2024 IEEE International Conference on Acoustics,
Speech and Signal Processing

DIFFEVENT: EVENT RESIDUAL DIFFUSION FOR IMAGE DEBLURRING

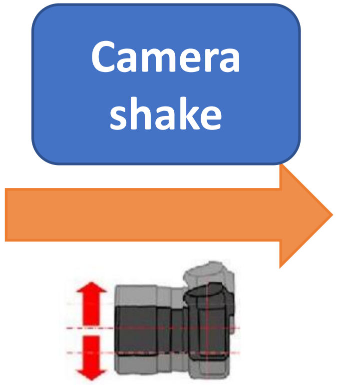
Pei Wang, Jiumei He, Qingsen Yan*, Yu Zhu, Jinqiu Sun, Yanning Zhang

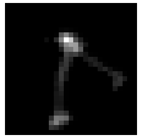
17 April, 2024

Image blurring process

- Conventional degradation model:

x: sharp image





k: blur kernel

y: blurry image

$$\mathbf{y} = \overset{?}{\mathbf{x}} * \overset{?}{\mathbf{k}} + \overset{?}{\mathbf{n}}$$

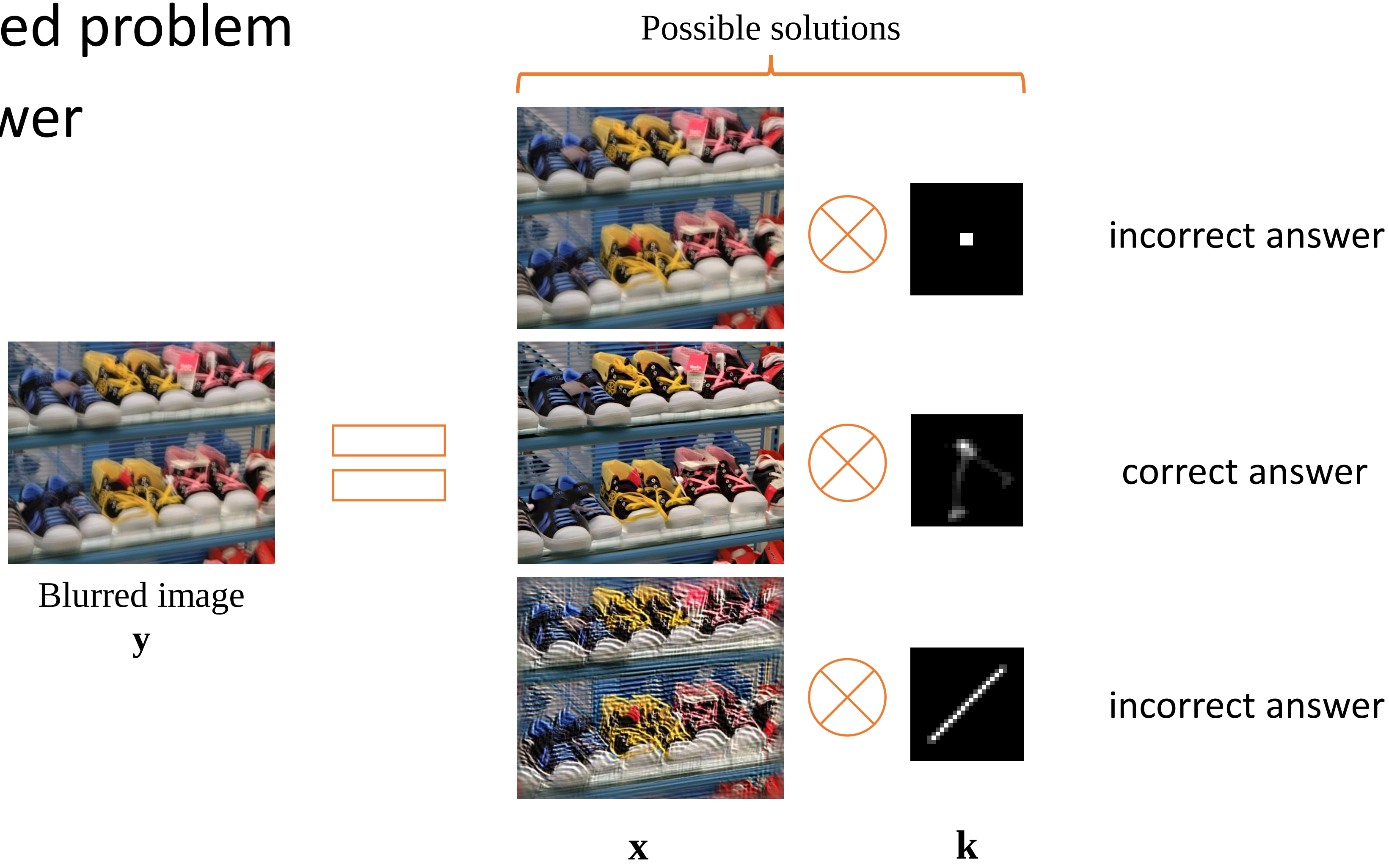
*: convolution operation

n: additive noise

Task description: From a blurred image to find a corresponding sharp image.

