artifacts



- Inverse tone mapping

  [Eilertsen *et al.*, SIGGRAPH Asia'17]
  [Liu *et al.*, CVPR'20]

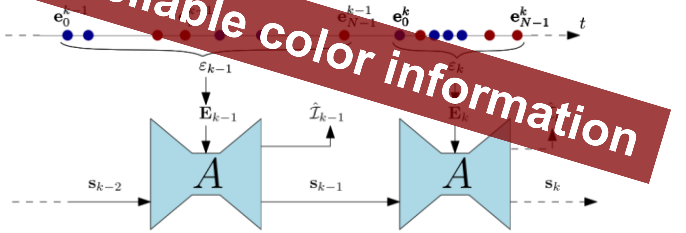
  No reliable HDR information



Event-based reconstruction

[Rebecq *et al.*, CVPR'19]
[Mostafavi *et al.*, IJCV'21]

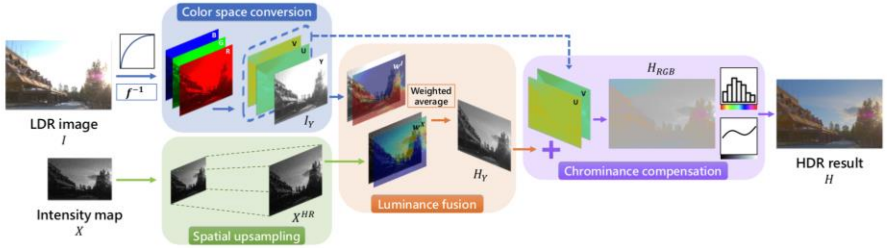
No reliable color information



DVS

Event-guided reconstruction

[Han *et al.*, CVPR'20]
[Wang *et al.*, ECCV'20]
[Shaw *et al.*, BMVC'22]

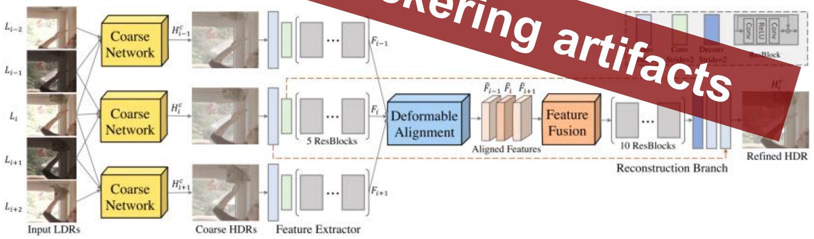


Related Works

Image-based reconstruction

- Merging multiple LDR images

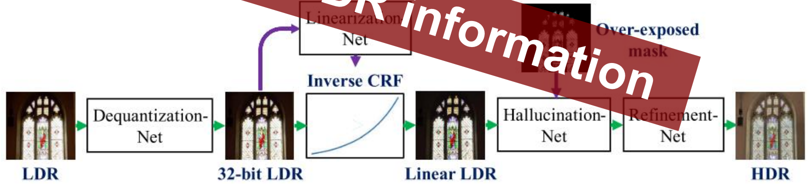
[Debevec *et al.*, SIGGRAPH'98]
[Li *et al.*, TIP'20]
[Chen *et al.*, ICCV'21]



Ghosting and flickering artifacts

- Inverse tone mapping

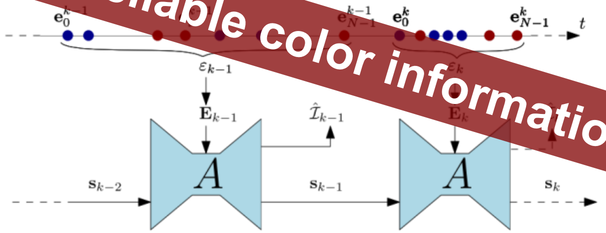
[Eilertsen *et al.*, SIGGRAPH Asia'17]
[Liu *et al.*, CVPR'20]



No reliable HDR information

Event-based reconstruction

[Rebecq *et al.*, CVPR'19]
[Mostafavi *et al.*, IJCV'21]

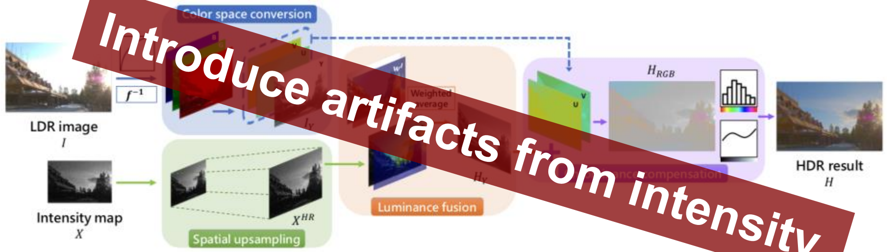


DVS

No reliable color information

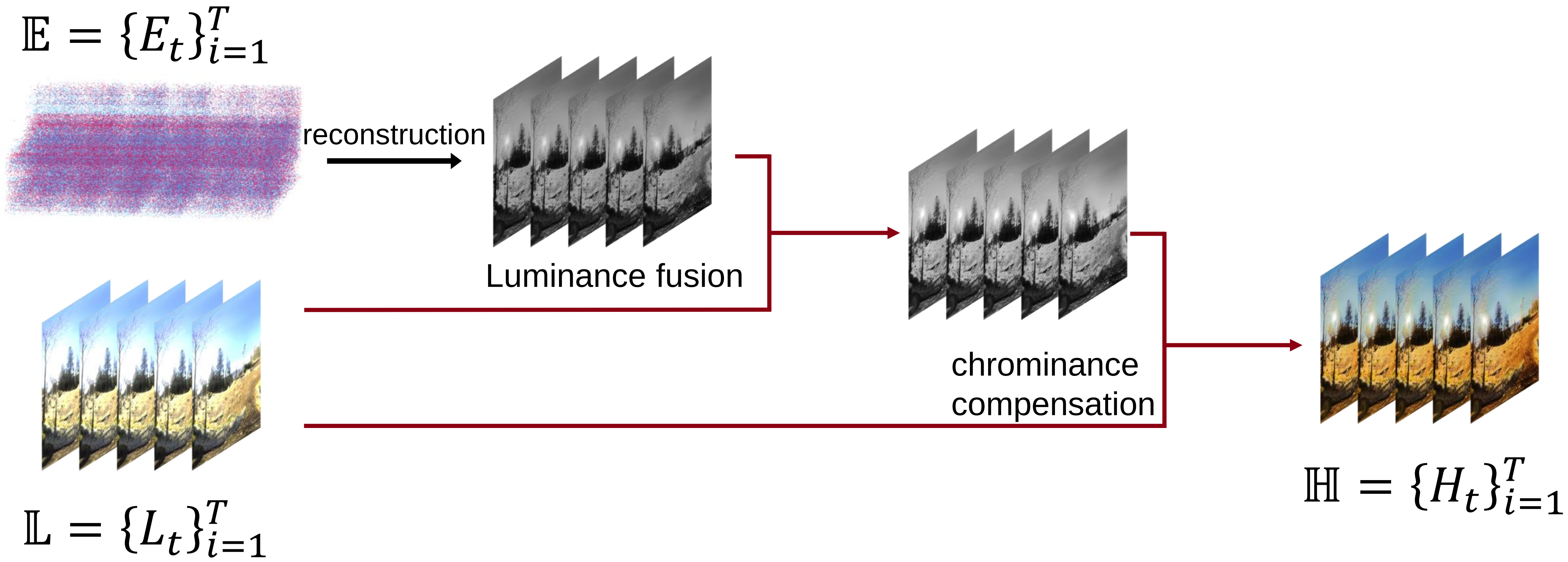
Event-guided reconstruction

[Han *et al.*, CVPR'20]
[Wang *et al.*, ECCV'20]
[Shaw *et al.*, BMVC'22]

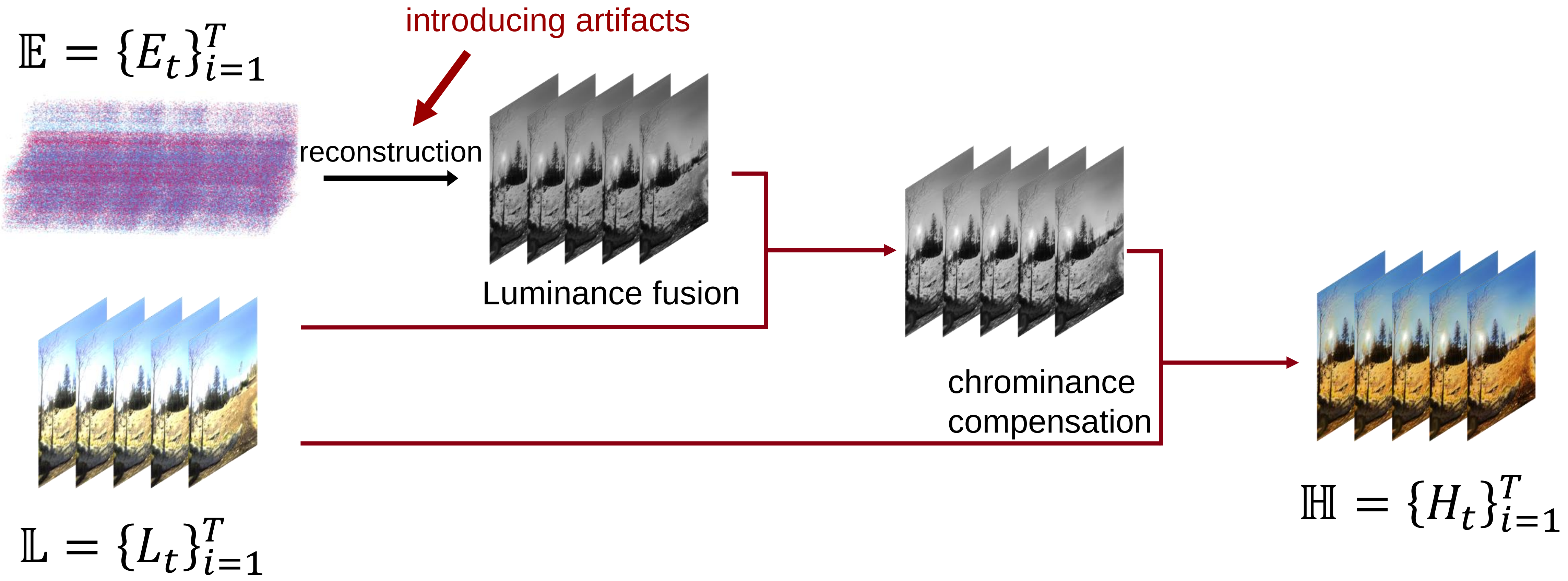


Introduce artifacts from intensity

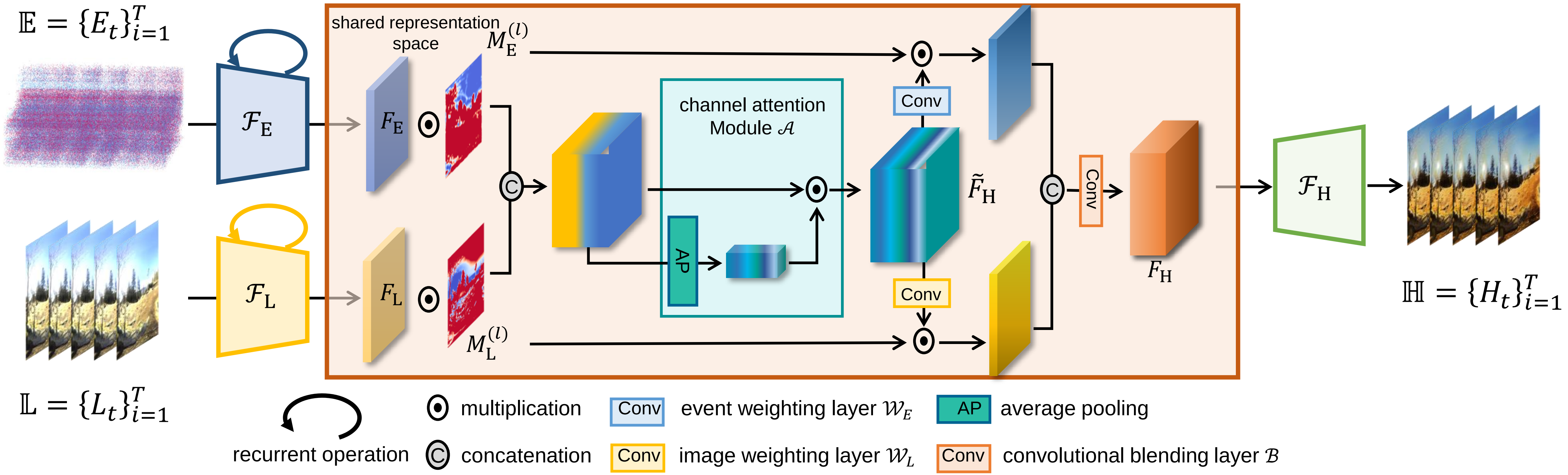
Problems in the Closest Work [Han et al., CVPR'20]



Problems in the Closest Work [Han et al., CVPR'20]

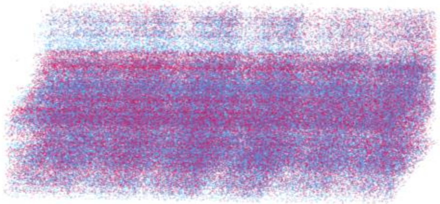


Solution: HDRev-Net Pipeline



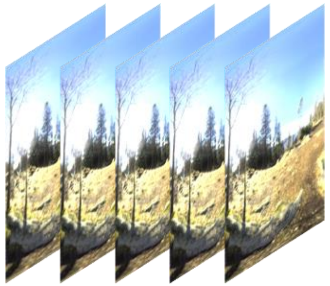
Motivation

$\mathbb{E} = \{E_t\}_{i=1}^{T}$



?

$\mathbb{H} = \{H_t\}_{i=1}^{T}$

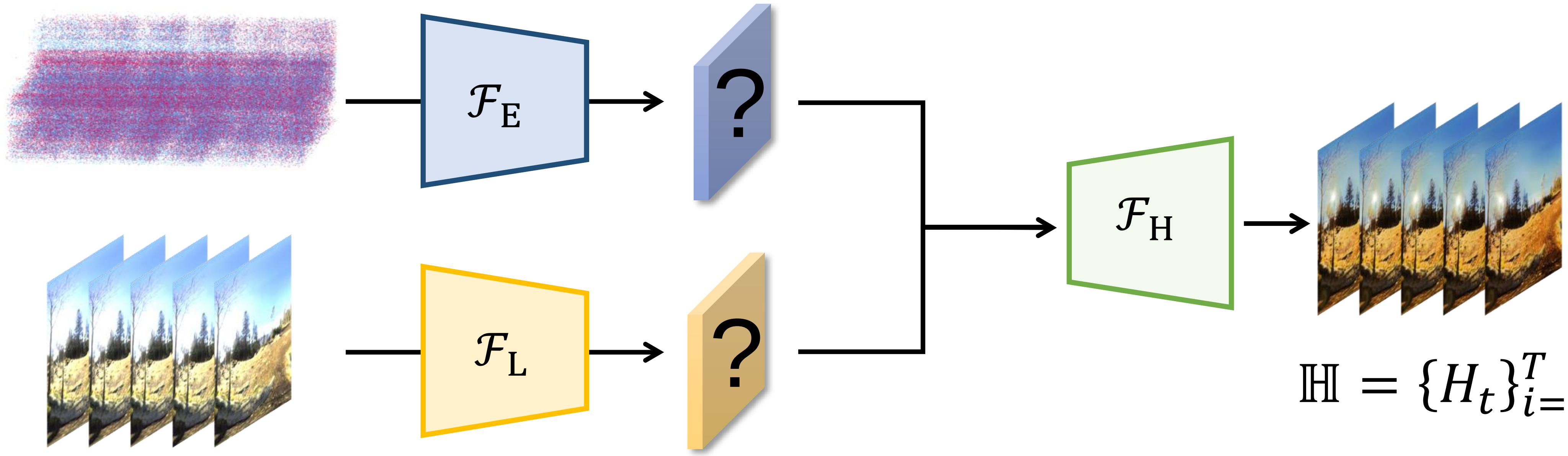


$\mathbb{L} = \{L_t\}_{i=1}^{T}$

- Can we avoid explicit reconstruction which introduces artifacts from event-to-image methods?

Motivation

$$\mathbb{E} = \{E_t\}_{i=1}^{T}$$

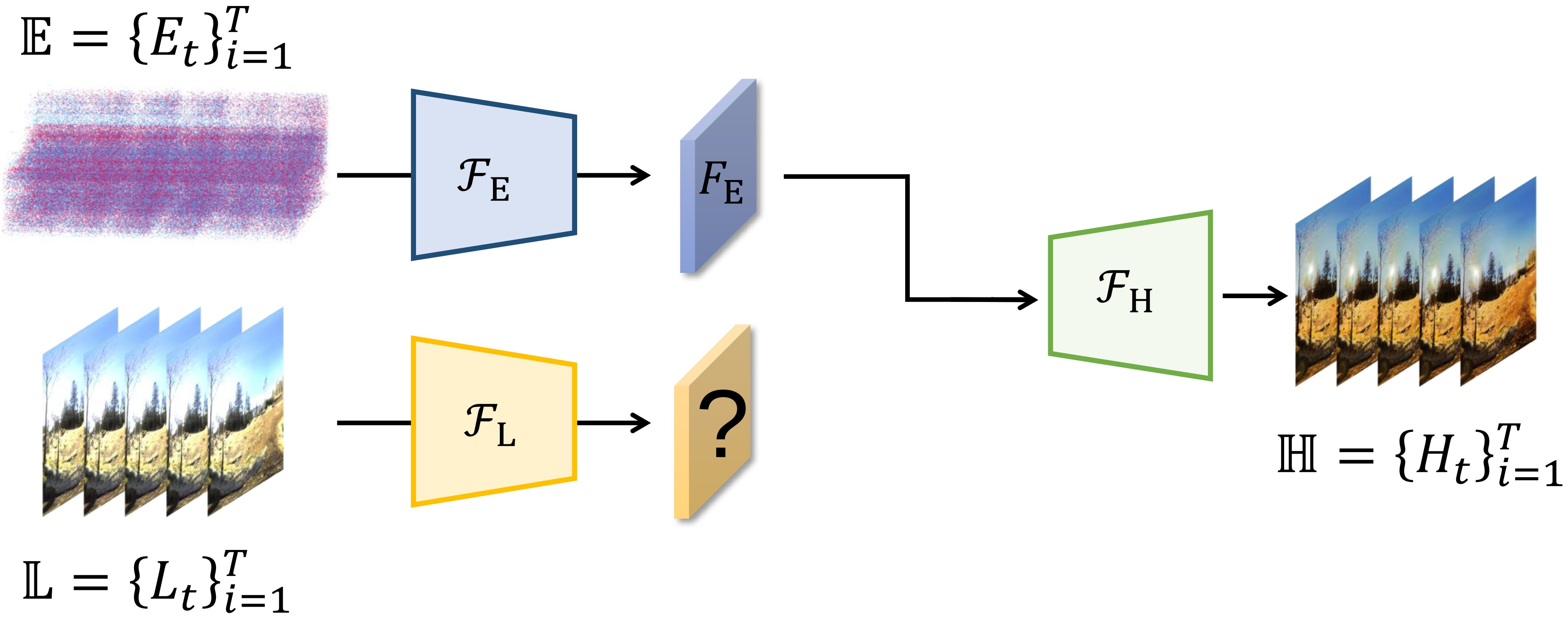


$$\mathbb{H} = \{H_t\}_{i=1}^{T}$$

$$\mathbb{L} = \{L_t\}_{i=1}^{T}$$

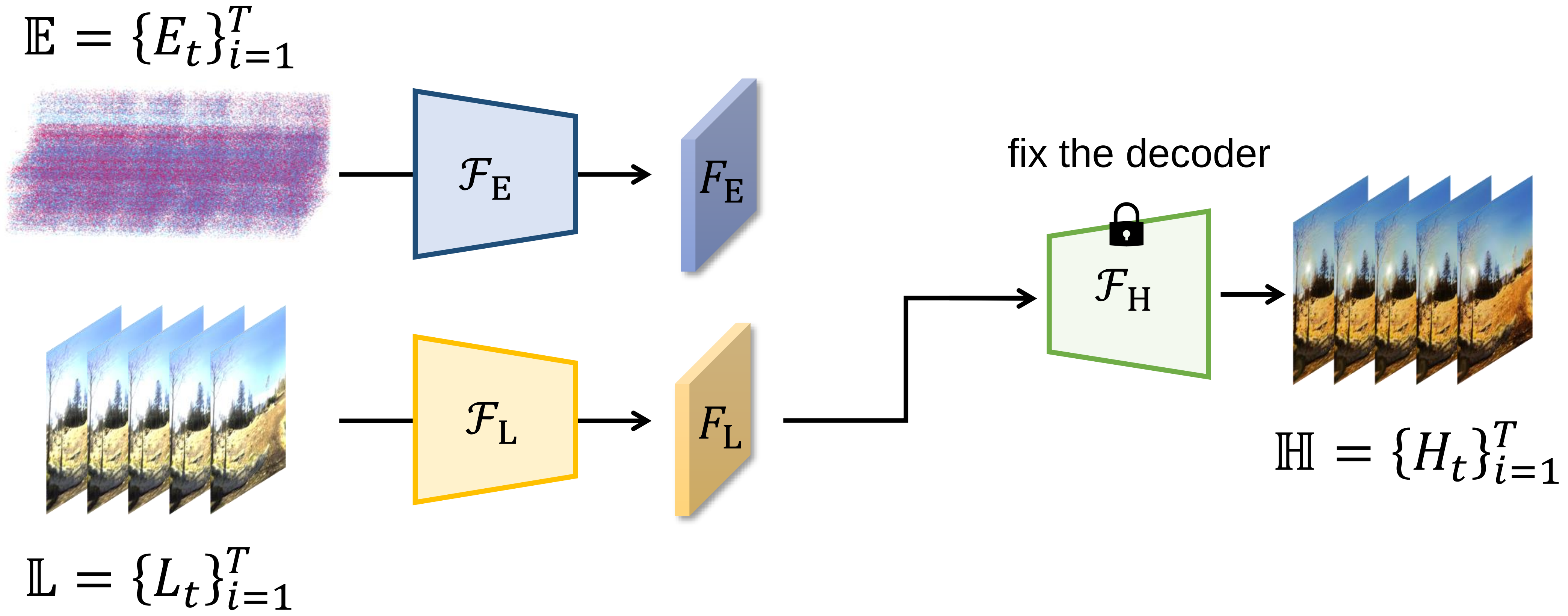
- How to represent events and LDR frames in a shared latent space by two modality specific encoders?
- How to extract the common and complementary scene information from different modalities?

Multimodal Representation Alignment



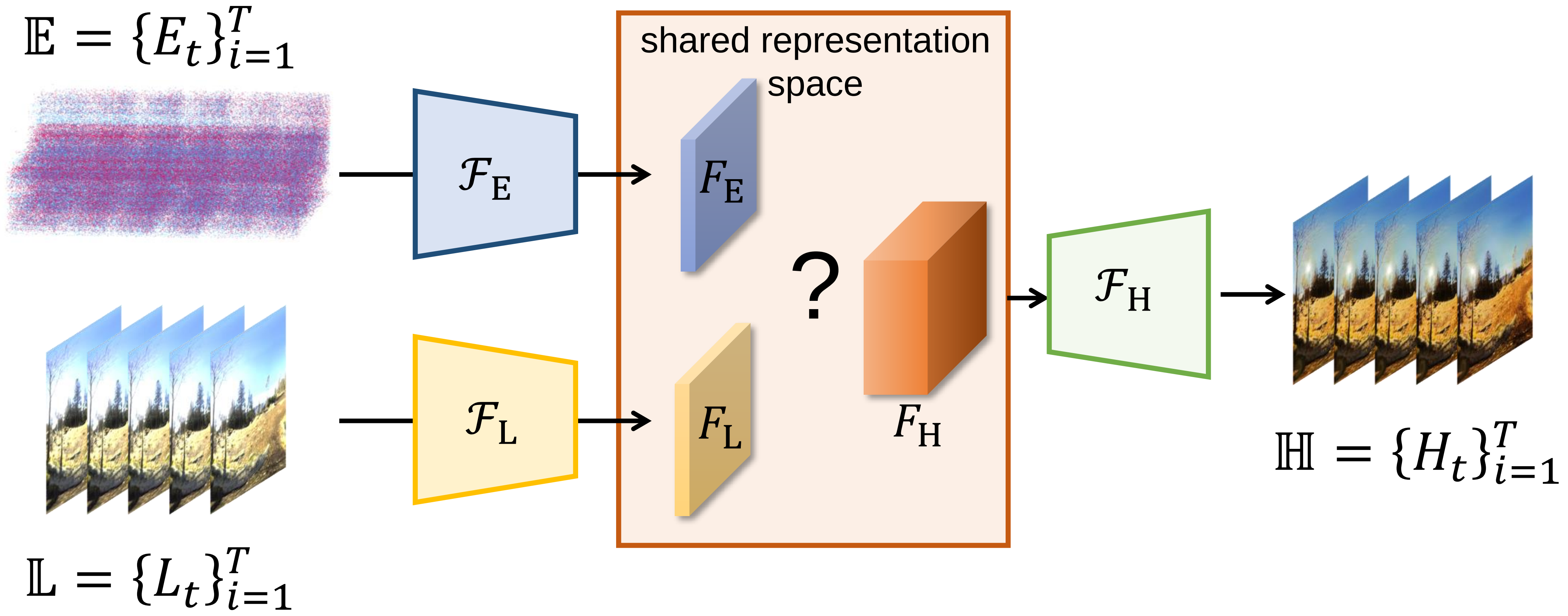
- Perform the inter-modality reconstruction (event to image)

Multimodal Representation Alignment



- Perform the intra-modality reconstruction (LDR image to HDR image)

Multimodal Fusion

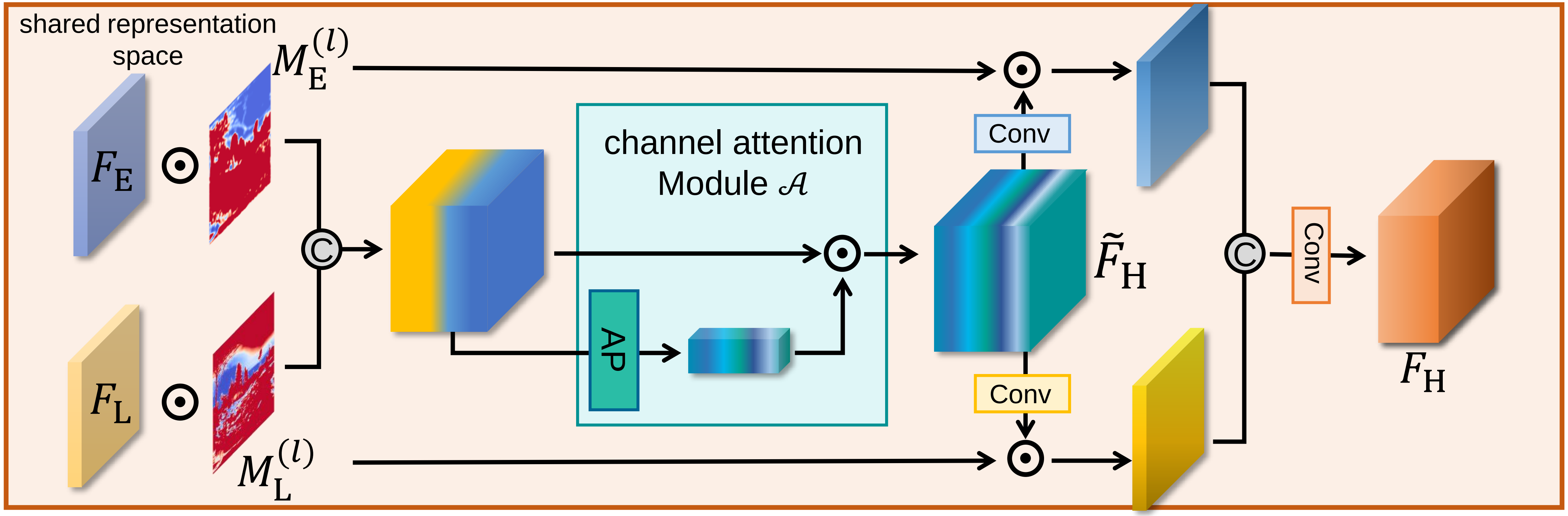


- How to provide a proper representation of HDR frames?

