南京航空航天大學
Nanjing University of Aeronautics and Astronautics

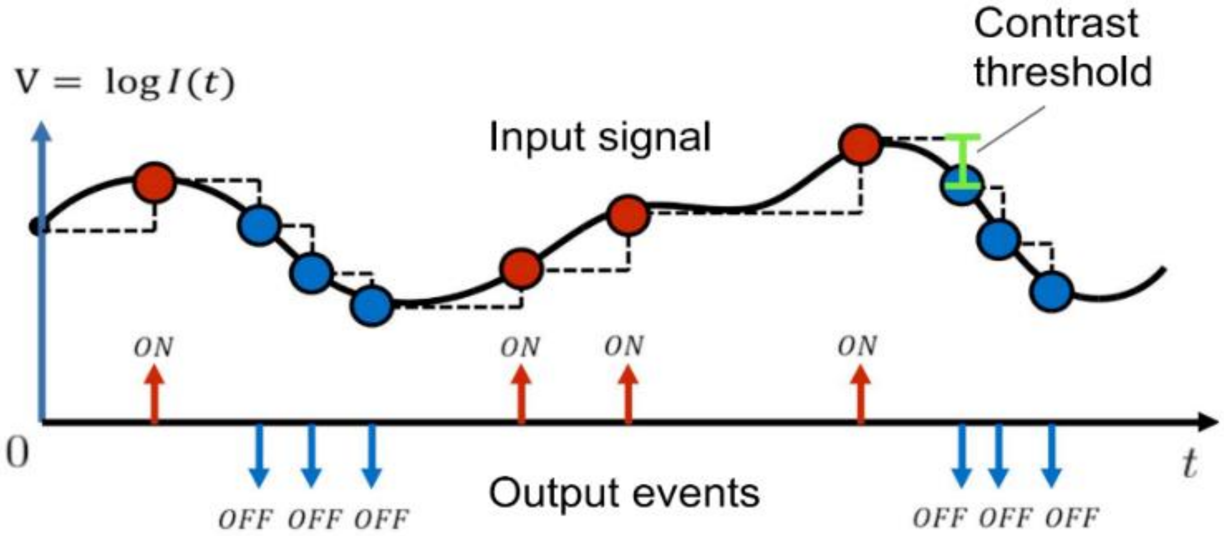
E-CIR: Event-Enhanced Continuous Intensity Recovery

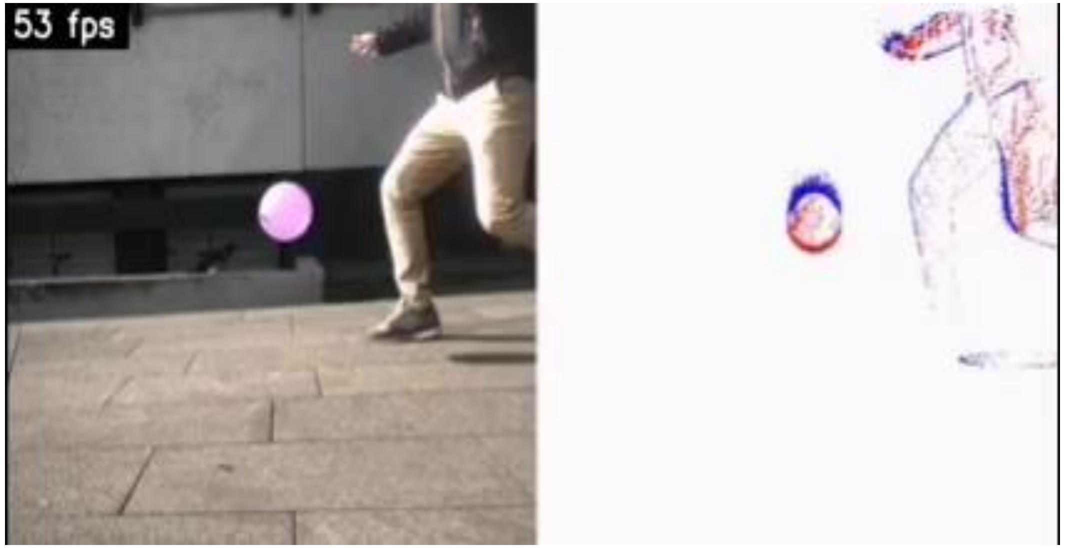
Chen Song Qixing Huang Chandrajit Bajaj
The University of Texas at Austin
{song, bajaj, huangqx}@cs.utexas.edu

