Supplementary Material – Taming Contrast Maximization for Learning Sequential, Low-latency, Event-based Optical Flow

Federico Paredes-Vallés[1,2] Kirk Y. W. Scheper[2] Christope De Wagter[1] Guido C. H. E. de Croon[1]

[1] Micro Air Vehicle Laboratory, Delft University of Technology

[2] Stuttgart Laboratory 1, Sony Semiconductor Solutions Europe, Sony Europe B.V.

A. DSEC-Flow: Sequence selection

The contrast maximization framework for motion compensation assumes constant illumination [6, 7]. Under this assumption, all the events captured with an event camera are generated by the apparent motion of objects in the image space. However, the constant illumination assumption is often violated in sequences recorded at night because the main source of light in these environments comes from flashing, artificial lights (e.g., street lamps). In addition, the signal-to-noise ratio of these sensors decreases under low light conditions, which means that a large percentage of the captured events are not triggered by motion but by sensor noise [5]. For these reasons, we remove any sequence recorded at night from the original DSEC-Flow training dataset [8, 9]. Specifically, we use all the sequences in the training dataset, from beginning to end, except for variants of those listed in Table S1.

zurich_city_00_*	zurich_city_01_*
zurich_city_09_*	zurich_city_10_*

Table S1: Sequences from the DSEC-Flow training dataset [8, 9] that were not used for the training of our models.

B. Additional results

B.1. Breakdown of DSEC-Flow results

A breakdown of the quantitative results on DSEC-Flow can be found in Table S3. Additionally, Fig. S1 shows the IWEs that correspond to the input events and pixel displacement predictions in Fig. 7 and discussed in Section 4.2.

B.2. Breakdown of MVSEC results

Fig. S3 shows a qualitative comparison of our self-supervised learning (SSL) method with state-of-the-art techniques on the outdoor_day1 sequence from MVSEC. As described in Section 4.2, for this evaluation we used all the events in between samples of the temporally-upsampled ground-truth data (provided at 45 Hz) as input for every forward pass. These results support the conclusions derived from Table 3.

	EPE↓	$\%_{3PE}$↓
dt = 0.1s*	3.48	34.72
dt = 0.05s*	3.24	32.45
dt = 0.01s*	15.85	90.52
dt = 0.005s*	28.45	97.27
dt = 0.002s*	15.28	94.79
dt = 0.05s	3.09	27.36
dt = 0.01s	**2.33**	**17.77**
dt = 0.005s	<u>2.34</u>	<u>17.92</u>
dt = 0.002s	2.66	21.83

Table S2: Quantitative evaluation of the impact of sequential processing on DSEC-Flow [9]. Best in bold, runner up underlined. ∗: Non-recurrent, volumetric event representation with 10 bins.

For completeness, a breakdown of the quantitative results on all MVSEC evaluation sequences can be found in Table S4. This includes not just the outdoor_day1 scenario (as discussed in the main text), but also the three indoor sequences. These indoor sequences present notably different statistics compared to the automotive dataset used for training our models [9], as they were recorded with a drone operating in an indoor environment. On these sequences, our transferred model demonstrates (on average) an improvement of $\downarrow 25\%$ in endpoint error (EPE, lower is better, $\downarrow$) compared to the architecturally-equivalent ConvGRU-EV-FlowNet model from Hagenaars *et al.* [10], while showing an error increase of $\uparrow 30\%$ compared to Shiba et al. [15]. However, note that the latter method is not learning-based, so it is not subject to generalization issues besides those inherent to contrast maximization (see main text). Lastly, it's worth noting that the benchmarking of event-based optical flow solutions is shifting from MVSEC [18] to the DSEC dataset [9] due to calibration issues [9] and the lack of a standardized training dataset in the former [3, 9–12, 15–17].