Confidence Guided Multimodal Fusion

$\odot$ multiplication

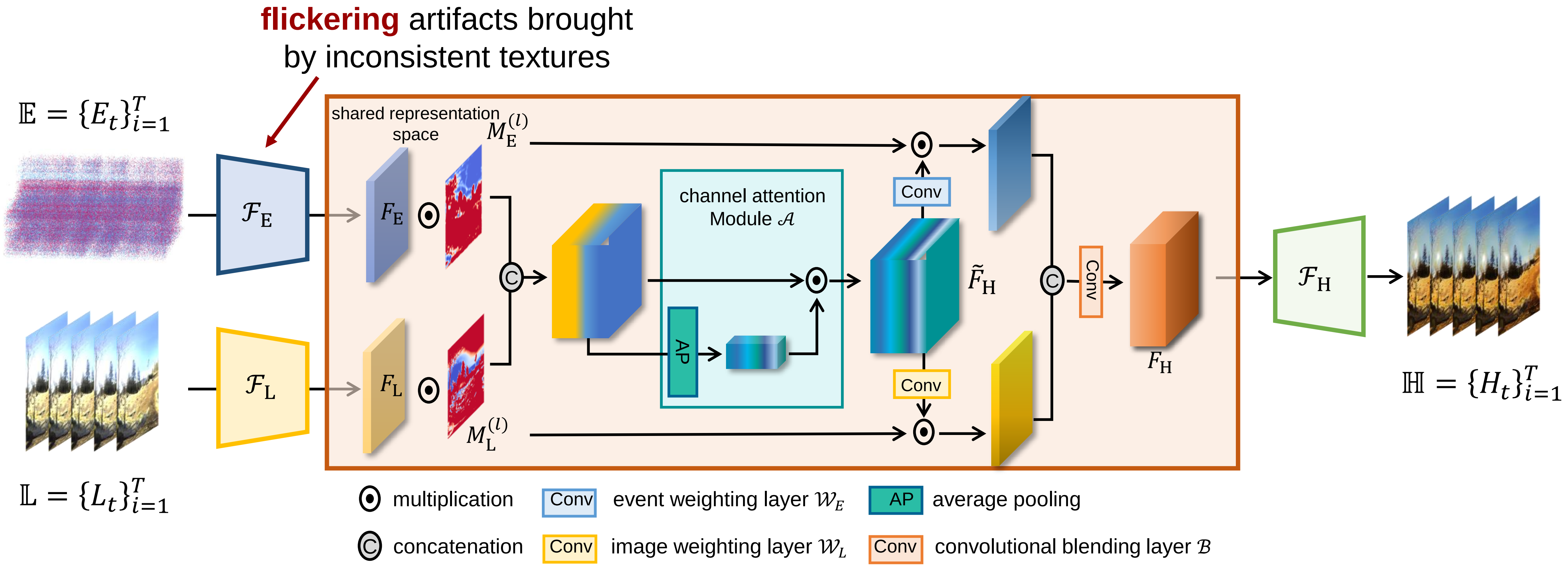
Confidence Guided Multimodal Fusion



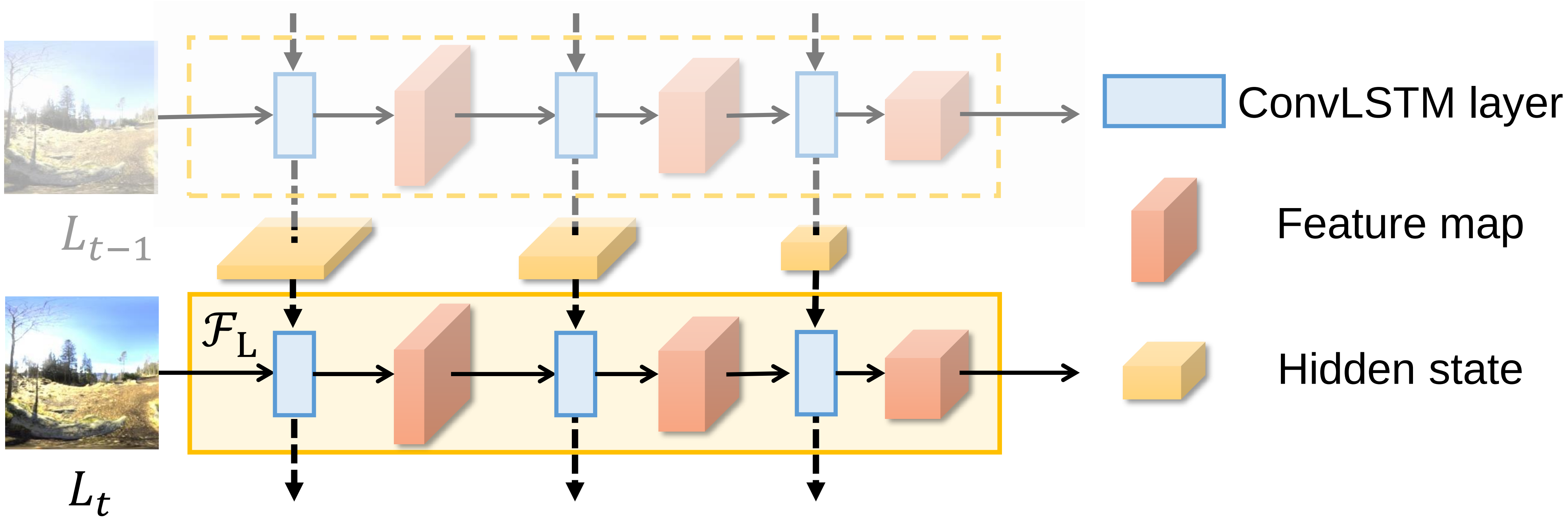
⊙ multiplication　Conv event weighting layer $\mathcal{W}_E$　AP average pooling

Ⓒ concatenation　Conv image weighting layer $\mathcal{W}_L$　Conv convolutional blending layer $\mathcal{B}$

Frame-based Pipeline

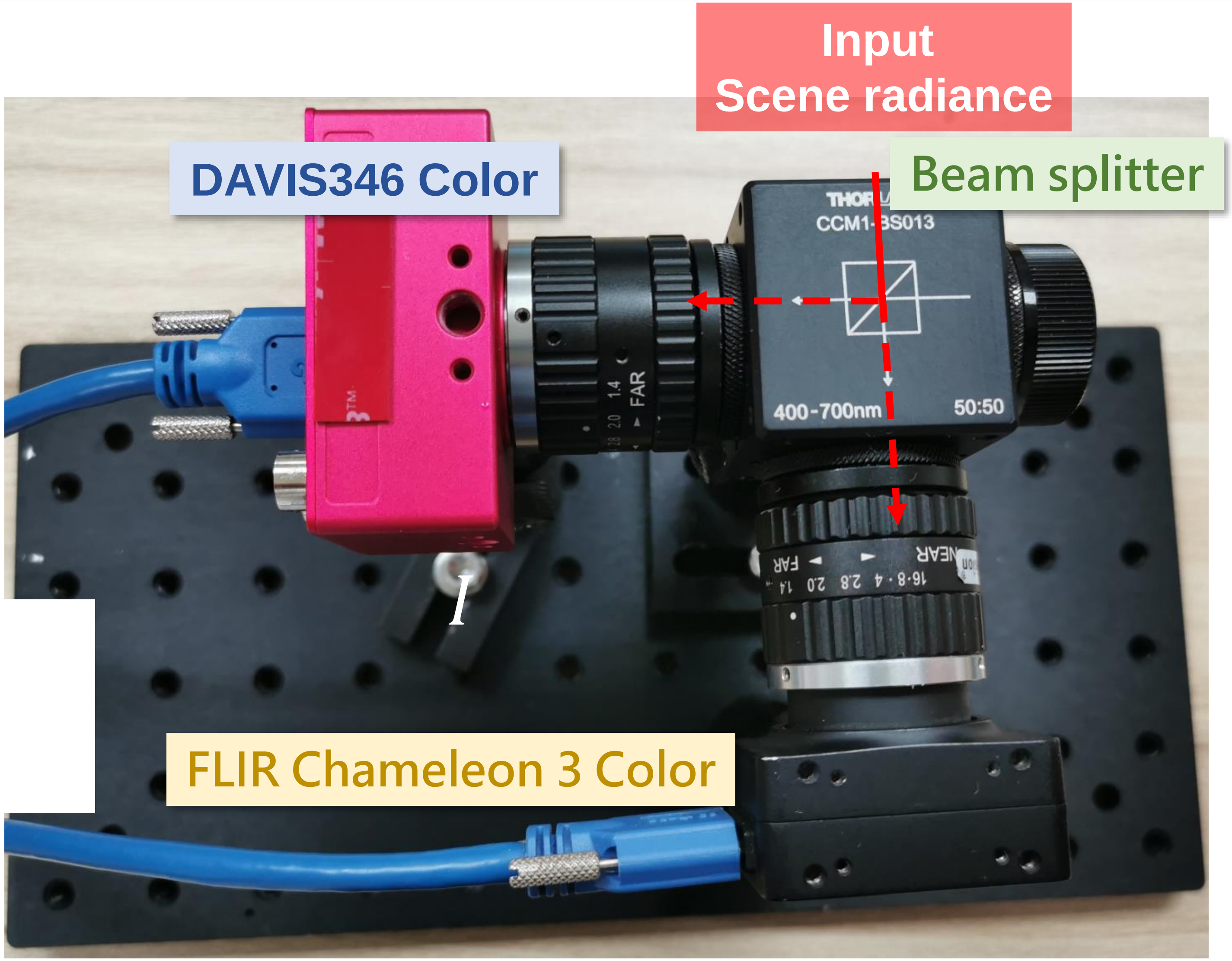


Temporal Context Encoding

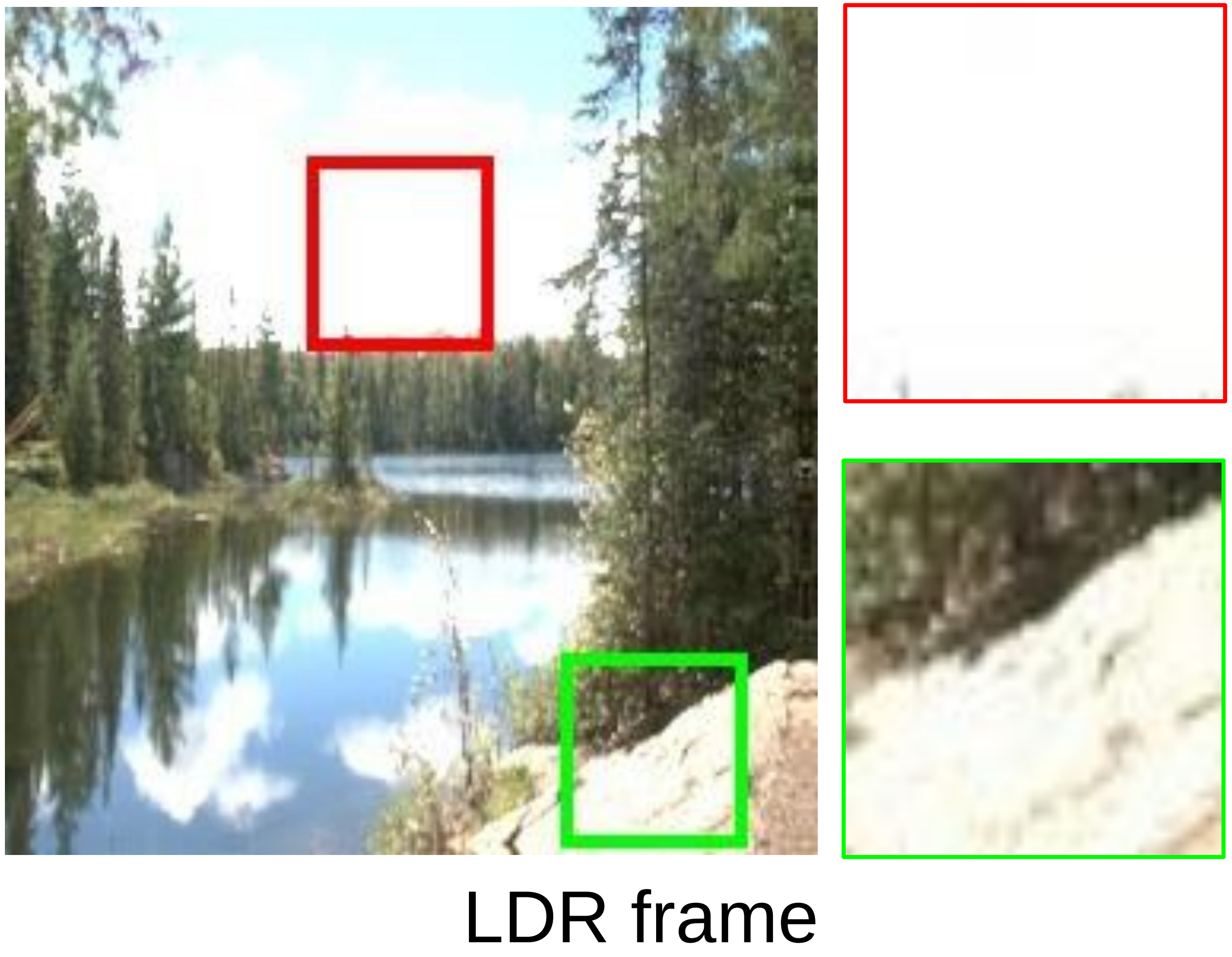


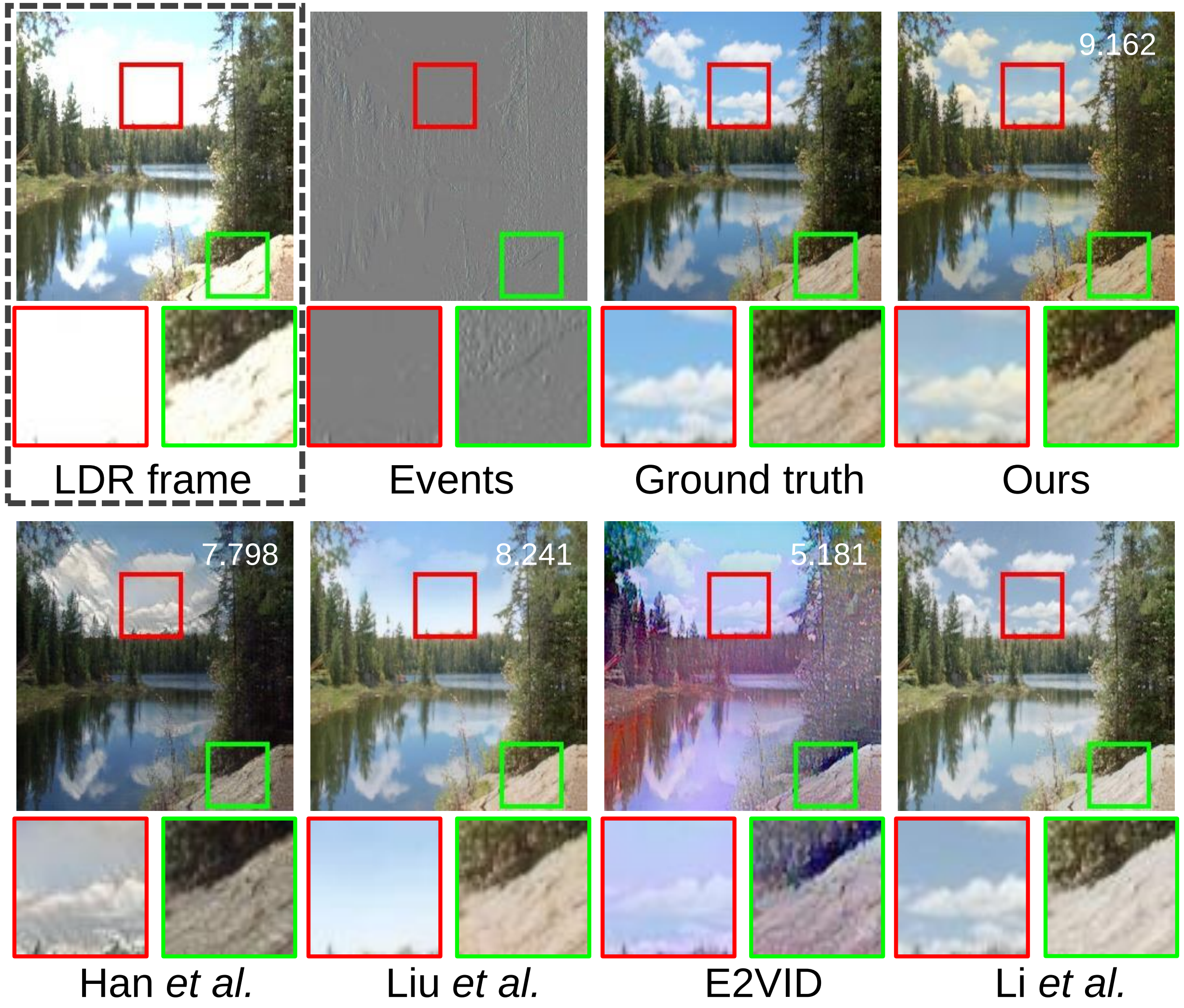
- Introduce temporal correlation to alleviate **flickering** and reduce noise

Hybrid Camera System



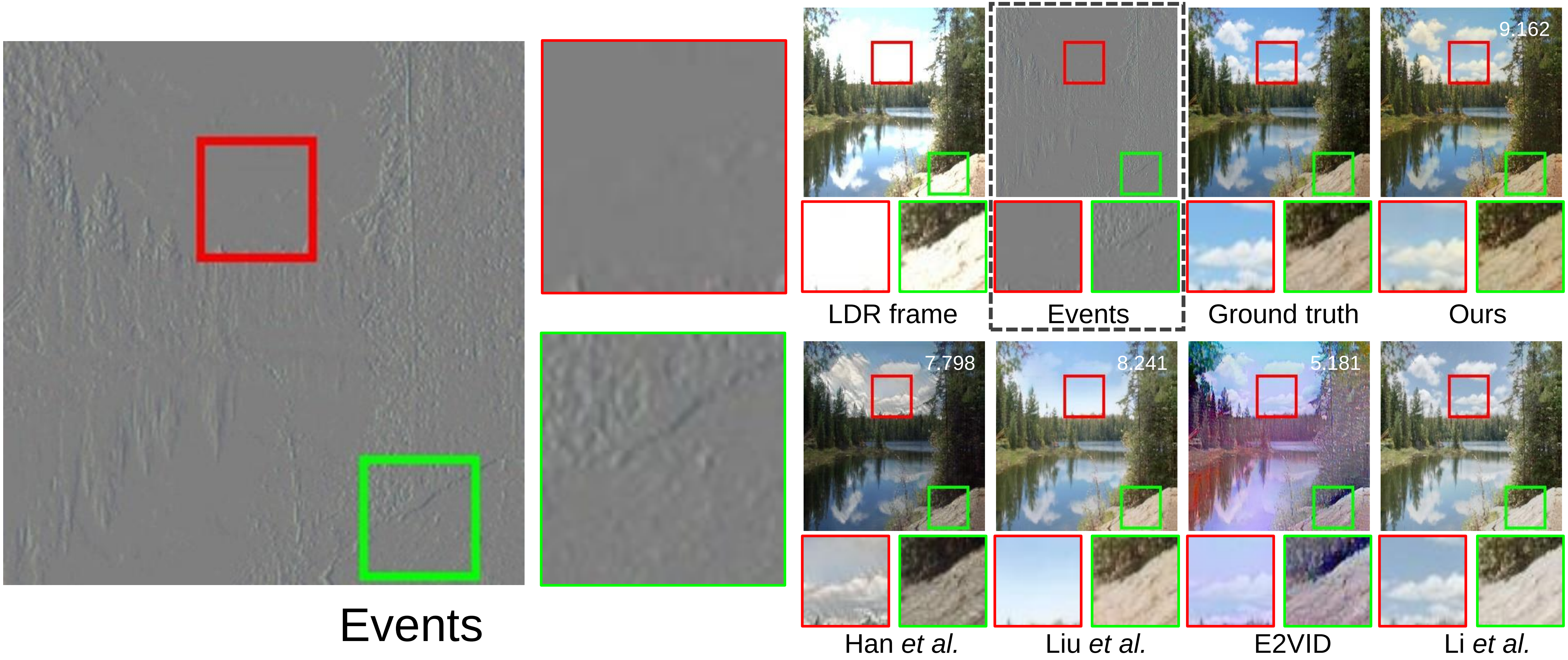
Evaluation on Synthetic Data





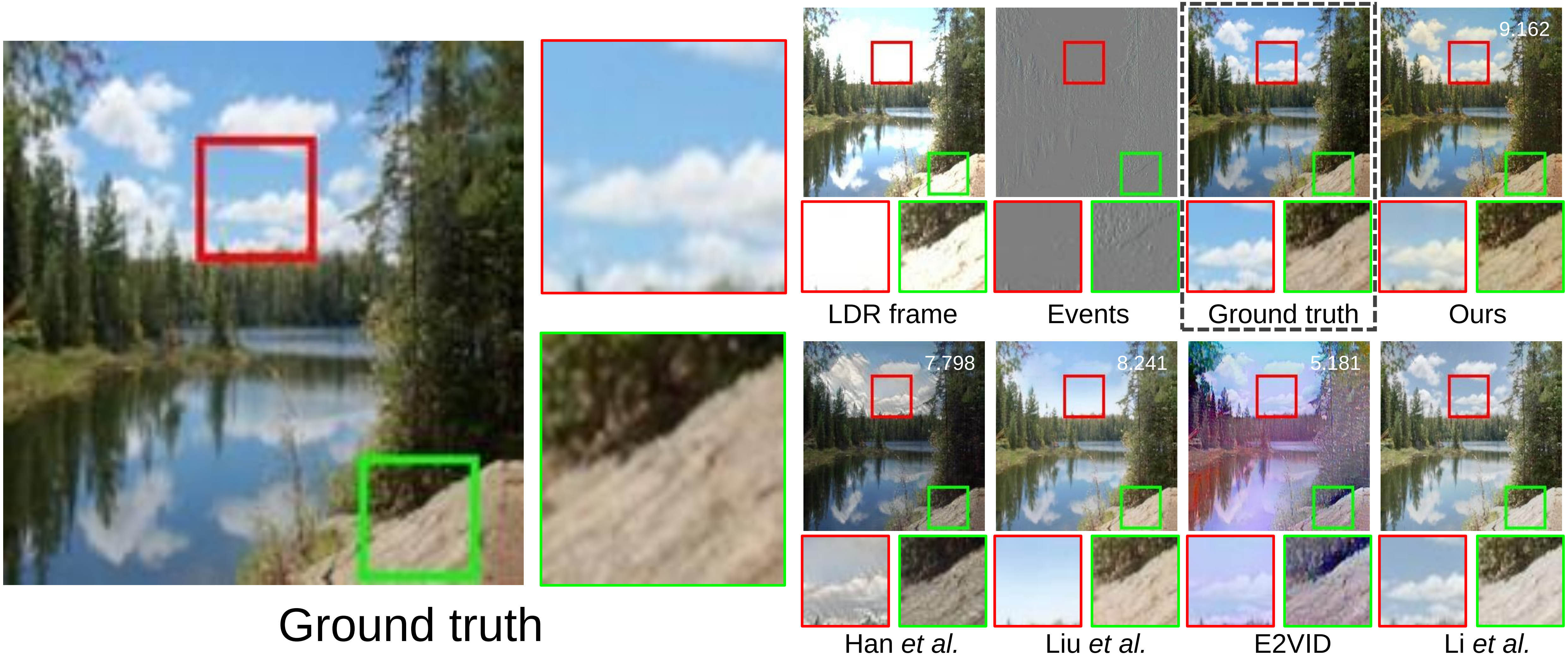
Single image based HDR: [*Liu et al.*, CVPR 2020] Two-exposure based HDR: [Li *et al.*, TIP2020]
Event based HDR: [E2VID, TPAMI2019] Hybrid HDR: [Han *et al.*, CVPR2020]

Evaluation on Synthetic Data



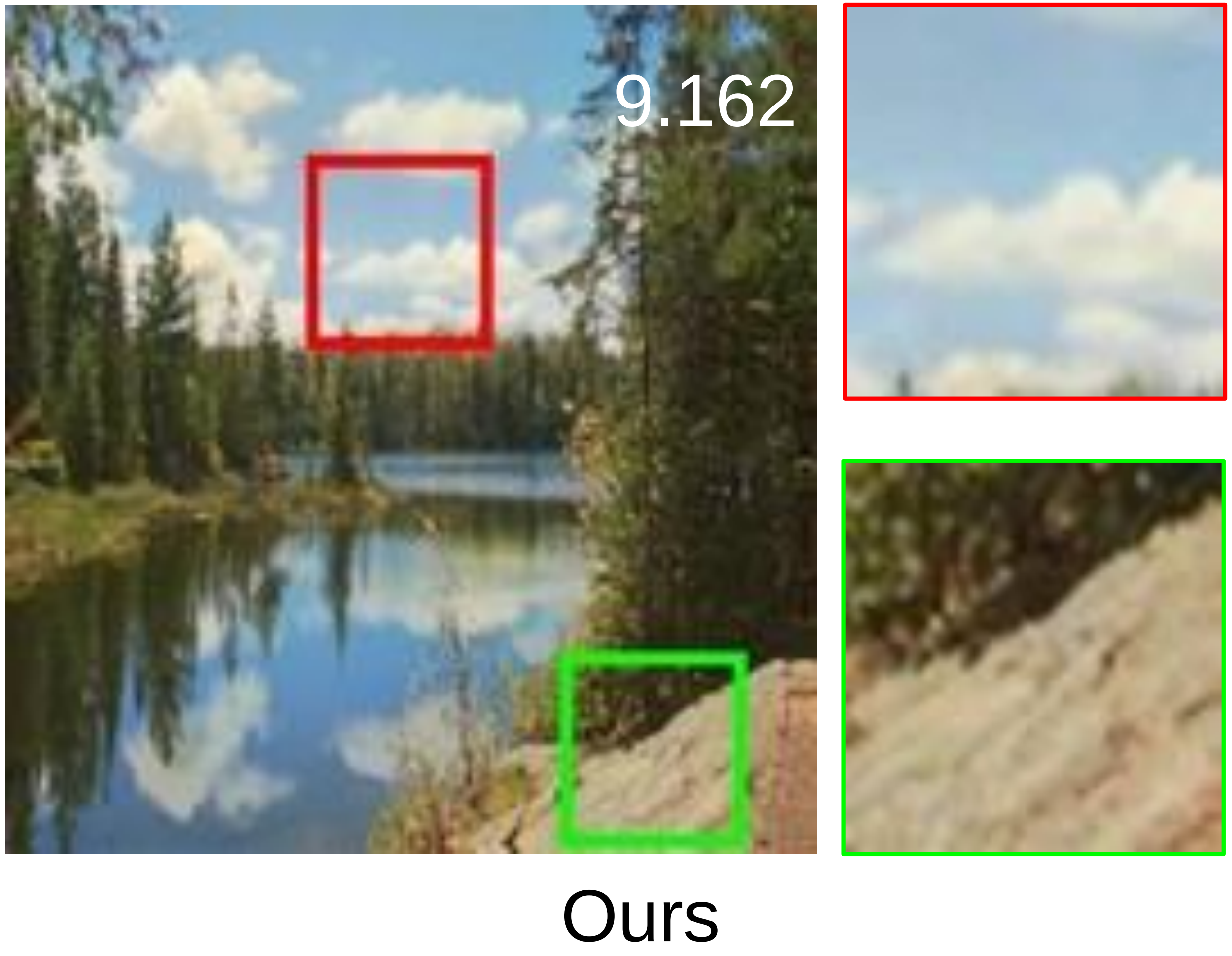
Single image based HDR: [*Liu et al.*, CVPR 2020] Two-exposure based HDR: [Li *et al.*, TIP2020]
Event based HDR: [E2VID, TPAMI2019] Hybrid HDR: [Han *et al.*, CVPR2020]

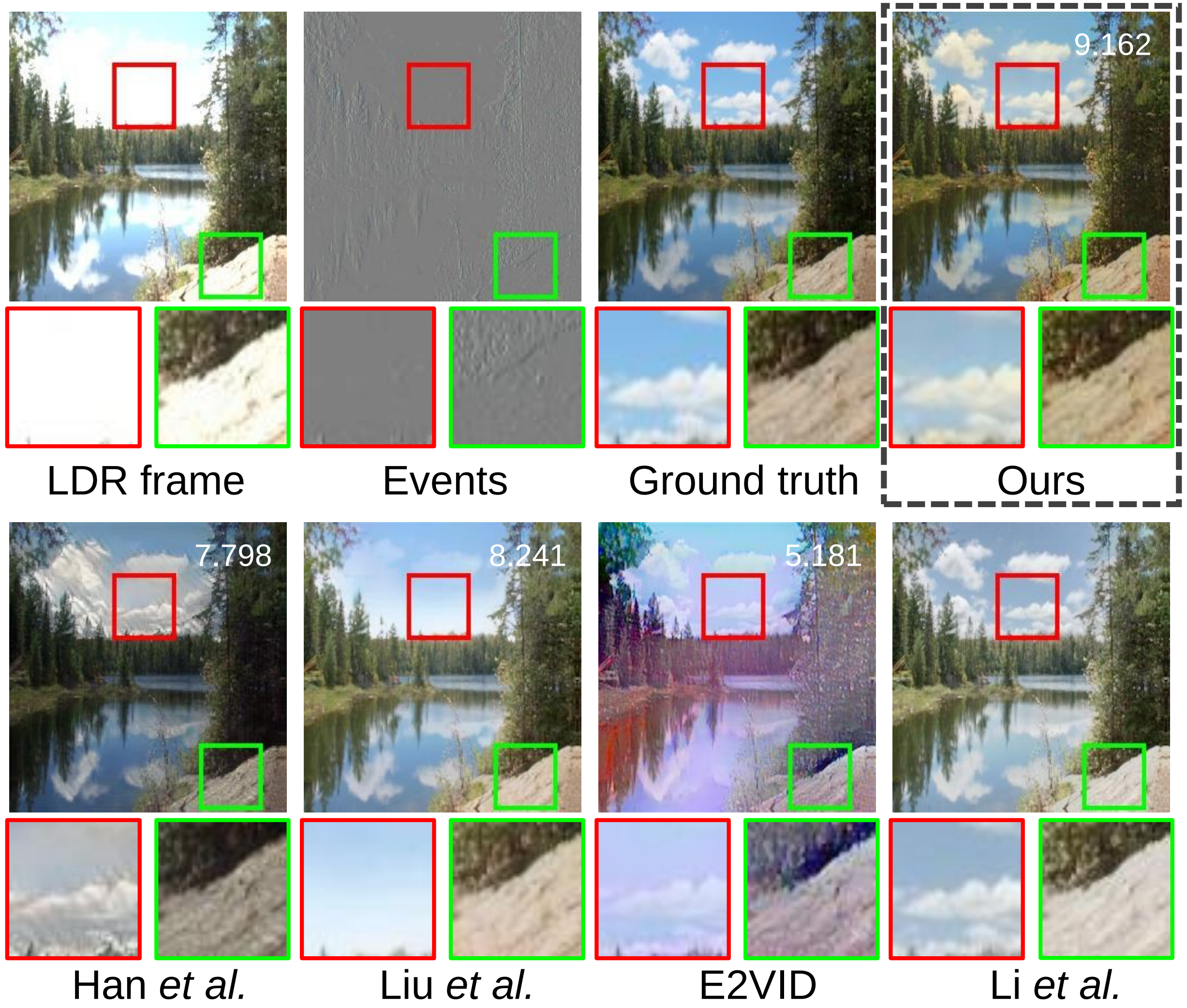
Evaluation on Synthetic Data



Single image based HDR: [*Liu et al.*, CVPR 2020] Two-exposure based HDR: [Li *et al.*, TIP2020]
Event based HDR: [E2VID, TPAMI2019] Hybrid HDR: [Han *et al.*, CVPR2020]

