Video Interpolation by Event-driven Anisotropic Adjustment of Optical Flow

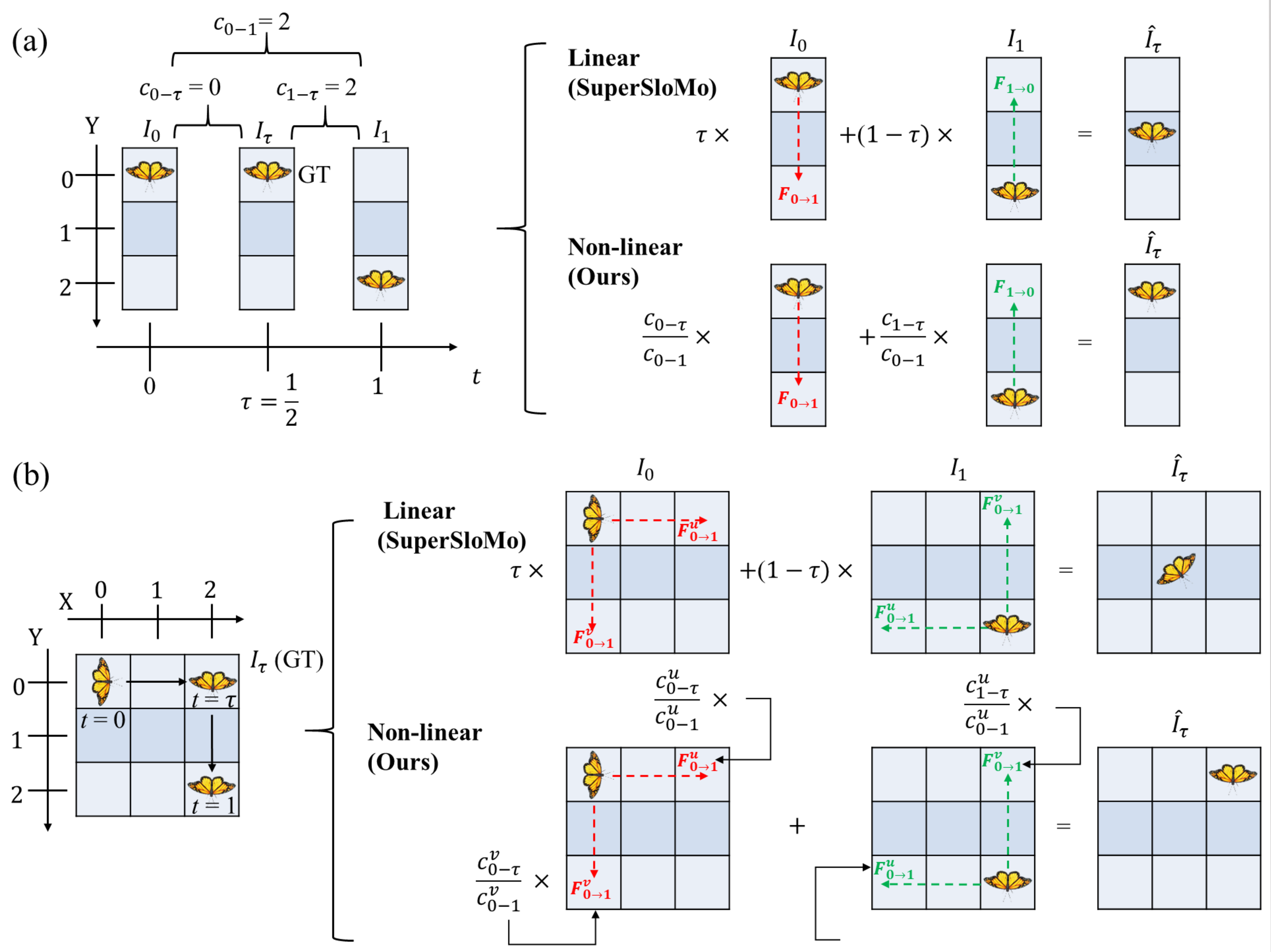
Song Wu*, Kaichao You*, Weihua He**, Chen Yang, Yang Tian, Yaoyuan Wang, Ziyang Zhang**, Jianxing Liao