		All				interlaken_00_b				interlaken_01_a				thun_01_a			
		EPE↓	%$_{3PE}$↓	FWL↑	RSAT↓	EPE↓	%$_{3PE}$↓	FWL↑	RSAT↓	EPE↓	%$_{3PE}$↓	FWL↑	RSAT↓	EPE↓	%$_{3PE}$↓	FWL↑	RSAT↓
SL	E-RAFT [9]	0.79	2.68	1.33	0.87	1.39	6.19	1.42	0.91	0.90	3.91	1.56	0.85	0.65	1.87	1.30	0.88
SL	EV-FlowNet, Gehrig *et al.* [9]	2.32	18.60	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SL	IDNet [17]	**0.72**	**2.04**	-	-	**1.25**	**4.35**	-	-	**0.77**	**2.60**	-	-	**0.57**	**1.47**	-	-
SL	TIDNet [17]	0.84	2.80	-	-	1.43	6.30	-	-	0.93	3.50	-	-	0.73	2.60	-	-
SL	TMA [12]	0.74	2.30	-	-	1.39	5.79	-	-	0.81	3.11	-	-	0.62	1.61	-	-
SL	Cuadrado *et al.* [3]	1.71	10.31	-	-	3.07	23.51	-	-	1.90	14.93	-	-	1.36	5.93	-	-
SL	E-Flowformer [11]	0.76	2.45	-	-	1.38	6.05	-	-	0.86	3.31	-	-	0.60	1.60	-	-
SSL_E	EV-FlowNet* [20]	3.86	31.45	1.30	**0.85**	6.32	47.95	1.46	**0.85**	4.91	36.07	1.42	**0.81**	2.33	20.92	1.32	**0.85**
SSL_E	ConvGRU-EV-FlowNet* [10]	4.27	33.27	1.55	0.90	6.78	46.77	1.74	0.92	5.21	32.26	1.92	0.88	2.15	17.50	1.49	0.91
SSL_E	dt = 0.01s, $R = 2$, $S = 1$ (Ours)	9.66	86.44	**1.91**	1.07	9.86	87.24	**1.89**	1.08	9.33	86.70	**2.07**	1.01	8.71	86.45	**1.81**	1.11
SSL_E	dt = 0.01s, $R = 5$, $S = 1$ (Ours)	4.05	52.22	1.58	0.97	5.18	61.14	1.62	0.98	3.86	53.19	1.91	0.91	3.34	44.37	1.46	0.99
SSL_E	dt = 0.01s, $R = 10$, $S = 1$ (Ours)	2.33	17.77	1.26	0.88	3.34	25.72	1.33	0.90	2.49	19.15	1.40	0.83	1.73	10.39	1.21	0.89
SSL_E	dt = 0.01s, $R = 20$, $S = 1$ (Ours)	16.63	33.67	1.06	1.10	30.98	46.26	0.86	1.18	7.43	33.94	1.17	0.98	14.41	21.36	0.74	1.20
SSL_E	dt = 0.01s, $R = 10$, $S = 3$ (Ours)	2.82	27.09	1.37	0.92	3.79	37.48	1.44	0.94	2.83	28.14	1.66	0.86	2.13	18.40	1.29	0.93
SSL_E	dt = 0.01s, $R = 20$, $S = 4$ (Ours)	2.73	23.73	1.24	0.90	3.31	29.97	1.34	0.91	2.73	25.92	1.45	0.86	1.81	13.19	1.21	0.92
MB	Shiba *et al.* [15]	3.47	30.86	1.37	0.89	5.74	38.93	1.46	0.90	3.74	31.37	1.63	0.88	2.12	17.68	1.32	0.89

		thun_01_b				zurich_city_12_a				zurich_city_14_c				zurich_city_15_a			
		EPE↓	%$_{3PE}$↓	FWL↑	RSAT↓	EPE↓	%$_{3PE}$↓	FWL↑	RSAT↓	EPE↓	%$_{3PE}$↓	FWL↑	RSAT↓	EPE↓	%$_{3PE}$↓	FWL↑	RSAT↓
SL	E-RAFT [9]	0.58	1.52	1.25	0.89	0.61	1.06	0.91	**0.93**	0.71	**1.91**	1.47	0.83	0.59	1.30	1.40	0.84
SL	EV-FlowNet, Gehrig *et al.* [9]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SL	IDNet [17]	**0.55**	1.35	-	-	0.60	1.16	-	-	0.76	2.74	-	-	**0.55**	**1.02**	-	-
SL	TIDNet [17]	0.65	1.70	-	-	0.67	1.30	-	-	0.80	4.60	-	-	0.65	1.30	-	-
SL	TMA [12]	**0.55**	**1.31**	-	-	**0.57**	**0.87**	-	-	**0.66**	1.99	-	-	**0.55**	1.08	-	-
SL	Cuadrado *et al.* [3]	1.41	6.38	-	-	1.24	3.85	-	-	1.58	9.96	-	-	1.24	4.89	-	-
SL	E-Flowformer [11]	0.57	1.50	-	-	0.58	0.91	-	-	0.67	2.09	-	-	0.56	1.15	-	-
SSL_E	EV-FlowNet* [20]	3.04	25.41	1.33	**0.87**	2.62	25.80	1.03	0.94	3.36	36.34	1.24	**0.82**	2.97	25.53	1.33	**0.82**
SSL_E	ConvGRU-EV-FlowNet* [10]	3.25	25.31	1.51	0.92	3.67	40.15	0.97	**0.93**	3.47	40.98	1.60	0.87	3.21	27.99	1.61	0.89
SSL_E	dt = 0.01s, $R = 2$, $S = 1$ (Ours)	9.38	86.68	**1.66**	1.08	11.54	85.35	1.40	1.10	10.18	86.39	**2.50**	1.03	8.54	86.30	**2.01**	1.06
SSL_E	dt = 0.01s, $R = 5$, $S = 1$ (Ours)	3.51	47.33	**1.66**	1.00	4.76	51.82	1.14	0.99	4.23	57.26	1.72	0.93	3.42	50.40	1.54	0.96
SSL_E	dt = 0.01s, $R = 10$, $S = 1$ (Ours)	1.66	9.34	1.25	0.91	2.72	26.65	1.04	0.94	2.64	23.01	1.38	0.85	1.69	9.98	1.23	0.86
SSL_E	dt = 0.01s, $R = 20$, $S = 1$ (Ours)	9.09	22.53	1.09	1.08	27.19	44.78	**1.49**	1.17	20.06	41.65	1.16	1.05	15.97	25.16	0.90	1.05
SSL_E	dt = 0.01s, $R = 10$, $S = 3$ (Ours)	2.03	17.19	1.40	0.94	3.53	33.77	1.08	0.97	2.95	32.75	1.43	0.88	2.26	21.95	1.29	0.89
SSL_E	dt = 0.01s, $R = 20$, $S = 4$ (Ours)	1.85	13.71	1.21	0.93	4.19	35.65	0.91	0.95	2.53	27.97	1.32	0.85	1.93	15.50	1.24	0.87
MB	Shiba *et al.* [15]	2.48	23.56	1.28	0.89	3.86	43.96	1.08	0.95	2.72	30.53	1.44	0.85	2.35	20.99	1.39	0.87

