Learning Dense and Continuous Optical Flow from an Event Camera - Supplementary Materials

Zhexiong Wan, Yuchao Dai, *Member, IEEE,* and Yuxin Mao

APPENDIX A
THE STRUCTURE OF PYRAMID NETWORK

In recent years, with the influence of PWC-Net [1], the coarse-to-fine learning pipeline through pyramid structure has become the mainstream of optical flow estimation. In addition to our proposed iterative network, we also designed a pyramid-based Event-Image optical flow network to further compare the performance of our model. The structure is shown in Fig. 1, which consists of four main modules: pyramid feature encoder, event-image correlation module, flow estimation decoder, and the last refinement layer.

Pyramid Feature Encoder. For an input single image I_1 and an event volume E, their L level pyramid feature maps $\{P^l_{I_1}\}^L_{l=1}$ and $\{P^l_E\}^L_{l=1}$ are generated by two pyramid feature encoders with the same structure but do not share weights. Each encoder has six stages with the output of different feature maps, and each stage consists of two convolution layers. The first layer increases the number of feature channels and halves the feature map size by setting the convolution $stride = 2$. The second layer can keep the number of feature channels and the feature map size by setting $stride = 1$. Using these two pyramid encoders, we can get two groups of feature maps for image and event, respectively, as shown on the left side of Fig. 1. The number of feature channels is $\{16, 32, 64, 96, 128, 196\}$ and the feature size relative to the input is $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}\}$. In the training stage, we further use the siamese image pyramid encoder to get another group of second image feature maps $\{P^l_{I_2}\}^L_{l=1}$.

Event-Image Local Correlation Module. As mentioned in the iterative structure, we first generate the pseudo second image feature P^l_{pseudo} by fusing the first image feature $P^l_{I_1}$ and event feature P^l_E. Then we use the coarse flow from the previous pyramid layer to get the warped pseudo second image feature and build a local correlation with the first image feature. In the training stage, the real second image is input into the network at the same time, and get the second image features at each pyramid layer from the encoder network, which is shared with the first image. We use this real feature to constrain the added pseudo second image features. Then we can use the upsampled coarse flow from the previous pyramid layer to warp the pseudo second image features, build correlation volume to compute their correlation, and further use it to estimate optical flow.

For the event feature P^l_E at the l-th pyramid level, we construct an event-image local correlation volume as follows.

$$C^l_{EI}(\boldsymbol{x}, \delta_{uv}) = P^l_{I_1}(\boldsymbol{x}) \cdot Warp\{P^l_{pesudo}, \boldsymbol{F}^{(l-1)}\}(\boldsymbol{x} + \delta_{uv}), \quad (1)$$

where $Warp(P^l_{pesudo}, \boldsymbol{F}^{(l-1)})$ represents the warped feature of pseudo second image feature P^l_{pesudo}, using the coarse flow $\boldsymbol{F}^{(l-1)}$ from previous pyramid layer.

In this correlation volume construction for the first image with the event feature, the pseudo second image feature is introduced as an intermediary. It does not need to add additional network modules or parameters. The second image