Challenges

- Severely ill-posed problem
- No unique answer



$$\overset{?}{} \quad \mathbf{y} = \overset{?}{\mathbf{x}} * \overset{?}{\mathbf{k}} + \overset{?}{\mathbf{n}}$$

General solution pipeline

Traditional optimization methods

Blurry image

Kernel estimation

Sharp image estimation

neural network

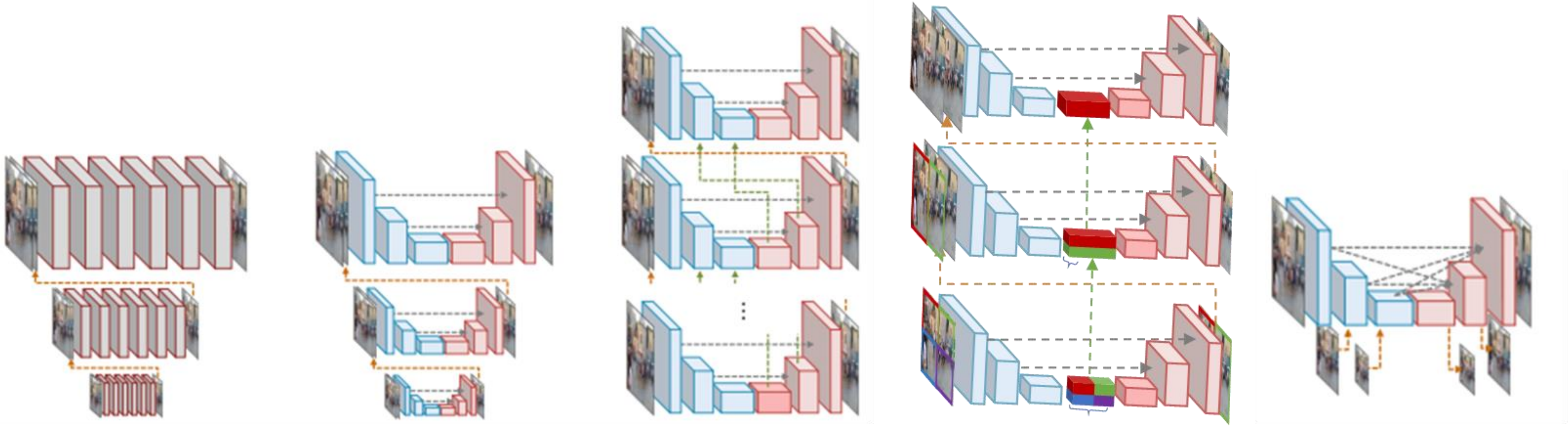
Deep learning methods

Sharp image

With the development of deep learning, more and more researches try to use the neural network to replace the traditional optimization process and achieve great progress.

Conventional deep learning methods

- Based on conventional degradation process, researchers developed various networks for alleviating the blurriness in synthetic scenarios



(a) multi-scale | (b) encoder-decoder multi-scale | (c) multi-temporal | (d) multi-patch/multi-stage | (e) multi-input-multi-output

[1] Dynamic Scene Deblurring With Parameter Selective Sharing and Nested Skip Connections, CVPR 2019
[2] Deep multi-scale convolutional neural network for dynamic scene deblurring, CVPR 2017
[3] Scale-recurrent network for deep image deblurring, CVPR 2018
[4] Rethinking Coarse-to-Fine Approach in Single Image Deblurring, ICCV 2021

Problems in conventional deblurring

- Conventional degradation model is constructed based on several assumptions: convolution process, Gaussian noise, motion kernels…
- Real-world scenarios include complicated degradation processes, even unknown degradation form

$$y = \overset{?}{k} * \overset{?}{x} + \overset{?}{n}$$

Conventional degradation model:
convolution form, unknown parameters

$$y = \overset{?}{H}[\overset{?}{x}]$$

Real-world degradation:
complicated and unknown form

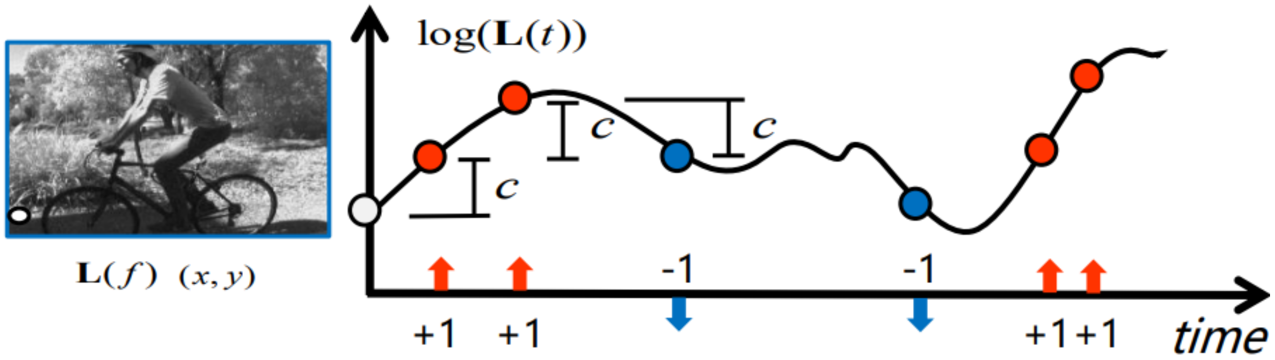
Conventional degradation model didn't work in the real-world degradation scenarios!