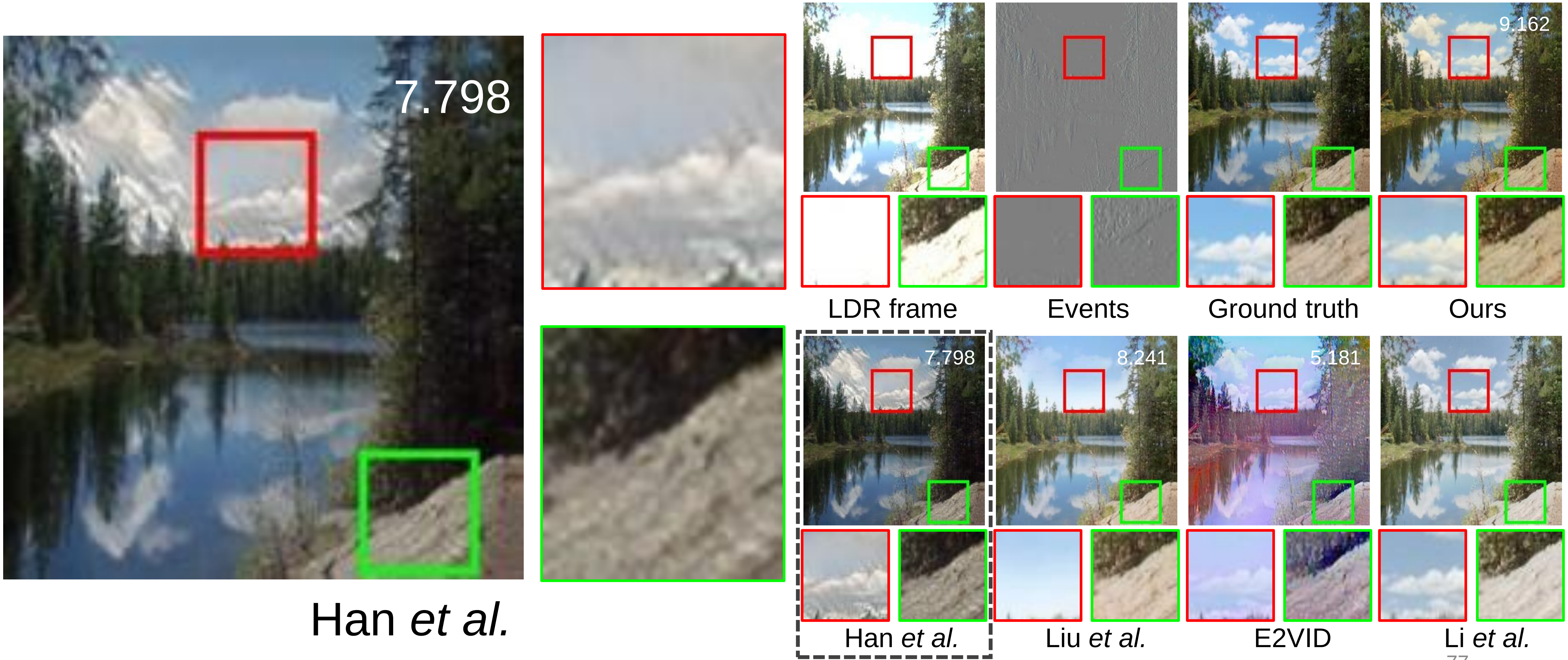
Evaluation on Synthetic Data





Single image based HDR: [*Liu et al.*, CVPR 2020] Two-exposure based HDR: [Li *et al.*, TIP2020]
Event based HDR: [E2VID, TPAMI2019] Hybrid HDR: [Han *et al.*, CVPR2020]

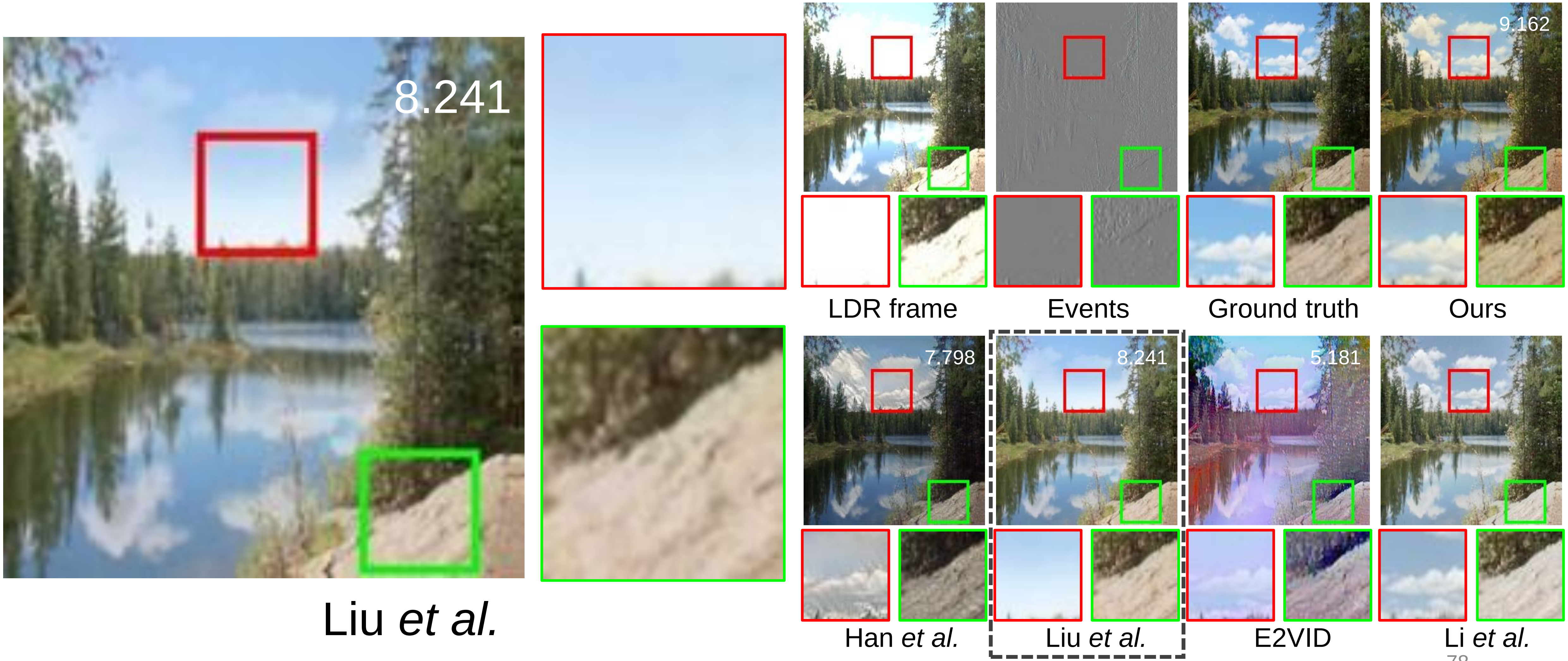
Evaluation on Synthetic Data



Han *et al.*

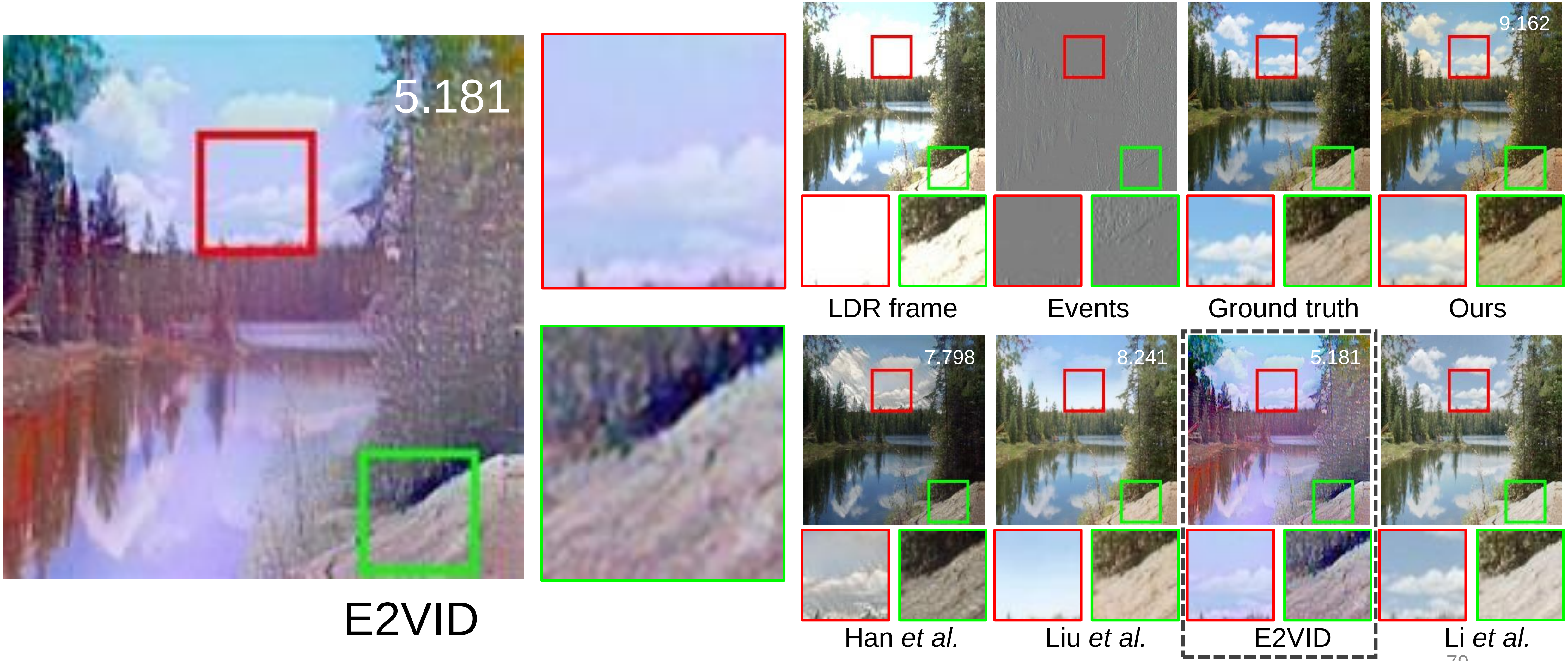
Single image based HDR: [*Liu et al.*, CVPR 2020] Two-exposure based HDR: [Li *et al.*, TIP2020]
Event based HDR: [E2VID, TPAMI2019] Hybrid HDR: [Han *et al.*, CVPR2020]

Evaluation on Synthetic Data



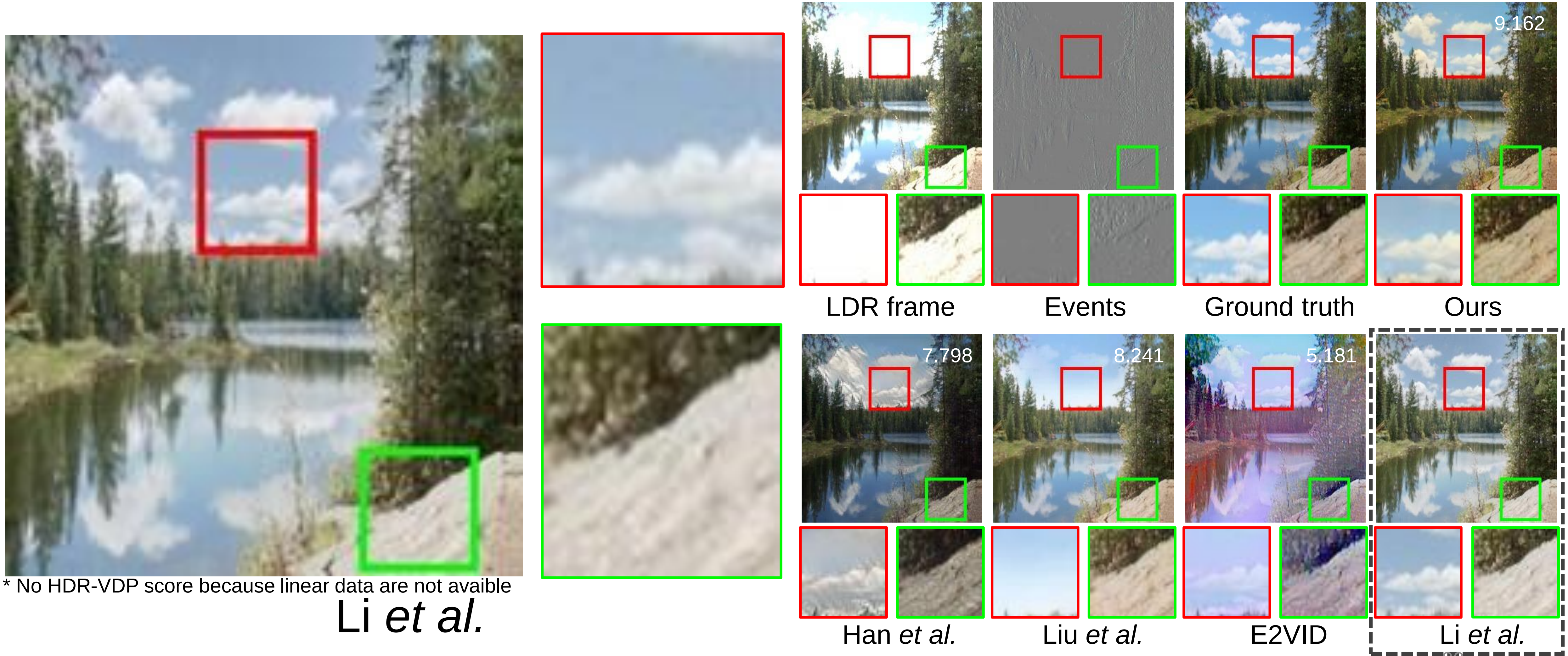
Single image based HDR: [*Liu et al.*, CVPR 2020] Two-exposure based HDR: [Li *et al.*, TIP2020]
Event based HDR: [E2VID, TPAMI2019] Hybrid HDR: [Han *et al.*, CVPR2020]

Evaluation on Synthetic Data



Single image based HDR: [*Liu et al.*, CVPR 2020] Two-exposure based HDR: [Li *et al.*, TIP2020]
Event based HDR: [E2VID, TPAMI2019] Hybrid HDR: [Han *et al.*, CVPR2020]

Evaluation on Synthetic Data

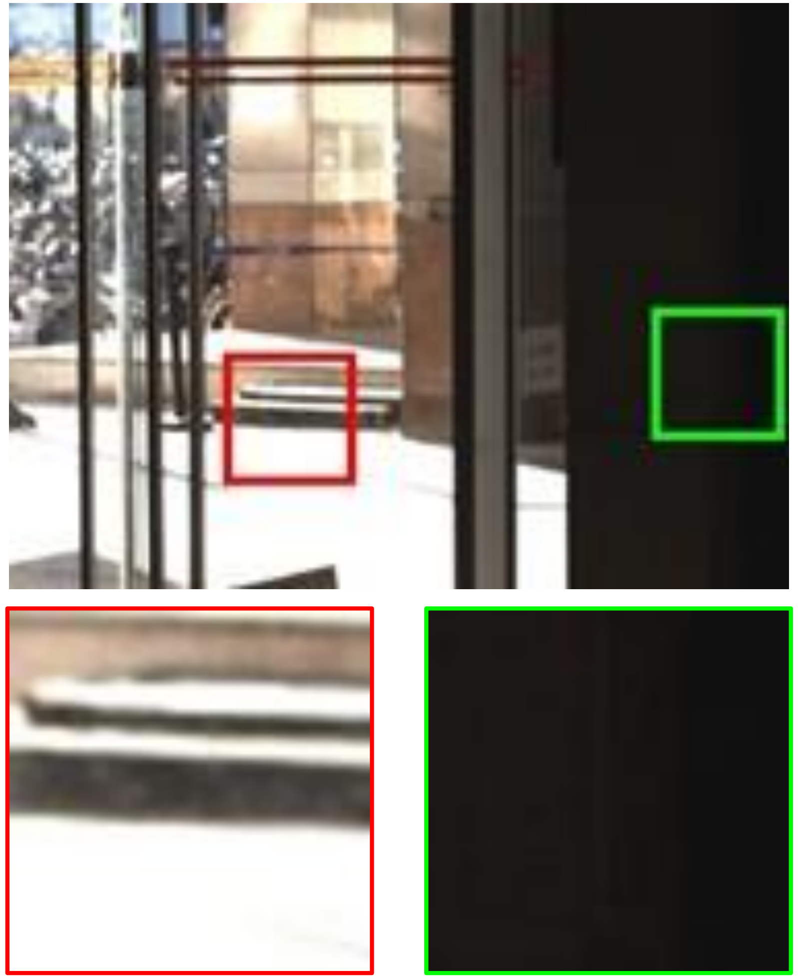


* No HDR-VDP score because linear data are not avaible

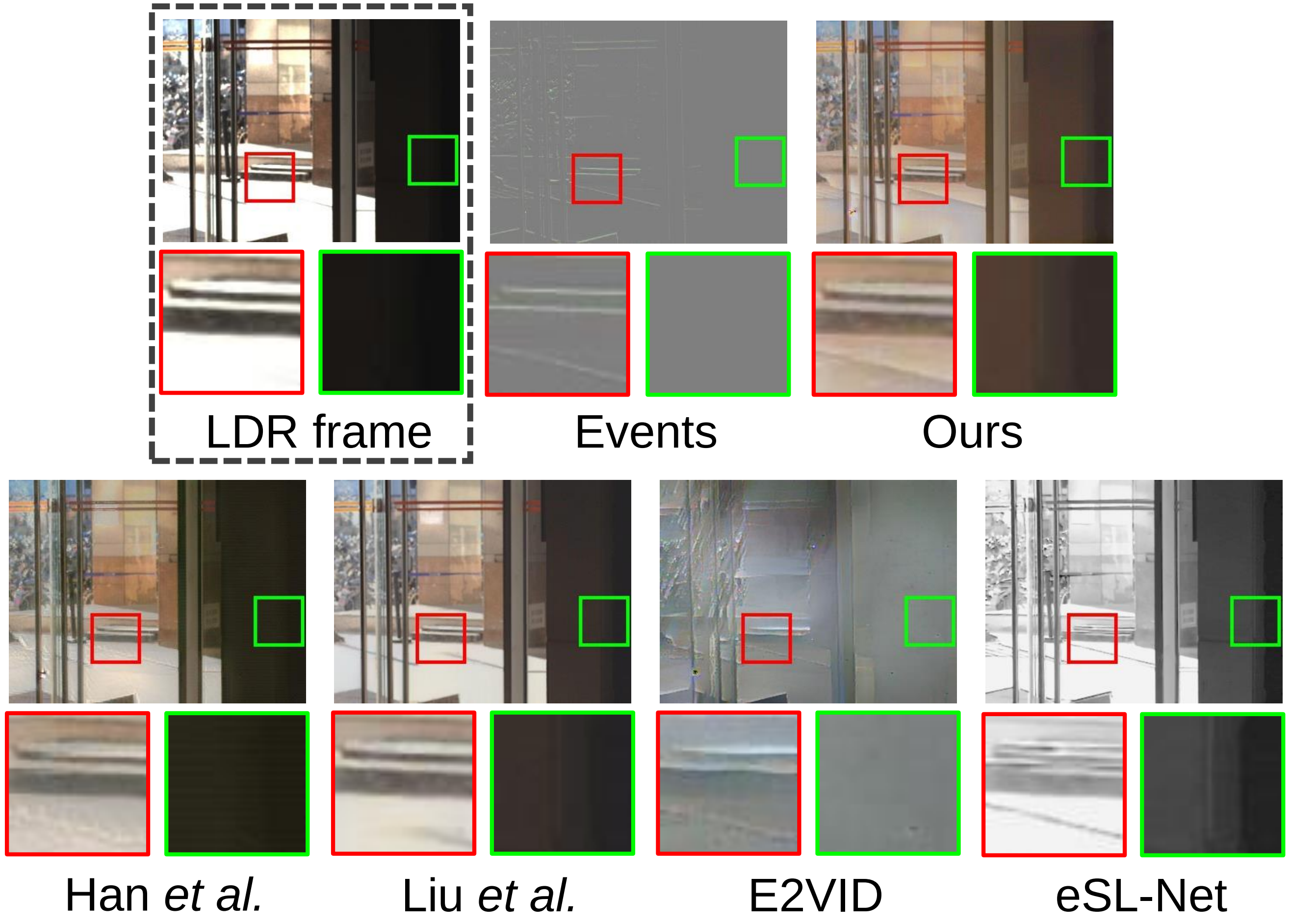
Li *et al.*

Single image based HDR: [*Liu et al.*, CVPR 2020] Two-exposure based HDR: [Li *et al.*, TIP2020]
Event based HDR: [E2VID, TPAMI2019] Hybrid HDR: [Han *et al.*, CVPR2020]

Evaluation on Real Data

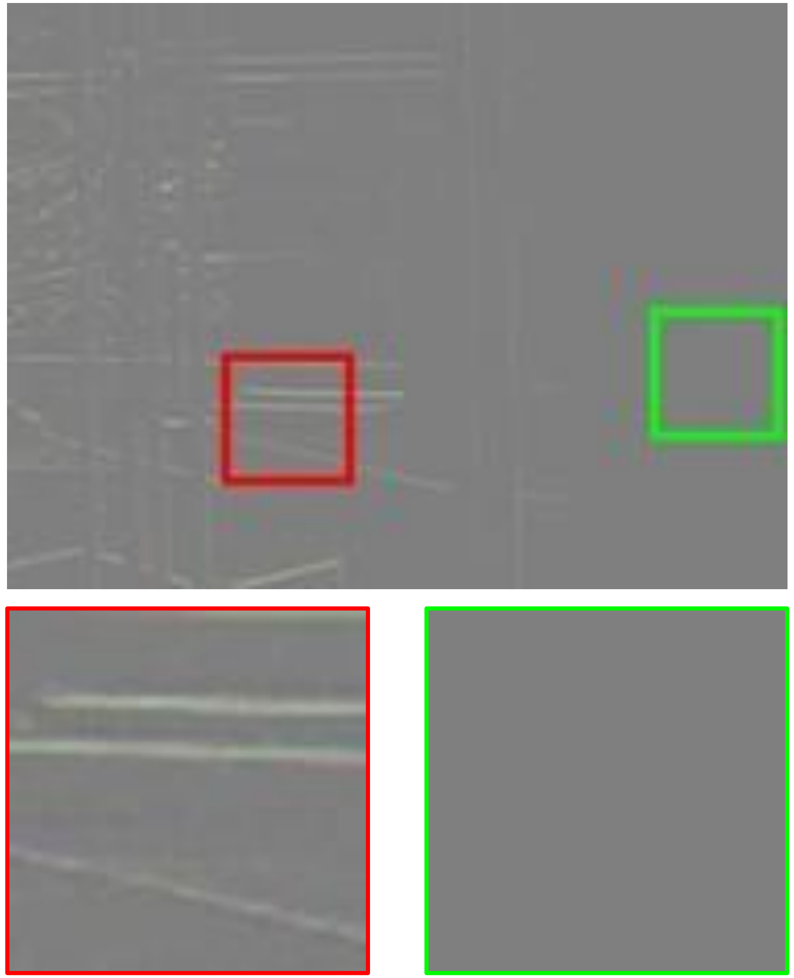


LDR frame

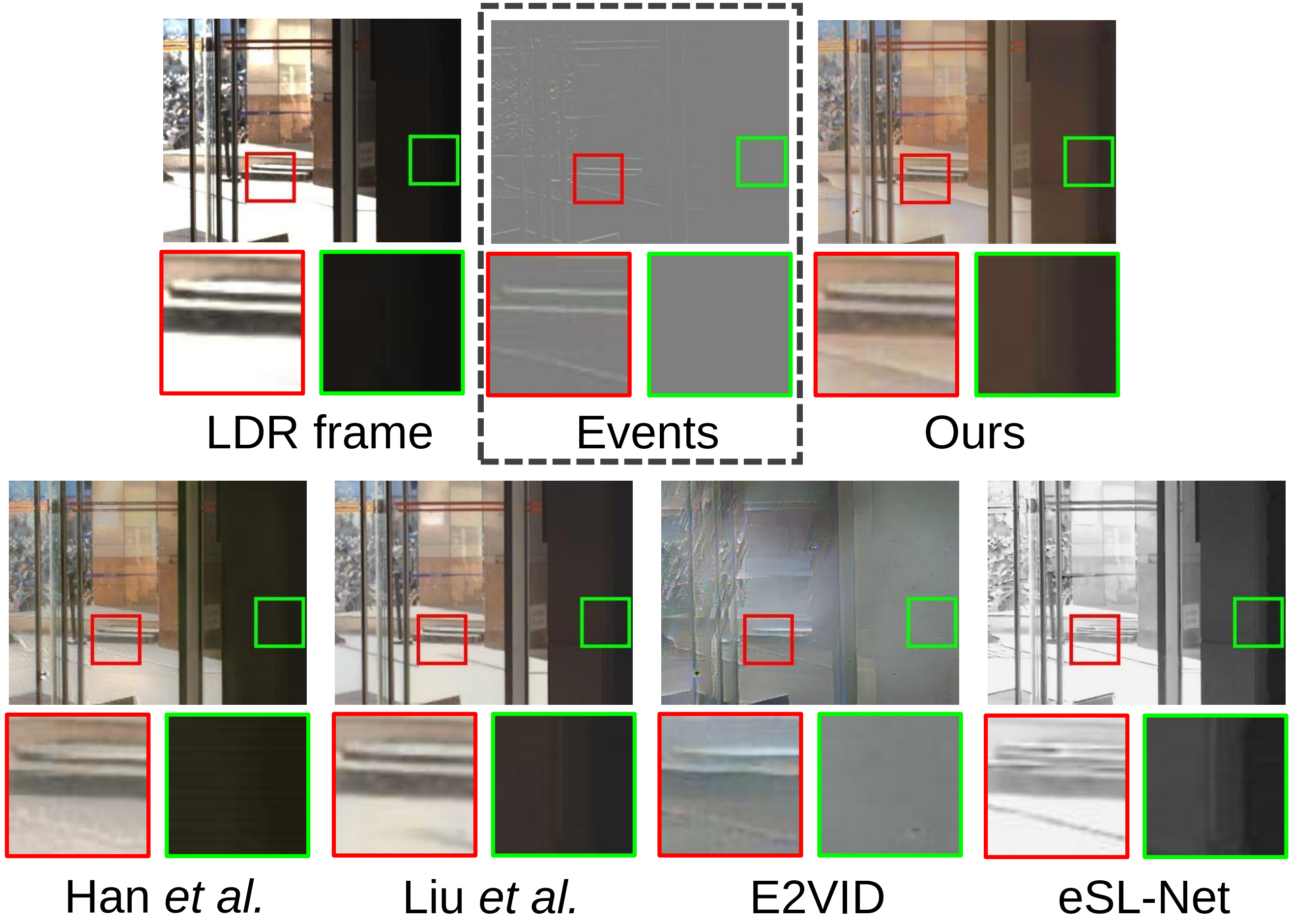


Single image based HDR: [*Liu et al.*, CVPR 2020] Event based HDR: [E2VID, TPAMI2019]
Hybrid HDR: [Han *et al.*, CVPR2020] [eSL-Net, CVPR 2020]