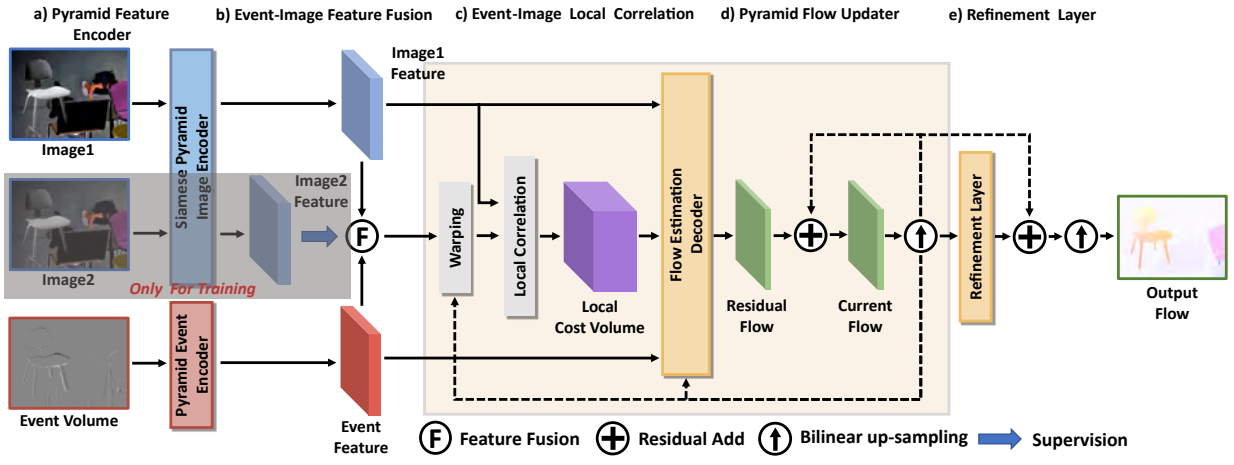


Fig. 1. **The pyramid-based network structure of ablation experiment (h).** We use the pyramid feature extractor (left) to get pyramid event and image features, and compute the local correlation using our proposed event-image fusion and correlation construction module (middle). Then we feed them into the pyramid flow updater (right) to estimate flow from coarse to fine. The structures enclosed by the orange box need to be repeated from coarse to fine. The refinement layer is used to get better prediction after the last pyramid layer.

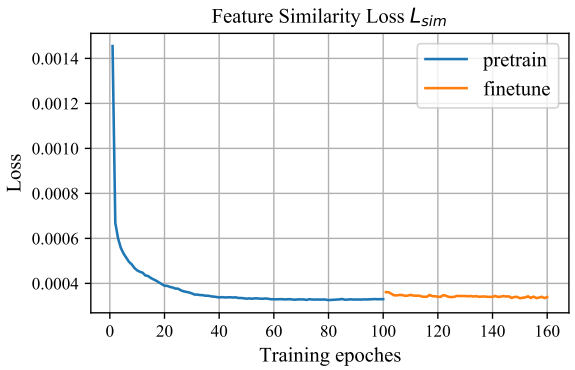


Fig. 2. **The line chart of image feature similarity loss.** The pre-trained model on FlyingChairs2 only needs a few epochs to converge when finetuning on MVSEC.

is required as input in the training stage. Only the first image and event are needed as the input in the inference stage. In the training stage, the input real second image feature $P^l_{I_2}$ is used to constrain the pseudo feature by using the feature similarity loss.

Flow Estimation Decoder. We concatenate the first image feature $P^l_{I_1}$, the event feature P^l_E, the upsampled coarse flow from the previous pyramid layer, and the built event-image correlation volume. Then we input them into the flow decoder and output the fine flow of this pyramid layer as the coarse flow of the next layer.

Refinement Layer. As in PWC-Net [1], there is a refinement layer behind the flow decoder at the last pyramid layer. The input of the refinement layer is the estimated flow and penultimate convolution feature map from the flow decoder, and get the fine flow by using a multi-layer dilated convolution network. Finally, the final full resolution optical flow is obtained through bilinear up-sampling.

APPENDIX B
ADDITIONAL EXPERIMENTS

A. Loss adaptability across different datasets

For the similarity loss function, Fig. 2 shows the trend of feature similarity loss in the total training stage. The pre-trained model on the FlyingChairs2 dataset at the last 100th epoch converges quickly to finetuning on the MVSEC dataset with a small loss gap. Thus the feature similarity loss can effectively converge and adapt across different datasets. Since the existing event-based approaches do not attempt to finetune the pre-trained model on the MVSEC dataset, we do not report the results of finetuning for a fair comparison.