Our works:

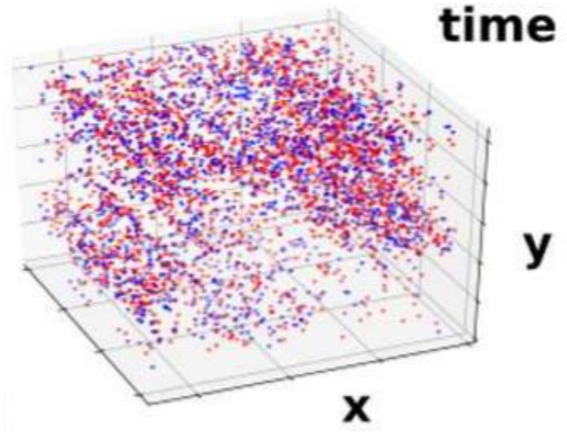
Introduce the event camera data into conventional image deblurring task to assist the scene restoration

Event camera

- **Event cameras** record an asynchronous stream of per-pixel brightness changes, called "events".
- Event data consists of a stream that encode the time, location, and polarity (-1,+1) of the brightness changes.



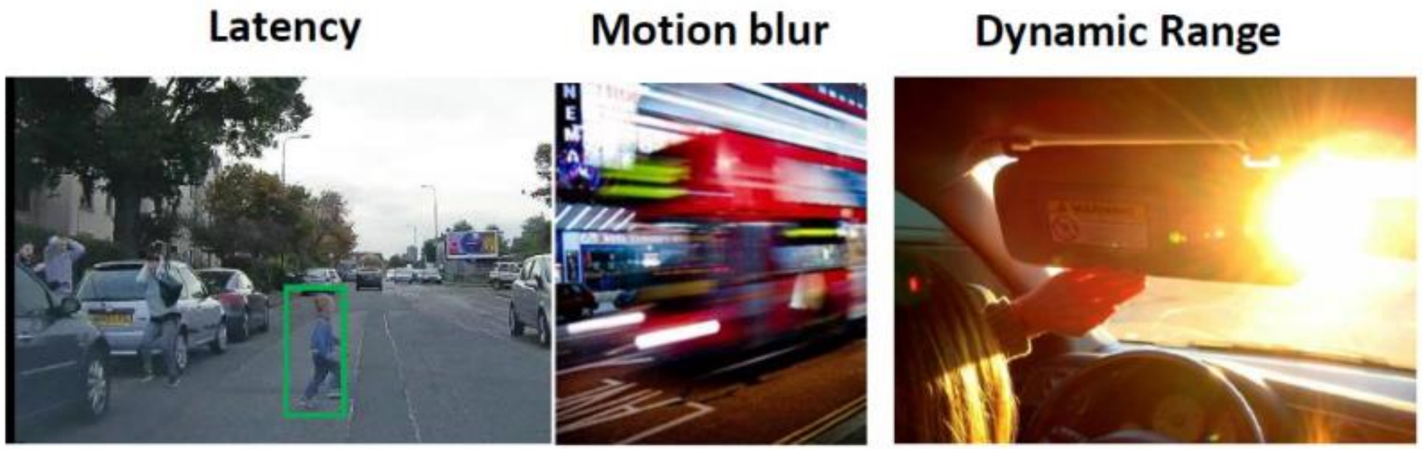
$\mathbf{L}$ is the intensity image, f is the reference timestamp.
The event is triggered when a change in the **log intensity** exceeds a given **threshold** c .



Event data

Advantages of event camera for deblurring

- High temporal resolution, low latency ($< 1\mu s$ vs. 1ms)
- High dynamic range (140dB vs. 60dB)
- No motion blur
- Low power consumption (20mW vs. 1.5W)



Event cameras do not suffer from blur and saturation issues

Correlation between blur image - event data

- Let $e(s)$ refer to current event, then event data from τ to t is $E(\tau, t)$, image at t is marked as:

$$x(t) = x(\tau)\exp(cE(\tau,t)) = x(\tau)\exp(c\int_{\tau}^{t} e(s)ds) \qquad (1)$$

- For a blurry image y obtained by a conventional camera during exposure period $[\tau - \frac{T}{2}, \tau + \frac{T}{2}]$ with event data:

EDI model[5]: $$y = \frac{1}{T}\int_{\tau-T/2}^{\tau+T/2} x(t)dt = \frac{1}{T}\int_{\tau-T/2}^{\tau+T/2} x(\tau)\exp(c\int_{\tau}^{t} e(s)ds)dt \qquad (2)$$

- However, EDI is an ideal model due to the fixed threshold c and noise free assumptions.