*Retrained by us on DSEC-Flow, linear warping.

Table S3: Breakdown of the quantitative evaluation on the DSEC-Flow dataset [9]. Best in bold, runner up underlined. The results of our best performing model are highlighted in red. SL: supervised learning; SSL_F: SSL trained with grayscale images; SSL_E; SSL trained with events; MB: model-based methods.

B.3. Impact of sequential processing

Here, we study the impact of the proposed contrast maximization framework for sequential event-based optical flow estimation (i.e., short input partitions, longer training partitions; see Section 3) and compare it to the non-sequential pipeline from Zhu *et al.* [20] (i.e., input and training partitions are of the same length). To do this, we trained multiple models on DSEC-Flow with different dt_{input}, but with $dt_{train} = 0.1s$ for the sequential models and $dt_{train} = dt_{input}$ for the non-sequential. Quantitative results in Table S2 confirm the claims made in Section 3.1 about the fact that, for contrast maximization to be a robust supervisory signal, the training event partition used for the computation of the supervisory signal needs to contain enough motion information (i.e., blur) so it can be compensated for. As shown, non-sequential models converge to worse solutions the shorter the input window. On the other hand, our sequential pipeline allows us to shorten the input window without compromising the performance, as discussed in Section 4.2.

B.4. Linear vs. iterative event warping

Here, we examine the effect of the type of event warping (linear [10] vs. iterative) on the performance of the sequential, stateful architecture introduced in Section 3.4 when it is trained on the DSEC-Flow dataset. To do this, we trained four models in total: two variants (with and without image border compensation, see Section B.5) of ConvGRU-EV-FlowNet [10], which is trained with linear warping; and another two variants of the *same architecture*, but trained with the proposed iterative warping module. Quantitative and qualitative results are presented in Table S5 and Fig. S4, respectively. In both cases (with and without image-border compensation), the models trained with iterative warping (i.e., ours) outperform those trained with linear warping (EPE dropped by 28% without compensation, and 62% with it), despite using the same architecture. This is expected, as the iterative warping module is able to better capture the trajectory of scene points over time, as explained in Section 3.2. The impact of the image-border compensation mechanism is presented and discussed in Section B.5.

To support the arguments presented in Section 3.2 and Fig. 2 regarding the limitations of linear warping, we also conducted an experiment in which we deployed models