As described in IRR-PWC [2], because the bi-directional training strategy is realized by exchanging the input orders on the unidirectional flow estimation network, the network can also be trained normally when the dataset has only one-direction ground-truth optical flow. The loss line chart in Supp.Fig. 2 also illustrates that the network can be trained normally even if only one-direction supervision is used in the finetune stage.

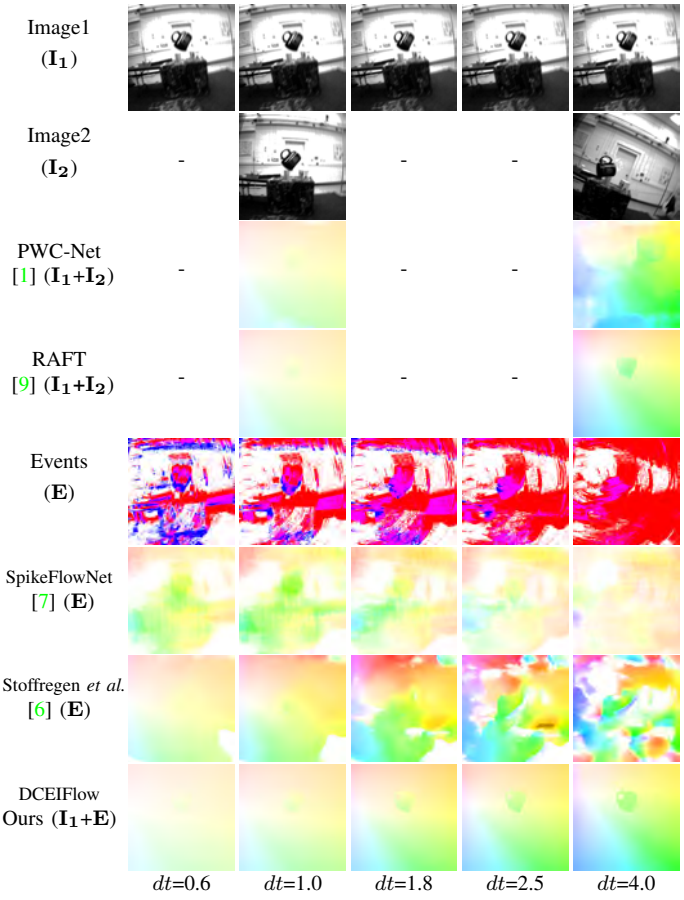
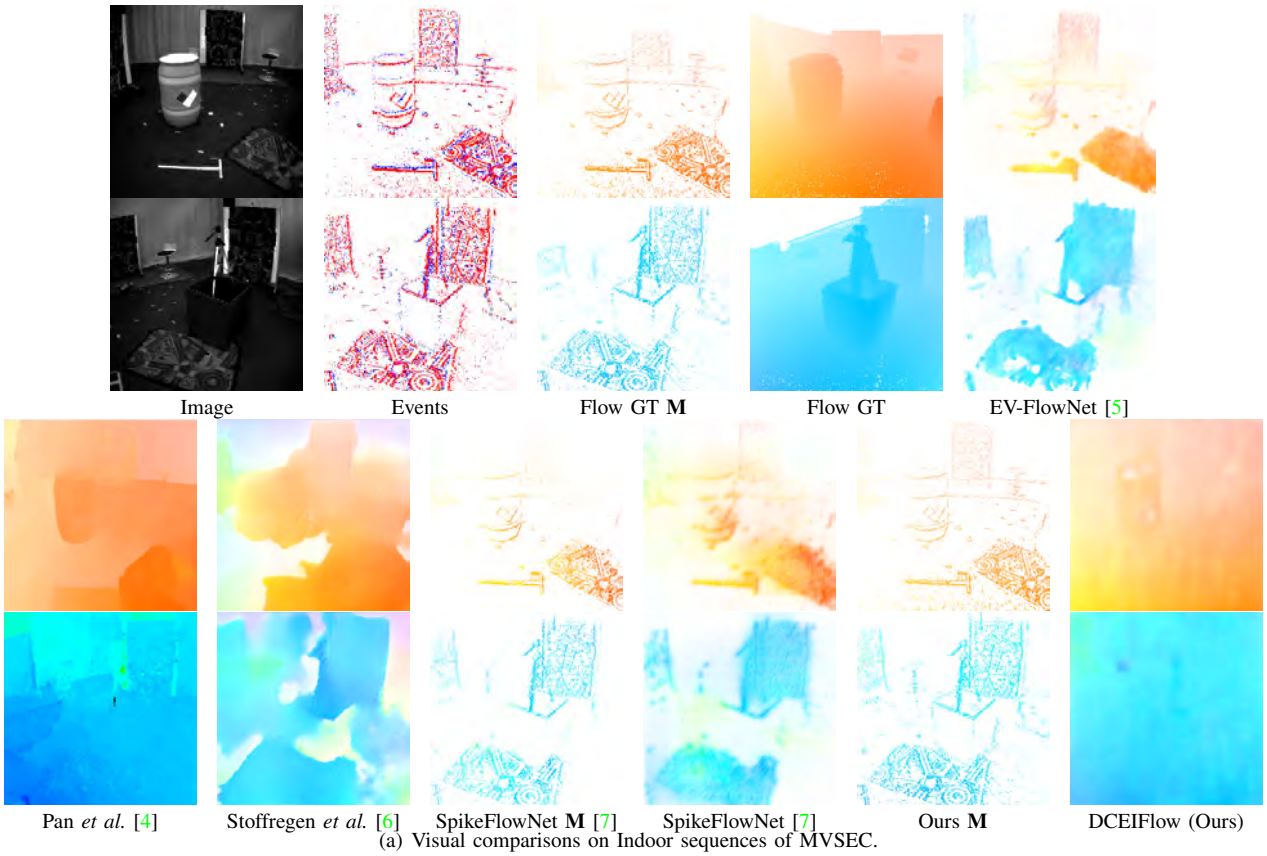


