Unifying Motion Deblurring and Frame Interpolation with Events - Supplementary Material -

Xiang Zhang, Lei Yu†
Wuhan University, Wuhan, China.
{xiangz, ly.wd}@whu.edu.cn

1. Proof in Feasibility Analysis

In this section, we prove the feasibility of applying Eq. (2) to interpolate the latent images outside the exposure time of blurry frames. For readability, we copy the formulations in the main text here.

$$B = \frac{1}{T}\int_{t\in\mathcal{T}} L(t)dt, \tag{1}$$

$$L(f) = \frac{B}{E(f,\mathcal{T})}, \tag{2}$$

$$E(f,\mathcal{T}) = \frac{1}{T}\int_{t\in\mathcal{T}} \exp(c\int_{f}^{t} e(s)ds)dt. \tag{3}$$

We assume three overlapped blurry frames B_{01}, B_{02} and B_{12}, as shown in Fig. 1, of which the exposure time is $\mathcal{T}_{01}$, $\mathcal{T}_{02}$ and $\mathcal{T}_{12}$, respectively. Since t_0 locates inside both $\mathcal{T}_{01}$ and $\mathcal{T}_{02}$, the following equations can be directly derived using Eq. (2):

$$\begin{aligned} B_{01} &= L(t_0)\cdot E(t_0,\mathcal{T}_{01}),\\ B_{02} &= L(t_0)\cdot E(t_0,\mathcal{T}_{02}). \end{aligned} \tag{4}$$

Therefore one can obtain

$$B_{02} - B_{01} = L(t_0)\cdot(E(t_0,\mathcal{T}_{02}) - E(t_0,\mathcal{T}_{01})). \tag{5}$$

Based on Eq. (1) and Eq. (3), we also have

$$\begin{aligned} B_{12} &= B_{02} - B_{01},\\ E(t_0,\mathcal{T}_{12}) &= E(t_0,\mathcal{T}_{02}) - E(t_0,\mathcal{T}_{01}), \end{aligned} \tag{6}$$

and thus can easily derive

$$L(t_0) = \frac{B_{12}}{E(t_0,\mathcal{T}_{12})}, \tag{7}$$

which means Eq. (2) can be also applied to recover the latent images, *e.g.*, $L(t_0)$, located outside the exposure period of the blurry frames, *e.g.*, B_{12}.

†Corresponding author

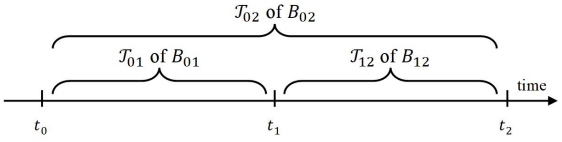

Figure 1. Example of three overlapped blurry frames, where the target timestamp t_0 locates inside the exposure time of B_{01} and B_{02}, but outside the exposure time of B_{12}.

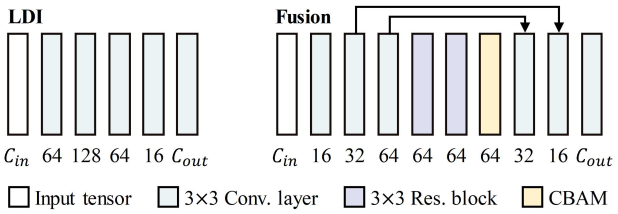

Figure 2. Detailed architecture of our LDI and fusion networks, where C_{in}, C_{out} indicate the channel of input and output tensors, respectively.

2. Network Details

The detailed architecture of our LDI and fusion networks is displayed in Fig. 2. We set $C_{in} = 32$, $C_{out} = 1$ for the LDI network and $C_{in} = 5$, $C_{out} = 1$ for the fusion network, where the input tensor of fusion network is obtained by concatenating $L_i(f), L_{i+1}(f), L^i_{i+1}(f), E(f,\mathcal{T}_i), E(f,\mathcal{T}_{i+1})$. Since the double integral of events satisfies $E(f,\mathcal{T}) > 0$ according to the physical model Eq. (3), we also apply the activation composed of $\mathrm{Sigmoid}(\cdot)$ and $\mathrm{ReLU}(\cdot)$ to the estimated $E(f,\mathcal{T})$ in our network before further processing, *i.e.*,

$$E(f,\mathcal{T}) := \mathrm{Sigmoid}(E(f,\mathcal{T})) + \mathrm{ReLU}(E(f,\mathcal{T})). \tag{8}$$

For color image processing, we modify the network by setting $C_{out} = 3$ in the LDI network and $C_{in} = 15$, $C_{out} = 3$ in the fusion network, where the total network parameters increase to 0.396M.