CVPR 2022

Event Camera Model

$\log I(x, y, t) - \log I(x, y, t\text{-}\Delta t) = \pm C$





(x, y, t, p) denote an event, where $p \in \{-1, +1\}$

Event set: $\mathcal{E} = \{\mathbf{e}_k\}_{k=1}^{N}$

Problem Description

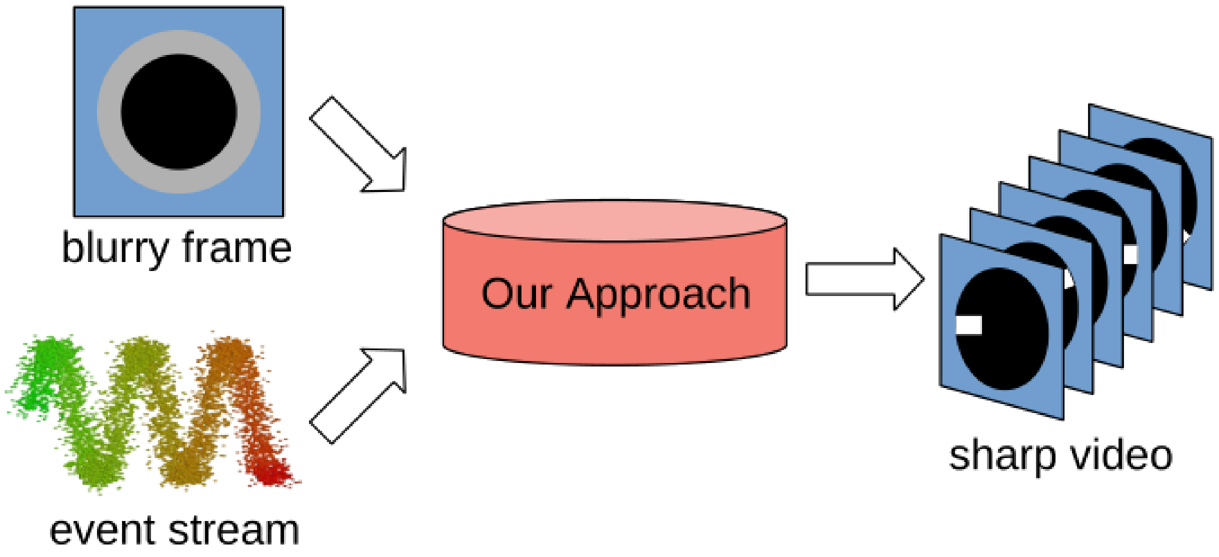


Figure 1. Problem Description. In this imaginary scene, we place a white square along the edge of a black disk. The image taken by the conventional camera is blurry because the disk rotates at a fast speed. It is as if the perimeter of the disk somehow grows into a gray collar. During the exposure interval, the event sensor produces a spiral of events. Our approach takes the blurry frame and the events as input and produces a sharp video sequence as output. The output video explains the motion blur by entailing the complete motion trajectory of the rotating disk.

Task Description

Input

1. The blurry intensity B_{xy} for all pixels (x, y);
2. A collection of events during the exposure interval

$$\{e_i = (x_i, y_i, t_i, p_i) | -\frac{T}{2} \leq t_i \leq \frac{T}{2}\}$$

Output

$$\mathbf{L}_{h \times w}(t) = \{\mathbf{L}_{xy}(t)\}$$

Idea

$$\mathbf{B}_{xy} = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \mathbf{L}_{xy}(t)dt \qquad (3)$$

The events associated with pixel (x, y) provide a set of timestamps where the intensity change considerably.