[5] Pan et.al. , Bringing a blurry frame alive at high frame-rate with an event camera, CVPR 2019

Issues of EDI model in real-world

$$y = \frac{1}{T}\int_{\tau-T/2}^{\tau+T/2} x(\tau)\exp(c\int_{\tau}^{t} e(s)ds)dt$$

- Spatial uncertainty thresholds c
- Neglected motion information by thresholds c
- Sensor noise in raw event data

$$p_i = \begin{cases} +1, if \log(\frac{\tau_t(x_i, y_i)}{\tau_{t-\Delta t}(x_i, y_i)}) > c \\ -1, if \log(\frac{\tau_t(x_i, y_i)}{\tau_{t-\Delta t}(x_i, y_i)}) < -c \end{cases}$$

Event data in $[-c, c]$ has been neglected

(a) Frame 1 (b) Frame 2 (c) Events (d) Average Threshold

[6] Songnan Lin et.al., Learning event-driven video deblurring and interpolation, ECCV 2020

Our Method: DiffEvent

Contributions:

- First method to introduce the diffusion into event deblurring task.
- **Event Blurry Residual Degradation** (EBRD) model for real-world blur degradation construction.
- Event Residual Diffusion for Image Deblurring (DiffEvent)
  - **Alternative sampling strategy** of clear images and residual images
  - **Pseudo-inverse guidance** to generate high-quality images

Event Blurry Residual Degradation model

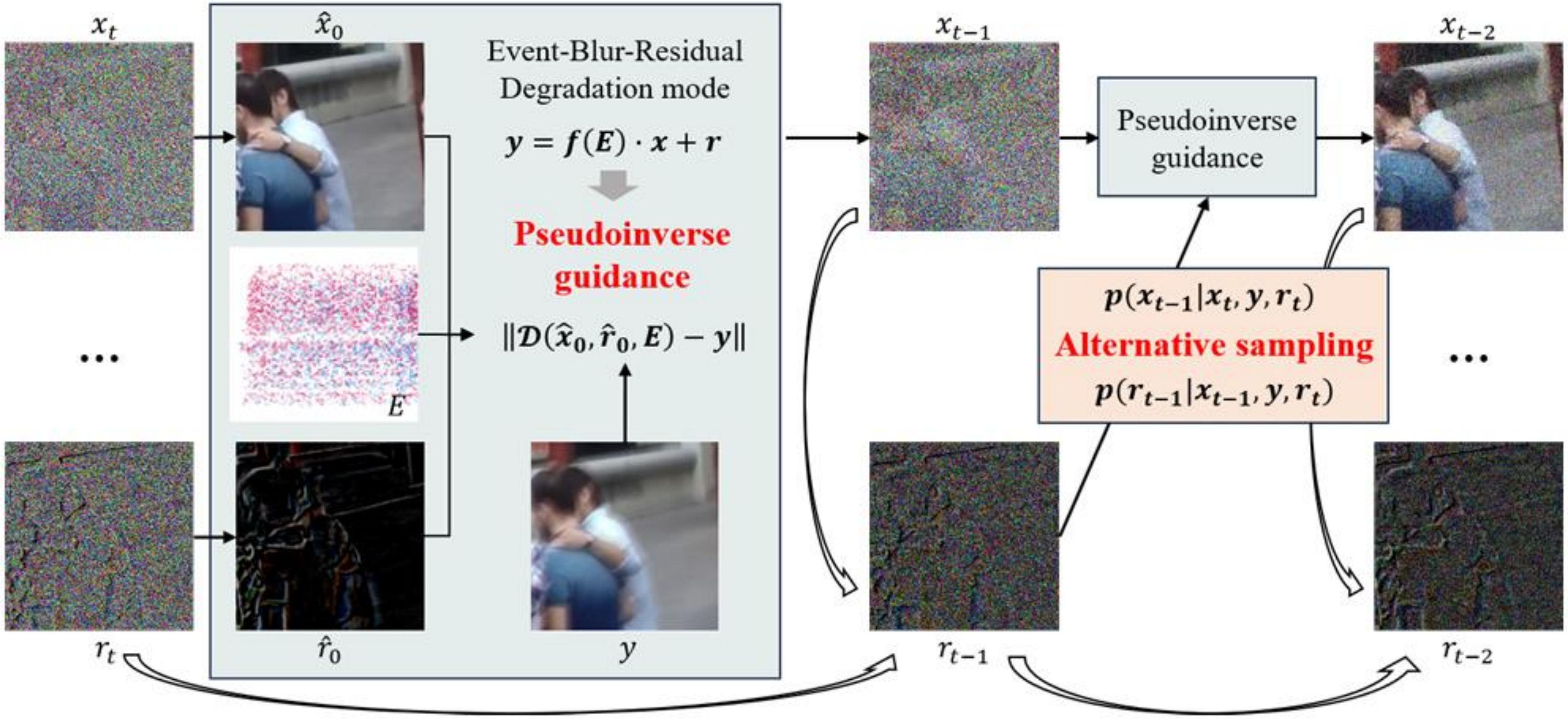
- We propose EBRD model:

$$E = \frac{1}{T}\int_{\tau-\frac{T}{2}}^{\tau+\frac{T}{2}} \exp(c\int_{\tau}^{t} e(s)ds)dt \quad (3)$$

$$y = (E + \Delta E) \cdot x = f(E) \cdot x + r = \mathcal{D}(x, E, r) \quad (4)$$

  - E: raw event data
  - ΔE: residuals of the raw event data
  - Function $f(E)$: alleviate the noise in the raw event data
  - Residual image r: supplement the motion clues ignored by threshold