Evaluation on Real Data

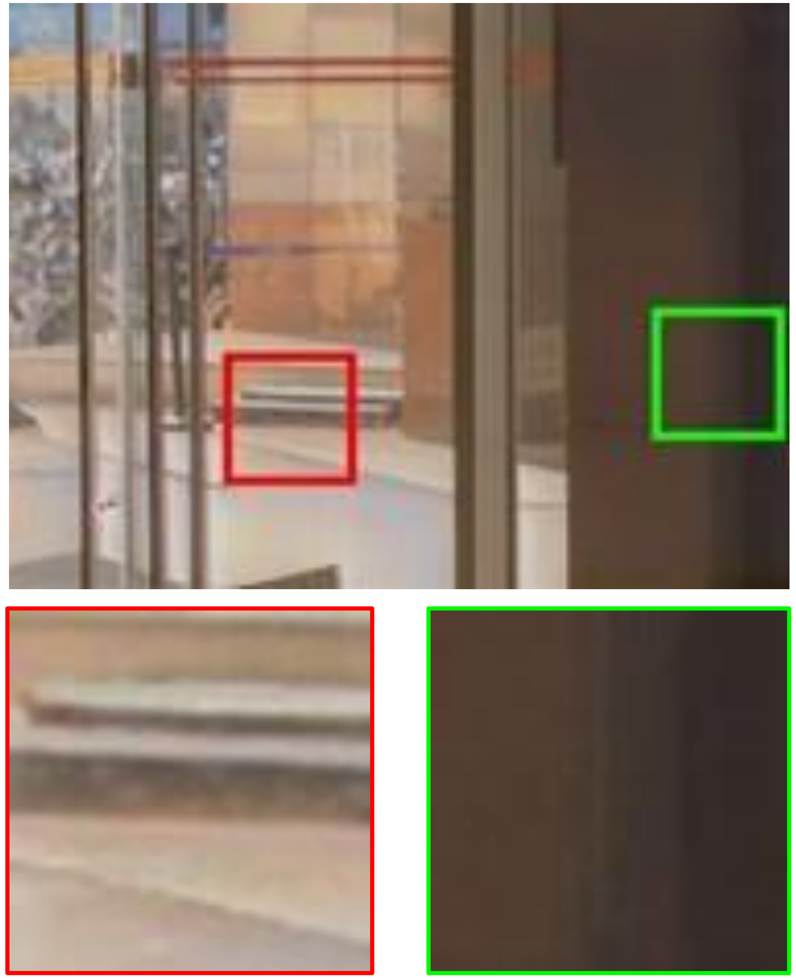


Events

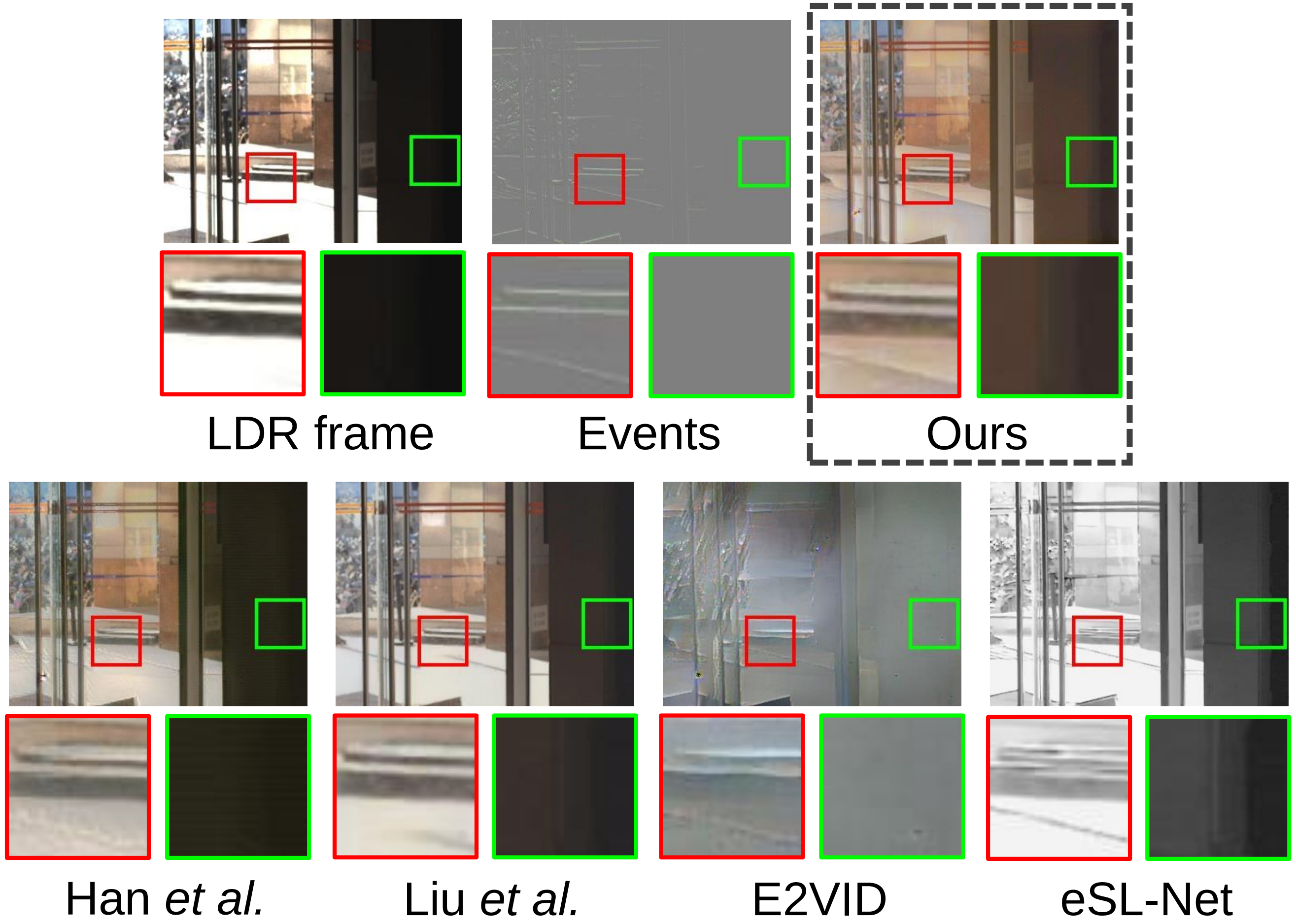


Single image based HDR: [*Liu et al.*, CVPR 2020] Event based HDR: [E2VID, TPAMI2019]
Hybrid HDR: [Han *et al.*, CVPR2020] [eSL-Net, CVPR 2020]

Evaluation on Real Data

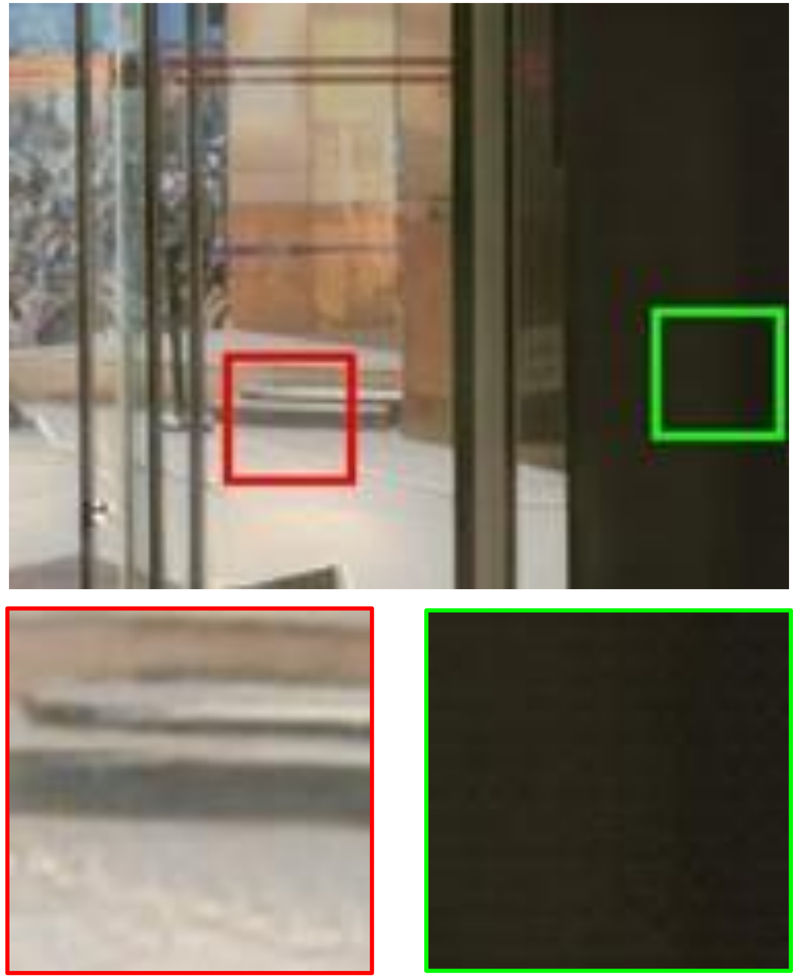


Ours

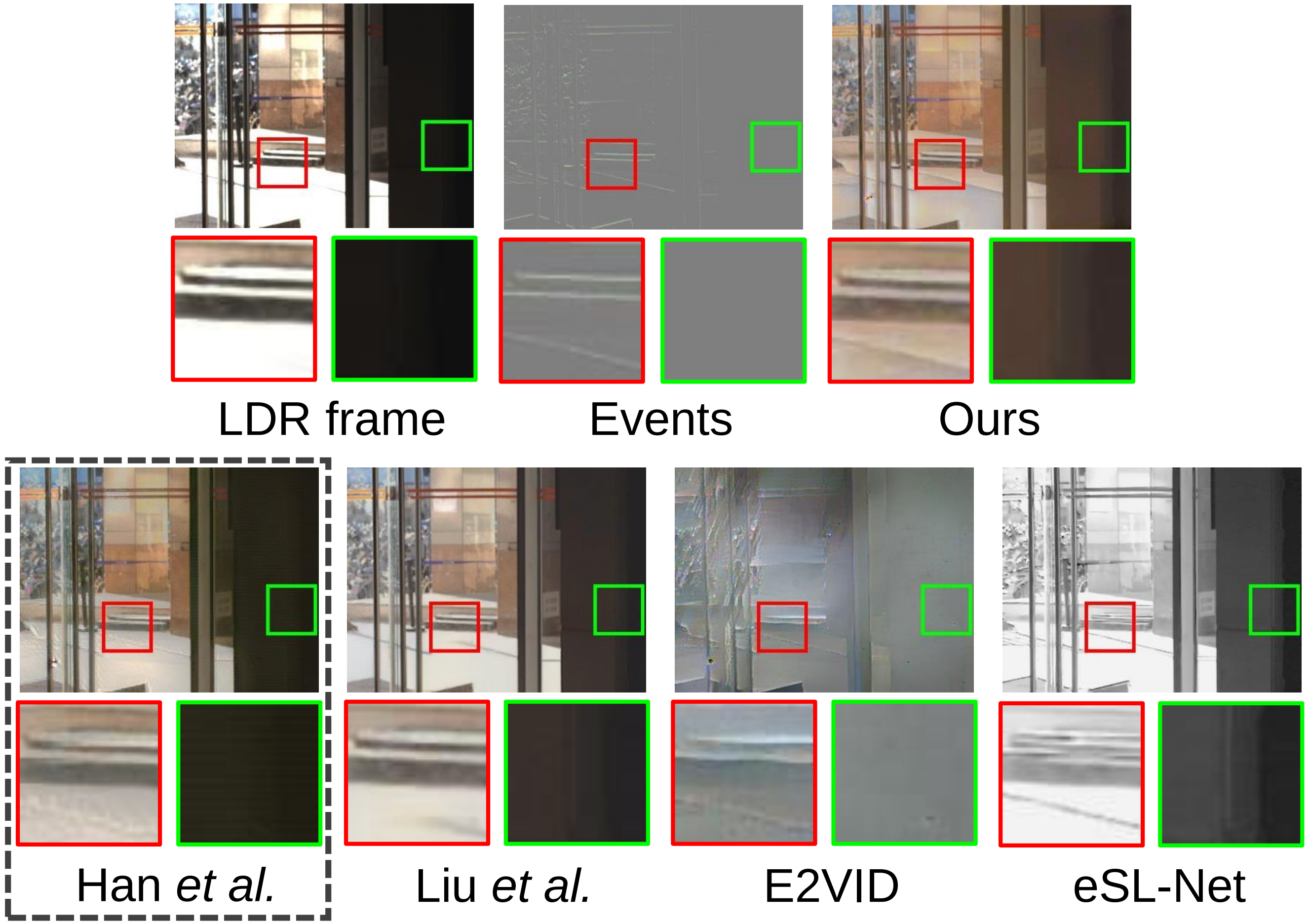


Single image based HDR: [*Liu et al.*, CVPR 2020] Event based HDR: [E2VID, TPAMI2019]
Hybrid HDR: [Han *et al.*, CVPR2020] [eSL-Net, CVPR 2020]

Evaluation on Real Data

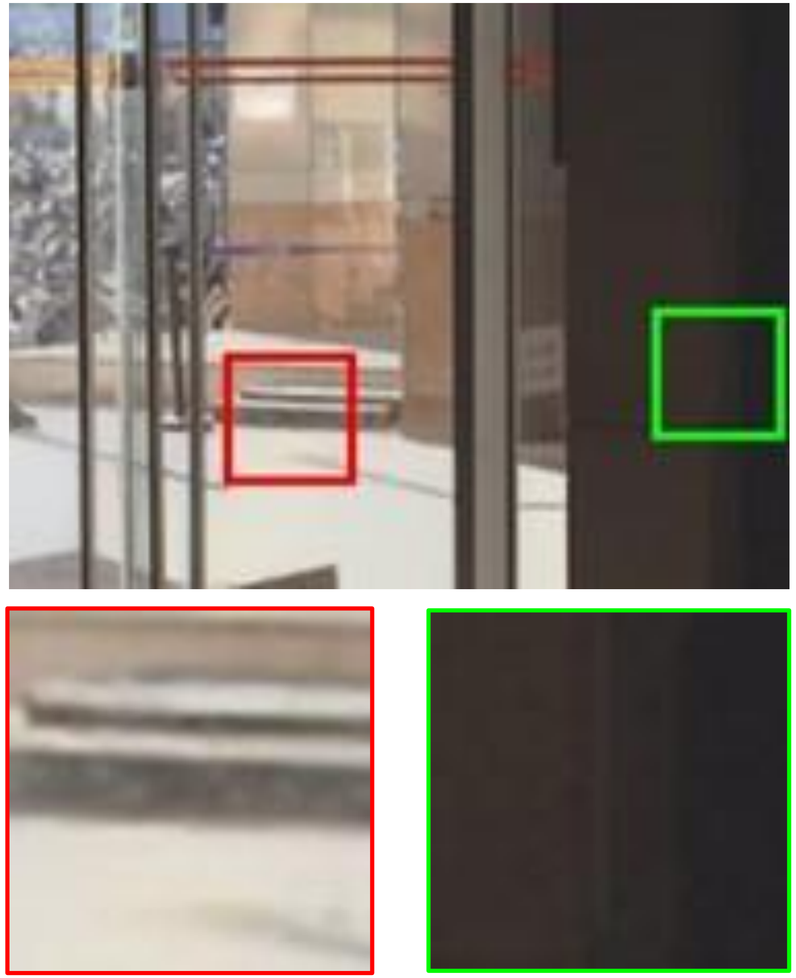


Han *et al.*

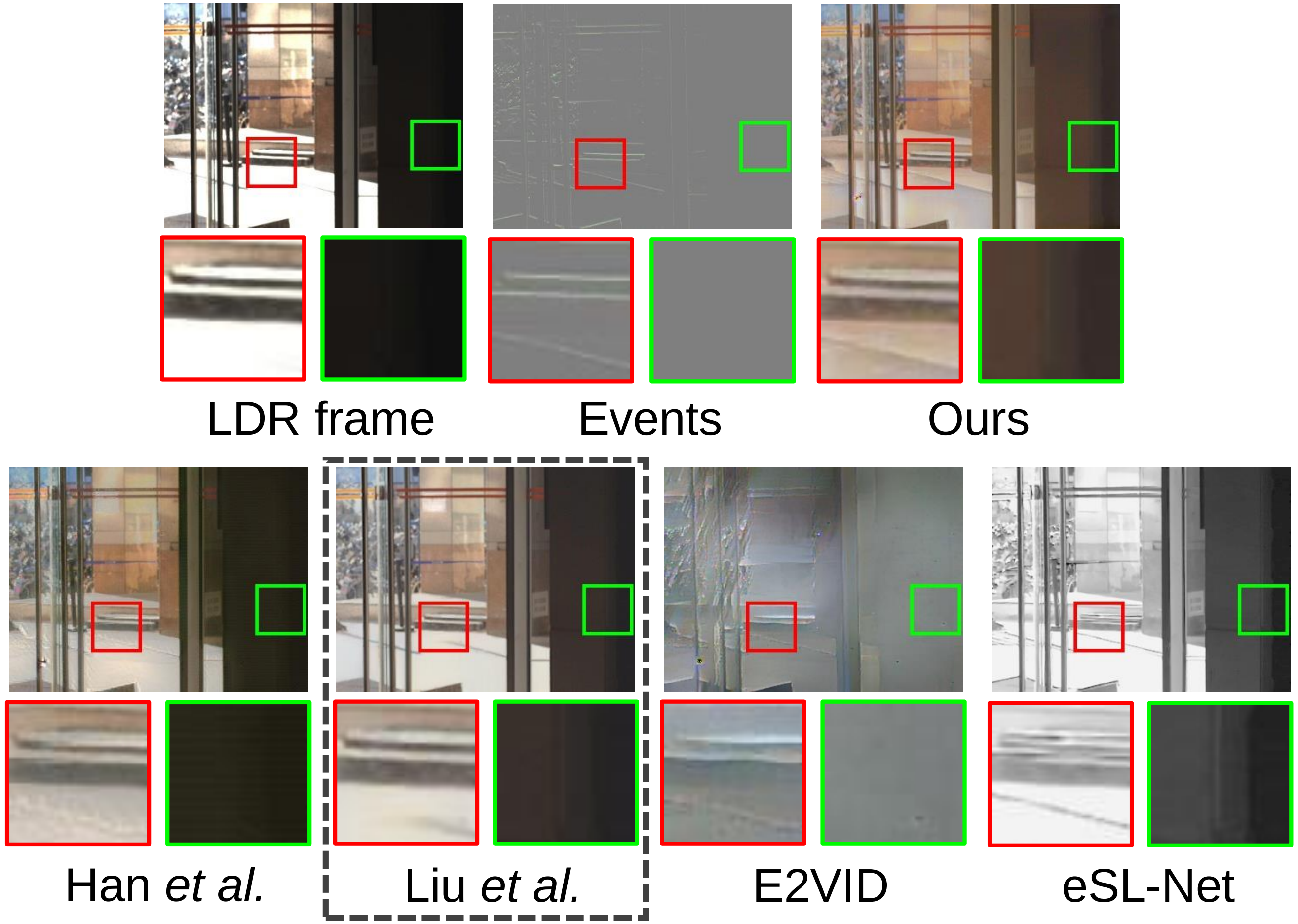


Single image based HDR: [*Liu et al.*, CVPR 2020] Event based HDR: [E2VID, TPAMI2019]
Hybrid HDR: [Han *et al.*, CVPR2020] [eSL-Net, CVPR 2020]

Evaluation on Real Data

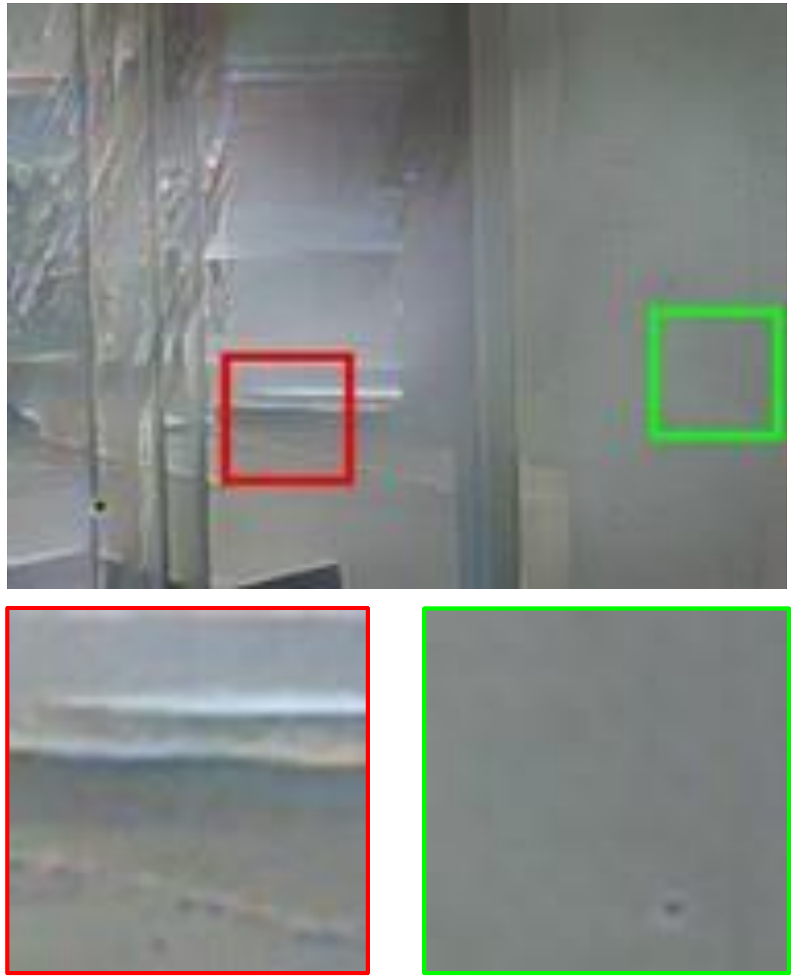


Liu *et al.*

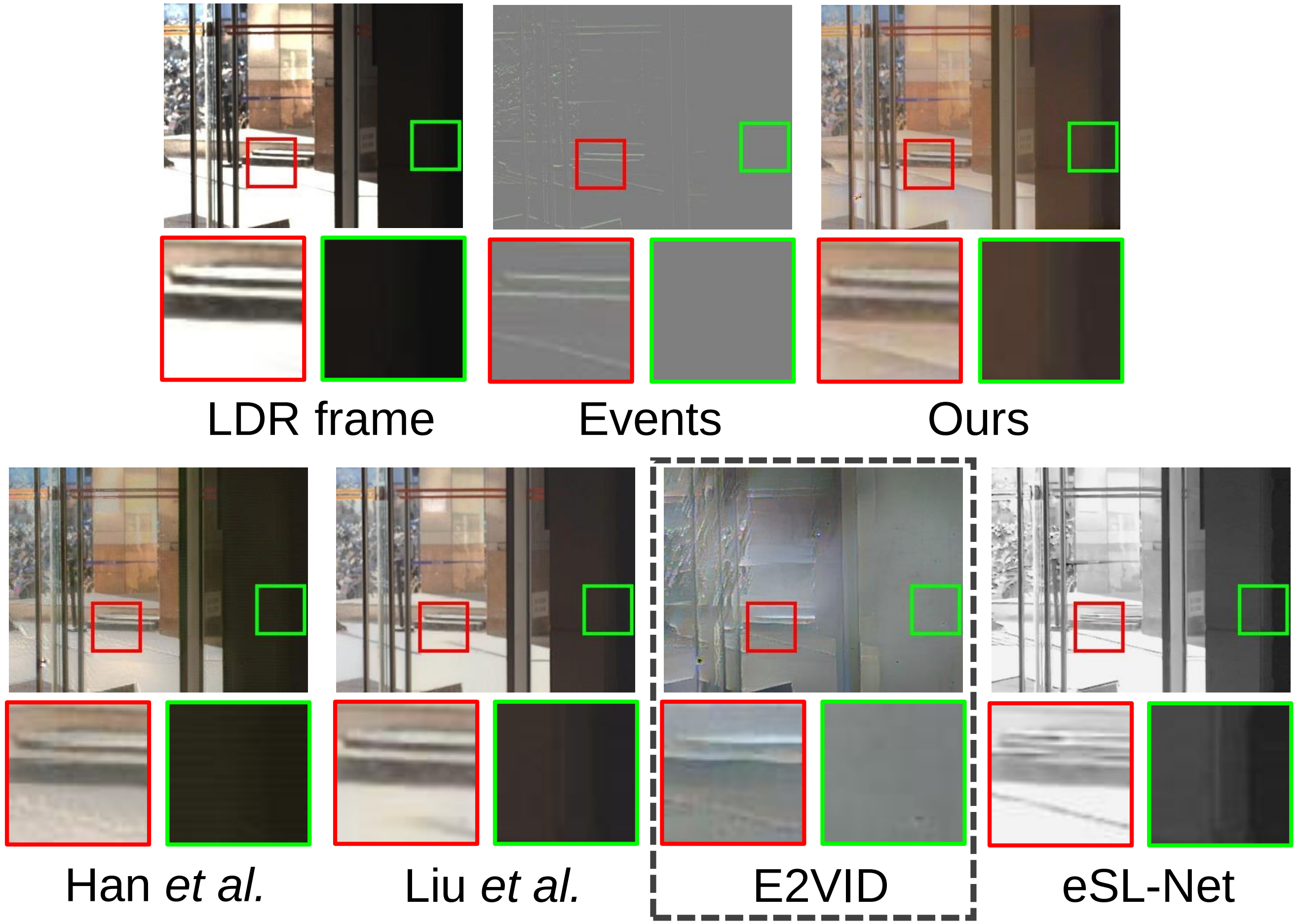


Single image based HDR: [*Liu et al.*, CVPR 2020] Event based HDR: [E2VID, TPAMI2019]
Hybrid HDR: [Han *et al.*, CVPR2020] [eSL-Net, CVPR 2020]

Evaluation on Real Data

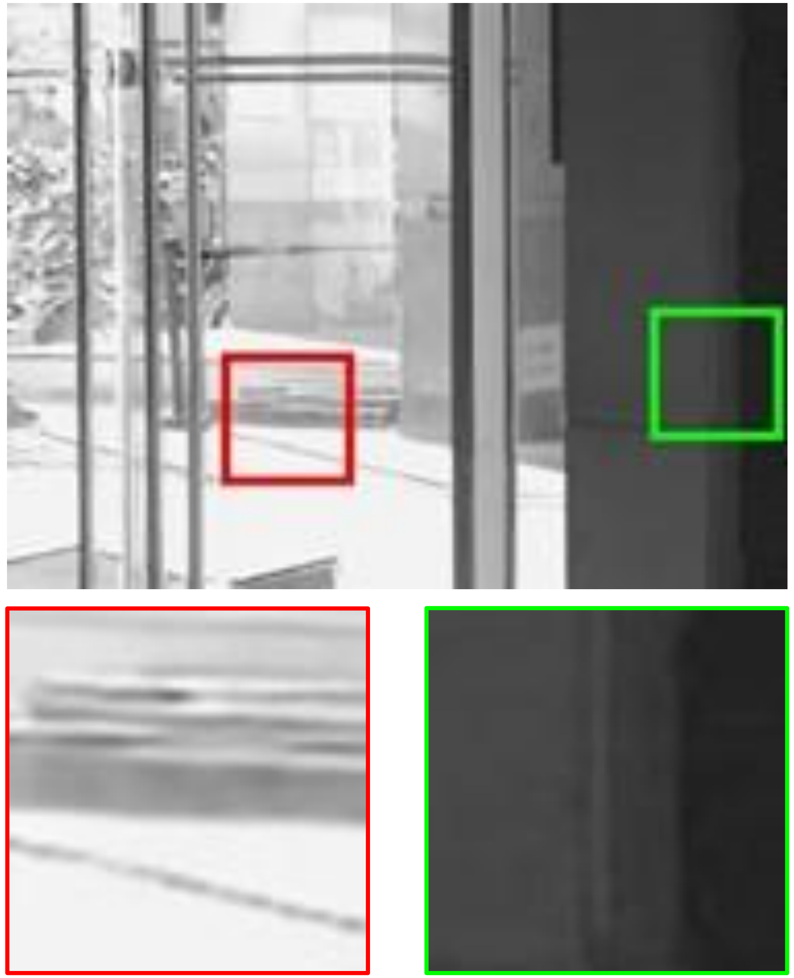


E2VID

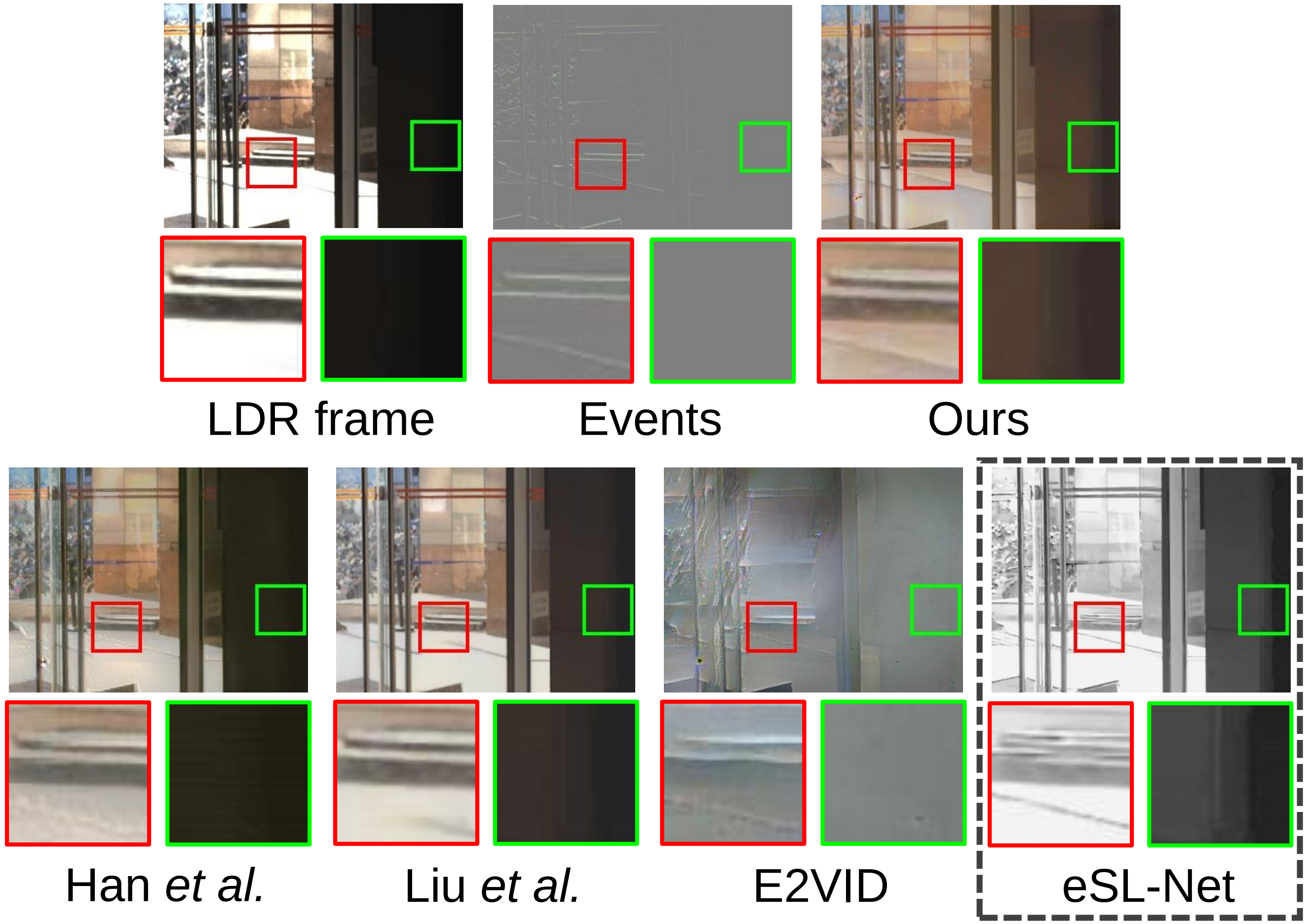


Single image based HDR: [*Liu et al.*, CVPR 2020] Event based HDR: [E2VID, TPAMI2019]
Hybrid HDR: [Han *et al.*, CVPR2020] [eSL-Net, CVPR 2020]