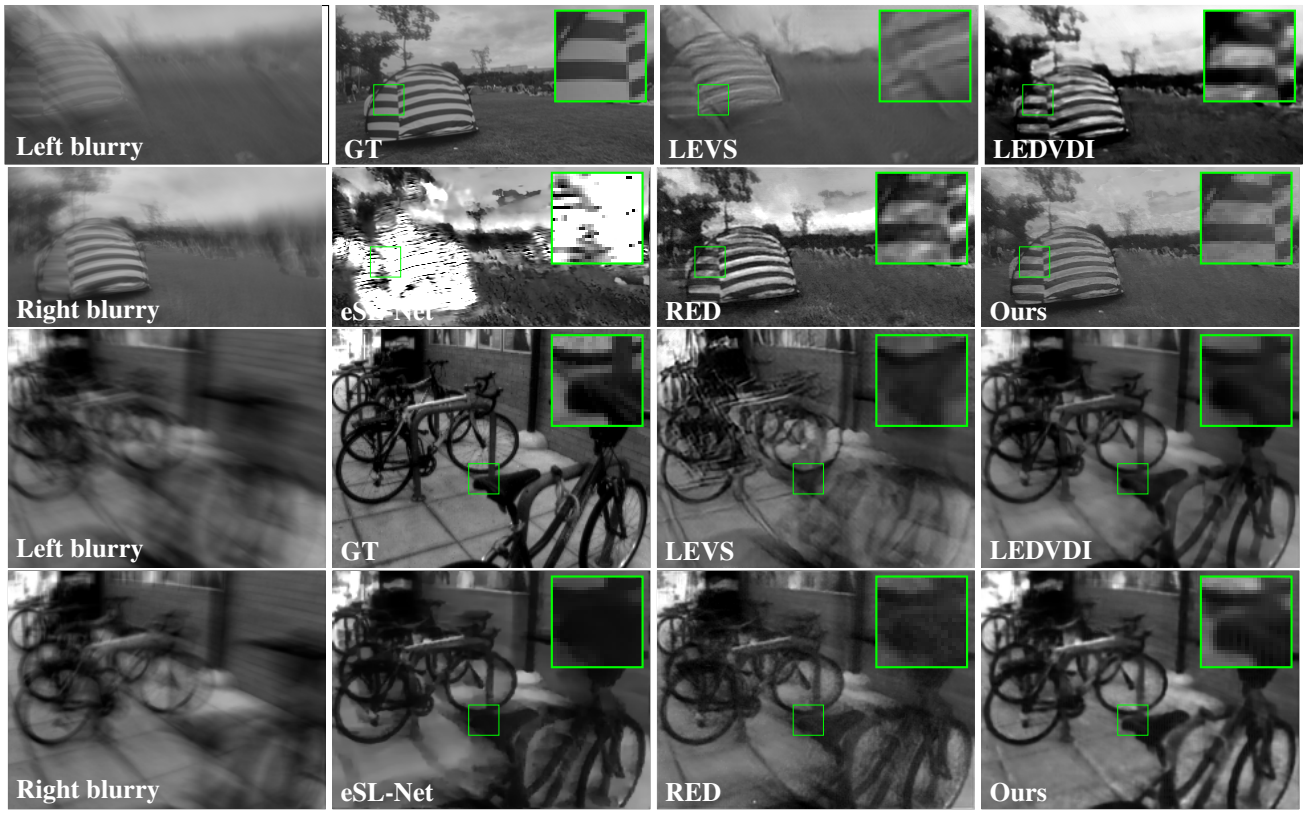

Figure 3. Qualitative comparisons of the deblurring task on the GoPro (row 1-2) and HQF (row 3-4) datasets. Details are zoomed in for a better view. Ground truth (GT) images are also provided as reference.