Fig. 5
ANOTHER VISUAL COMPARISONS OF CONTINUOUS FLOW PREDICTION WITH DIFFERENT TIME INTERVALS.

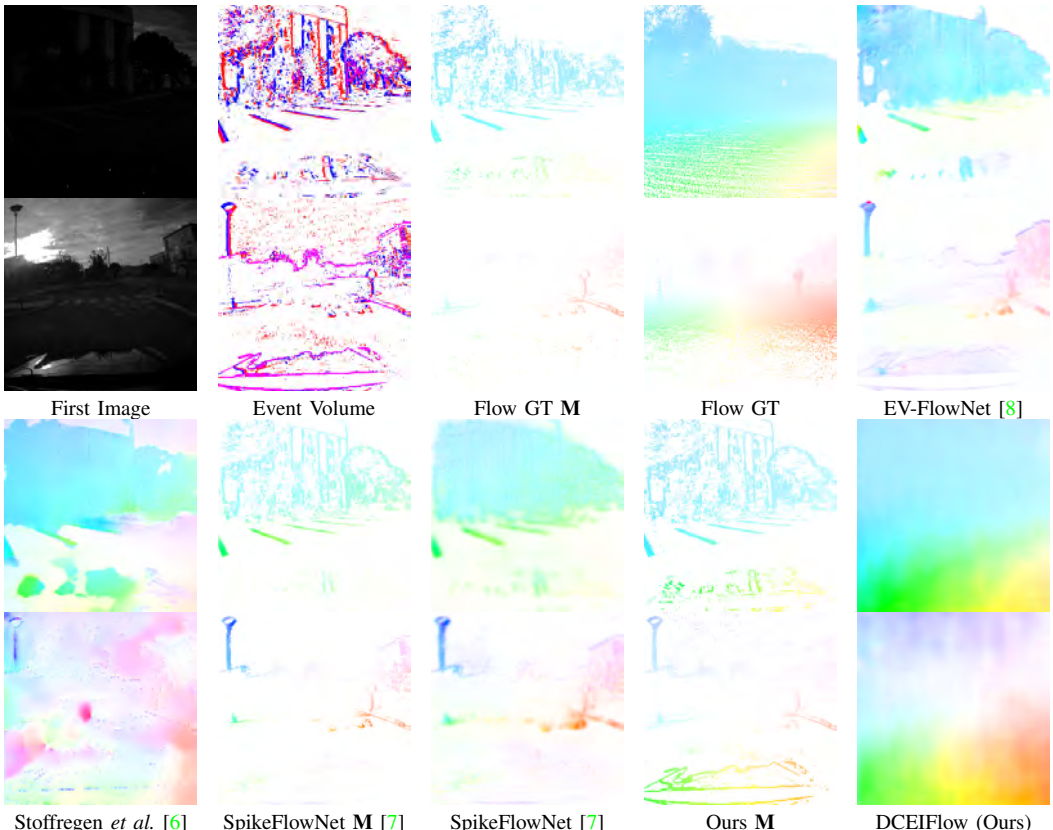
B. More visual comparisons

We make more visual comparisons on the MVSEC [3] Indoor and Outdoor sequences in Fig. 3. In particular, [4] is an optimization algorithm without open source code and requires parameter tuning, which inputs the image with events. However, there are many incorrect prediction areas compared with the ground truth flow, which is easy to cause visual misunderstanding.

In the manuscript, we only compared the quantitative results of our proposed baselines on the Sintel dataset. Here we add two visual comparisons on the Sintel test set in Fig. 4. Because the test set has no ground truth flow, we use the result of RAFT [9] as a reference. Pan *et al.* [10] also uses this dataset. The results are provided by the authors. In addition, we also compare the open source event-based method Stoffregen *et al.* [6], which claims to predict the dense optical flow. Our model not only achieves better predictions compared with the proposed baselines but also has better density and accuracy compared with the existing event-based method Stoffregen *et al.* [6] and with the single image method Pan *et al.* [10]. In addition, we make another continuous flow prediction comparisons on the EV-IMO [11] dataset in Fig. 5.



(a) Visual comparisons on Indoor sequences of MVSEC.



(b) Visual comparisons on the Outdoor sequences of MVSEC.

Fig. 3. **Visual comparisons on the MVSEC [3] dataset.** Our model achieves better visual results in both indoor and outdoor scenes. Best viewed on screen.