Overview

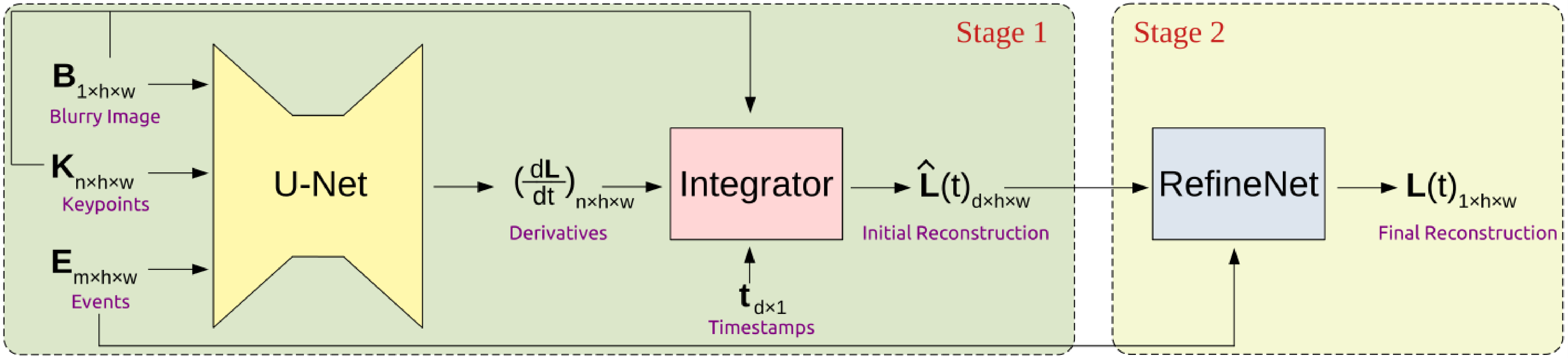


Figure 3. The overall pipeline. We use the U-Net [34] model to regress the polynomial coefficients. The network takes three inputs: the blurry frame $\mathbf{B}$, the keypoint timestamps $\mathbf{K}$, and the event histogram $\mathbf{E}$. The network then outputs the intensity derivatives $\frac{d\mathbf{L}}{dt}$. Given an arbitrary timestamp t, the integrator follows Equation (7) and calculates the initial frame reconstruction $\hat{\mathbf{L}}(t)$. The refinement module takes $\hat{\mathbf{L}}(t)$ and $\mathbf{E}$ as input and outputs the final frame reconstruction $\mathbf{L}(t)$. In this figure, m, n, d, h, w represent the number of histogram bins, the number of keypoints, the number of frames in the output video, the frame height, and the frame width, respectively.

Initialization: Polynomial Coefficient Regression

南京航空航天大学 Nanjing University of Aeronautics and Astronautics

$$\mathbf{L}_{xy}(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_n t^n \quad (4)$$

$$|\alpha_0| \approx |\alpha_1| > |\alpha_2| > \cdots \gg |\alpha_n|$$

$$\mathcal{K}_{xy} = \{(t_i, \frac{d\mathbf{L}_{xy}}{dt}(t_i))|1 \leq i \leq n\} \quad (5)$$