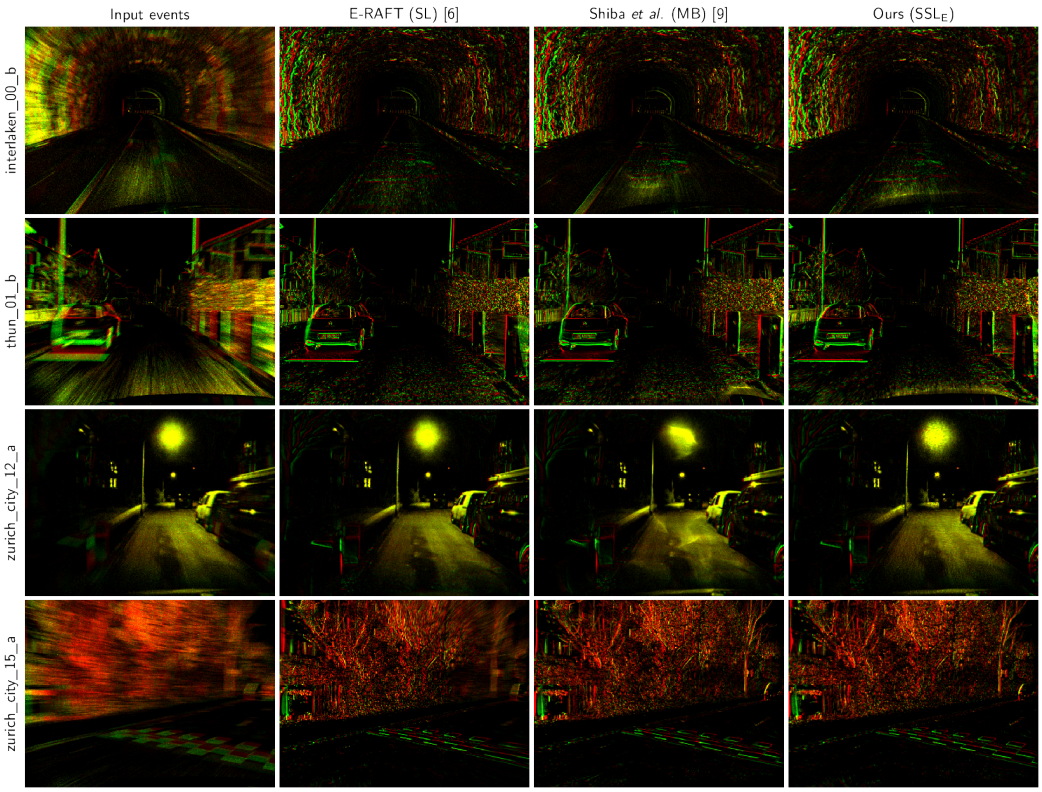
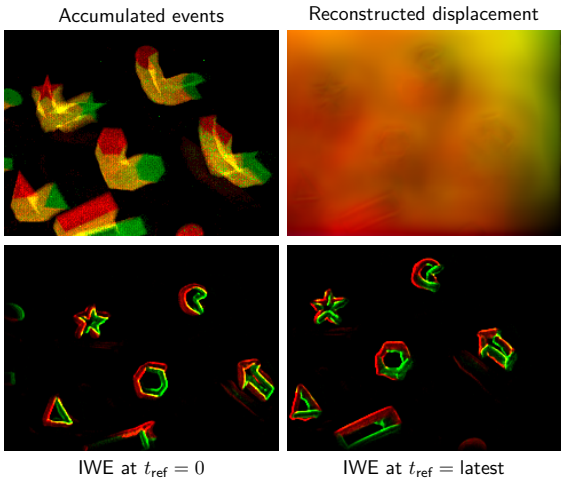


Figure S1: IWEs corresponding to the qualitative comparison of our method with the state-of-the-art E-RAFT architecture [9] and the model-based approach from Shiba *et al*. [15] on sequences from the test partition of the DSEC-Flow dataset [9] (see Fig. 7).

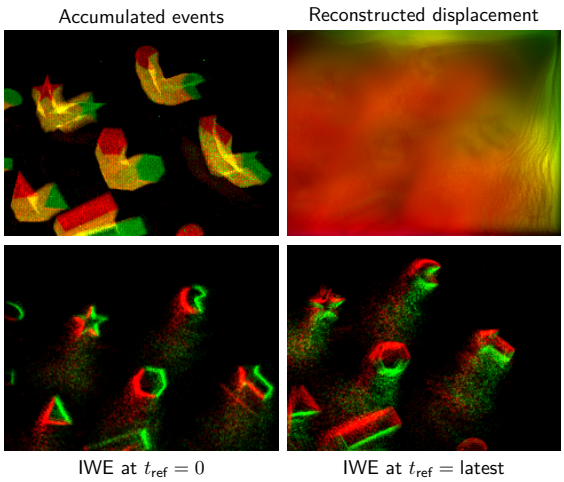
		outdoor_day1		indoor_flying1		indoor_flying2		indoor_flying3	
		EPE↓	%$_{3PE}$↓	EPE↓	%$_{3PE}$↓	EPE↓	%$_{3PE}$↓	EPE↓	%$_{3PE}$↓
SL	EV-FlowNet+ [16]	0.68	0.99	0.56	1.00	0.66	1.00	0.59	1.00
SL	E-RAFT [9]	**0.24**	1.70	-	-	-	-	-	-
SL	EV-FlowNet [9]	0.31	**0.00**	-	-	-	-	-	-
SL	TMA [12]	0.25	0.07	1.06	3.63	1.81	27.29	1.58	23.26
SL	Cuadrado *et al*. [3]	0.85	-	0.58	-	0.72	-	0.67	-
SSL$_F$	EV-FlowNet [19]	0.49	0.20	1.03	2.20	1.72	15.1	1.53	11.9
SSL$_F$	Ziluo *et al*. [4]	0.42	**0.00**	0.57	0.10	0.79	1.60	0.72	1.30
SSL$_E$	EV-FlowNet [20]	0.32	**0.00**	0.58	**0.00**	1.02	4.00	0.87	3.00
SSL$_E$	EV-FlowNet [14]	0.92	5.40	0.79	1.20	1.40	10.9	1.18	7.40
SSL$_E$	EV-FlowNet [15]	0.36	0.09	-	-	-	-	-	-
SSL$_E$	ConvGRU-EV-FlowNet [10]	0.47	0.25	0.60	0.51	1.17	8.06	0.93	5.64
SSL$_E$	Ours $dt = 0.005$	0.27	0.05	0.44	**0.00**	0.88	4.51	0.70	2.41
MB	Akolkar *et al*. [1]	2.75	-	1.52	-	1.59	-	1.89	-
MB	Brebion *et al*. [2]	0.53	0.20	0.52	0.10	0.98	5.50	0.71	2.10
MB	Shiba *et al*. [15]	0.30	0.11	**0.42**	0.09	**0.60**	**0.59**	**0.50**	**0.29**