ECCV TEL AVIV 2022

Motivations

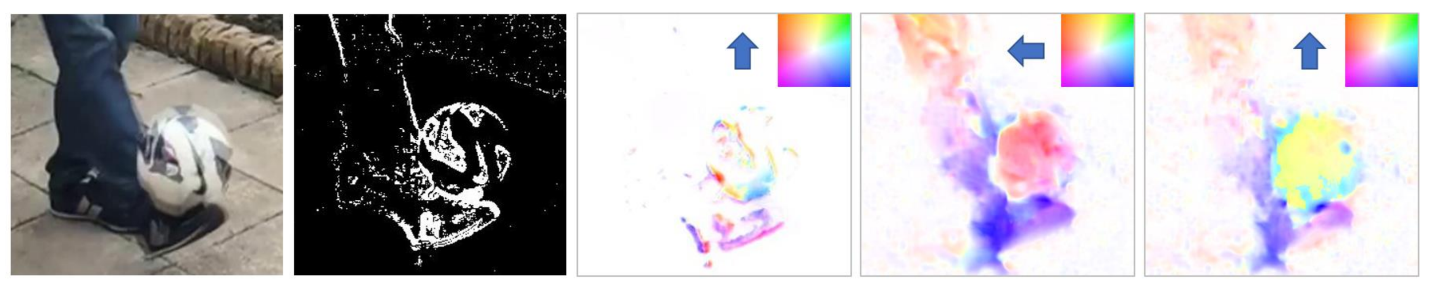
Optical flow is a common tool in VFI. Previous methods often calculate the bi-directional optical flows and then predict the intermediate optical flows under the linear motion assumptions, leading to isotropic intermediate flow generation.

A toy example



We take SuperSloMo as an example model. I_0 records a butterfly locates in Y_0, while I_2 records a butterfly locates in Y_2. For SuperSloMo model, the synthesized frame I_τ shows the butterfly locates in Y_τ when $\tau = \frac{1}{2}$. However, the actual coordinates of the butterfly are at Y_0. For the 2d motion, the butterfly is predicted at location $Y_{(1,1)}$ while the actual coordinates of the butterfly are at $Y_{(0,2)}$.

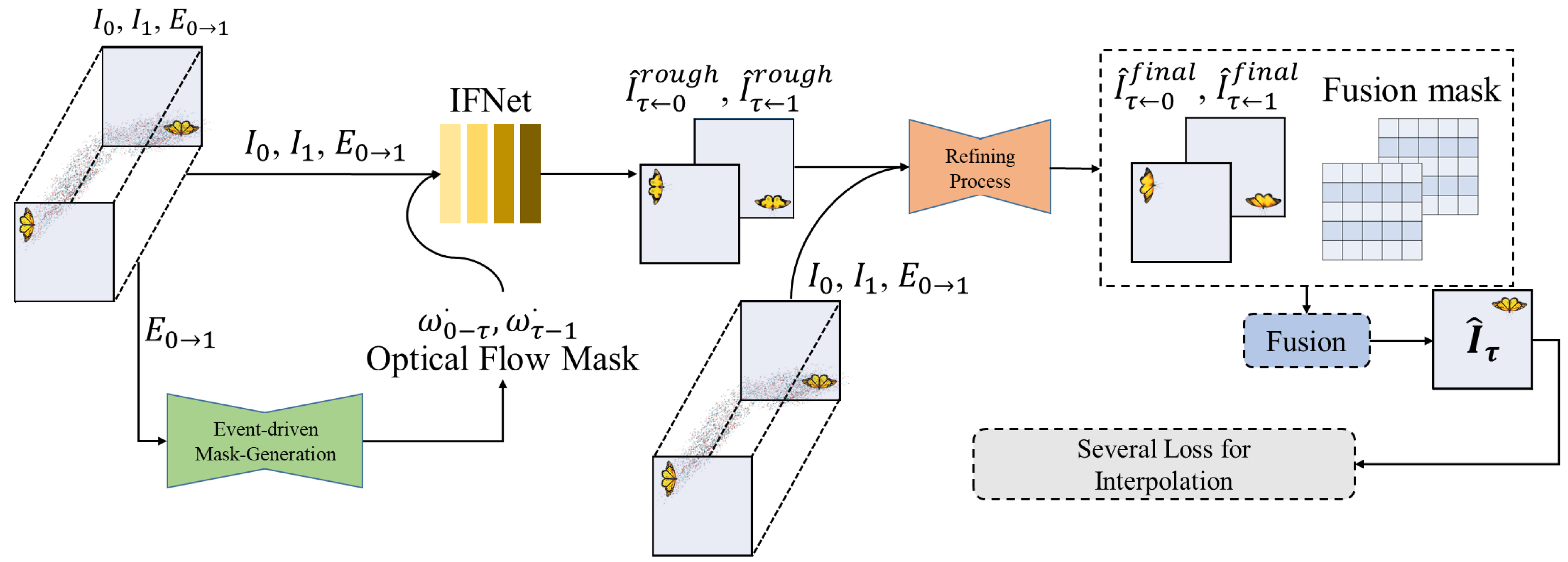
The reason for this phenomenon is that they synthesize the intermediate optical flow in an isotropic way. So we introduce the event data, which record the real motion information, to adjust optical flow in an anisotropic way.