$$-\frac{T}{2} \leq t_1 < t_2 < \cdots < t_n \leq \frac{T}{2}$$

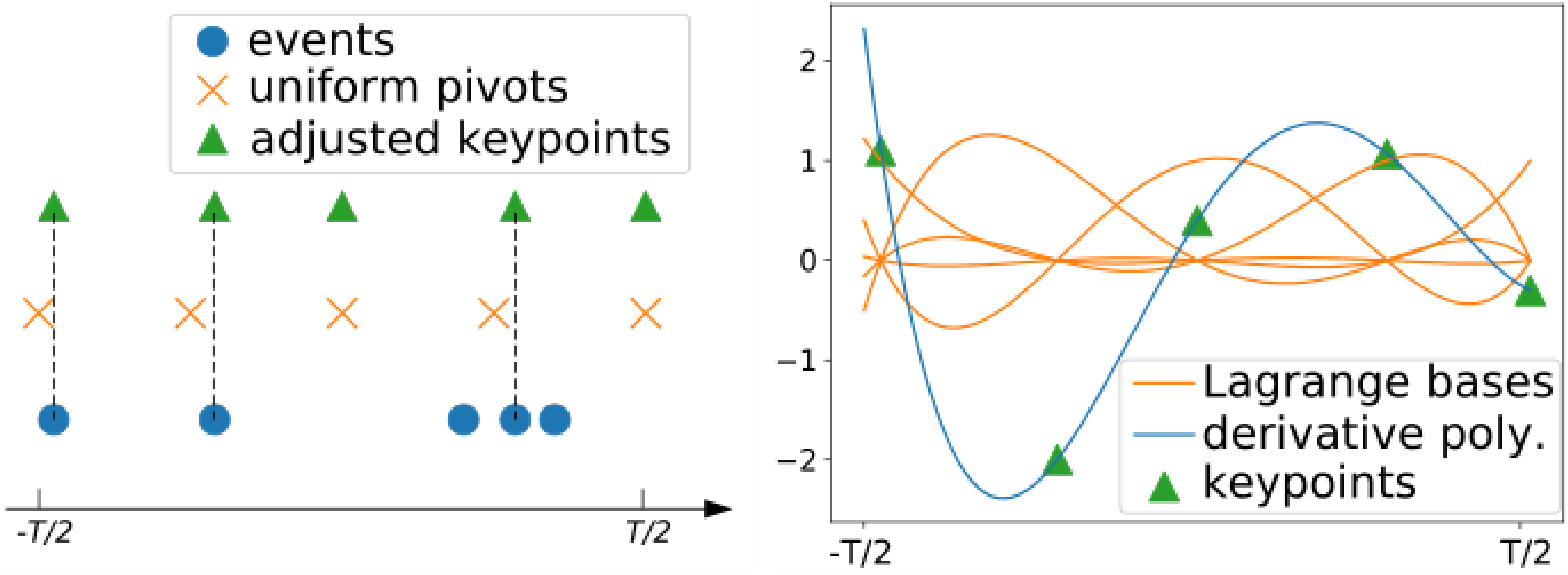
$$\frac{d\mathbf{L}_{xy}}{dt}(t) = \sum_{i=1}^{n} \frac{d\mathbf{L}_{xy}}{dt}(t_i) \cdot \beta_{xyi}(t) \quad (6)$$

$$\beta_{xyi}(t) = \frac{(t-t_1)\cdots(t-t_{i-1})(t-t_{i+1})\cdots(t-t_n)}{(t_i-t_1)\cdots(t_i-t_{i-1})(t_i-t_{i+1})\cdots(t_i-t_n)}$$

$$\mathbf{L}_{xy}(t) = \int \frac{d\mathbf{L}_{xy}}{dt}(t)dt + a_{xy} \quad (7)$$

$$\mathbf{B}_{xy} = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} \mathbf{L}_{xy}(t)dt \quad (3)$$

Keypoint Selection Algorithm.



(a) (b)

Polynomial interpolation

南京航空航天大学 Nanjing University of Aeronautics and Astronautics

$$l_0(x) = \frac{x - x_1}{x_0 - x_1}, \quad l_1(x) = \frac{x - x_0}{x_1 - x_0},$$

$$L_1(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0). \longrightarrow L_1(x) = \frac{x - x_1}{x_0 - x_1} y_0 + \frac{x - x_0}{x_1 - x_0} y_1, \longrightarrow$$

$$L_1(x) = y_0 l_0(x) + y_1 l_1(x) = \sum_{i=0}^{1} y_i l_i(x),$$

$l_i(x)$ is called Lagrange bases

$$L_n(x_i) = y_i = f(x_i), \quad i = 0, 1, \ldots, n. \longrightarrow L_n(x) = \sum_{i=0}^{n} y_i l_i(x) = \sum_{i=0}^{n} \Big(\prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j} \Big) y_i.$$

$$l_i(x) = \frac{(x - x_0) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_n)}{(x_i - x_0) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n)} = \prod_{\substack{j=0 \\ i \neq j}}^{n} \frac{x - x_j}{x_i - x_j}.$$

From a geometric point of view, it is to find an n-th degree curve $y=L_n(x)$, so that it passes through (n+1) points$(x_0,y_0),(x_1,y_1),\ldots,(x_n,y_n)$