Table S4: Quantitative evaluation on all MVSEC sequences [18]. Best in bold, runner up underlined. SL: supervised learning; SSL$_F$: SSL trained with grayscale images; SSL$_E$; SSL trained with events; MB: model-based methods.

trained on DSEC-Flow with linear and iterative warping on a sequence from the Event Camera Dataset [13] with strong nonlinearities in the trajectories of scene points. Note that this sequence, known as shapes_6dof, was recorded with a different event camera and that its statistics are significantly different from those of DSEC-Flow (i.e., hand-held camera looking at a planar scene [13] vs. automotive scenario [8]). Qualitative results are presented in Fig. S2. In addition to showing that the models generalize (to some extent) to this new sequence, these results demonstrate that only the models trained with iterative event warping are able to produce sharp IWEs at multiple reference times.



(a) Ours.



(b) ConvGRU-EV-FlowNet*† [10].

Figure S2: Qualitative results of the ablation study with respect to the type of event warping. Inference settings: dt $= 0.01$s, accumulation window of 0.5s. Both models were trained on the DSEC-Flow dataset [9], with dt $= 0.01$s, $R = 10$, $S = 1$. $*$: Retrained by us on DSEC-Flow [9]. †: Without border compensation. The optical flow color coding can be found in Fig. 2 (top).

B.5. Optical flow at the image borders

As discussed in Section 3.2, for a given temporal scale, we mask the events that are transported outside the image space at any time during the warping process from the computation of the loss to prevent learning incorrect optical flow at the image borders. Here we study the impact of this masking mechanism on the performance of not only the proposed SSL framework but also of two other literature methods: EV-FlowNet [20] and ConvGRU-EV-FlowNet [10]. For this experiment, we trained two versions of each model, one with and one without the proposed image-border compensation technique, on the DSEC-Flow dataset. Note that EV-FlowNet is a stateless model trained with a volumetric event representation with 10 bins, and hence processes all the input events in between ground-truth samples at once.

	EPE↓	$\%_{\text{3PE}}$↓
EV-FlowNet*† [20]	3.86	31.45
EV-FlowNet* [20]	3.48	34.72
ConvGRU-EV-FlowNet*† [10]	4.27	33.27
ConvGRU-EV-FlowNet* [10]	6.09	36.36
dt $= 0.01$s, $R = 10$, $S = 1^{\dagger}$ (Ours)	3.08	21.38
dt $= 0.01$s, $R = 10$, $S = 1$ (Ours)	2.33	17.77

*Retrained by us on DSEC-Flow, linear warping.
†Without border compensation.

Table S5: Quantitative results of the ablation study on the DSEC-Flow dataset [9] with respect to the effectiveness of the proposed warping module and image-border compensation mechanism.

Quantitative and qualitative results are presented in Table S5 and Fig. S4, respectively. These results highlight that, for both EV-FlowNet and our model, adding the proposed image-border compensation improves performance (EPE dropped by 10% and 24%, respectively). However, the performance degraded when adding it to the training pipeline of ConvGRU-EV-FlowNet (EPE went up by 43%). We believe that the reason for this drop in performance is the event warping method used during training. While the proposed iterative warping allows for the error to propagate through all the pixels covered in the warping process, the linear warping used to train ConvGRU-EV-FlowNet only propagates the error through pixels with input events [10]. Therefore, if events are removed from the computation of the loss, the error is not propagated through the corresponding pixels, and then the spatial coherence of the resulting optical flow maps degrades. Despite sharing the same warping methodology, this is less of an issue for EV-FlowNet since it processes the events from longer temporal windows in a single forward pass, producing a single optical flow map per loss. The longer this window, the more likely it is that a pixel contains events triggered by multiple moving objects (i.e., reflected as events with different timestamps), and hence the higher the probability that the error is propagated through that pixel.

B.6. Visualizing the endpoint error

To support the hypothesis in Section 3.3 that the length of the training partition R has a significant impact on the quality of the training, here we study the distribution of the EPE of our models in Table 1 as a function of the ground truth optical flow magnitude in the thun_00_a[1] sequence

[1]Note that, since ground truth is required for this experiment, this sequence belongs to the training partition of DSEC-Flow [9]. Consequently, this means that our models have had access to a randomly cropped version of it during training.