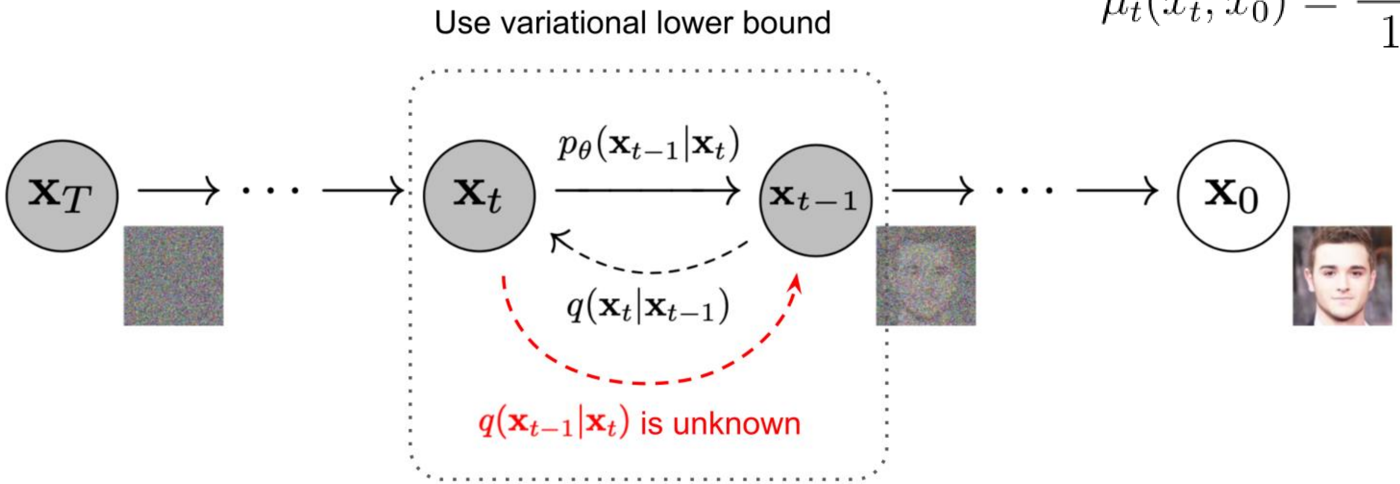
Event Residual Diffusion for Image Deblurring



Noise diffusion model

$$\beta_t = 1 - \alpha_t \qquad \sigma_t^2 = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$$

$$\mu_t(x_t, \hat{x}_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{x}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t$$



- Forward diffusion: $q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t}x_{t-1}, (1-\alpha_t)I)$ (5)
- Inverse sampling: $p(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2(x_t, t)I)$ (6)
- Sample by predicted x_0: $x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$ (7)

$$p(x_{t-1}|x_t, \hat{x}_0) = \mathcal{N}(x_{t-1}; \mu_t(x_t, \hat{x}_0), \sigma_t^2 I) \quad (8)$$

[7] Jonathan Ho et.al., Denoising diffusion probabilistic models, NeurIPS 2020

Alternative sampling strategy

- Step x: sampling x_{t-1} from x_t and r_t

$$p(x_{t-1}|x_t, y, r_t) = p(x_{t-1}|x_t)\frac{p(y|x_{t-1}, r_t)}{p(y|x_t, r_t)}\frac{p(r_t|x_{t-1})}{p(r_t|x_t)}$$

$$= p(x_{t-1}|x_t)e^{\log p(y|x_{t-1},r_t)-\log p(y|x_t,r_t)}e^{\log p(r_t|x_{t-1})-\log p(r_t|x_t)}$$

$$\approx p(x_{t-1}|x_t)exp[(x_{t-1} - x_t) \cdot \nabla_{x_t} \log(p(y|x_t, r_t) + p(r_t|x_t))]$$

$$p(x_{t-1}|x_t, y, r_t) \propto \mathcal{N}(x_{t-1}; \mu_t + \sigma_t^2\nabla_{x_t} \log \boxed{p(y|x_t, r_t)} + \sigma_t^2\nabla_{x_t} \log \boxed{p(r_t|x_t)}, \sigma_t^2 I) \quad (9)$$

- Step r: sampling r_{t-1} from x_{t-1} and r_t

$$p(r_{t-1}|x_{t-1}, y, r_t) \propto \mathcal{N}(r_{t-1}; \mu_t' + \sigma_t'^2\nabla_{r_t} \log \boxed{p(y|x_{t-1}, r_t)} + \sigma_t'^2\nabla_{r_t} \log \boxed{p(x_{t-1}|r_t)}, \sigma_t'^2 I) \quad (10)$$

Pseudo-inverse guidance

- Based on EBRD, we introduce blurry image y into sampling process:

Sample x

$$p(y|x_t, r_t) = exp(-s\mathcal{L}(\mathcal{D}(x_t, E, r_t), y)) = exp(-s\left\|y - (f(E)x_t + r_t)\right\|_2^2) \quad (11)$$

$$p(r_t|x_t) = \frac{p(x_t|r_t)p(r_t)}{p(x_t)} = exp(-\mathcal{L}(y - r_t, f(E)x_t) - \|r_t\|_1 - \lambda\left\|x_t\right\|_{TV}) \quad (12)$$

$$\boxed{\mathcal{N}(x_{t-1}; \mu_t - \sigma_t^2 \nabla_{x_t}[s\left\|y - f(E)x_t - r_t\right\|_2^2 + \lambda\left\|x_t\right\|_{TV}], \sigma_t^2 I)} \quad (13)$$

