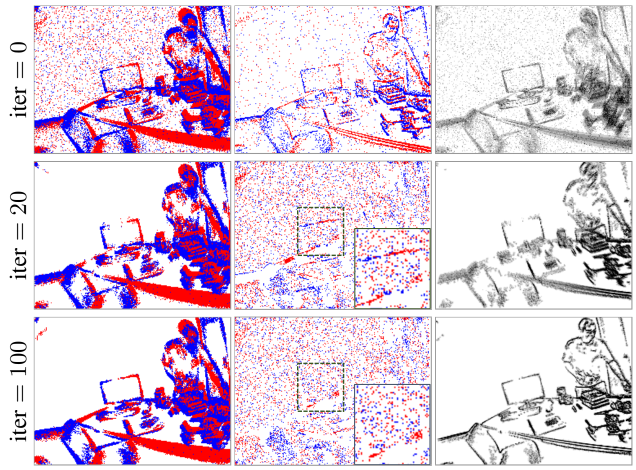
7. Supplementary

7.1. Video

We encourage readers to inspect the attached video, which summarizes the method and results.



(a) Signal (b) Noise (c) IWE

Figure 11. Evolutions of signal, noise, and motion (IWE) during optimization. The edge structure (e.g., green boxes in (b) Noise) converges to move to signal events, while CMax converges to the sharp IWE (i.e., expected motion parameters).

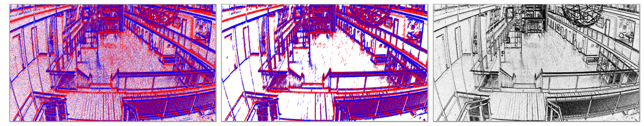
7.2. Experiment details

Hyper-parameters. For the denoising experiments, we test various values $\tau = \{0.9, \dots, 0.1\}$ to calculate the ROC on DND21 and $\tau = \{0.9, \dots, 0.7\}$ for the RMS on ECD. For E-MLB benchmarking, we fix the number of signal events to follow the prior work [10]. To analyze the proposed pipeline in the CMax framework, we use model-based rotational motion estimation [15] and tile-based optical flow estimation [48] approaches. We use the magnitude of the IWE gradient [18] as the CMax objective function.

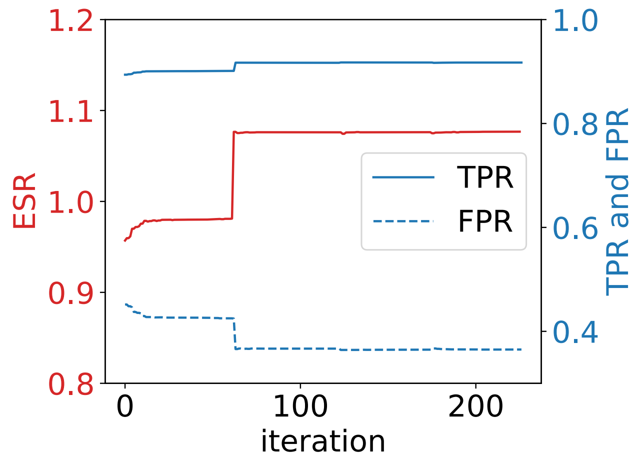
7.3. Denoising Convergence During Optimization

To further validate the proposed joint estimation approach, we analyze the convergence during the joint estimation using the ECD dataset in Fig. 11 (see also results in Fig. 5). In the first iteration (i.e., *initialization*), signal and noise events are randomly classified. As optimization proceeds, signal events evolve towards keeping edge structures in the scene, while noise events evolve towards dropping such edge structures (see second and third rows). Also, the IWEs converge to sharp edges with correct motion parameters. This example confirms the efficacy of joint estimation.