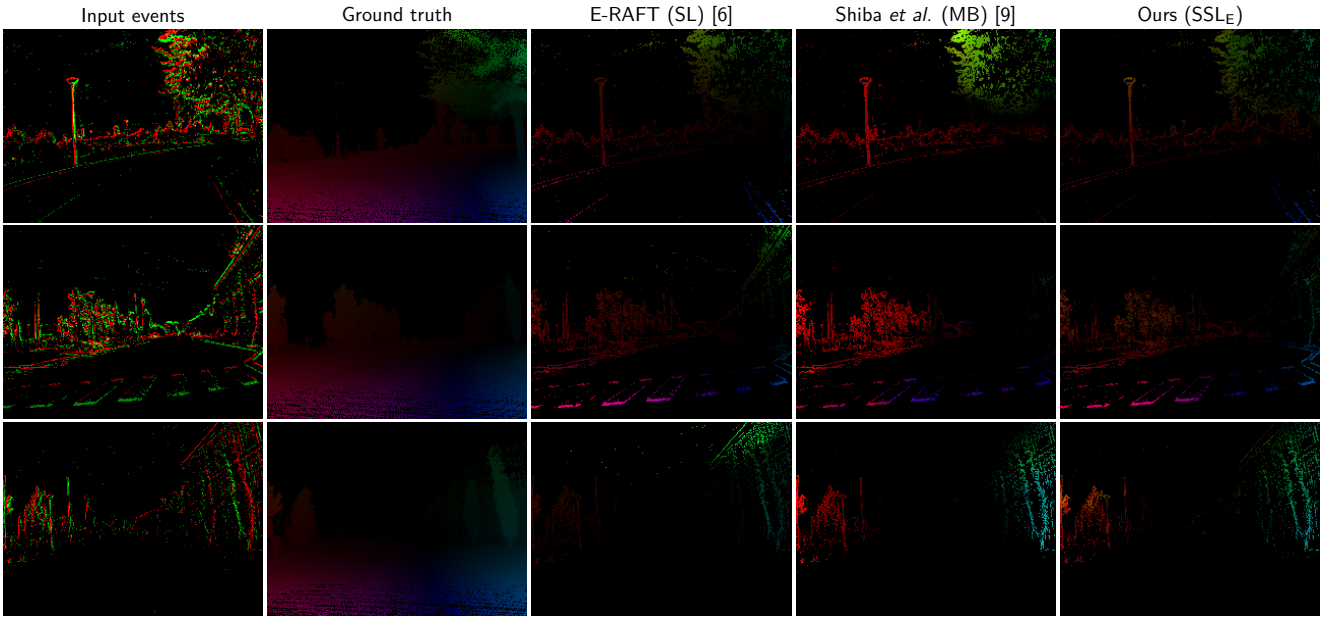


Figure S3: Qualitative comparison of our method with the state-of-the-art E-RAFT architecture [9] and the model-based approach from Shiba *et al*. [15] on the outdoor_day1 sequence from the MVSEC dataset [18]. Optical flow predictions are masked with the input events to be consistent with the evaluation proposed in [19]. The optical flow color coding can be found in Fig. 2 (top).