Sample r

$$p(y|x_{t-1}, r_t) = exp(-s'\mathcal{L}(\mathcal{D}(x_{t-1}, E, r_t), y)) = exp(-s'\left\|y - (f(E)x_{t-1} + r_t)\right\|_2^2)) \quad (14)$$

$$p(x_{t-1}|r_t) = \frac{p(r_t|x_{t-1})p(x_{t-1})}{p(r_t)} = exp(-\mathcal{L}(y - f(E)x_{t-1}, r_t) - \|r_t\|_1 - \lambda'\left\|x_{t-1}\right\|_{TV}) \quad (15)$$

$$\boxed{\mathcal{N}(r_{t-1}; \mu_t' - \sigma_t'^2 \nabla_{r_t}[s'\left\|y - f(E)x_{t-1} - r_t\right\|_2^2 + \lambda'\left\|r_t\right\|_1], \sigma_t'^2 I)} \quad (16)$$

Sampling algorithm

- Similar to original diffusion process[7], we sample images from $\hat{x}_0, \hat{r}_0$
- $\hat{x}_0$ and $\hat{x}_0$ are predicted from x_t and r_t
- $\hat{x}_0^*$ refers to prediction from x_{t-1}

$$x_{t-1} = \mu_t - \sigma_t^2 \nabla_{\hat{x}_0}[s\mathcal{L}(\mathcal{D}(\hat{x}_0, E, \hat{r}_0), y) + \lambda \|\hat{x}_0\|_{TV}] + \sigma_t \epsilon$$

$$r_{t-1} = \mu'_t - \sigma_t'^2 \nabla_{\hat{r}_0}[s'\mathcal{L}(\mathcal{D}(\hat{x}_0^*, E, \hat{r}_0), y) + \lambda' \|\hat{r}_0\|_1] + \sigma'_t \epsilon$$

Sampling models

Algorithm 1 Alternative sampling of images x, r.

Require: Sampling steps T, blurry image y, event data E, scaling factors s, s', λ, λ', pre-trained diffusion models $(\mu(x_t), \sigma(x_t)), (\mu'(r_t), \sigma'(r_t))$, distance metrics $\mathcal{L}$

$x_T, r_T \sim \mathcal{N}(0, I)$
for t **from** T **to** 1 **do**
  $\hat{x}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(x_t, t))$
  $\hat{r}_0 = \frac{1}{\sqrt{\bar{\alpha}'_t}}(r_t - \sqrt{1-\bar{\alpha}'_t}\epsilon_\theta(r_t, t))$
  $\mathcal{L}_{\hat{x}_0}^{total} = s\mathcal{L}(\mathcal{D}(\hat{x}_0, E, \hat{r}_0), y) + \lambda \|\hat{x}_0\|_{TV}$
  Sample x_{t-1} *by* $\mathcal{N}(\mu_t - \sigma_t^2 \nabla_{\hat{x}_0}\mathcal{L}_{\hat{x}_0}^{total}, \sigma_t^2 I)$
  $\hat{x}_0^* = \frac{1}{\sqrt{\bar{\alpha}_{t-1}}}(x_{t-1} - \sqrt{1-\bar{\alpha}_{t-1}}\epsilon_\theta(x_{t-1}, t-1))$
  $\mathcal{L}_{\hat{r}_0}^{total} = s'\mathcal{L}(\mathcal{D}(\hat{x}_0^*, E, \hat{r}_0), y) + \lambda' \|\hat{r}_0\|_1$
  Sample r_{t-1} *by* $\mathcal{N}(\mu'_t - \sigma_t'^2 \nabla_{\hat{r}_0}\mathcal{L}_{\hat{r}_0}^{total}, \sigma_t'^2 I)$
end
return x_0, r_0

[7] Jonathan Ho et.al., Denoising diffusion probabilistic models, NeurIPS 2020

Experiments

- **Dataset:**

✓Synthetic event data: GoPro dataset[8] with ESIM simulating

✓Real-world event data: REBlur dataset[9] with real event camera

- **Compared methods:**

✓Generation-based methods, CNN/RNN-based methods

✓Conventional camera methods, Event camera degradation methods

- **Implementation details:**

✓Pytorch & NVIDIA Tesla A100

✓Finetune the pre-trained diffusion models on deblurring datasets

[8] Nah et.al., Deep multi-scale convolutional neural network for dynamic scene deblurring, CVPR 2017
[9] Lei Sun et.al., Event-based fusion for motion deblurring with cross-modal attention, ECCV 2022

Experiments

Quantitative comparison results:

Our method performs better compared to various methods, including generation and CNN/RNN-based methods, as well as some methods with conventional degradation model and event degradation model.

Table 1. Evaluation results on GoPro/REBlur[6]. * means the event-enhanced version and **bold** indicates the best results.