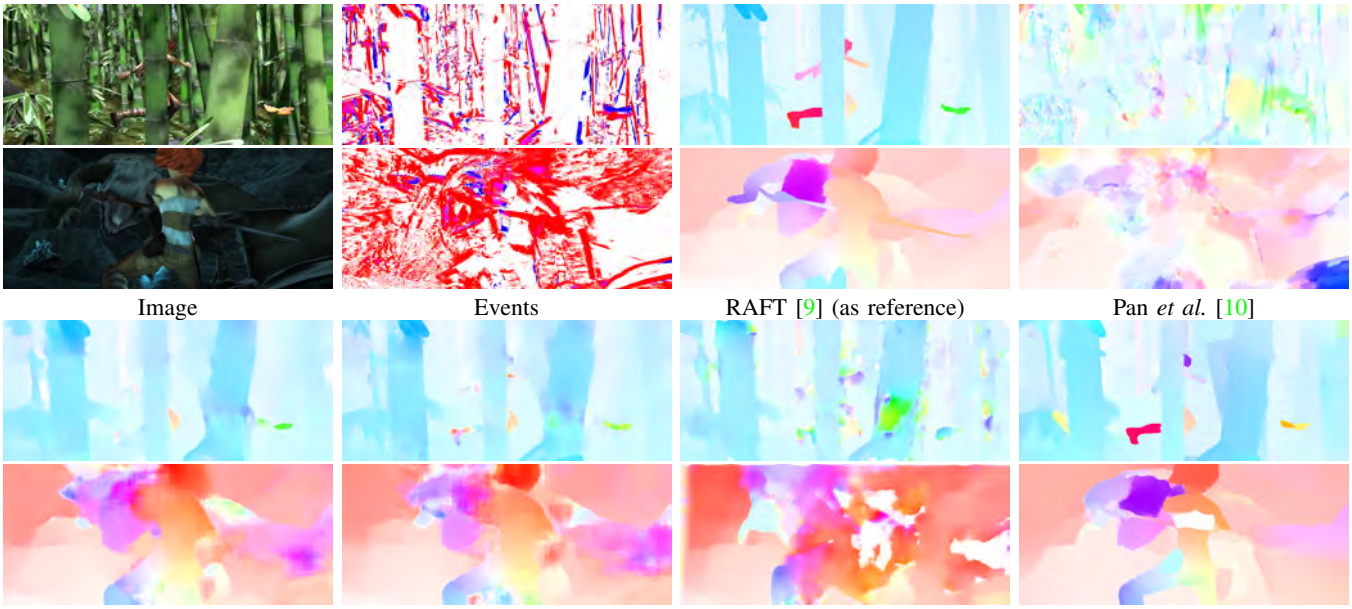


Fig. 4. **Visual comparisons on the Sintel [12] test set**. The models used for inference are all the pre-trained models that have not been trained on the Sintel dataset, except for the results of Pan *et al.* [10] are provided by the authors.

C. More baseline comparisons

1) Two-frame-based baselines: Our DCEIFlow model is based on the RAFT framework that removes the second image and introduces event data to achieve dense and continuous optical flow estimation. This raises a natural question, if the continuous optical flow is not needed, can the model achieve better performance by simultaneously inputting the two images with event data? To compare with the two-frame setting, we evaluate the results of four two-frame-based methods by using the open-sourced pre-trained models in Table II in the main paper. We also retrain the RAFT model on FlyingChairs2 using the bidirectional training loss function, and the results also show a significant improvement, as shown in Table I.

In addition, we experiment with introducing the event data with two images. We first conduct a simple experiment by removing the fusion module and the pseudo feature fusion module. We construct the correlation by using the two image features and concatenate the event features into the GRU-based flow updater. This is equivalent to inputting the event feature into the flow updater of RAFT additionally. We call it **Bl-RAFTE2I**. On this basis, we use our proposed fusion module to fuse the event features with image1 and image2 features, respectively, and then feed the fused two features into the correlation module. We call it **Bl-RAFTE2IFusion**.

As shown in Table I and Table II, the Bl-RAFTE2I model only inputs events into the updater, which improves the performance compared with RAFT. Then Bl-RAFTE2IFusion introduces the fusion module we proposed to utilize event information and improve performance on multiple datasets. We think this illustrates that the introduction of events can help achieve better optical flow estimation, and our proposed fusion module can utilize the motion information from events.

2) Event-based baselines: To further illustrate the effectiveness of our proposed event-image fusion module, we incorporate our proposed event-image fusion module to Baseline-EI, feed the fused features into the decoder along with the first frame image and event features, and use the similarity loss L_{sim} in training to make a fair comparison with our model. We call this model Baseline-EIFusion, and the experimental results are shown in Table II and Table III. Both baselines and our model follow the same pre-training settings and training data to achieve a fair performance comparison. The experimental comparisons show that introducing our proposed event-image fusion module into Baseline-EI leads to further performance improvements. We think this illustrates that our proposed fusion module can further utilize the motion information from events and improve the optical flow estimation performance.