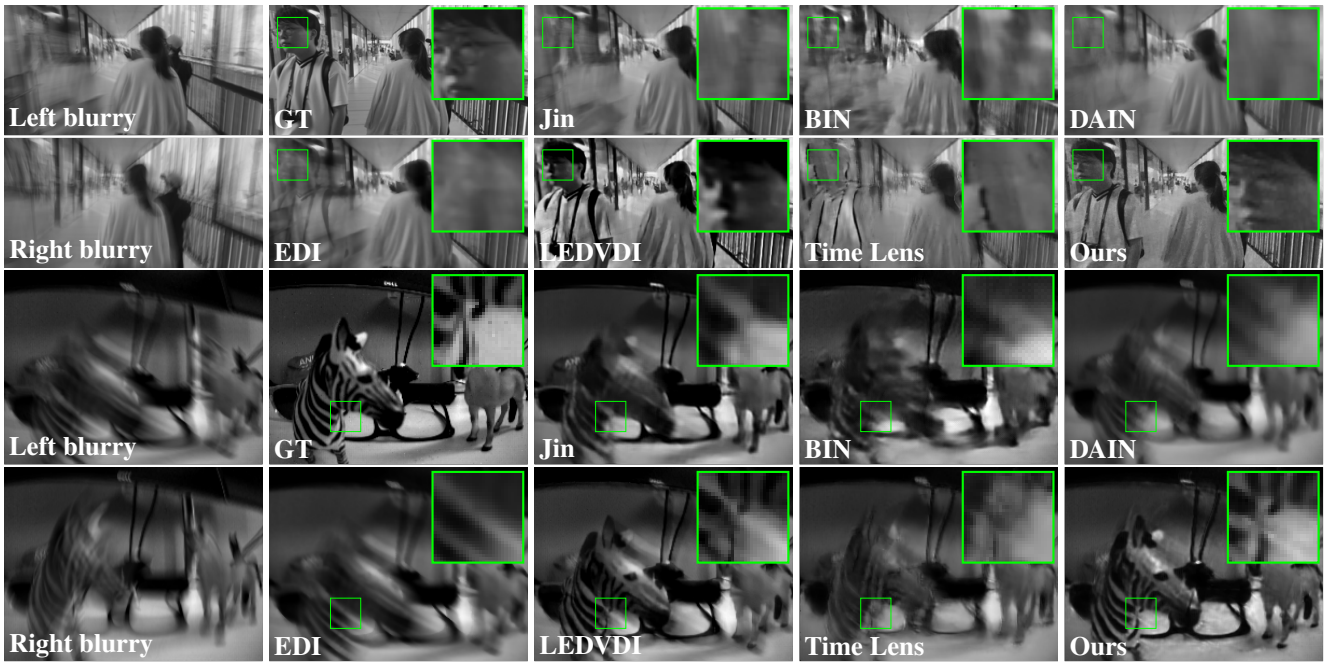

Figure 4. Qualitative comparisons of the interpolation task on the GoPro (row 1-2) and HQF (row 3-4) datasets. Details are zoomed in for a better view. Ground truth (GT) images are also provided as reference.

3. Additional Experiment Results

We provide more qualitative results to verify the effectiveness of EVDI. In the deblurring task, the frame-based method LEVS often suffers from motion ambiguity under highly dynamic scenes, resulting in incorrect reconstruction as shown in Fig. 3. Although the event-based methods, *e.g.*, LEDVDI and RED, are able to produce accurate latent images using the precise motion inside events, their results are often degraded by halo artifacts or noises due to data inconsistency. For the interpolation task, all the frame-based methods struggle to recover sharp textures under fast or non-linear motions, as displayed in Fig. 4. Among event-based methods, LEDVDI outperforms EDI by employing neural networks to learn better deblurring features, and beats Time Lens via designing a pre-deblurring stage to handle the motion blur in reference frames. However, LEDVDI is developed within a supervised learning framework and often meets data inconsistency when inferring on different datasets, *e.g.*, the brightness inconsistency shown in Fig. 4. The proposed EVDI method tackles the inconsistency issue by exploiting the self-supervised learning framework, and produces results with natural visual effects on both deblurring and interpolation tasks.