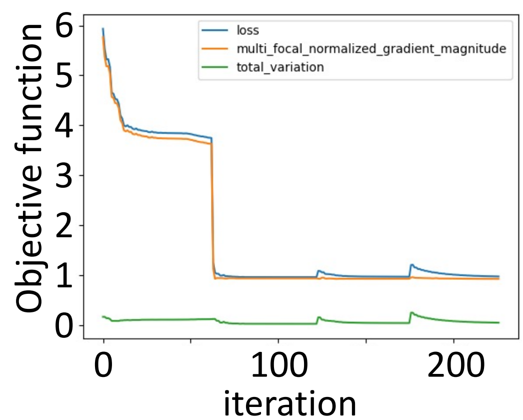
The optimization process on an HD-resolution (1280×720 px) real-world dataset, TUM-VIE [33], is shown in



(a) Signal (first iter) (b) Signal (last iter) (c) IWE (last iter)



(d) Evolution of denoising metrics



(e) Evolution of loss function

Figure 12. Results on HD real-world data and intermediate denoising values (TPR, FPR and ESR metrics) during optimization.

Fig. 12. The intermediate progress (d)–(e) demonstrates how motion estimation converges and the denoising performance improves, simultaneously.

7.4. Full Results on Denoising DND21 data

Table 5 is the full version of Tab. 2 in the main paper (including added noise at rates of 3 and 7 Hz).

7.5. Full Results on Angular Velocity Estimation

While the quantitative evaluation on angular velocity estimation is summarized in Fig. 6, here, we report the detailed results with different target ratio parameters, also compared with other baselines such as BA Filter [9]. The original CMax degrades due to noise, as reported in previous work

		1Hz		3Hz		5Hz		7Hz		10Hz	
		hotel	*driving*	*hotel*	*driving*	*hotel*	*driving*	*hotel*	*driving*	*hotel*	*driving*
Model-based	BAF [9]	0.954	0.848	0.920	0.816	0.892	0.793	0.866	0.773	0.837	0.748
	TS [34]	<u>0.972</u>	<u>0.931</u>	**0.972**	**0.926**	<u>0.961</u>	**0.927**	**0.965**	**0.924**	**0.962**	**0.920**
	KNoise [31]	0.677	0.630	0.652	0.623	0.670	0.624	0.658	0.616	0.641	0.614
	Ynoise [14]	0.969	**0.941**	0.952	<u>0.924</u>	0.923	<u>0.909</u>	0.918	<u>0.897</u>	0.899	<u>0.880</u>
	DWF [24]	0.927	0.741	0.893	0.710	0.862	0.690	0.834	0.675	0.796	0.656
	Ours	**1.014**	0.882	<u>0.968</u>	0.851	**0.963**	0.855	<u>0.951</u>	0.847	<u>0.961</u>	0.836
Learning	EDnCNN [4]	0.957	0.887	0.937	0.877	0.937	0.875	0.925	0.865	0.901	0.874
	MLPF [24]	<u>0.970</u>	<u>0.889</u>	<u>0.972</u>	<u>0.887</u>	<u>0.970</u>	<u>0.885</u>	<u>0.969</u>	<u>0.882</u>	<u>0.963</u>	<u>0.876</u>
	EDformer [29]	**0.993**	**0.954**	**0.989**	**0.947**	**0.985**	**0.942**	**0.979**	**0.934**	**0.970**	**0.926**

Table 5. The AUC↑ of ROC on the two DND21 sequences (hotel and driving) at different noise rates.