REFERENCES

[1] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8934–8943. 1, 2

[2] J. Hur and S. Roth, "Iterative residual refinement for joint optical flow and occlusion estimation," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 5754–5763. 2

[3] A. Z. Zhu, D. Thakur, T. Özaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, "The multivehicle stereo event camera dataset: An event camera dataset for 3d perception," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2032–2039, 2018. 2, 3, 6

[4] L. Pan, M. Liu, and R. Hartley, "Single image optical flow estimation with an event camera," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 1669–1678. 2, 3

[5] A. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Ev-flownet: Self-supervised optical flow estimation for event-based cameras," in *Proceedings of Robotics: Science and Systems*, Pittsburgh, Pennsylvania, June 2018. 3

[6] T. Stoffregen, C. Scheerlinck, D. Scaramuzza, T. Drummond, N. Barnes, L. Kleeman, and R. Mahony, "Reducing the sim-to-real gap for event cameras," in *European Conf. on Computer Vision (ECCV)*, 2020, pp. 534–549. 2, 3, 4, 7

TABLE I
PERFORMANCE COMPARISON WITH SEVERAL BASELINES AND OUR PROPOSED MODEL ON THE FLYINGCHAIRS2 [13] VALIDATION SET, AND THE SINTEL [12] TRAINING SET.

Input Data	Method Name	Event Fusion	Bi. Train.	Param. Num. (M)	FlyingChairs2 EPE	FlyingChairs2 %Out	Sintel EPE	Sintel %Out
$\mathbf{I_1}$+$\mathbf{I_2}$	RAFT(C) [9]	-	×	5.26	0.83	3.80	3.20	9.19
	Retrain-RAFT(C2)	-	✓	5.26	0.77	3.39	3.01	9.01
$\mathbf{I_1}$+$\mathbf{I_2}$ +$\mathbf{E}$	Bl-RAFTE2I	Concat	✓	6.56	0.66	2.80	2.95	8.83
	Bl-RAFTE2IFusion	E-2I Fusion	✓	8.52	0.57	2.35	2.84	8.28
$\mathbf{E}$	Baseline-EV	-	✓	18.05	2.25	12.24	8.89	43.28
$\mathbf{I_1}$+$\mathbf{E}$	Baseline-EI	Concat	✓	34.54	1.86	11.40	8.14	35.53
	Baseline-EIFusion	E-I Fusion	✓	43.73	1.75	10.02	7.92	34.54
	DCEIFlow (Ours)	E-I Fusion	✓	**7.51**	**1.58**	**7.88**	**6.47**	**32.23**

[7] C. Lee, A. K. Kosta, A. Z. Zhu, K. Chaney, K. Daniilidis, and K. Roy, "Spike-flownet: event-based optical flow estimation with energy-efficient hybrid neural networks," in *European Conf. on Computer Vision (ECCV)*, 2020, pp. 366–382. 2, 3, 7

[8] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Unsupervised event-based learning of optical flow, depth, and egomotion," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 989–997. 3

[9] Z. Teed and J. Deng, "Raft: Recurrent all-pairs field transforms for optical flow," in *European Conf. on Computer Vision (ECCV)*, 2020, pp. 402–419. 2, 4, 5, 7

[10] L. Pan, C. Scheerlinck, X. Yu, R. Hartley, M. Liu, and Y. Dai, "Bringing a blurry frame alive at high frame-rate with an event camera," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 6820–6829. 2, 4

[11] A. Mitrokhin, C. Ye, C. Fermüller, Y. Aloimonos, and T. Delbruck, "Ev-imo: Motion segmentation dataset and learning pipeline for event cameras," in *IEEE Int. Conf. on Intelligent Robots and Systems (IROS)*, 2019, pp. 6105–6112. 2

[12] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *European Conf. on Computer Vision (ECCV)*, 2012, pp. 611–625. 4, 5

[13] E. Ilg, T. Saikia, M. Keuper, and T. Brox, "Occlusions, motion and depth boundaries with a generic network for disparity, optical flow or scene flow estimation," in *European Conf. on Computer Vision (ECCV)*, 2018, pp. 614–630. 5