Methods	DeblurGAN[14]	GDP[15]	BlindDPS[16]	SRN*[17]	HINet*[18]	EFNet[6]	DiffEvent
Structure	GAN	Diffusion	Diffusion	CNN+event	CNN+Event	CNN+Event	Diff+Event
PSNR↑	29.55/30.21	19.93/21.38	11.37/10.43	31.02/36.87	33.69/37.68	35.46/38.12	**35.55/38.23**
SSIM↑	0.934/0.945	0.583/0.603	0.197/0.274	0.936/0.970	0.961/0.973	0.972/0.973	**0.972/0.974**
LPIPS↓	0.117/0.115	0.312/0.366	0.705/0.770	0.124/0.119	0.088/0.076	0.064/0.023	**0.060/0.021**

Generation-based methods
Conventional camera degradation model

CNN/RNN-based methods
Event camera degradation model

Experiments

Qualitative comparison results:

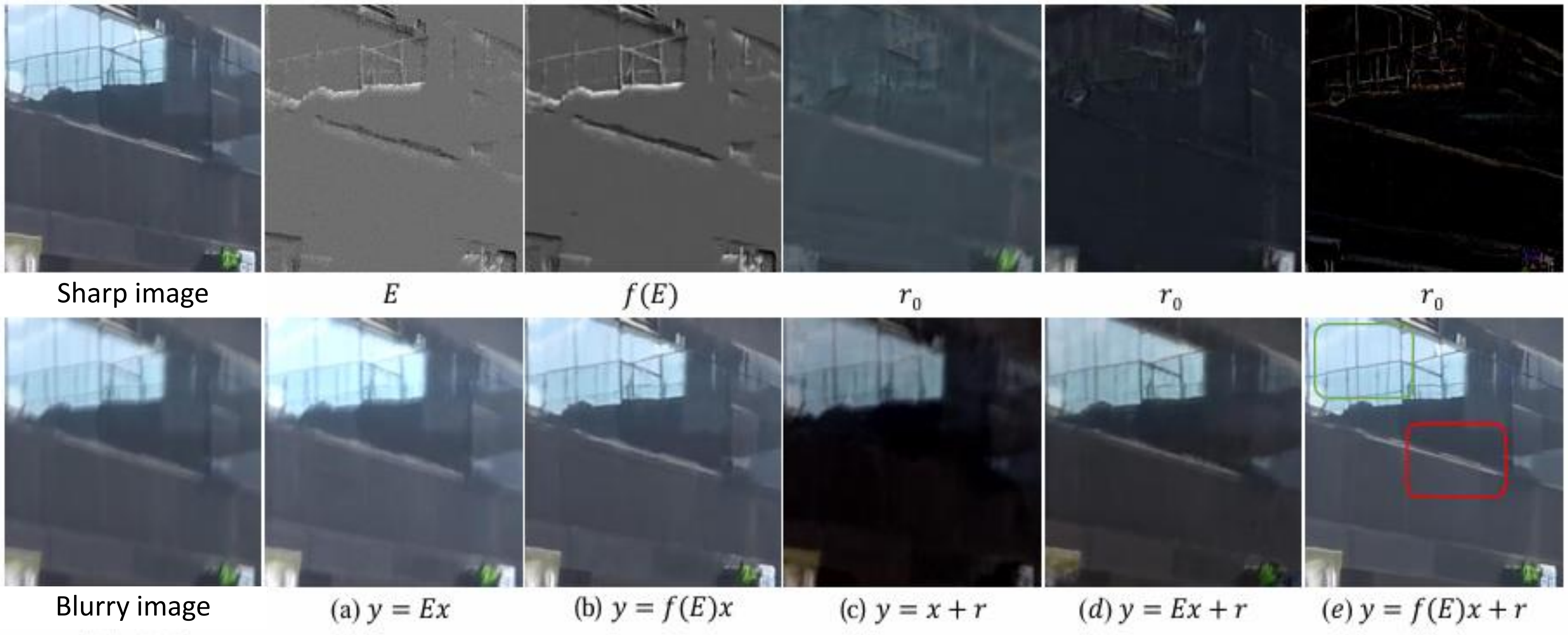
Our method can effectively handle real-world deblurring problems, generate vivid and detailed structures because of the proposed pseudoinverse guidance from the EBRD mode and alternative sampling strategy during diffusion.

blurry GDP BlindDPS EFNet ours sharp

Ablation study

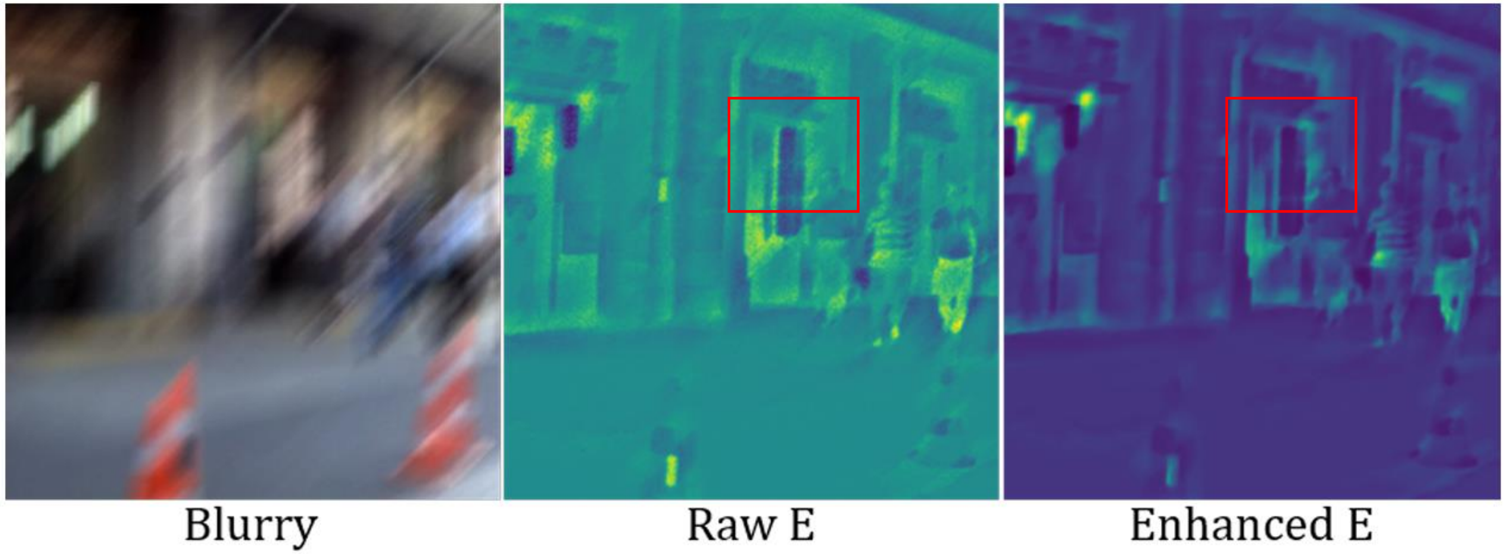
The effect of different parts on EBRD model