		dynamic_rot		*boxes_rot*	
		RMS ↓	FWL ↑	RMS ↓	FWL ↑
5 Hz Noise	CMax [17]	20.001	1.259	124.641	1.129
	– w/ Init.	8.275	1.274	20.659	1.214
	Downsampling 90%	8.808	1.273	124.869	1.108
	– w/ Init.	8.226	1.274	20.619	<u>1.214</u>
	Downsampling 80%	9.965	1.271	161.811	1.061
	– w/ Init.	8.244	1.274	20.798	1.214
	Downsampling 70%	14.399	1.262	184.511	1.030
	– w/ Init.	8.231	1.274	23.679	1.211
	Ours 90%	8.522	1.273	97.585	1.141
	– w/ Init.	<u>8.189</u>	<u>1.274</u>	<u>20.604</u>	**1.214**
	Ours 80%	8.511	1.273	131.356	1.103
	– w/ Init.	**8.170**	**1.274**	20.862	1.214
	Ours 70%	9.180	1.272	163.676	1.062
	– w/ Init.	8.086	1.274	21.151	1.214
	BAF	19.675	1.260	125.028	1.127
	– w/ Init.	8.253	1.274	**19.550**	1.214
1 Hz	CMax [17]	19.395	1.276	117.440	1.144
	– w/ Init.	8.254	1.290	20.628	1.223
	Downsampling 90%	8.676	1.289	110.568	1.130
	– w/ Init.	<u>8.184</u>	1.290	<u>20.620</u>	<u>1.223</u>
	Ours 90%	8.506	**1.290**	87.775	1.159
	– w/ Init.	**8.177**	<u>1.290</u>	**20.569**	**1.223**
	BAF	19.569	1.276	117.554	1.143
	– w/ Init.	8.189	1.290	20.713	1.223

Table 6. Angular velocity estimation on ECD dataset [40].

(e.g., [3]). The signal-to-noise (S/N) target ratio τ affects accuracy: the closer it is to the actual value of noise injection, the better the results of the proposed method. The amount of artificial noise injected is around 15 % for 5 Hz and 3 % for 1 Hz conditions. Although the "true" noise level is unknown due to the original noise in the ECD sequences, our method constantly produces better accuracy and FWL values than the baselines. Please refer to Sec. 4.3 for more discussions about the dependency on initialization and comparison with other baselines. The AUCs for the conditions that we test ($\tau = \{0.9, \dots, 0.7\}$) are 0.70 ("Ours") and 0.67 ("Downsampling").

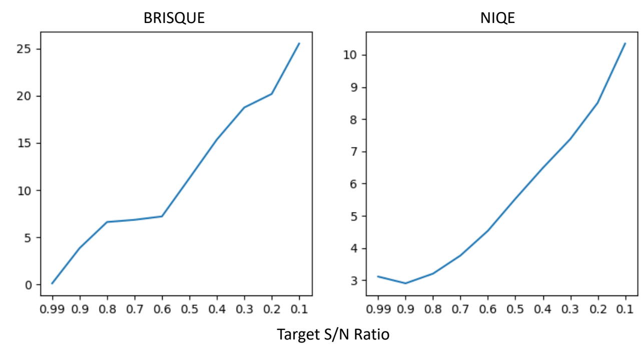


Figure 13. Results of the non-reference image quality indices for image reconstruction. See Fig. 8 for details.

7.6. Quantitative Evaluation of Intensity Reconstruction

In Secs. 4.4 and 4.5 we show qualitative results of the intensity reconstruction application. Here, we discuss possible quantitative evaluation. The challenge of the quantitative evaluation lies in the quality of reference frames (i.e., "GT") in the existing dataset as shown in Figs. 2 and 7: the frames become underexposed or blurry due to their limited dynamic range, when event data suffer from more BA noise (i.e., in dark scenes).

Nonetheless, we report non-reference image quality indices for different S/N ratios τ. Figure 13 reports the scores of Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [38] and Naturalness Image Quality Evaluator (NIQE) [39], using *Bicycle-ND64-2* sequence (same as Fig. 8). These scores indicate the perceptual quality of images, and smaller is better. Although BRISQUE monotonically increases as the target ratio decreases (i.e., more events are removed), NIQE scores the lowest at $\tau = 0.9$, indicating the best quality of the reconstructed image. Although the results potentially suggest that it could estimate the "true" noise ratio in the data using the non-reference indices, which is useful for image reconstruction applications, we leave further evaluation and discussion as future work.