(a)Overlap of Inputs and GT (b)Events (c)Sparse Flow Field (d)Linear Flow Field (e)Adjusted Flow Field

The visualization of the generated optical flow.

Model Architecture



The pipeline of our proposed event-driven Anisotropic Adjustment of Optical Flows(A^2OF).

The event is input to Event-driven Mask Generation network to get optical flow masks $\omega_{0-\tau}$ and $\omega_{\tau-1}$. We modify the structure of the IFNet so that it can input event and use anisotropic flow mask. Two consecutive frames, corresponding event and optical flow mask are input to tailored IFNet to get intermediate optical flows and warp the original frames. The pre-warped frames $\hat{I}^{rought}_{\tau\to 0}$ and $\hat{I}^{rought}_{\tau\to 1}$ are fused to final intermediate frame $\hat{I}_\tau$ under the supervision of a series of loss.

Event-driven Optical Flow Mask

In order to better analyze the motion in two-dimensional space and model the motion between two frames, we calculate different weights in the orthogonal directions. The final optical flow in the horizontal direction with the bi-directional optical flow $F^u_{0\to 1}$ and $F^u_{1\to 0}$ can be represented as:

$$F^u_{\tau\to 0} = -(1-\tau)\cdot\omega^u_{0-\tau}\cdot F^u_{0\to 1} + \tau\cdot\omega^u_{1-\tau}\cdot F^u_{1\to 0}$$
$$F^u_{\tau\to 0} = (1-\tau)\cdot\omega^u_{1-\tau}\cdot F^u_{0\to 1} - \tau\cdot\omega^u_{0-\tau}\cdot F^u_{1\to 0},$$

where $\omega^u_{0-\tau}$ and $\omega^u_{1-\tau}$ denote the event-driven optical flow mask for anisotropic optical flow adjustment. This optical flow mask can be calculated with CNNs-processed event counts according to

$$\omega^u_{0-\tau} = \frac{c^u_{0-\tau}}{c^u_{0-1}} \; and \; \omega_{1-\tau} = \frac{c^u_{1-\tau}}{c^u_{0-1}}.$$

We can also obtain intermediate optical flow in the vertical direction through the above equations.