Degradation model	Raw event E	Enhanced E $f(E)$	Residual r	PSNR	SSIM
(a) $y = Ex$	✓	✗	✗	25.95	0.901
(b) $y = f(E)x$	✓	✓	✗	33.51	0.967
(c) $y = x + r$	✗	✗	✓	10.96	0.269
(d) $y = Ex + r$	✓	✗	✓	16.53	0.729
(e) $y = f(E)x + r$	✓	✓	✓	35.55	0.973



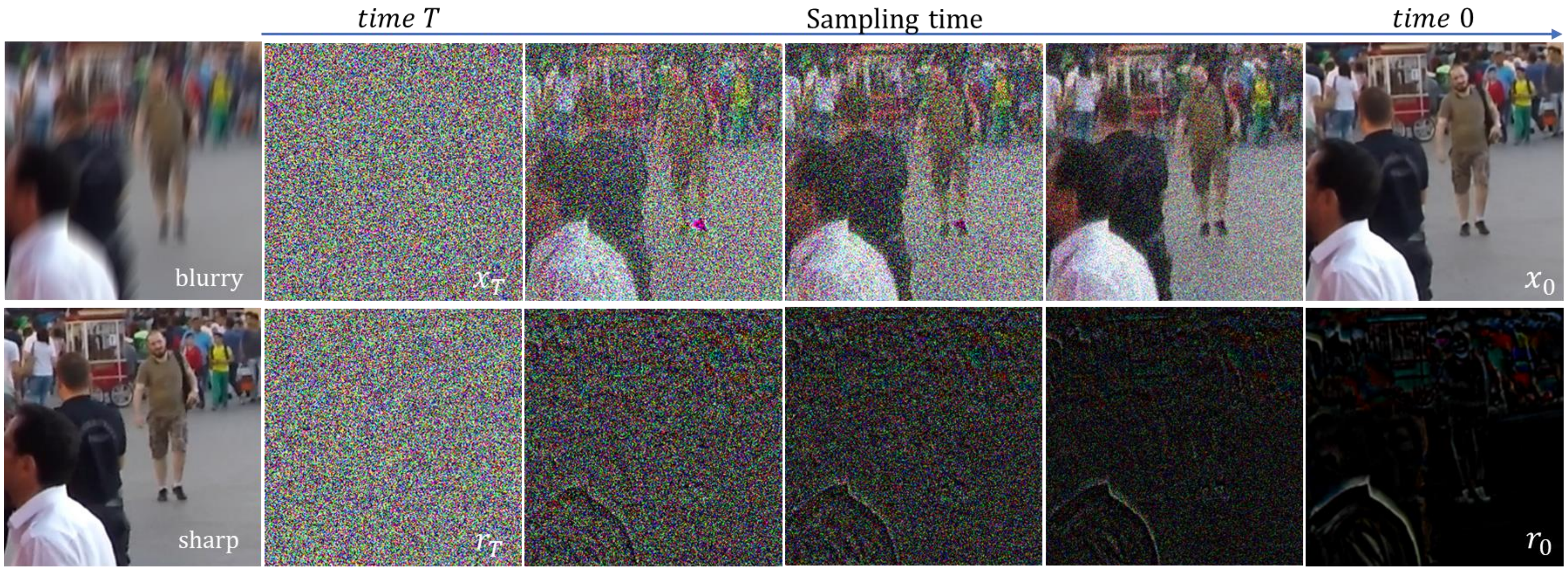
Discussion

- The necessity of event enhancement module
  - Raw event still contains some non-continuous pixels, and the sensor noise can be found everywhere
  - To prevent negative impact, the enhanced event module $f(E)$ effectively alleviates the noise in raw events, resulting in smoother signals.



Discussion

- Effectiveness of alternating sampling strategy



Conclusion

- This paper tackles the challenge of real-world blur in conventional cameras by harnessing the potential of event cameras with high temporal resolution.
- Our approach modifies the original EDI model to EBRD with additional residual and enhanced event components for event deblurring .
- We present an alternative diffusion sampling strategy that jointly estimates clear and residual images to ensure high-quality results. A pseudo-inverse guidance module has enhanced fine details.
- Experiments on benchmark event datasets underscore the effectiveness of our approach.

Thank you for your attention!

Q&A