Refinement: Temporal Feature Propagation

Algorithm 1 Refinement

Input: Initial reconstruction: $\{\hat{\mathbf{L}}(t_i) | 1 \le i \le d\}$
Input: All events: $\{(x, y, t, p) | -\frac{T}{2} \le t \le \frac{T}{2}\}$
Output: Final reconstruction: $\{\mathbf{L}(t_i) | 1 \le i \le d\}$

loop I_{max} iterations
  for $i = 1$ **to** $d - 1$ **do** ▷ Residual Prediction
    $\mathbf{E} \leftarrow$ VOXELIZE(events from t_i to t_{i+1})
    if i == 1 **then**
      $\mathbf{R}_i \leftarrow g^{\mathbf{R}}_{\theta_1}(\mathbf{E}, \hat{\mathbf{L}}(t_i), \hat{\mathbf{L}}(t_{i+1}), \nabla\hat{\mathbf{L}}(t_i), \nabla\hat{\mathbf{L}}(t_{i+1}))$
    else
      $\mathbf{R}_i \leftarrow g^{\mathbf{R}}_{\theta_2}(\mathbf{R}_{i-1}, \mathbf{E}, \hat{\mathbf{L}}(t_i), \hat{\mathbf{L}}(t_{i+1}), \nabla\hat{\mathbf{L}}(t_i), \nabla\hat{\mathbf{L}}(t_{i+1}))$
    end if
  end for
  for $i = 1$ **to** d **do** ▷ Apply Updates
    $\mathbf{A_i} \leftarrow g^{\mathbf{A}}_{\phi}(\hat{\mathbf{L}}(t_i))$
    $\mathbf{D_i} \leftarrow \frac{\partial f(\mathbf{R}_1, \cdots, \mathbf{R}_{d-1}, \hat{\mathbf{L}}(t_1), \cdots, \hat{\mathbf{L}}(t_d), \mathbf{L}(t_1), \cdots, \mathbf{L}(t_d))}{\partial \mathbf{L}(t_i)}$
    $\hat{\mathbf{L}}(t_i) \leftarrow \hat{\mathbf{L}}(t_i) - \mathbf{A_i} \odot \mathbf{D_i}$
  end for
end loop
for $i = 1$ **to** d **do** ▷ Final Polishing
  $\mathbf{L}(t_i) \leftarrow g^{\mathbf{L}}_{\gamma}(\hat{\mathbf{L}}(t_i))$
end for

$$f = \sum_{i=1}^{d-1} \mathcal{L}(\mathbf{L}(t_i) + \mathbf{R}_i, \mathbf{L}(t_{i+1})) + \lambda \sum_{i=1}^{d} \mathcal{L}(\mathbf{L}(t_i), \hat{\mathbf{L}}(t_i))$$

L2-distance as L(·, ·)

$g^{\mathbf{R}}_{\theta_1}$, $g^{\mathbf{R}}_{\theta_2}$: Recurrent Neural Network

$g^{\mathbf{A}}_{\phi}$, $g^{\mathbf{L}}_{\gamma}$: Convolutional Neural Network

Training Objective

Derivative Loss
$$\mathcal{L}_{\text{d}} = |(\frac{d\mathbf{L}}{dt})_{\text{gt}} - (\frac{d\mathbf{L}}{dt})_{\text{pred}}|_1 \quad (9)$$

Primitive Loss
$$\mathcal{L}_{\text{p}} = |\mathbf{L}_{\text{gt}} - \mathbf{L}_{\text{pred}}|_1 \quad (10)$$

Refinement Loss
$$\mathcal{L}_{\text{ref}} = \sum_t |\mathbf{L}_{\text{gt}}(t) - \mathbf{L}_{\text{pred}}(t)|_1 \quad (11)$$

Residual Loss
$$\mathcal{L}_{\text{res}} = \sum_i (exp(\rho \cdot |\mathbf{R}_{i_{\text{gt}}}|_1) \odot |\mathbf{R}_{i_{\text{gt}}} - \mathbf{R}_{i_{\text{pred}}}|_1) \quad (12)$$

Total Objective
$$\mathcal{L}_{\text{total}} = \lambda_{\text{d}}\mathcal{L}_{\text{d}} + \lambda_{\text{p}}\mathcal{L}_{\text{p}} + \lambda_{\text{ref}}\mathcal{L}_{\text{ref}} + \lambda_{\text{res}}\mathcal{L}_{\text{res}} \quad (13)$$

Experiments

Table 1. Quantitative evaluation on the REDS [22] dataset.