Evaluation on Real Data



eSL-Net



Single image based HDR: [*Liu et al.*, CVPR 2020] Event based HDR: [E2VID, TPAMI2019]
Hybrid HDR: [Han *et al.*, CVPR2020] [eSL-Net, CVPR 2020]

Video Evaluation on Synthetic Data


LDR video


Events


Ground truth


Ours


Debevec *et al.*


eSL-Net


E2VID

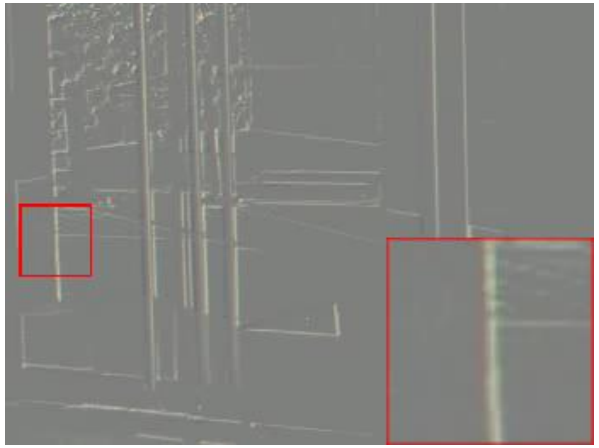

Liu *et al.*


Han *et al.*


Li *et al.*

Video Evaluation on Real Data

LDR video



Events

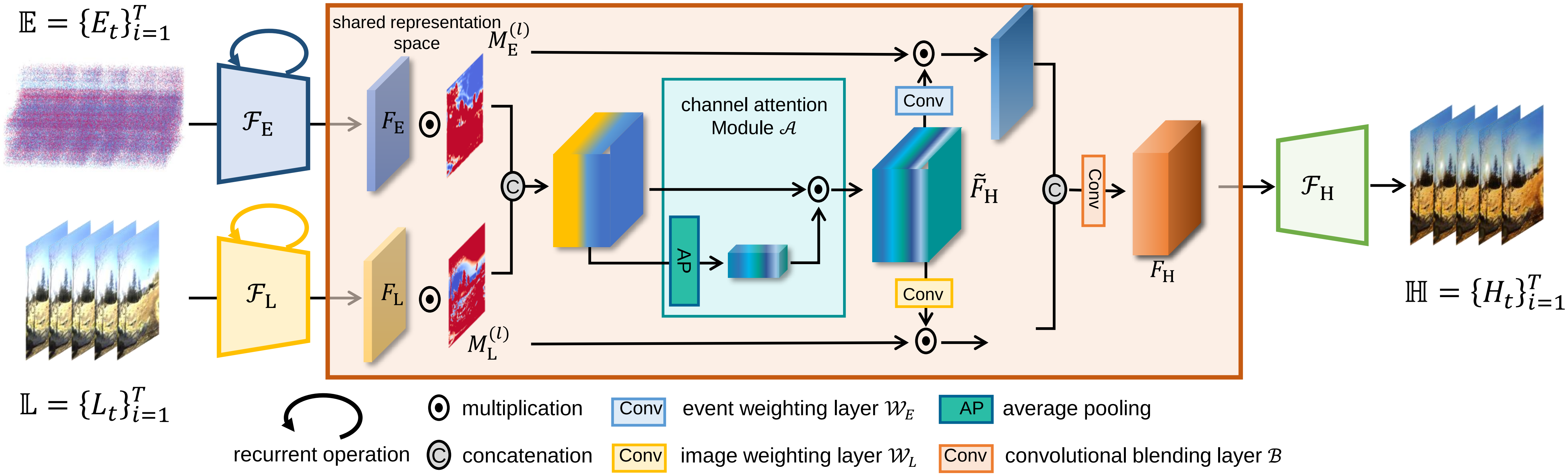
Ours

Liu *et al.*

E2VID

Han *et al.*

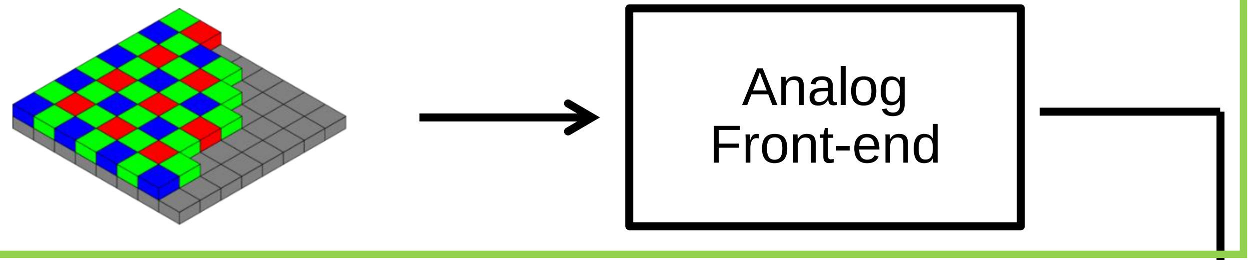
Conclusion



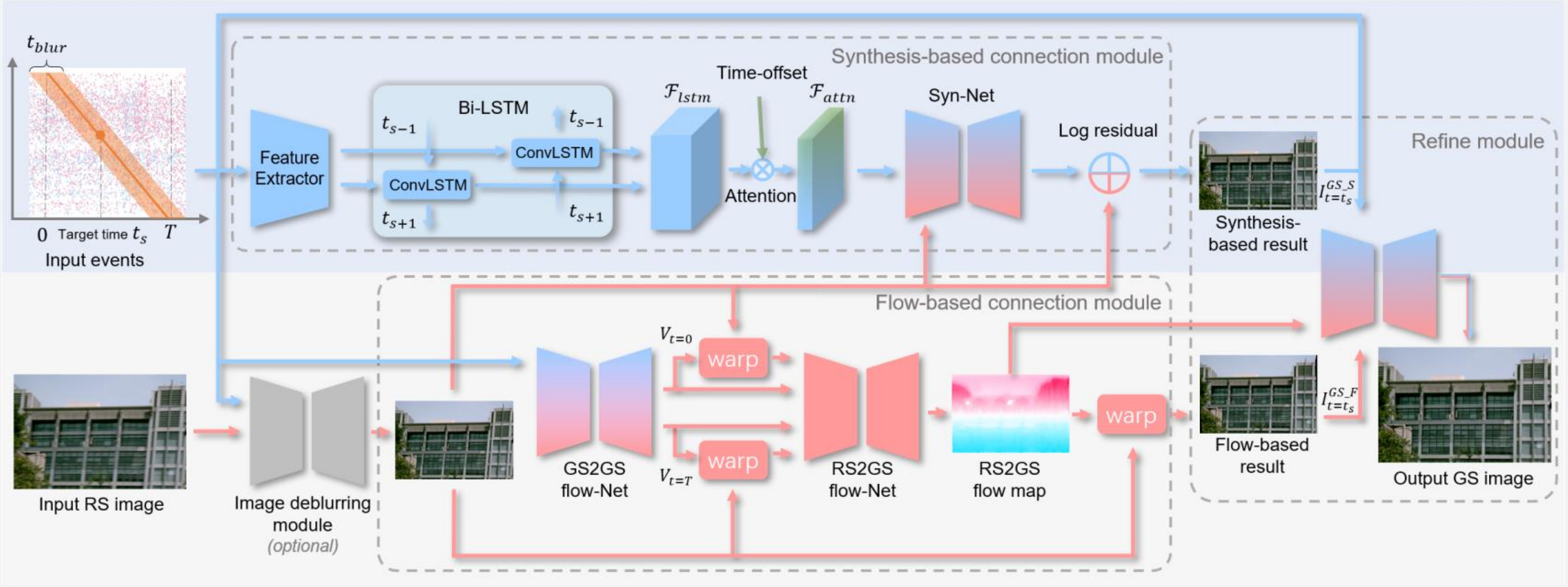
- We design a multimodal alignment strategy to bridge the gap between events and frames.
- We develop a confidence guided fusion module to complement events and LDR frames.
- We utilize the temporal correlation to alleviate the flickering effects for recovered HDR videos.

Camera Pipeline

Rolling Shutter [CVPR 2022]



GS Imaging



EvUnroll

CVPR 2022

EvUnroll: Neuromorphic Events based Rolling Shutter Image Correction

Xinyu Zhou[1#] Peiqi Duan[1#] Yi Ma[1] Boxin Shi[1,2,3]

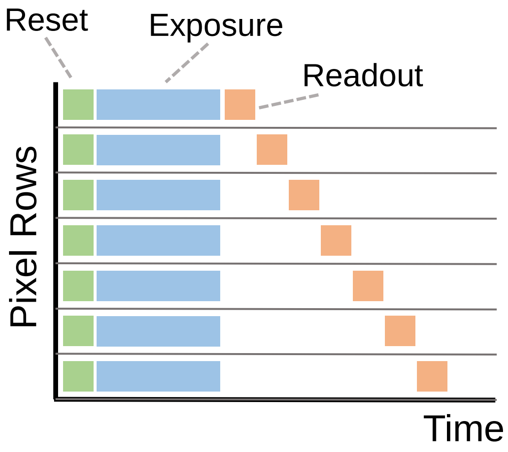
[1]National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

[2]Institute for Artificial Intelligence, Peking University

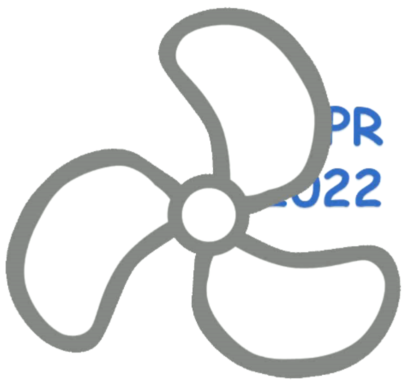
[3]Beijing Academy of Artificial Intelligence

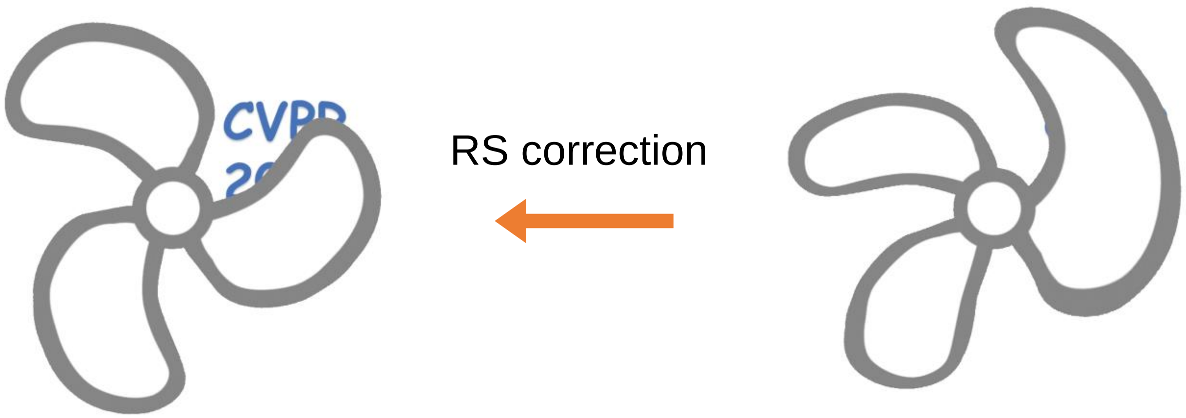
[#] Contributed equally to this work as first authors

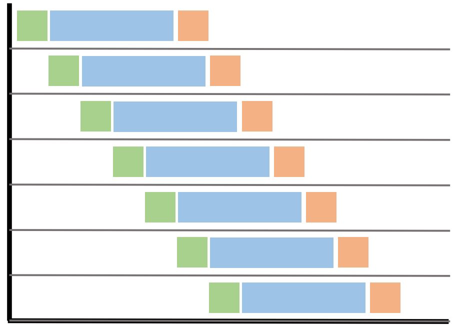
Rolling Shutter Effect



Global Shutter



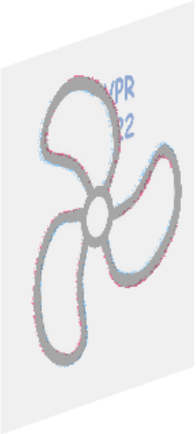




Rolling Shutter

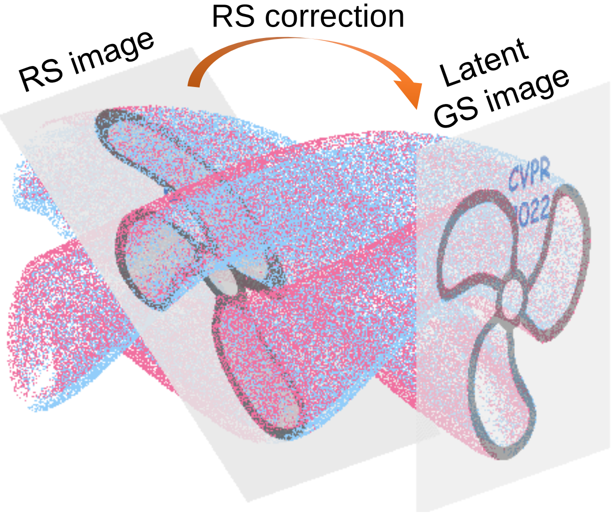
Motivation

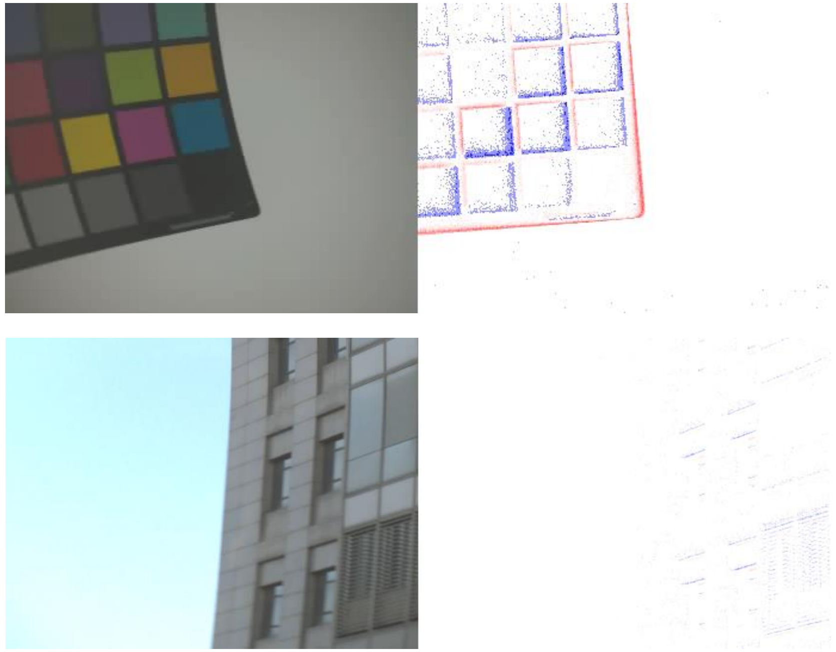
Event + RS Frame



Motivation

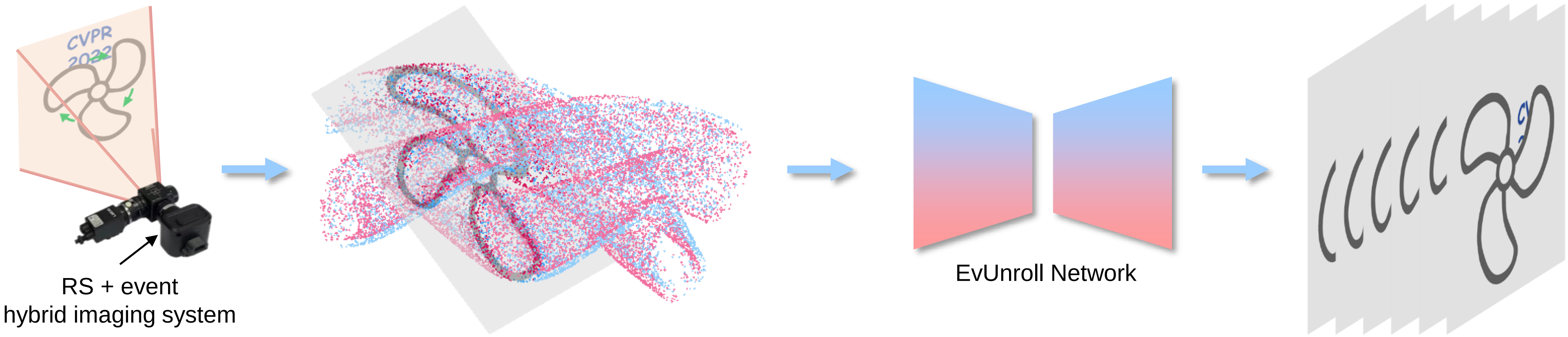
Event + RS Frame





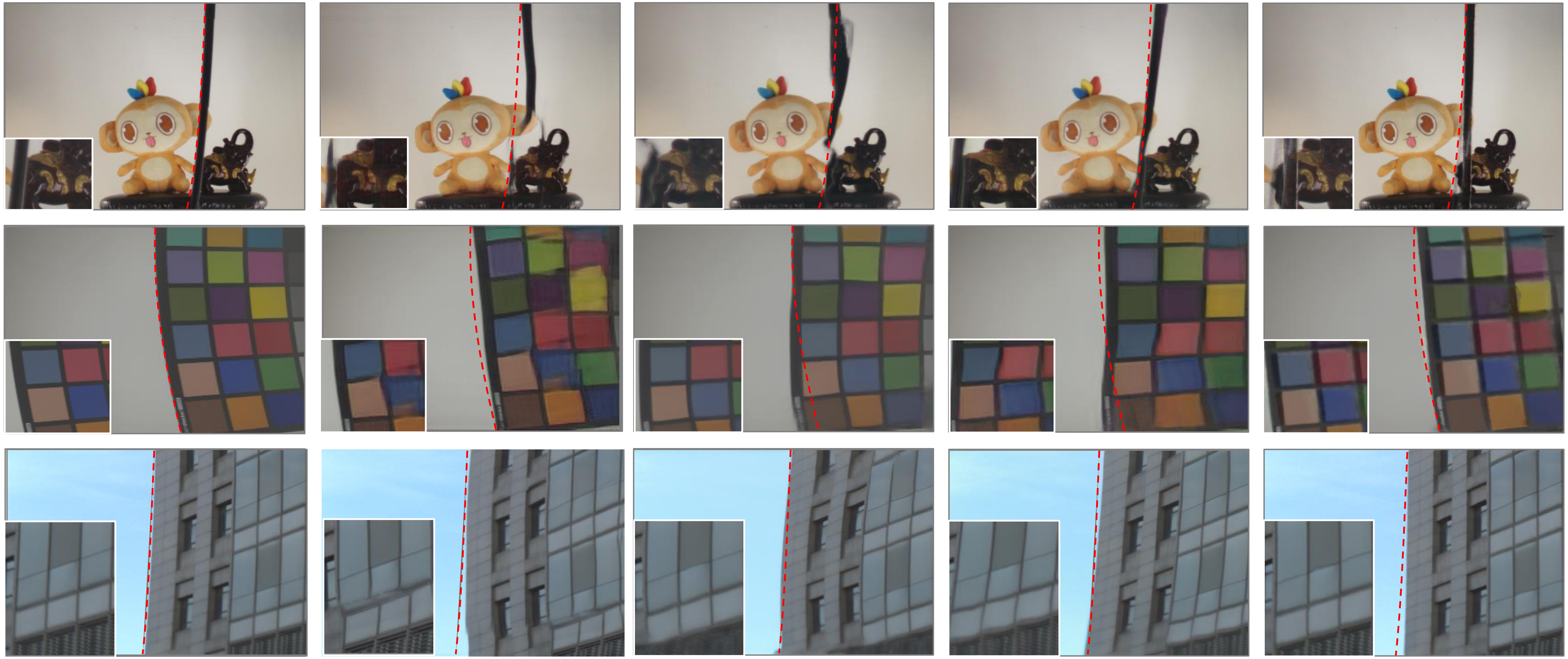
- **Motion Estimation**:
  High-speed events can address the challenge of fast and complicated motions.
- **Occlusion Region Restoration**:
  Brightness changes encoded in events can be utilized to restore occluded regions.

Contribution



- EvUnroll is the first trial to improve RS correction with motion estimation and occlusion region restoration by involving event signals.
- We build an RS-event hybrid camera to collect a real testing dataset.
- EvUnroll restores high-frame-rate GS videos, and outperforms state-of-the-art RS correction methods on commonly used datasets.

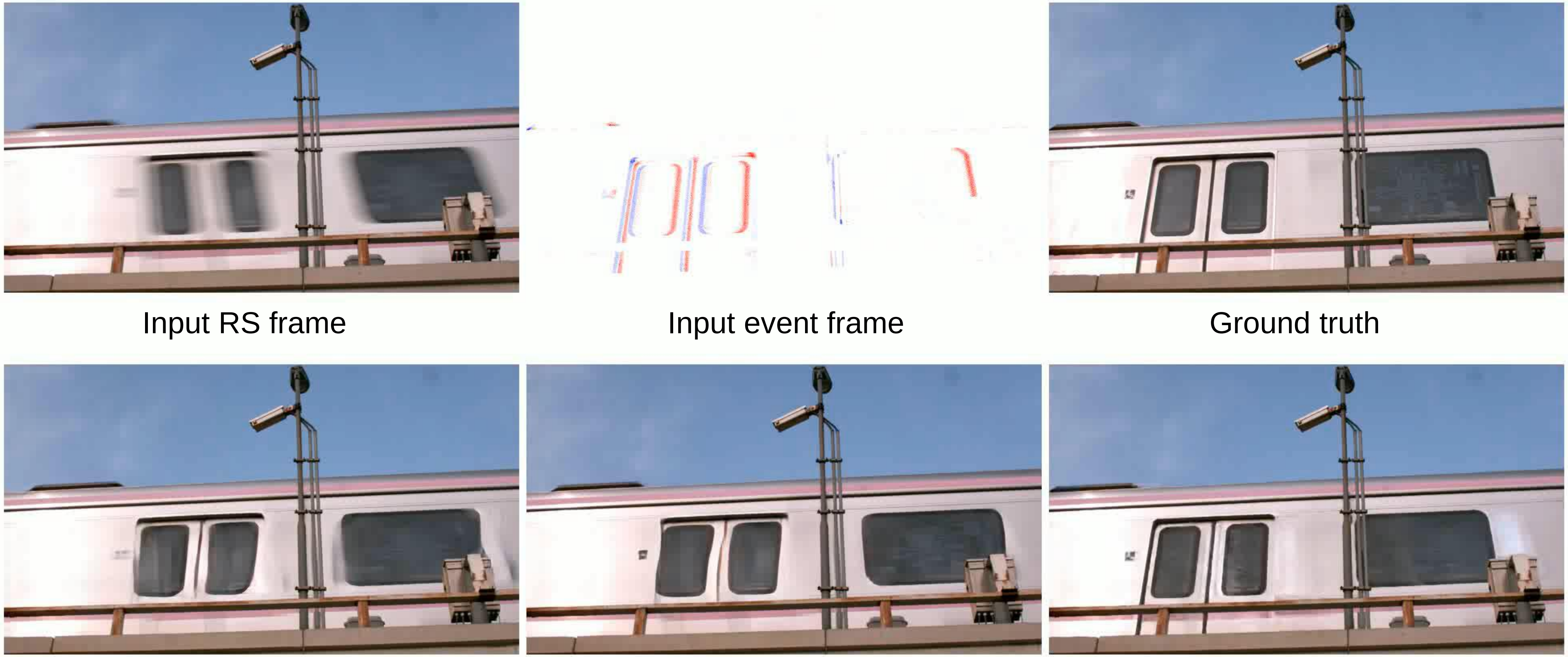
Results on Real Data



RS frame SUNet [8] DSUN [26] JCD [54] Ours

Image-based method: [DSUN, CVPR 2020], [JCD, CVPR 2021], [SUNet, ICCV 2021]

HFR GS Videos Results

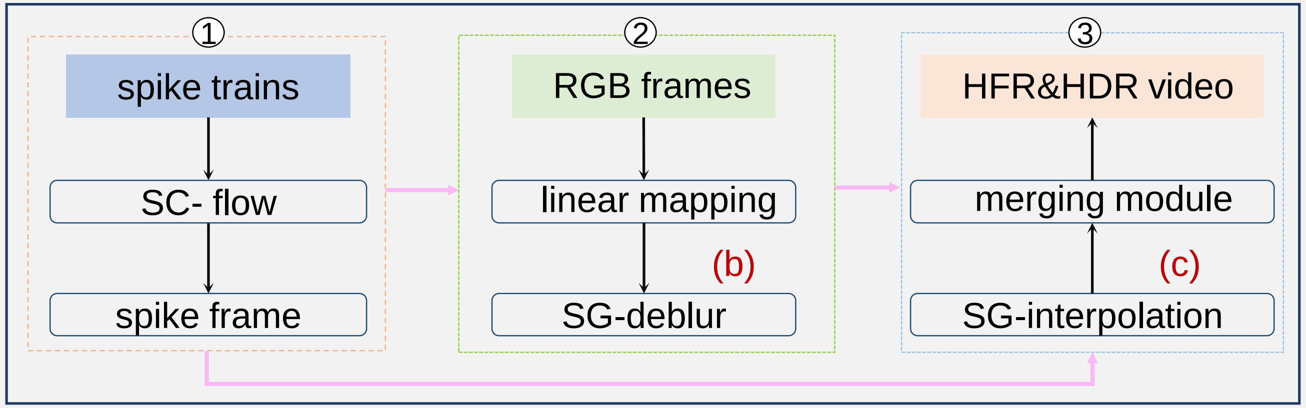


Input RS frame | Input event frame | Ground truth

DSUN [26] | JCD [54] | Ours

Image-based method: [DSUN, CVPR 2020], [JCD, CVPR 2021]

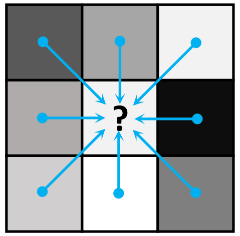
Camera Pipeline



Spike-RGB HDR HFR

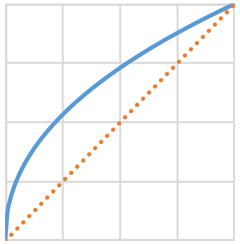
HFR HDR Video

LFR、LDR
[CVPR 2023]



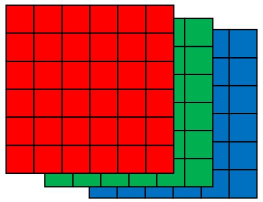
Denoise

color transforms



Tone Reproduction

compression



Final RGB Image
(nonlinear, 8-bit)

CVPR 2023

1000 FPS HDR Video with a Spike-RGB Hybrid Camera

Yakun Chang[1,2] Chu Zhou[3] Yuchen Hong[1,2] Liwen Hu[2] Chao Xu[3] Tiejun Huang[1,2] Boxin Shi[1,2]

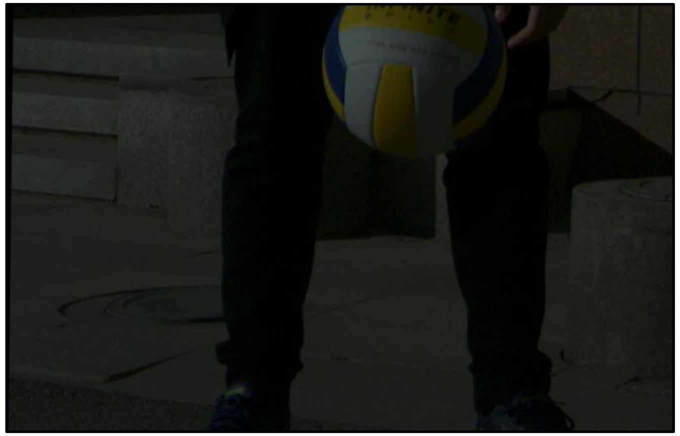
[1] National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University

[2] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

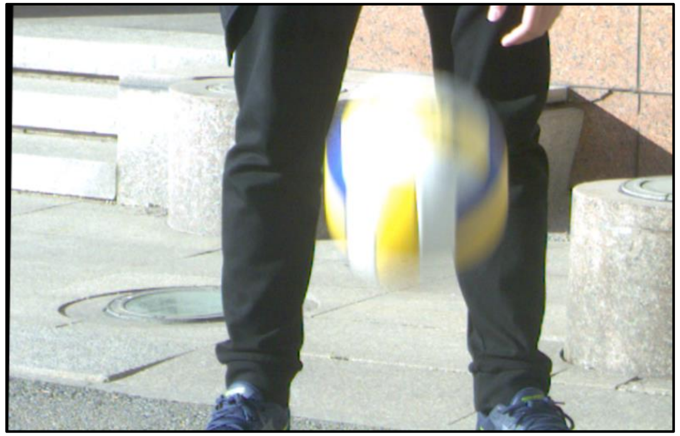
[3] National Key Laboratory of General AI, School of Intelligence Science and Technology, Peking University