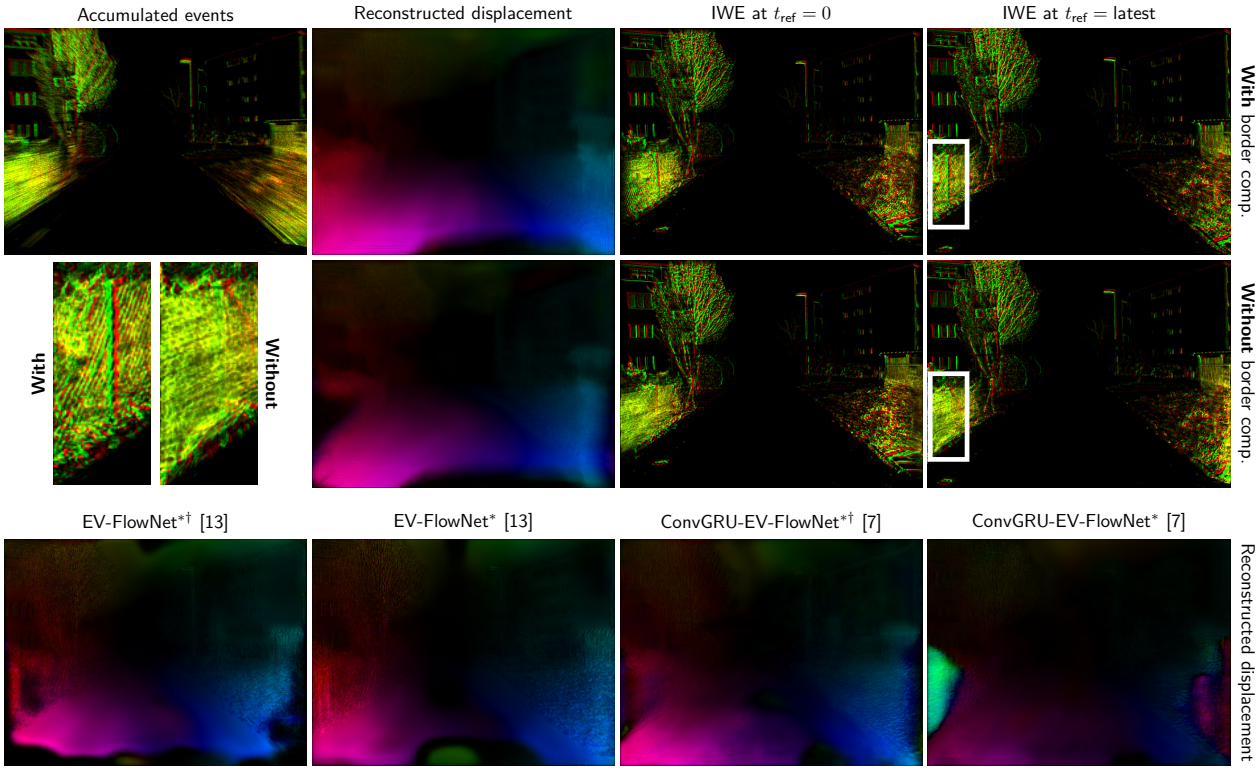


Figure S4: Qualitative results of the ablation study on the DSEC-Flow dataset [9] with respect to the effectiveness of the proposed event warping module and image-border compensation mechanism. *Top*: Models trained with the proposed SSL framework (dt $= 0.01$s, $R = 10$, $S = 1$). *Bottom*: Literature methods EV-FlowNet [20] and ConvGRU-EV-FlowNet [10]. $*$: Retrained by us on DSEC-Flow. †: Without border compensation. The optical flow color coding can be found in Fig. 2 (top).

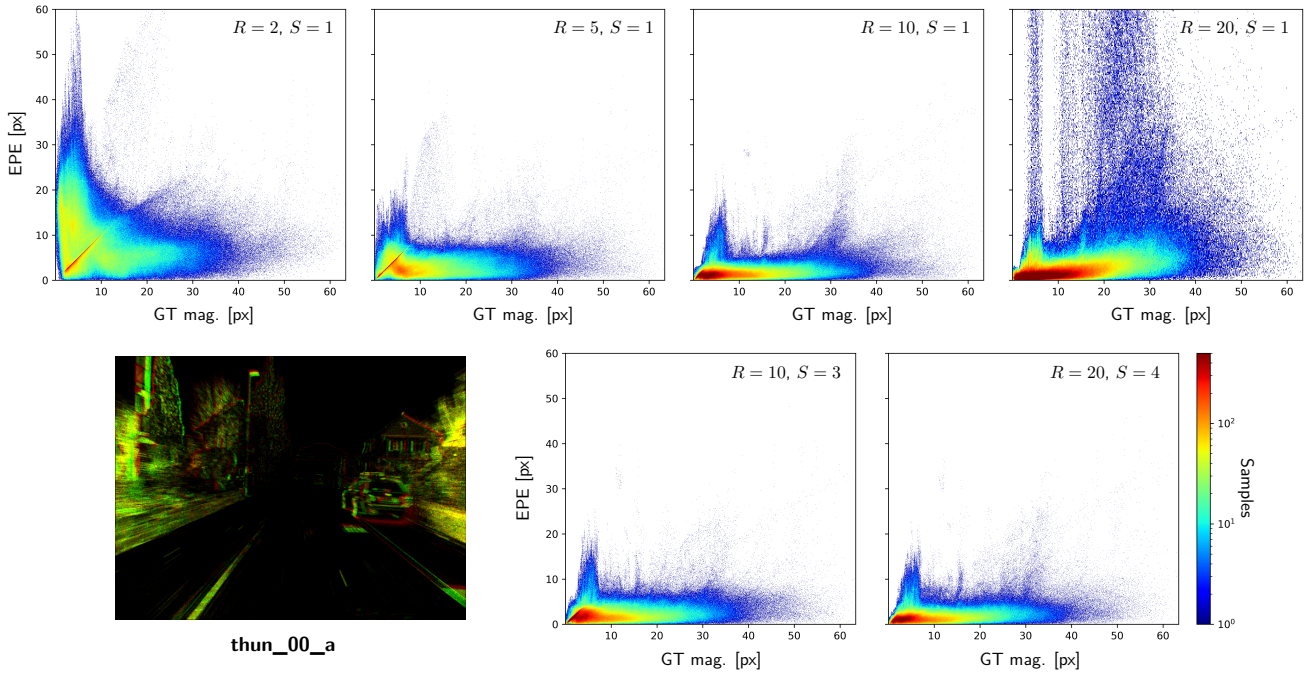


Figure S5: Distribution of the EPE of our models in Table 1 as a function of the ground truth magnitude in the thun_00_a sequence from DSEC-Flow [9]. For this experiment, and as in Table 1, all models were trained and deployed with $\text{dt}_{\text{input}} = 0.01$s.

from DSEC-Flow [9]. The error distributions are shown in Fig. S5 and confirm the conclusions derived from Table 1 in Section 4.2. Models trained with short training partitions (i.e., $R \in [2, 5]$) converge to solutions that are less accurate (i.e., high EPE) for low ground truth magnitudes, while long partitions (i.e., $R \geq 20$) do the same but for high ground truth magnitudes. The proposed multi-timescale approach (i.e., $S > 1$) to contrast maximization alleviates this issue and allows for the training of models that are accurate for all ground truth magnitudes without having to fine-tune the length of the training partition. As shown in this figure, the error distribution of the $S > 1$ models closely resembles that of our best performing solution: $R = 10$, $S = 1$.