Event-driven motion consistency loss

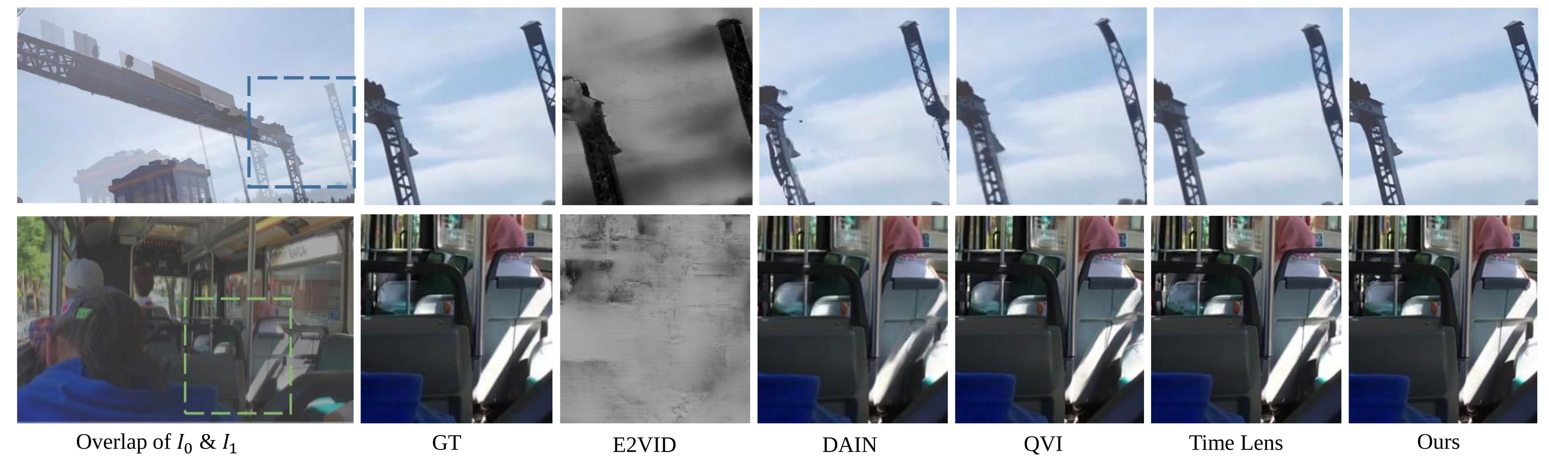
We design an extra event-driven loss based on the change of intensity.

$$\hat{E}^{count}_{0-\tau} = [\mathrm{sgn}\left(I_{diff} - N_{t_{positive}}\right), \mathrm{sgn}(N_{t_{negative}} - I_{diff})]$$

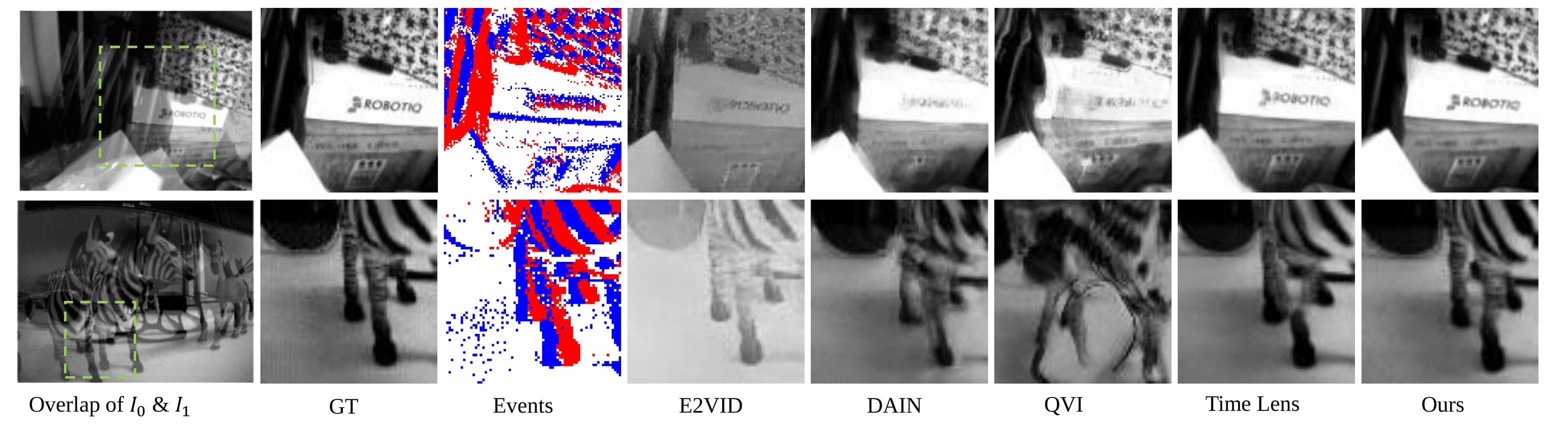
$$\mathcal{L}_{mc} = L_1\left(G\left(\hat{E}^{count}_{0\to\tau}, E^{count}_{0\to\tau}\right)\right) + L_1(G(\hat{E}^{count}_{\tau\to 1}, E^{count}_{\tau\to 1}))$$