High-Speed HDR Video is Challenging

Short exposure

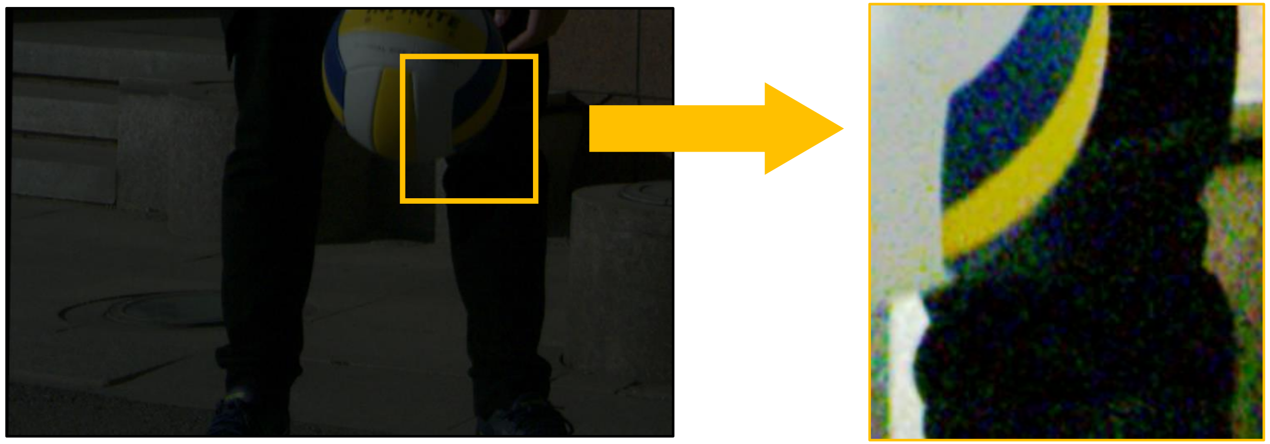


Long exposure

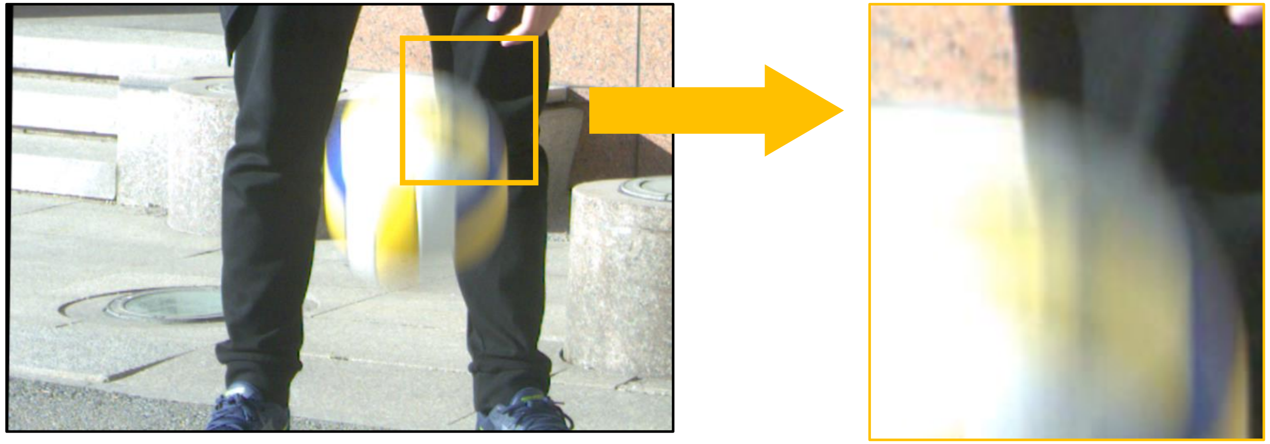


High-Speed HDR Video is Challenging

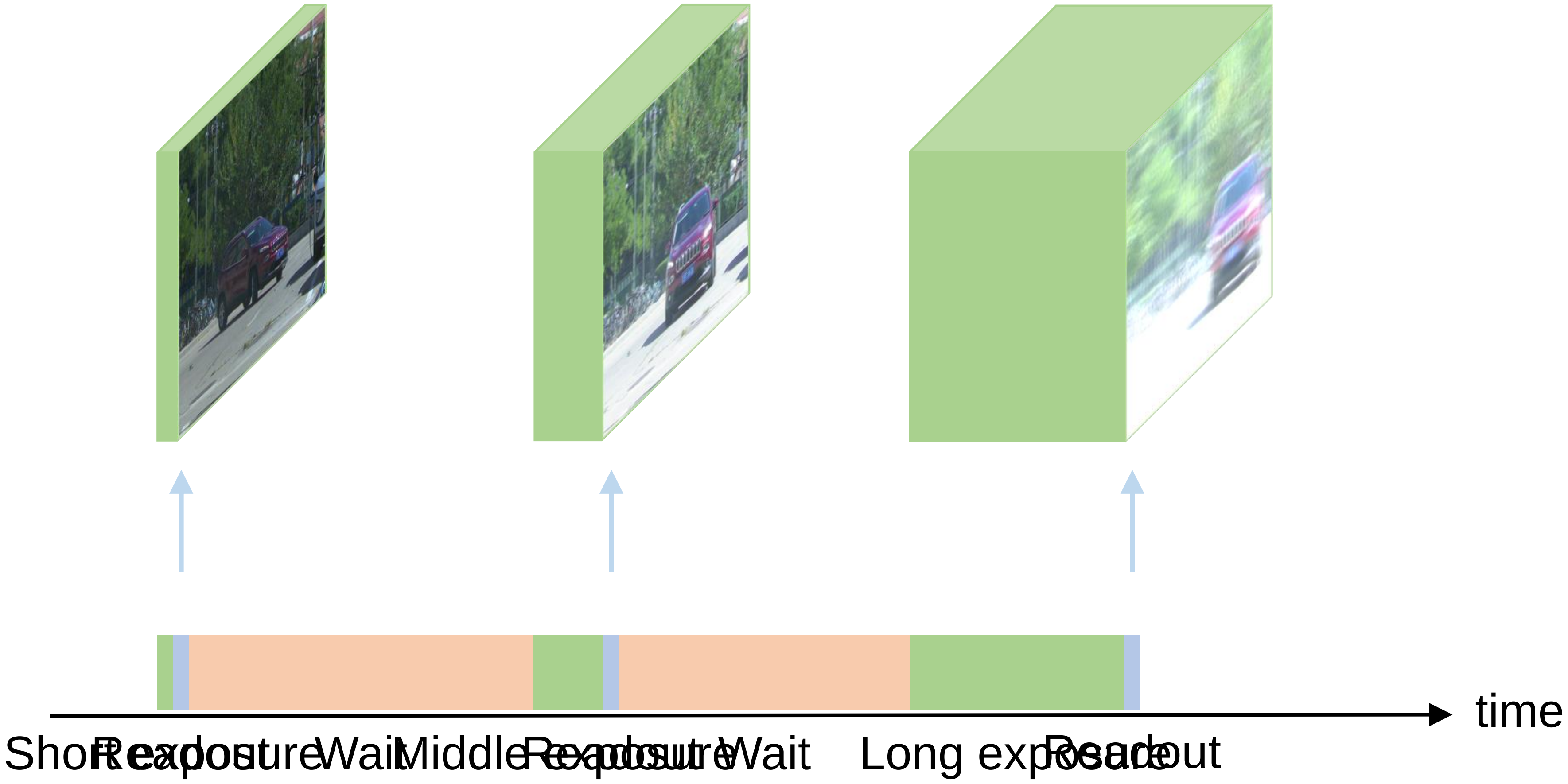
Short exposure



Long exposure



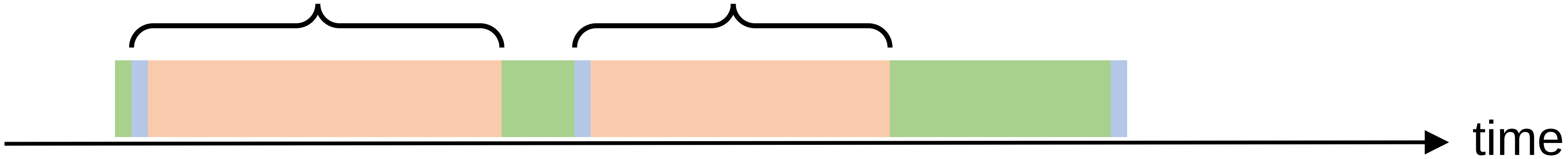
Exposure Bracketing



$$T = T_{exposure} + T_{readout} + T_{wait}$$

Exposure Bracketing

The colors are not recorded during the readout time and waiting time



Limited the frame rate to dozens of frames per second

Conventional Camera vs. Spiking Camera

Conventional camera: cannot capture fast motions

Two adjacent RGB frames

Conventional Camera vs. Spiking Camera

Conventional camera: cannot capture fast motions

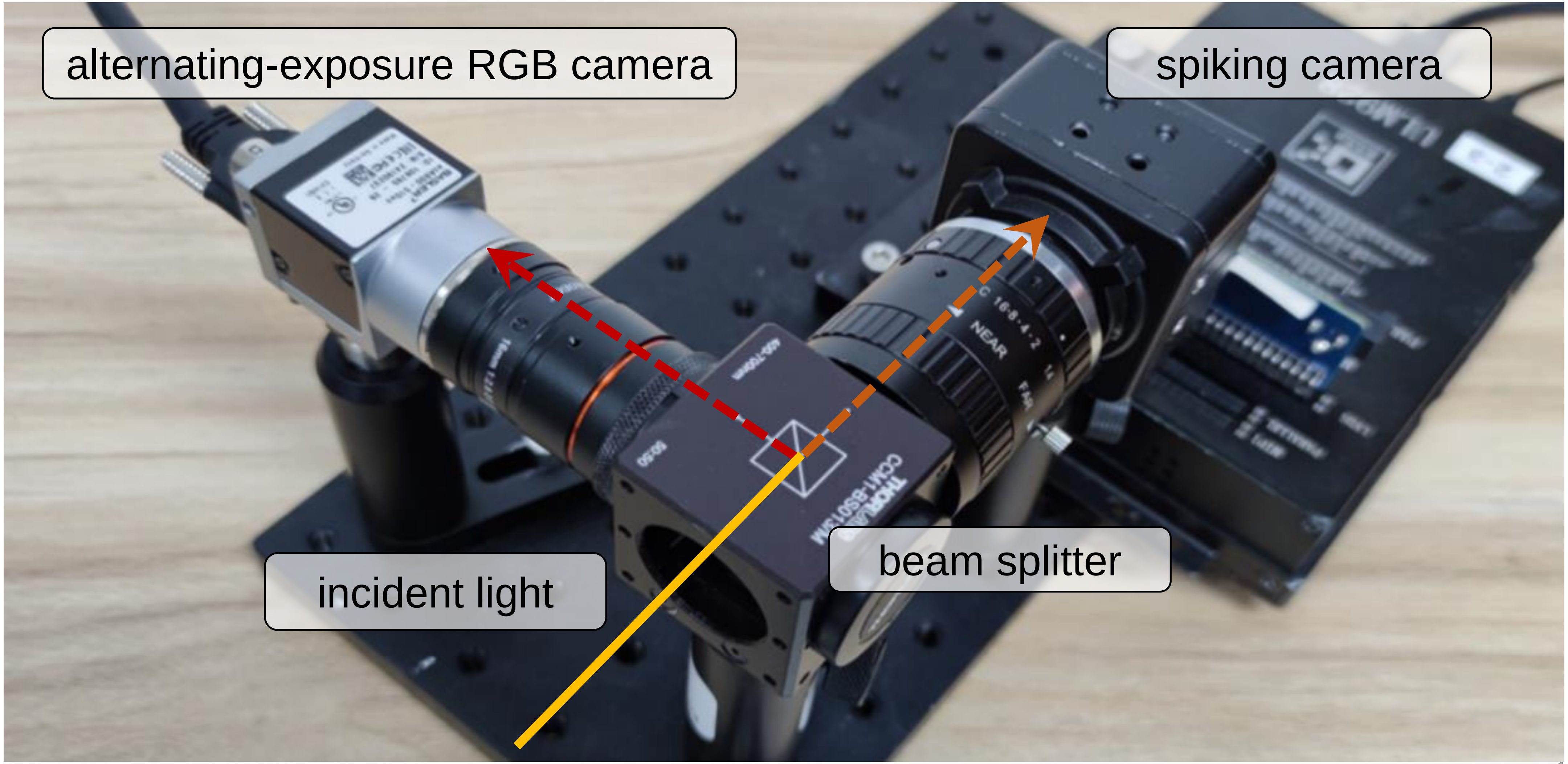
Two adjacent RGB frames

Spiking camera: capturing continuous fast motions in high dynamic range

Spike signal

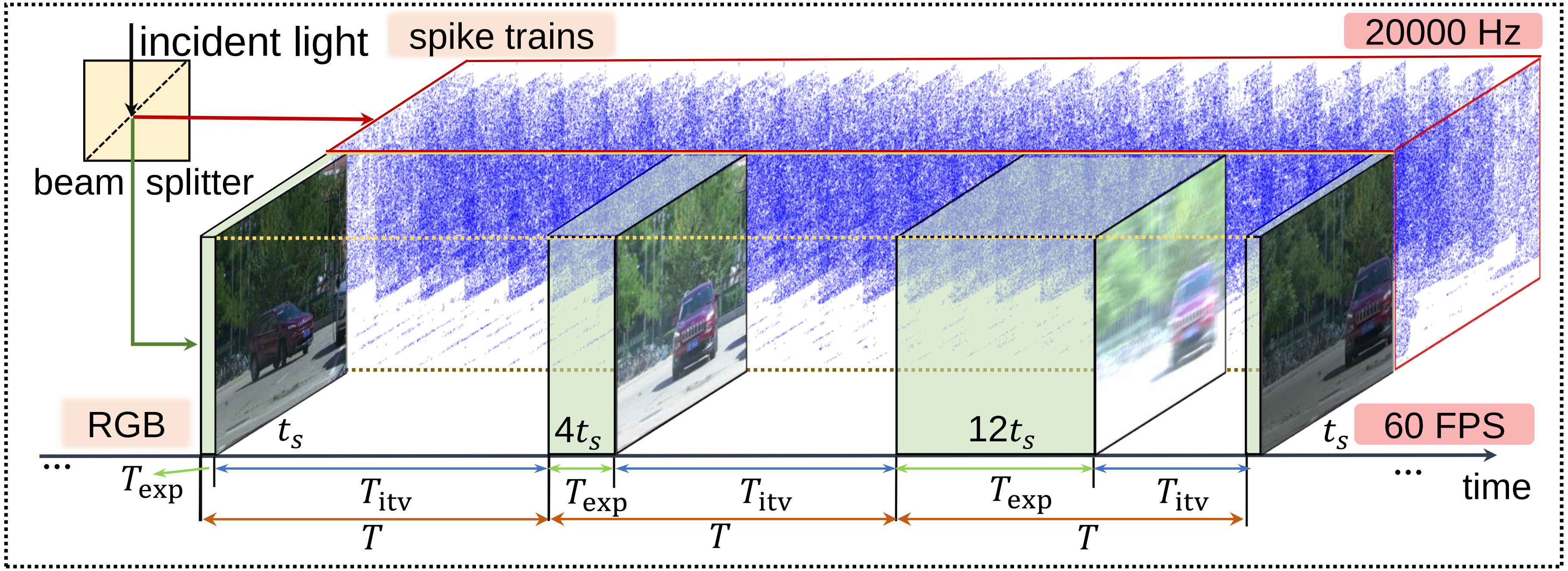
Potential information for recovering high frame rate (HFR) and high dynamic range (HDR) video

Hybrid Camera System

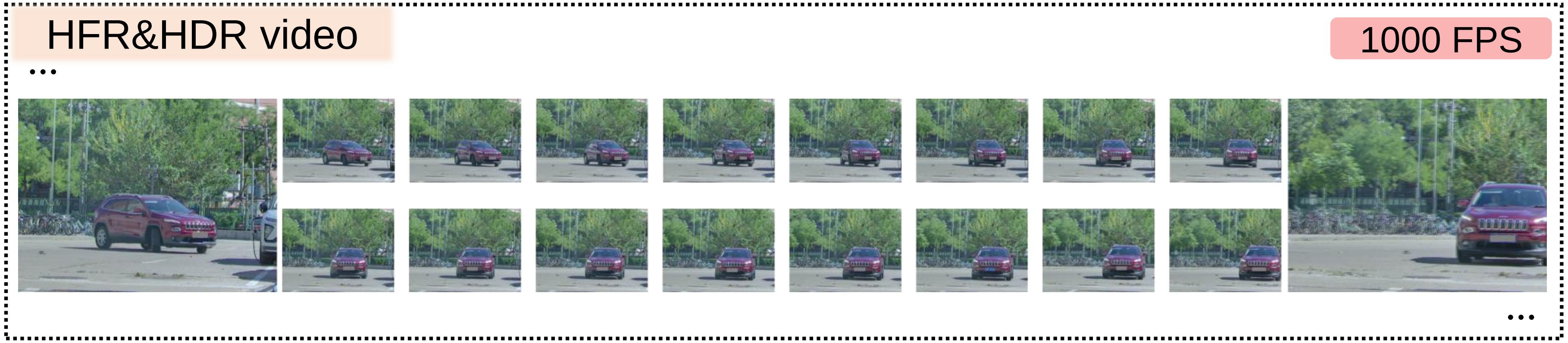


Target

Input



Output

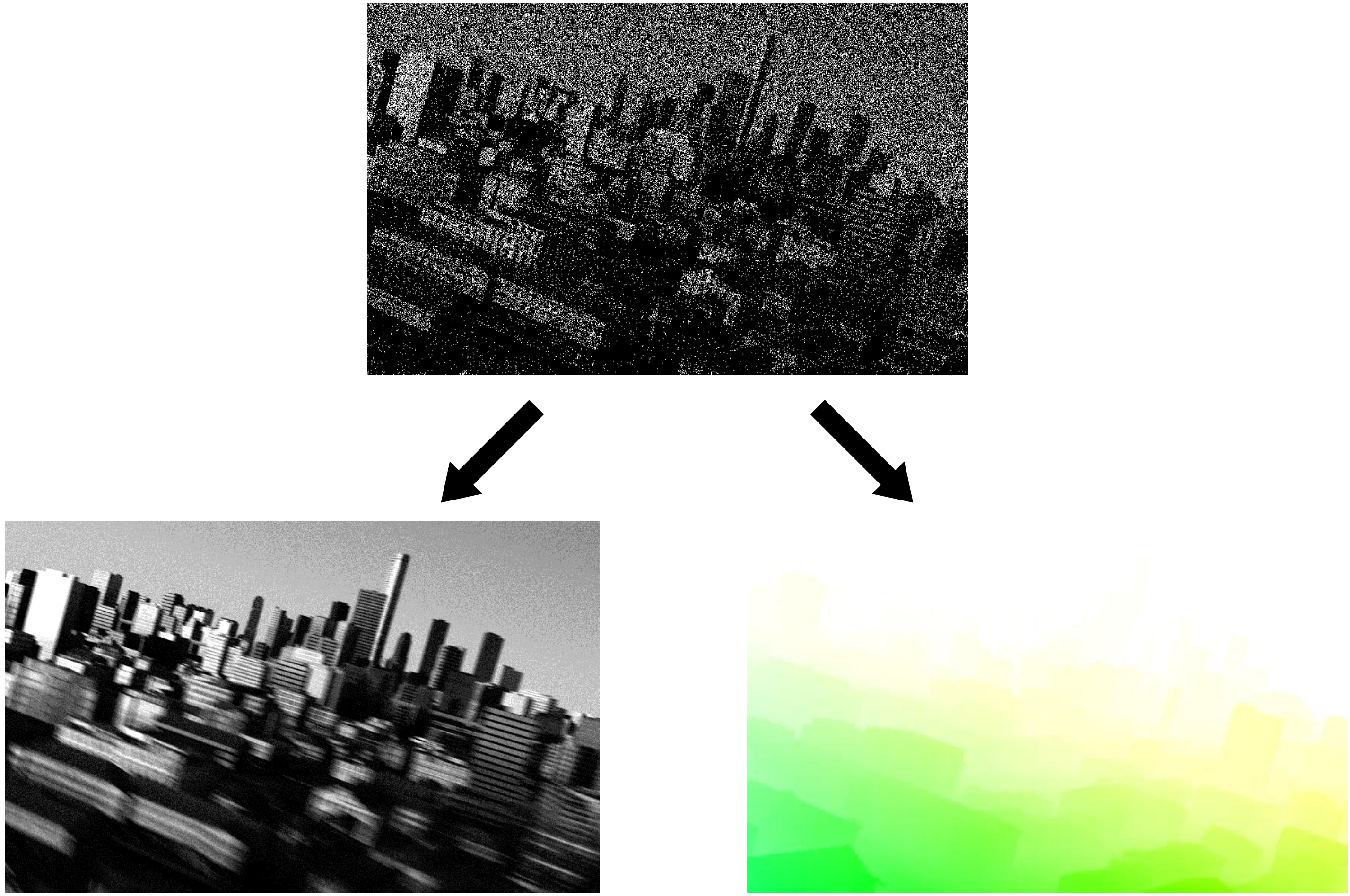


Methodology

Spike signal

Methodology

Spike signal



Spike frame

Optical flow

Methodology

Short exposure

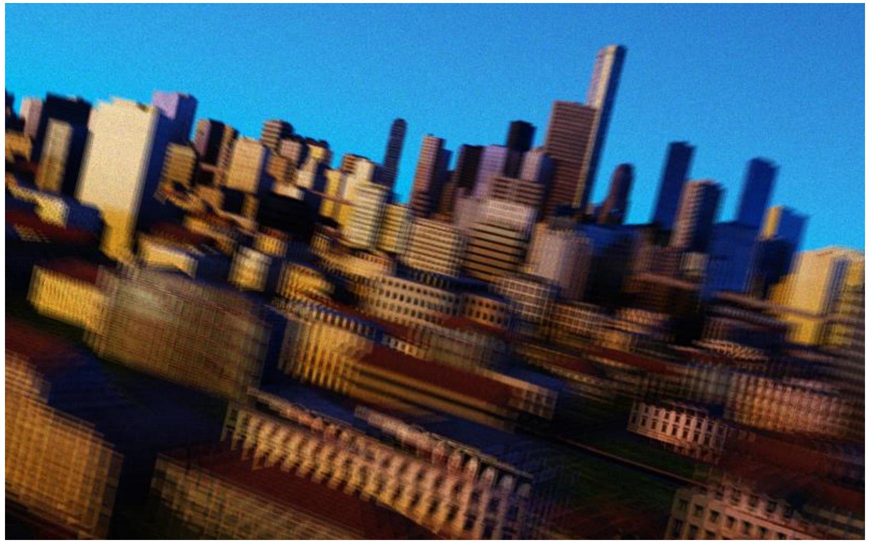
Spike frame

Optical flow

Reduce noise

Methodology

Middle exposure



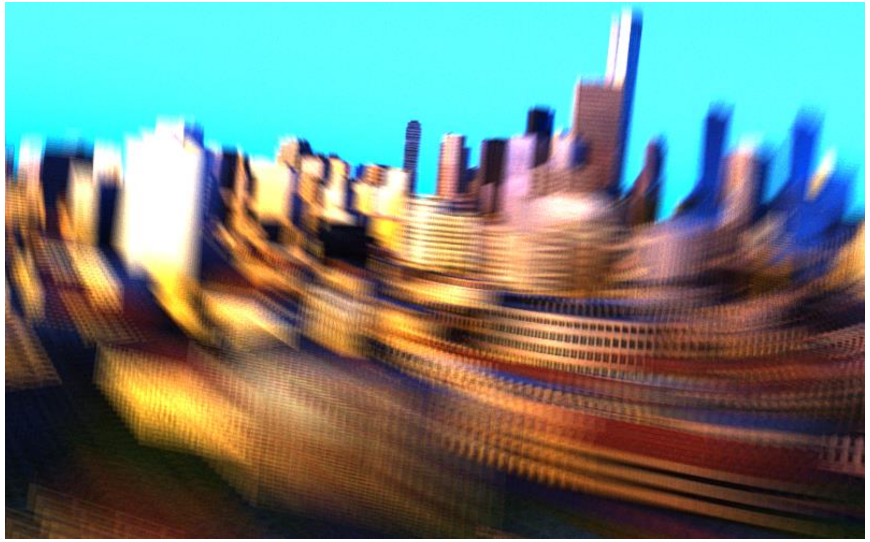
Spike frame

Optical flow

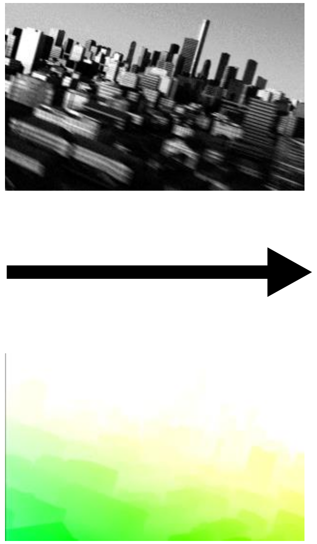
Motion deblurring

Methodology

Long exposure



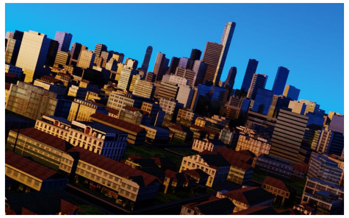
Spike frame



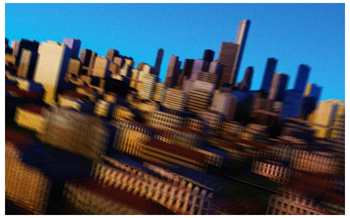
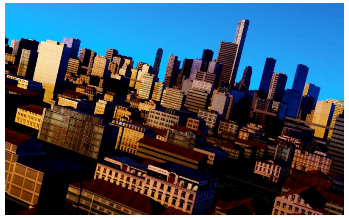
Optical flow

Motion deblurring
Compensate clipped regions

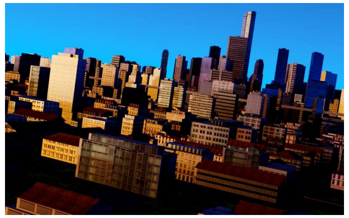
Methodology

Short exposure

Middle exposure

Long exposure

Alternating-exposure frames

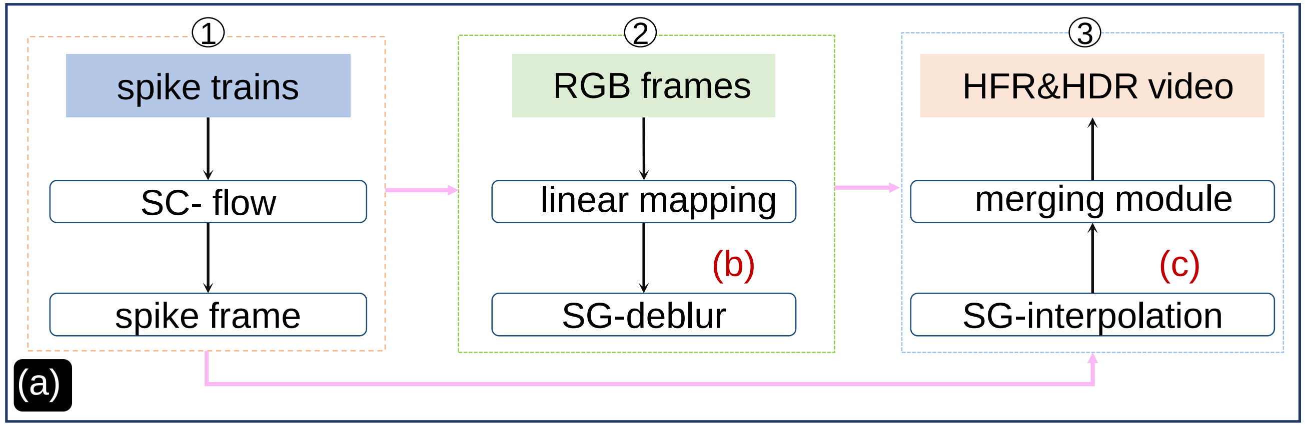
Spike frame

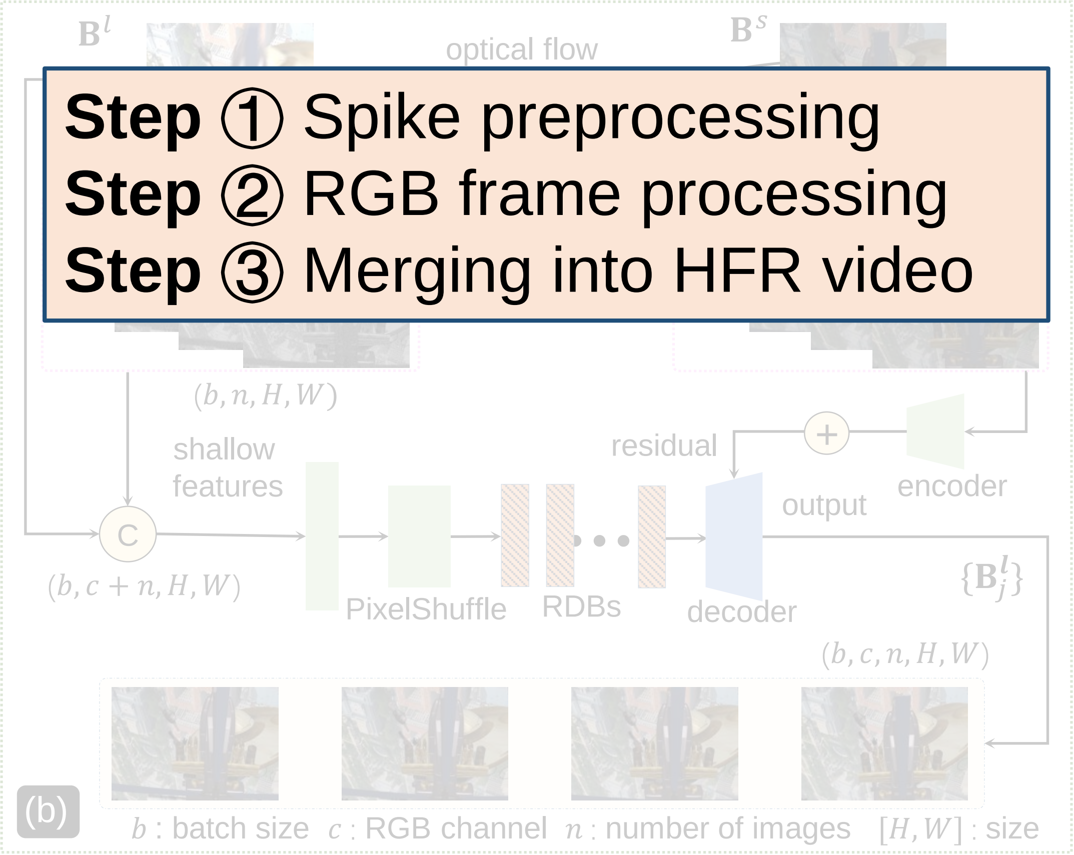
Optical flow

Recover 1000 FPS color video

1000 FPS HDR Video Pipeline

The pipeline of the proposed solution.