References

[1] Himanshu Akolkar, Sio-Hoi Ieng, and Ryad Benosman. Real-time high speed motion prediction using fast aperture-robust event-driven visual flow. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(1):361–372, 2020. 3

[2] Vincent Brebion, Julien Moreau, and Franck Davoine. Real-time optical flow for vehicular perception with low-and high-resolution event cameras. *IEEE Trans. on Intell. Transp. Systems*, 2021. 3

[3] Javier Cuadrado, Ulysse Rançon, Benoit R. Cottereau, Francisco Barranco, and Timothée Masquelier. Optical flow estimation from event-based cameras and spiking neural networks. *Frontiers in Neuroscience*, 17:1160034, 2023. 1, 2, 3

[4] Ziluo Ding, Rui Zhao, Jiyuan Zhang, Tianxiao Gao, Ruiqin Xiong, Zhaofei Yu, and Tiejun Huang. Spatio-temporal recurrent networks for event-based optical flow estimation. In *Proc. AAAI Conf. on Artificial Intell.*, volume 36, pages 525–533, 2022. 3

[5] Guillermo Gallego, Tobi Delbruck, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE Trans. Pattern Anal. and Mach. Intell.*, 2020. 1

[6] Guillermo Gallego, Mathias Gehrig, and Davide Scaramuzza. Focus is all you need: Loss functions for event-based vis. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 12280–12289, 2019. 1

[7] Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3867–3876, 2018. 1

[8] Mathias Gehrig, Willem Aarents, Daniel Gehrig, and Davide Scaramuzza. DSEC: A stereo event camera dataset for driving scenarios. *IEEE Robot. and Autom. Lett.*, 6(3):4947–4954, 2021. 1, 3

[9] Mathias Gehrig, Mario Millhäusler, Daniel Gehrig, and Davide Scaramuzza. E-RAFT: Dense optical flow from event cameras. In *Int. Conf. 3D Vis.*, pages 197–206. IEEE, 2021. 1, 2, 3, 4, 5, 6

[10] Jesse J. Hagenaars, Federico Paredes-Vallés, and Guido C. H. E. de Croon. Self-supervised learning of event-based optical flow with spiking neural networks. In *Advances in Neural Information Process. Syst.*, volume 34, pages 7167–7179, 2021. 1, 2, 3, 4, 5

[11] Yijin Li, Zhaoyang Huang, Shuo Chen, Xiaoyu Shi, Hongsheng Li, Hujun Bao, Zhaopeng Cui, and Guofeng Zhang. Blinkflow: A dataset to push the limits of event-based optical flow estimation. *arXiv:2303.07716*, 2023. 1, 2

[12] Haotian Liu, Guang Chen, Sanqing Qu, Yanping Zhang, Zhijun Li, Alois Knoll, and Changjun Jiang. TMA: Temporal motion aggregation for event-based optical flow. *arXiv:2303.11629*, 2023. 1, 2, 3

[13] Elias Mueggler, Henri Rebecq, Guillermo Gallego, Tobi Delbruck, and Davide Scaramuzza. The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM. *Int. J. Robot. Research*, 36(2):142–149, 2017. 3

[14] Federico Paredes-Vallés and Guido C. H. E. de Croon. Back to event basics: Self-supervised learning of image reconstruction for event cameras via photometric constancy. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2021. 3

[15] Shintaro Shiba, Yoshimitsu Aoki, and Guillermo Gallego. Secrets of event-based optical flow. In *Eur. Conf. Comput. Vis.*, 2022. 1, 2, 3, 5

[16] Timo Stoffregen, Cedric Scheerlinck, Davide Scaramuzza, Tom Drummond, Nick Barnes, Lindsay Kleeman, and Robert Mahony. Reducing the sim-to-real gap for event cameras. In *European Conf. Comput. Vis.*, 2020. 1, 3

[17] Yilun Wu, Federico Paredes-Vallés, and Guido C. H. E. de Croon. Rethinking event-based optical flow: Iterative deblurring as an alternative to correlation volumes. *arXiv:2211.13726*, 2022. 1, 2

[18] Alex Z. Zhu, Dinesh Thakur, Tolga Özaslan, Bernd Pfrommer, Vijay Kumar, and Kostas Daniilidis. The multivehicle stereo event camera dataset: An event camera dataset for 3D perception. *IEEE Robot. and Autom. Lett.*, 3(3):2032–2039, 2018. 1, 3, 5

[19] Alex Z. Zhu and Liangzhe Yuan. EV-FlowNet: Self-supervised optical flow estimation for event-based cameras. In *Robot.: Science and Syst.*, 2018. 3, 5

[20] Alex Z. Zhu, Liangzhe Yuan, Kenneth Chaney, and Kostas Daniilidis. Unsupervised event-based learning of optical flow, depth, and egomotion. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 989–997, 2019. 2, 3, 4, 5