Methods	MSE ↓	PSNR ↑	SSIM ↑
EDI [26]	0.182	21.663	0.664
eSL-Net [40] (official)	0.203	20.640	0.601
eSL-Net [40] (re-trained)	0.201	20.748	0.646
Ours	**0.114**	**25.531**	**0.819**

REDS is a standard deblurring benchmark dataset designed for conventional cameras. The dataset contains 240 training videos and 30 validation videos

Experiments

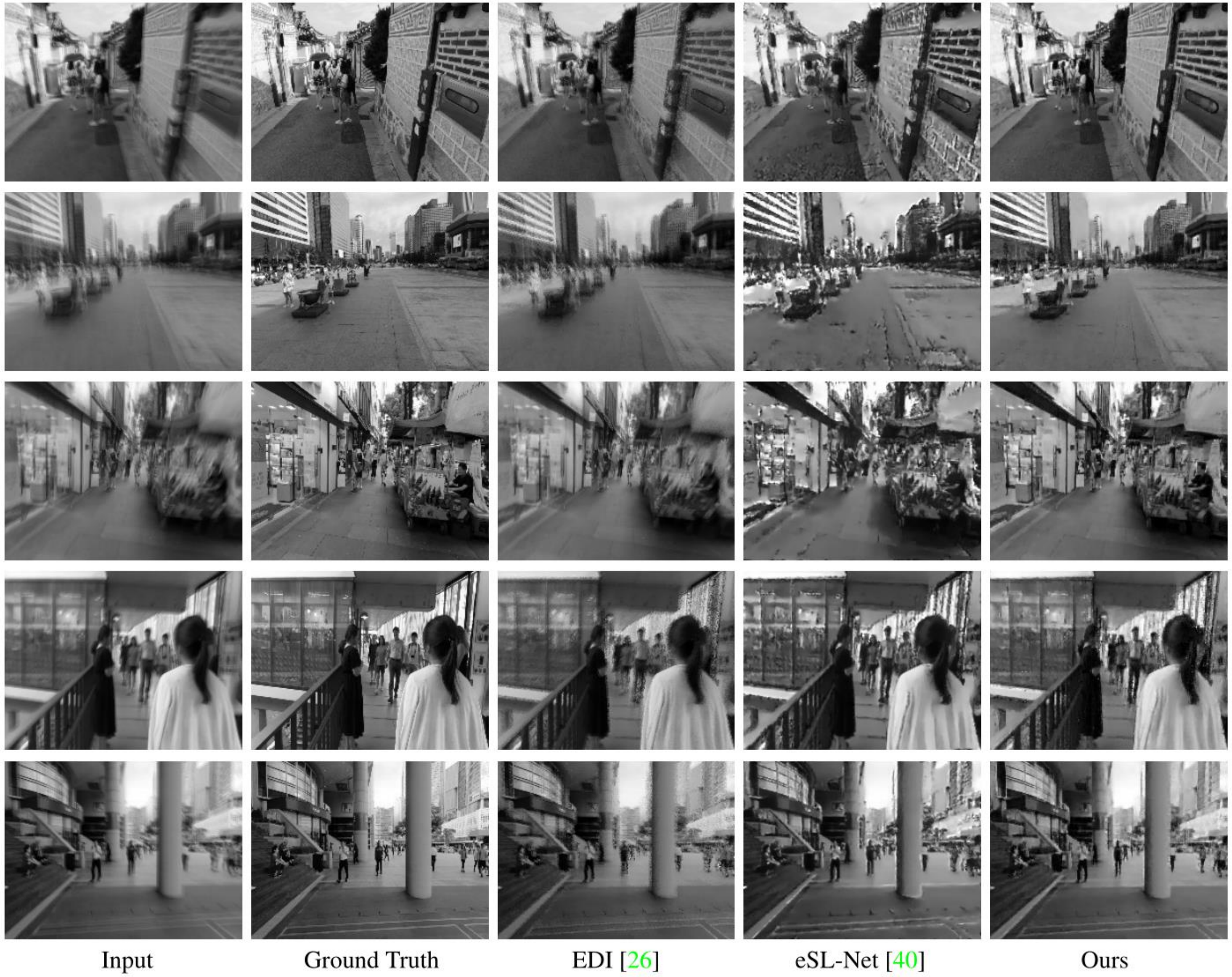
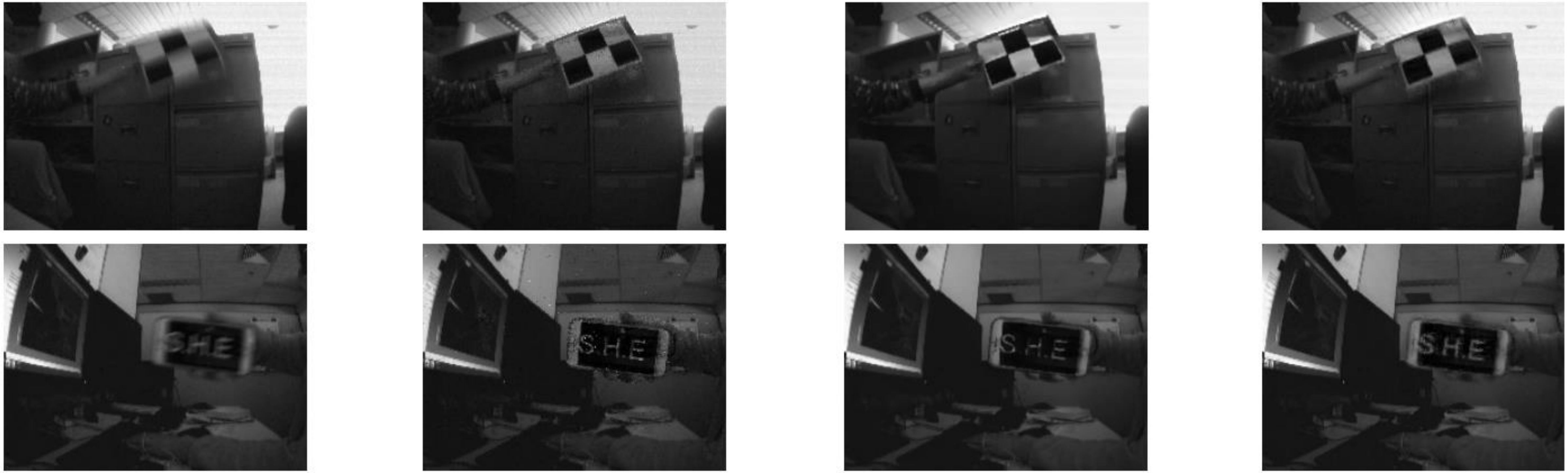


Figure 4. Qualitative visualization on the REDS [22] dataset.

Experiments



Input | EDI [26] | eSL-Net [40] | Ours

Figure 5. Qualitative visualization on real captures by Pan et al. [26].

Experiments

Row	Input Sources		Video Format		Stages		Performance Metrics		
	Frame	Events	Poly.	Frame	Initialization	Refinement	MSE ↓	PSNR ↑	SSIM ↑
1	✗	✓	✓	✗	✓	✗	0.134	24.210	0.767
2	✓	✗	✓	✗	✓	✗	0.180	21.721	0.654
3	✓	✓	✓	✗	✓	✗	0.125	24.807	0.787
4	✓	✓	✗	✓	✓	✗	0.136	23.504	0.723
5	✓	✓	✓	✗	✓	✓	**0.114**	**25.531**	**0.819**

Table 2. On the REDS [22] dataset, we use ablation studies to demonstrate the importance of using a dual-stream input, the power of the polynomial representation, and the strength of the refinement module.

南京航空航天大學

Nanjing University of Aeronautics and Astronautics

Thanks for Listening