where $G(\cdot)$ is denote Gaussian Blur Function.

Visual Results



Visual comparison on the Adobe240 datasets with synthetic events.



Visual comparison on the HQF dataset with real events.

Quantitative Results

Adobe240		All frames in 7 skips			Middle frame in 7 skips		
Method	Input	PSNR	SSIM	IE	PSNR	SSIM	IE
E2VID [23]	Event	10.40	0.570	75.21	10.32	0.573	76.01
SepConv	RGB	32.31	0.930	7.59	31.07	0.912	8.78
DAIN	RGB	32.08	0.928	7.51	30.31	0.908	8.94
SuperSloMo	RGB	31.05	0.921	8.19	29.49	0.900	9.68
QVI	RGB	32.87	0.939	6.93	31.89	0.925	7.57
Time Lens	RGB+E	35.47	0.954	5.92	34.83	0.949	6.53
Ours(A^2OF)	RGB+E	**36.59**	**0.960**	**5.58**	**36.21**	**0.957**	**5.96**
GoPro		All frames in 7 skips			Middle frame in 7 skips		
Method	Input	PSNR	SSIM	IE	PSNR	SSIM	IE
E2VID	Event	9.74	0.549	79.49	9.88	0.569	80.08
SepConv	RGB	29.81	0.913	8.87	28.12	0.887	10.78
DAIN	RGB	30.92	0.901	8.60	28.82	0.863	10.71
SuperSloMo	RGB	29.54	0.880	9.36	27.63	0.840	11.47
QVI	RGB	31.39	0.931	7.09	29.84	0.911	8.57
Time Lens	RGB+E	34.81	0.959	5.19	34.45	0.951	5.42
Ours(A^2OF)	RGB+E	**36.61**	**0.971**	**4.23**	**35.95**	**0.967**	**4.62**
Middlebury (other)		All frames in 3 skips			Middle frame in 3 skips		
Method	Input	PSNR	SSIM	IE	PSNR	SSIM	IE
E2VID	Event	11.26	0.427	69.73	11.12	0.407	70.35
SepConv	RGB	25.51	0.824	6.74	25.12	0.811	7.06
DAIN	RGB	26.67	0.838	6.17	25.96	0.793	6.54
SuperSloMo	RGB	26.14	0.825	6.33	25.53	0.805	6.85
QVI	RGB	26.31	0.827	6.58	25.72	0.798	6.73
Time Lens	RGB+E	32.13	0.908	4.07	31.57	0.893	4.62
Ours(A^2OF)	RGB+E	**32.59**	**0.916**	**3.92**	**31.81**	**0.903**	**4.13**

HQF		3 skips		1 skip	
Method	Input	PSNR	SSIM	PSNR	SSIM
E2VID	Event	6.70	0.315	6.70	0.315
RRIN	RGB	26.11	0.778	29.76	0.874
BMBC	RGB	26.32	0.781	29.96	0.875
DAIN	RGB	26.10	0.782	29.82	0.875
SuperSloMo	RGB	25.54	0.761	28.76	0.861
Time Lens	RGB+E	30.57	0.900	32.49	0.927
Ours(A^2OF)	RGB+E	**31.85**	**0.932**	**33.94**	**0.945**

Song Wu: wusong5533@gmail.com;
Kaichao You: ykc20@mails.tsinghua.edu.cn;
Weihua He: hwh20@mails.tsinghua.edu.cn
ZiyangZhang: zhangziyang11@huawei.com