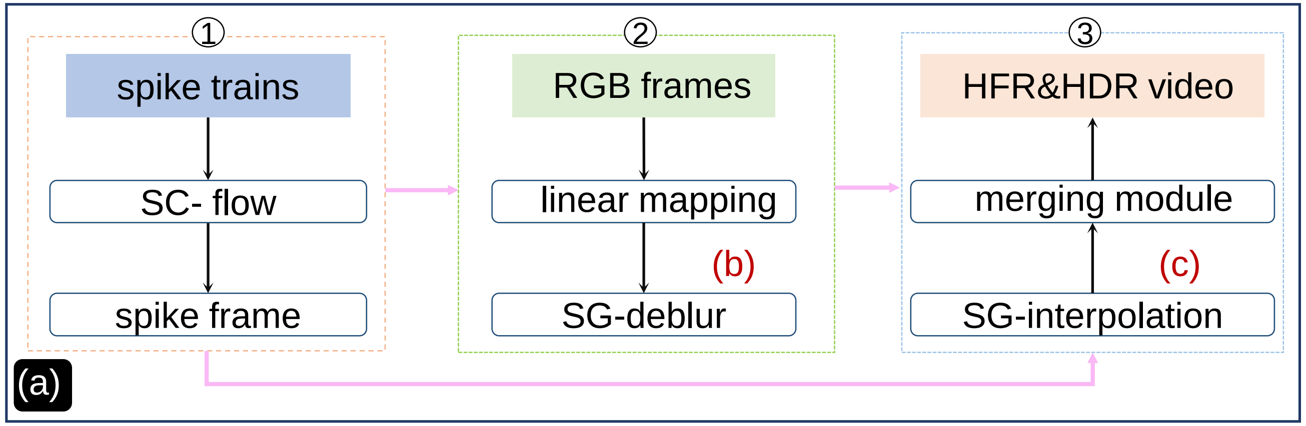


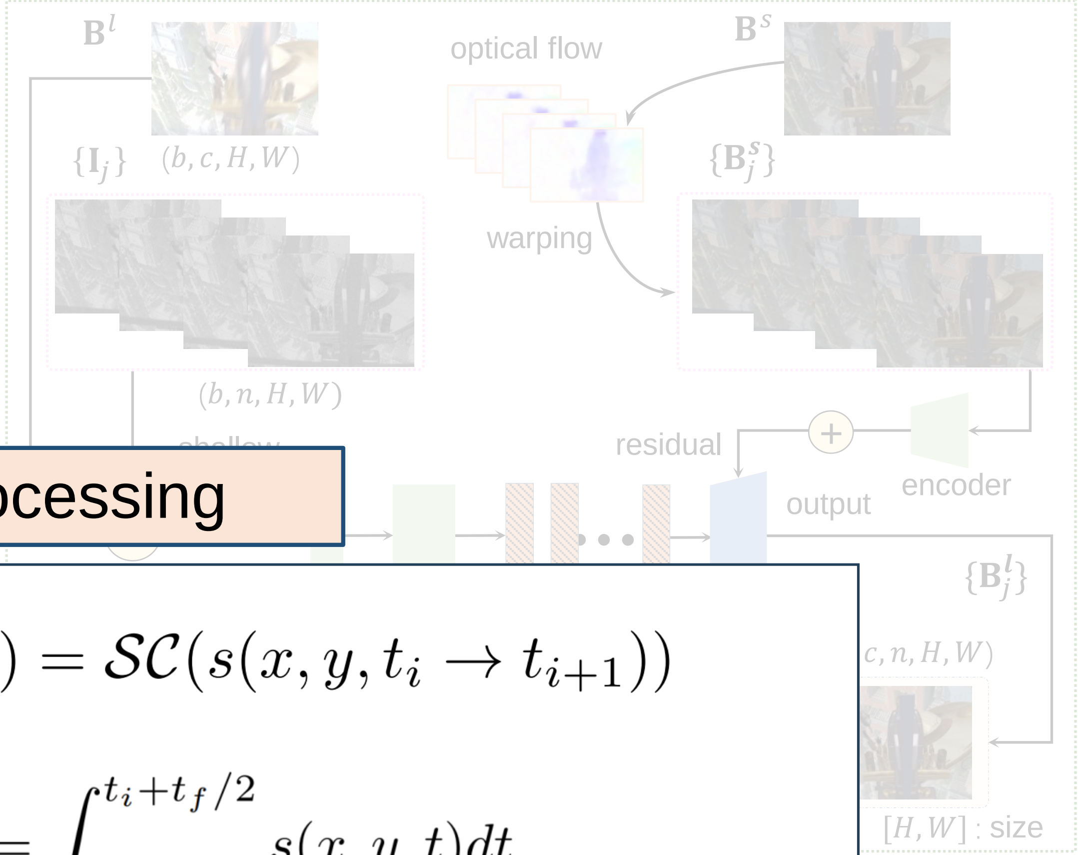


Step ① Spike preprocessing
Step ② RGB frame processing
Step ③ Merging into HFR video

1000 FPS HDR Video Pipeline

The pipeline of the proposed solution.



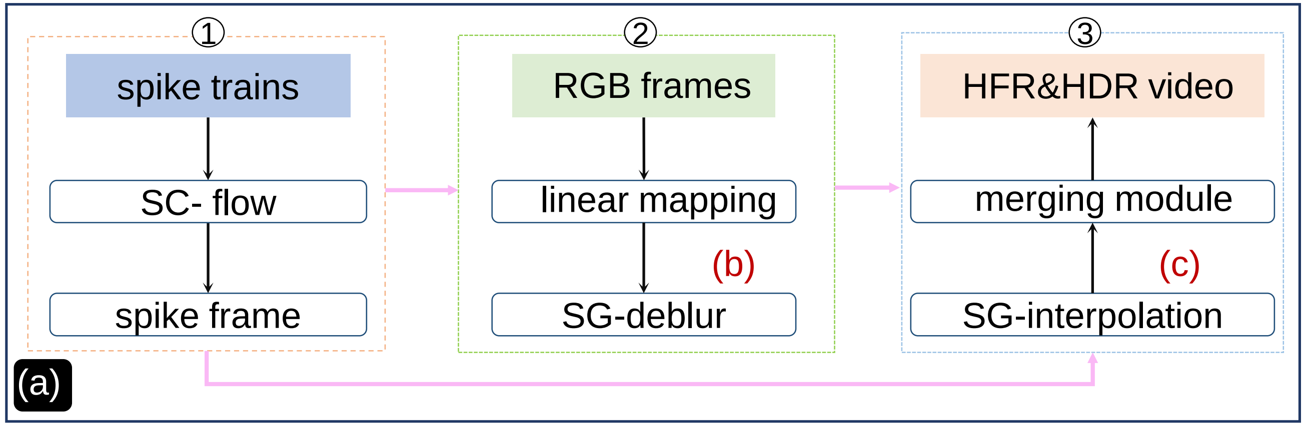


Step ① Spike preprocessing

- Estimate optical flow $$\mathbf{F}_i(x, y) = \mathcal{SC}(s(x, y, t_i \to t_{i+1}))$$
- Reconstruct spike frames $$\mathbf{I}_i(x, y) = \int_{t_i \quad t_f/2}^{t_i + t_f/2} s(x, y, t)dt$$

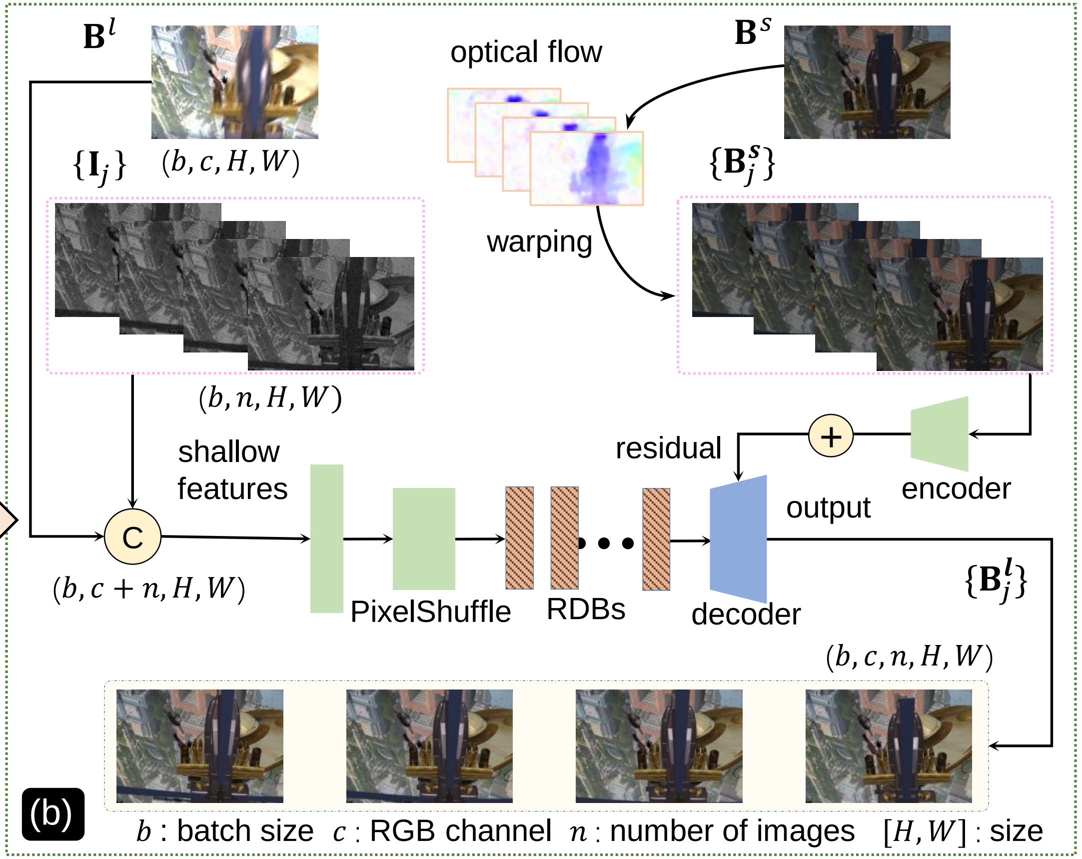
Deblurring for Long-Exposure Images

The pipeline of the proposed solution.



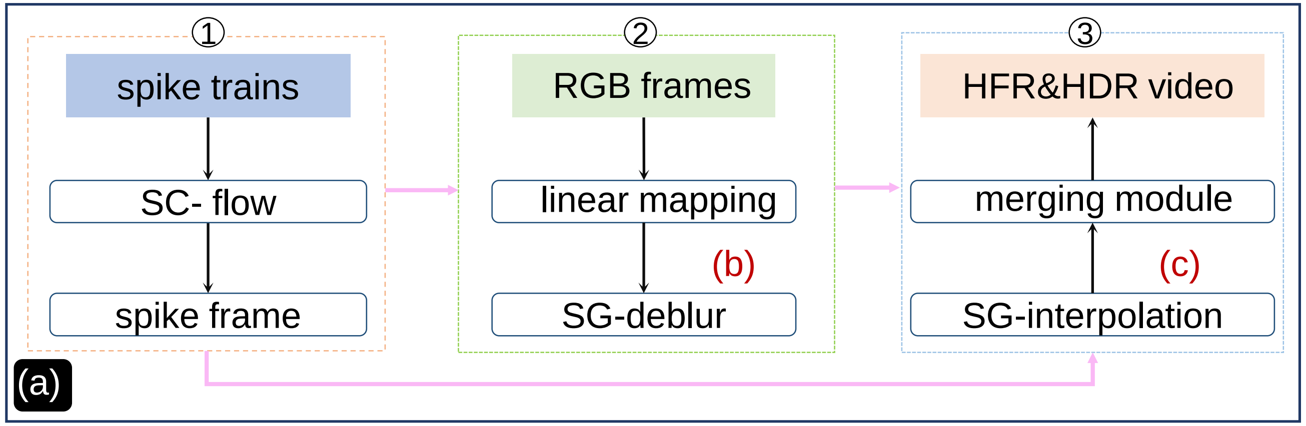
Step ② RGB frame processing

- How can the deblurring problem be resolved using the rich temporal information provided by spike trains?
- How many sharp images can be obtained from a blurry image?

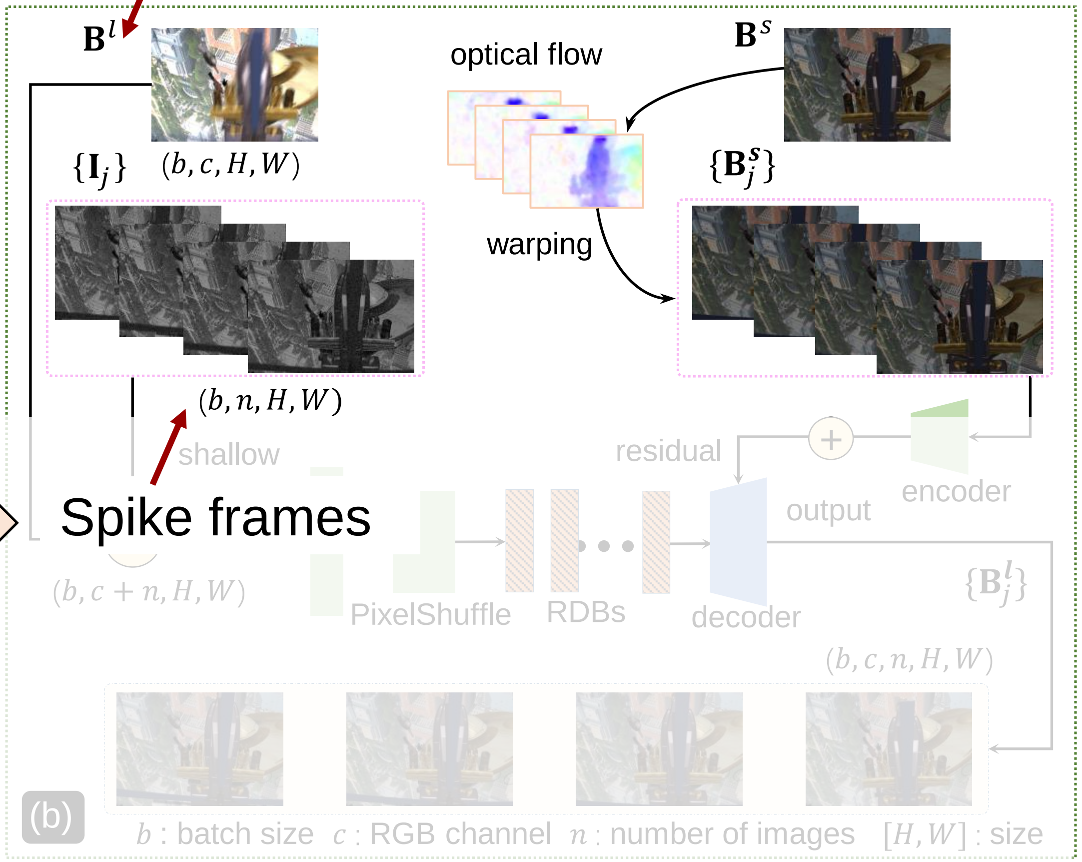


Deblurring for Long-Exposure Images

The pipeline of the proposed solution.



Middle-exposure image



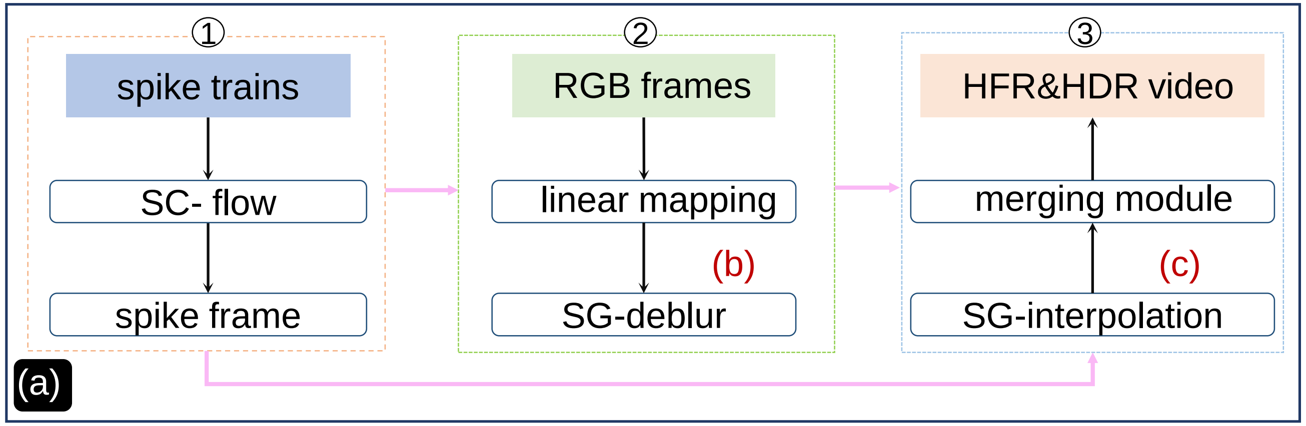
Spike frames

Step ② RGB frame processing

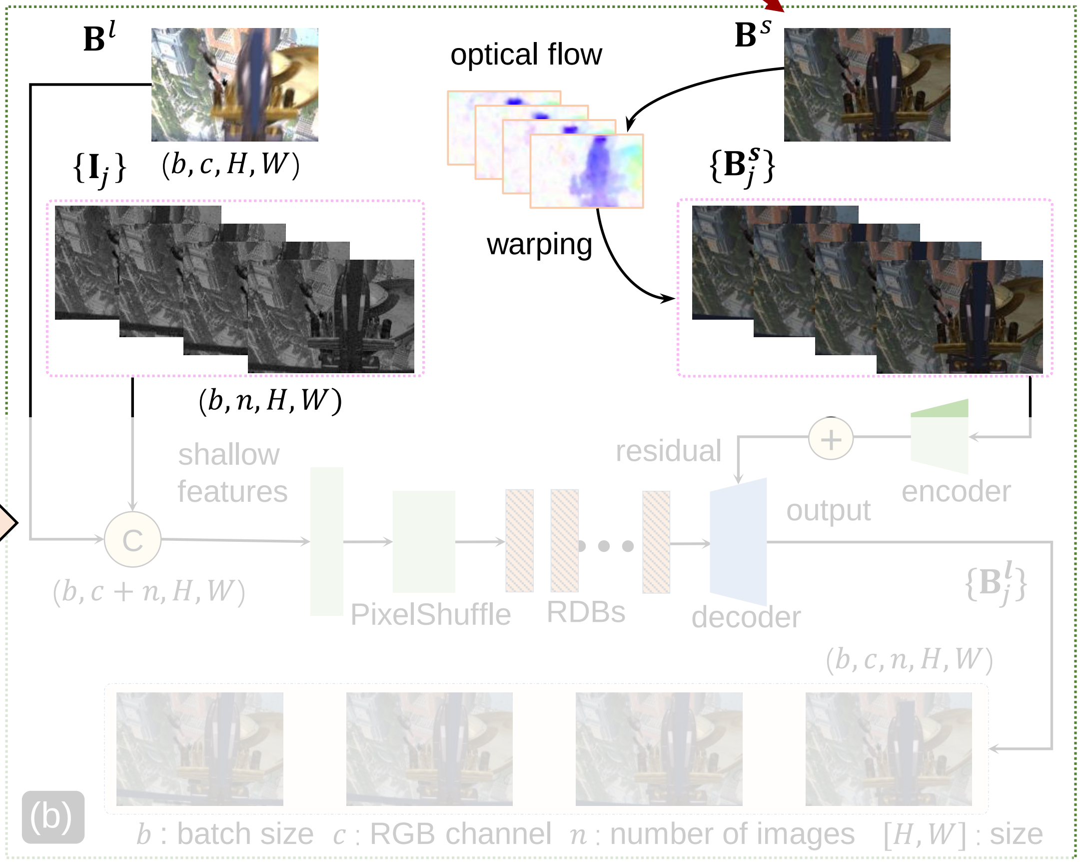
- How can the deblurring problem be resolved using the rich temporal information provided by spike trains?
  - Removing blurs in the middle-exposure image using blur-free spike fames.

Deblurring for Long-Exposure Images

The pipeline of the proposed solution.



Adjacent short-exposure image

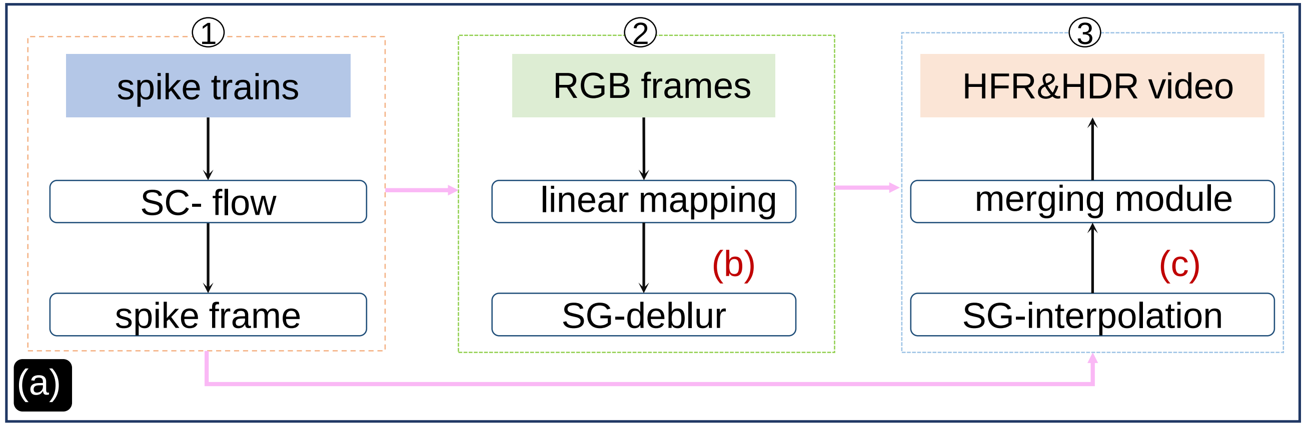


Step ② RGB frame processing

- How can the deblurring problem be resolved using the rich temporal information provided by spike trains?
  - Warping the adjacent short-exposure image to align with the spike frames guided by optical flow.

Deblurring for Long-Exposure Images

The pipeline of the proposed solution.



Step ② RGB frame processing

- How many sharp images can be obtained from a blurry image?
  - Middle-exposure ($4t_s$) and long-exposure ($12t_s$) images are recovered at each t_s time stamp.

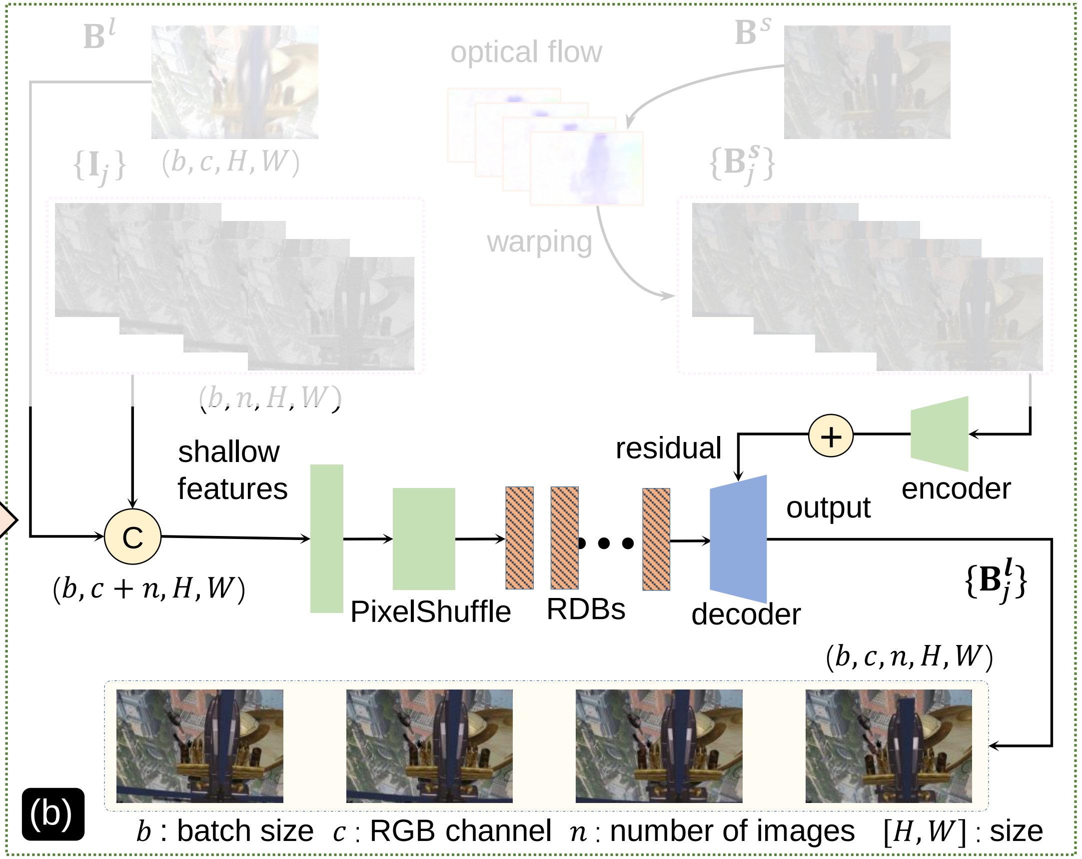
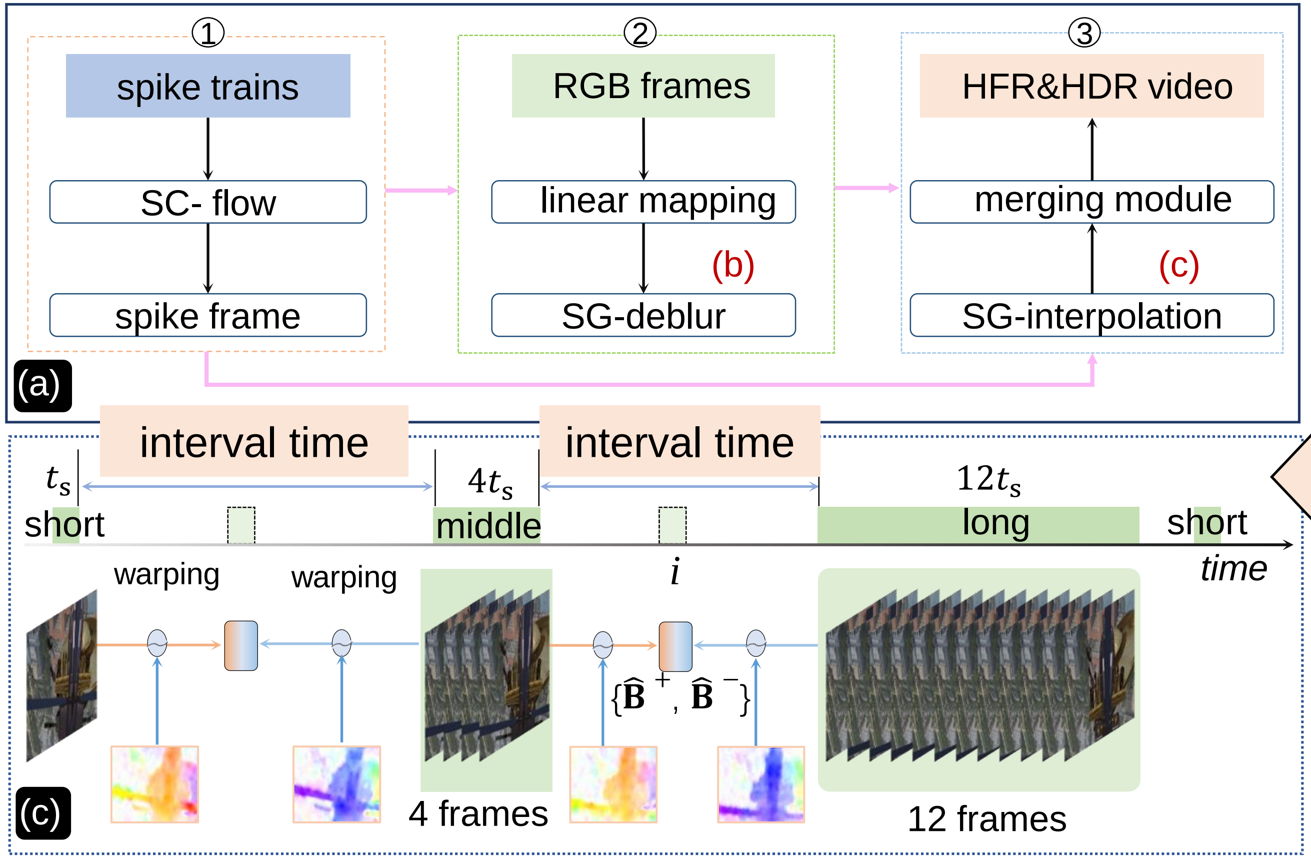
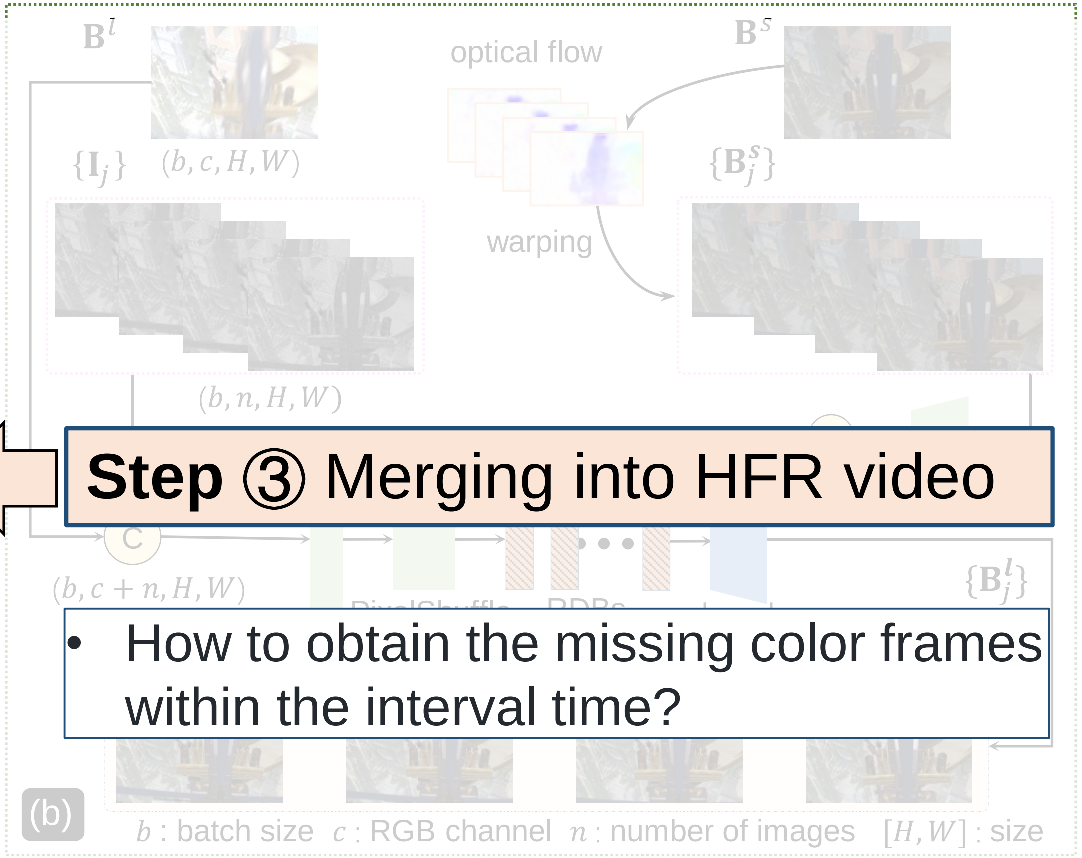


Image Interpolation with Optical Flows

The pipeline of the proposed solution.



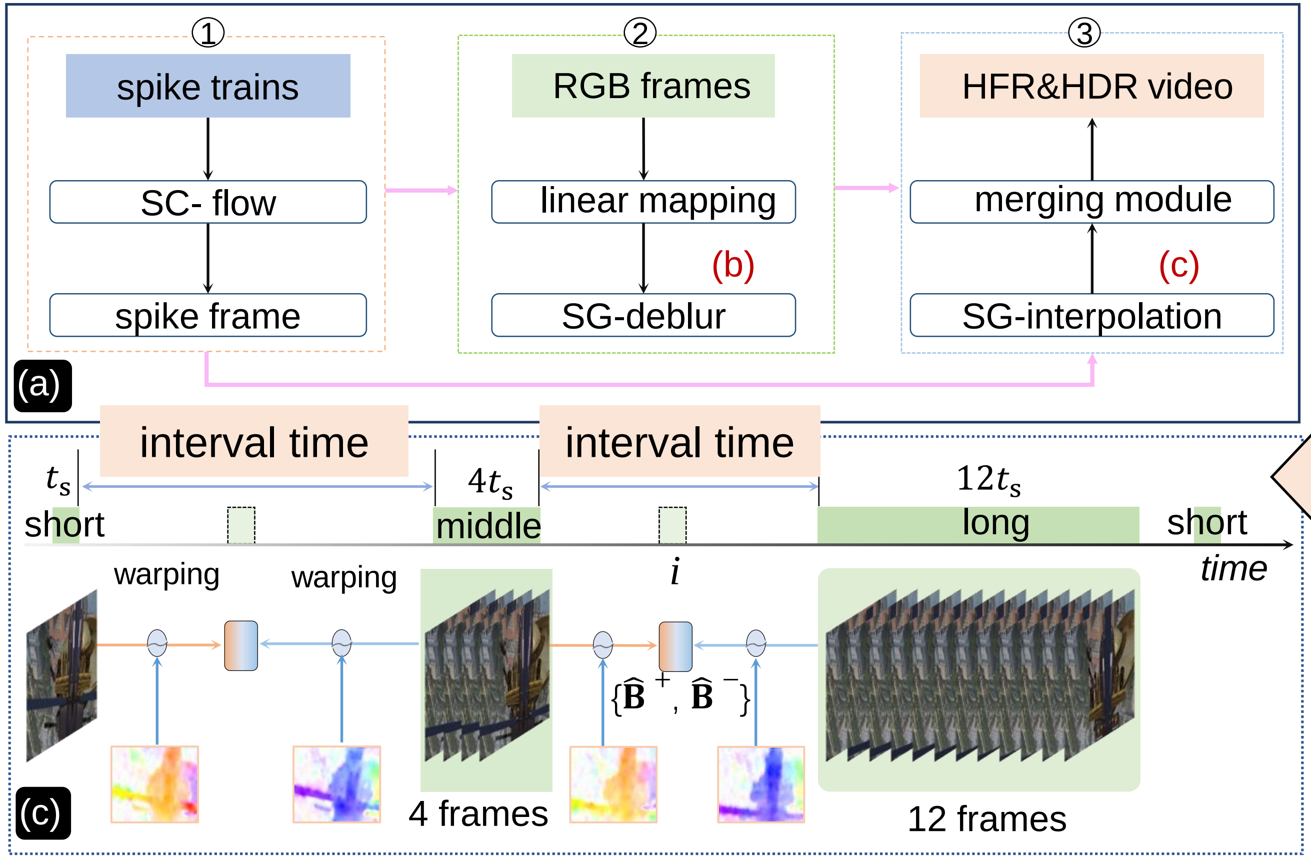


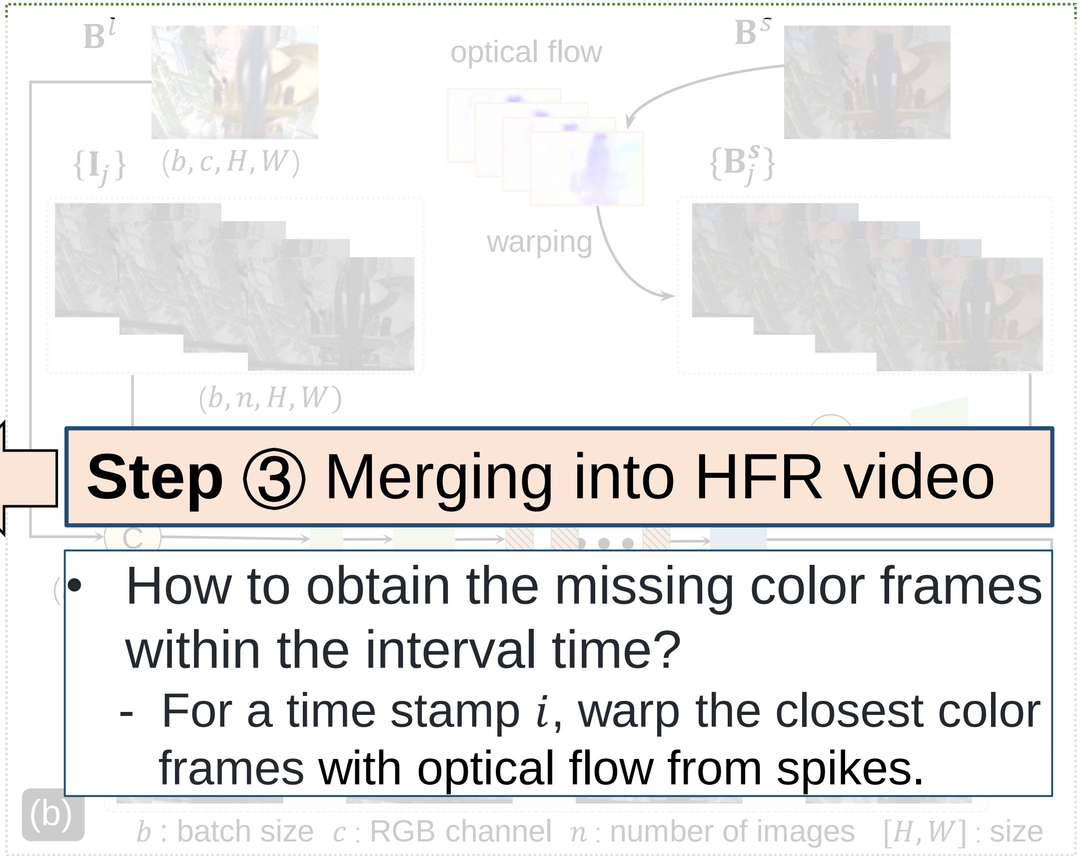
Step ③ Merging into HFR video

- How to obtain the missing color frames within the interval time?

Image Interpolation with Optical Flows

The pipeline of the proposed solution.

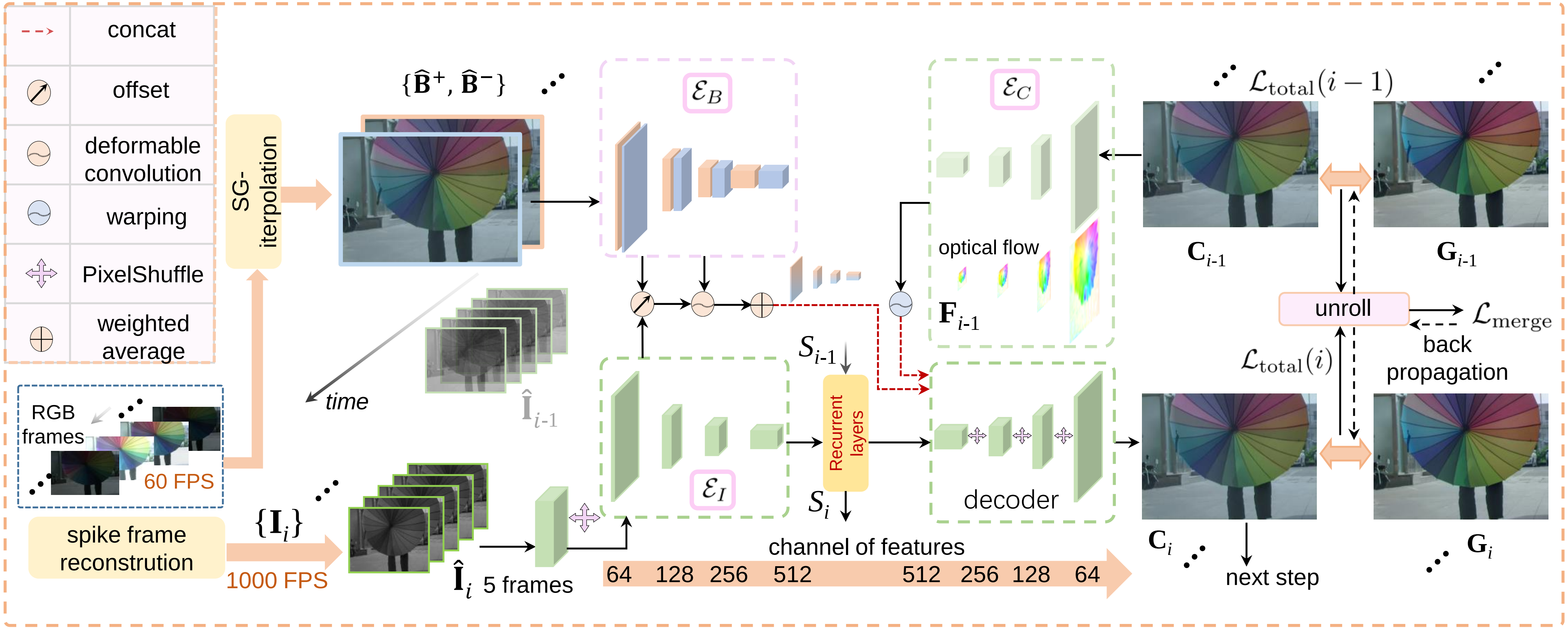




Step ③ Merging into HFR video

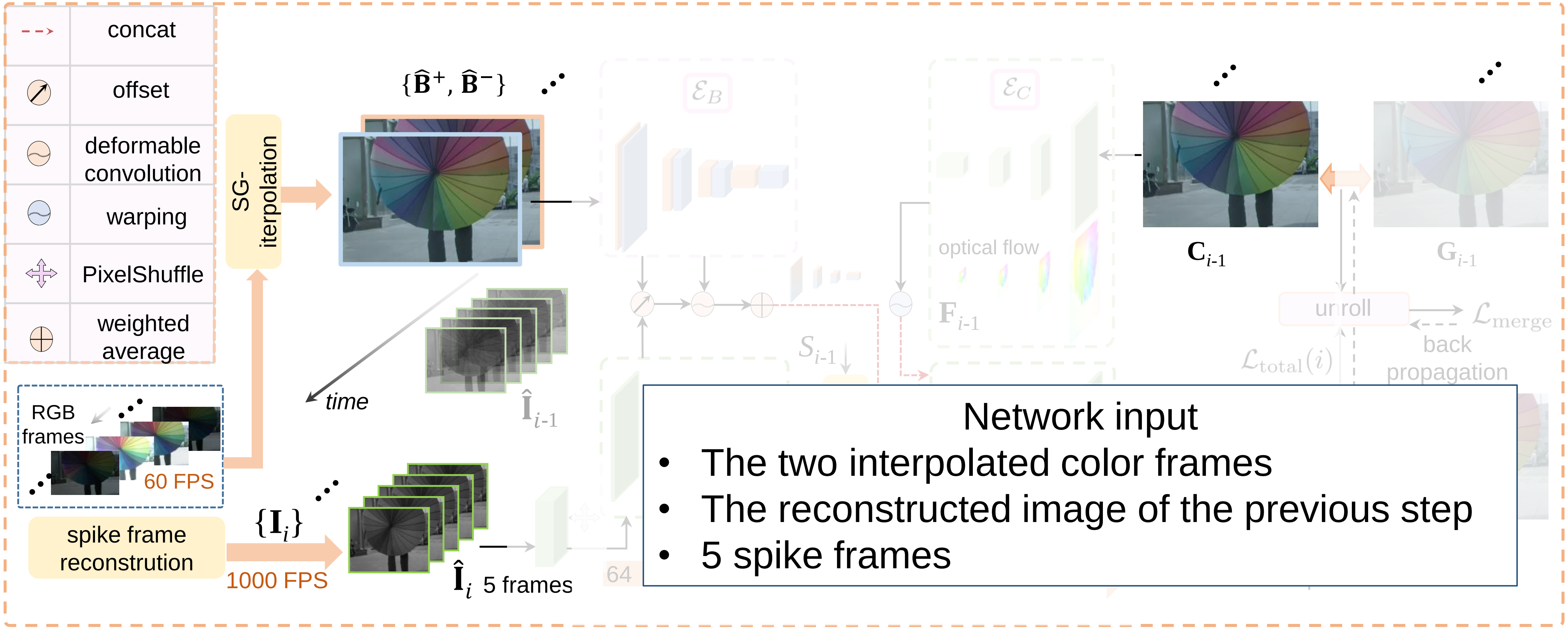
- How to obtain the missing color frames within the interval time?
  - For a time stamp i, warp the closest color frames with optical flow from spikes.

Merging Module



- Build a CNN-RNN based HFR&HDR video reconstruction network for refinement.

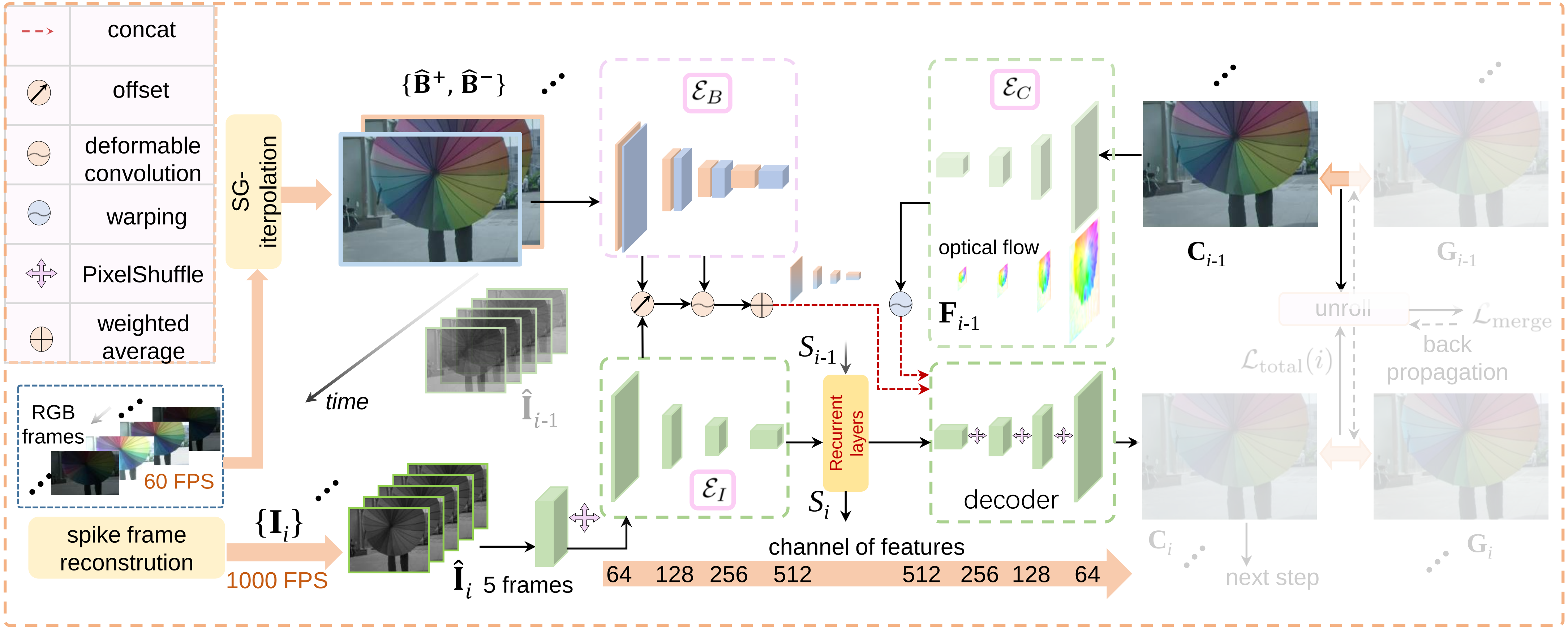
Merging Module



Network input

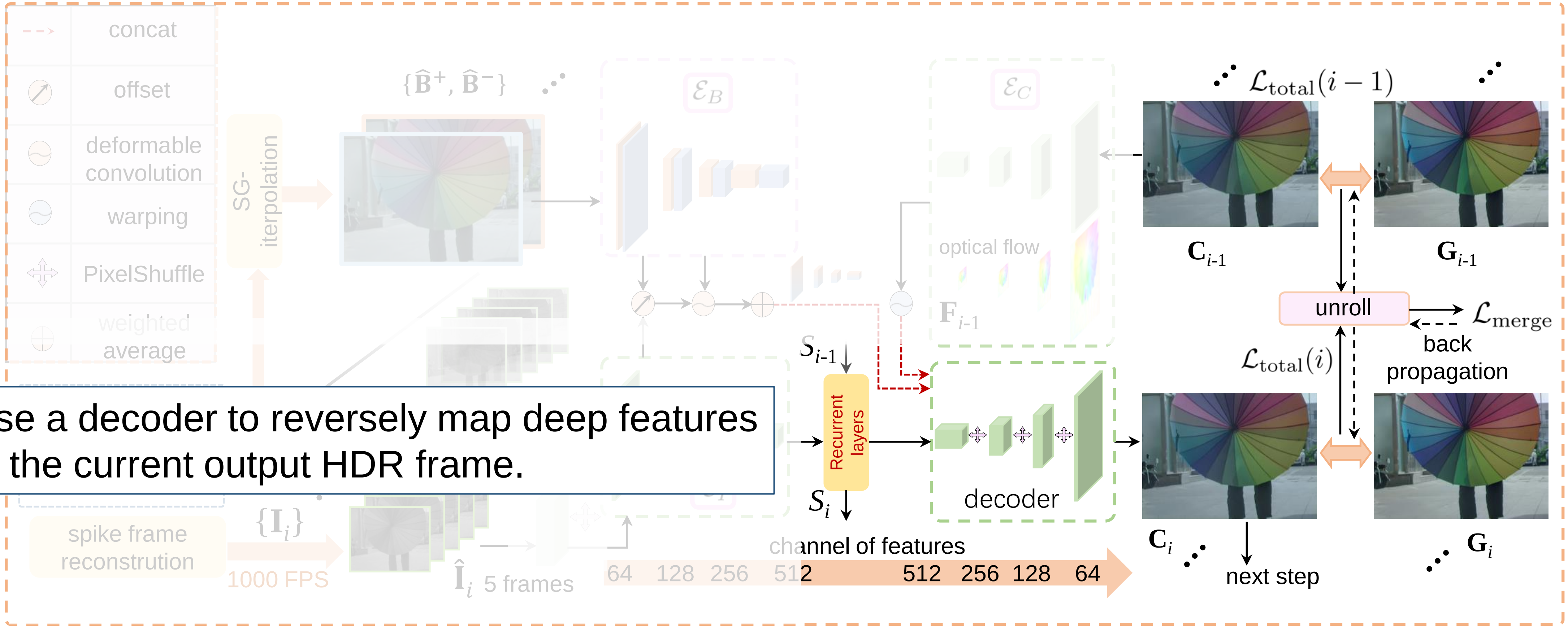
- The two interpolated color frames
- The reconstructed image of the previous step
- 5 spike frames

Merging Module



- The merging module consists of three encoders, which are respectively designed for feature extraction.

Merging Module

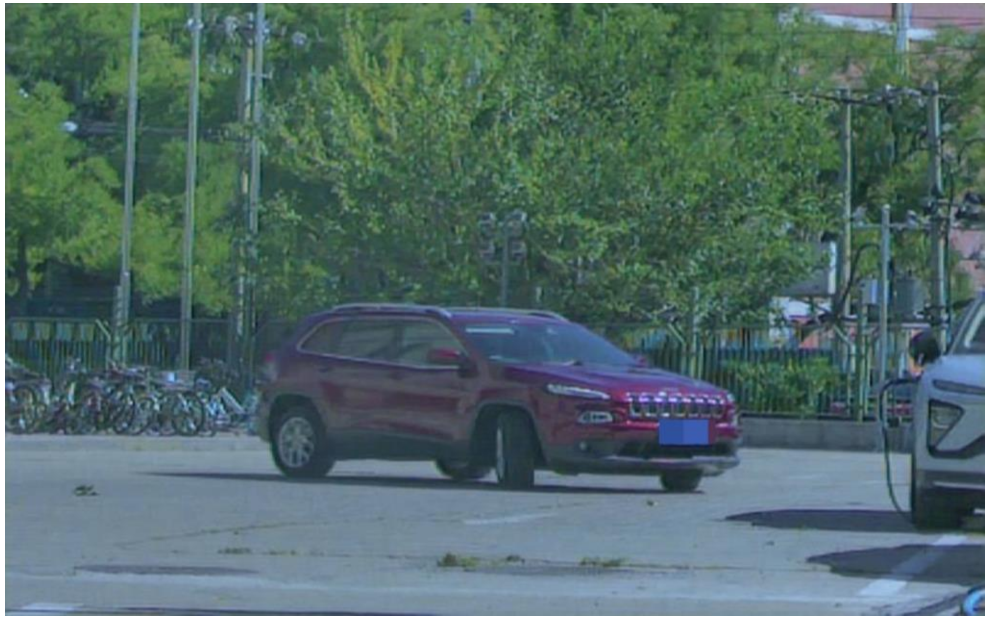


- Use a decoder to reversely map deep features to the current output HDR frame.

Evaluation on Real-Synthetic Data

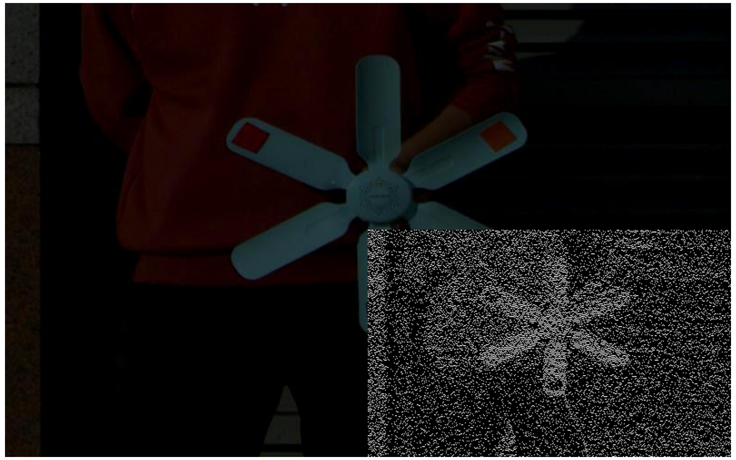
Input (RGB & spike)

Output

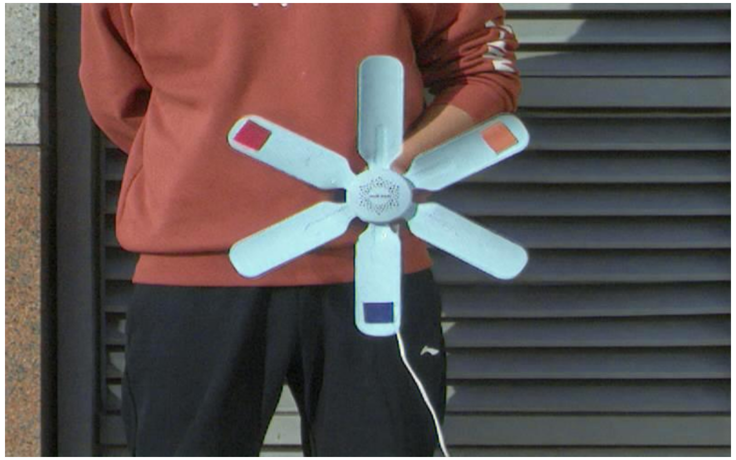


Evaluation on Real Data

Input (RGB & spike)



Ours

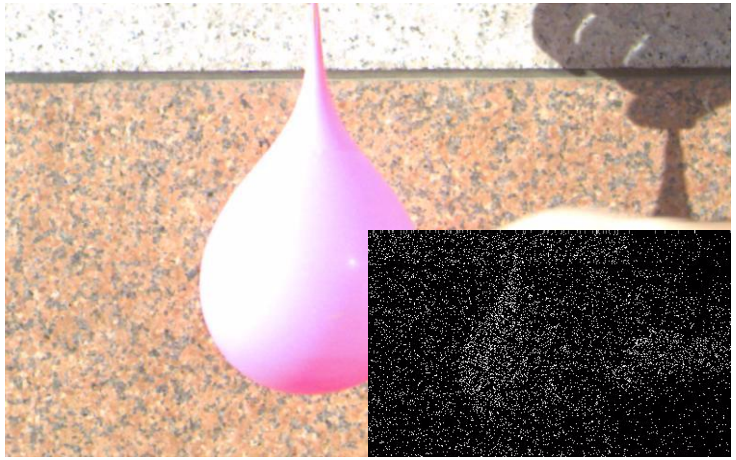


iPhone 13 (240 FPS)

Mi 10 (120 FPS)

Evaluation on Real Data

Input (RGB & spike)



Ours

iPhone 13 (240 FPS)

Mi 10 (120 FPS)

Evaluation on Real Data

Input (RGB & spike)

Ours

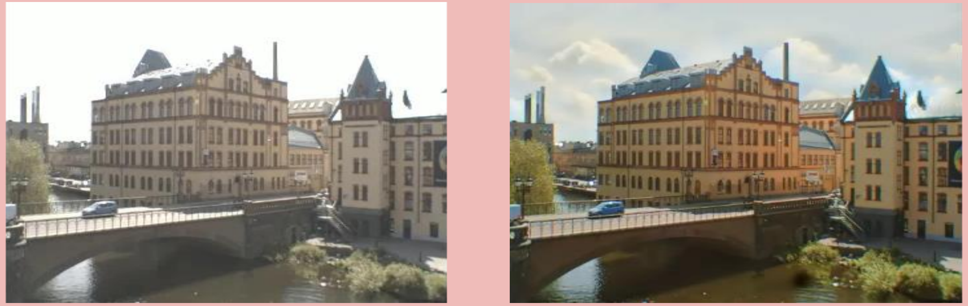
Phantom camera (1000 FPS)

Camera Pipeline

[CVPR 2023]

All-in-focus Imaging

[CVPR 2023]



HDR Imaging

[CVPR 2022]

GS Imaging

[CVPR 2023]

HDR HFR Imaging

北京大学计算机学院
School of Computer Science

A Computational Photography Laboratory at Peking University
http://camera.pku.edu.cn

NeuCAP

Neuromorphic Camera Aided Photography

(Super)

北京大学计算机学院
School of Computer Science

A Computational Photography
Laboratory at Peking University
http://camera.pku.edu.cn

Thank You!

Q and A

Boxin Shi
shiboxin@pku.edu.cn