[14] D. Sun, X. Yang, M. Liu, and J. Kautz, "Models matter, so does training: An empirical study of cnns for optical flow estimation," *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 42, no. 6, pp. 1408–1423, 2019. 7

[15] L. Liu, J. Zhang, R. He, Y. Liu, Y. Wang, Y. Tai, D. Luo, C. Wang, J. Li, and F. Huang, "Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 6489–6498. 7

[16] A. Stone, D. Maurer, A. Ayvaci, A. Angelova, and R. Jonschkowski, "Smurf: Self-teaching multi-frame unsupervised raft with full-image warping," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 3887–3896. 7

[17] M. Gehrig, M. Millhäusler, D. Gehrig, and D. Scaramuzza, "E-raft: Dense optical flow from event cameras," in *Int. Conf. on 3D Vision (3DV)*, 2021, pp. 197–206. 7

TABLE II

PERFORMANCE COMPARISON BETWEEN DIFFERENT BASELINES AND OUR PROPOSED MODEL ON THE MVSEC [3] DATASET. WE USE THE SAME PRE-TRAINED MODEL TO GET THE RESULTS OF EACH METHOD UNDER TWO INPUT INTERVAL SETTINGS ($dt=1$ AND $dt=4$ FRAMES).

Input $dt=1$	Method	Train Mann.	Train D.Type	Train D.Set	Eval. Metric	*indoor_flying1* EPE	*indoor_flying1* %Out	*indoor_flying2* EPE	*indoor_flying2* %Out	*indoor_flying3* EPE	*indoor_flying3* %Out	*outdoor_day1* EPE	*outdoor_day1* %Out	*outdoor_day2* EPE	*outdoor_day2* %Out
E	Baseline-EV	SL	E	C2	dense	0.91	1.06	0.98	1.04	1.02	1.60	1.09	1.56	0.99	4.07
	Baseline-EV	SL	E	C2	sparse	0.93	1.11	1.04	2.47	1.00	1.22	0.95	0.82	0.98	2.52
	Baseline-EV	SL	E	M	dense	0.80	0.65	0.95	2.04	0.89	1.56	0.44	0.03	×	×
	Baseline-EV	SL	E	M	sparse	0.89	1.17	1.14	4.31	0.97	2.50	0.47	0.10	×	×
I_1+E	Baseline-EI	SL	I_1,E	C2	dense	0.78	0.49	0.81	0.58	0.80	0.25	0.92	0.75	0.84	3.33
	Baseline-EI	SL	I_1,E	C2	sparse	0.82	0.60	0.89	1.37	0.82	0.40	0.80	0.54	0.83	2.69
	Baseline-EI	SL	I_1,E	M	dense	0.75	0.54	0.80	0.77	0.80	0.93	0.36	0.00	×	×
	Baseline-EI	SL	I_1,E	M	sparse	0.78	0.68	0.91	1.45	0.82	1.03	0.36	0.00	×	×
	Baseline-EIFusion	SL	I_1,I_2,E	C2	dense	0.72	0.62	0.74	0.47	0.74	0.23	0.90	0.98	0.86	3.56
	Baseline-EIFusion	SL	I_1,I_2,E	C2	sparse	0.76	0.91	0.82	0.98	0.76	0.33	0.86	1.01	0.90	3.05
	Baseline-EIFusion	SL	I_1,I_2,E	M	dense	0.78	0.31	0.80	0.55	0.77	0.37	0.40	0.02	×	×
	Baseline-EIFusion	SL	I_1,I_2,E	M	sparse	0.80	0.29	0.91	1.23	0.81	0.52	0.38	0.09	×	×
	DCEIFlow (Ours)	SL	I_1,I_2,E	C2	dense	**0.56**	**0.28**	**0.64**	**0.16**	**0.57**	**0.12**	0.91	0.71	**0.79**	2.59
	DCEIFlow (Ours)	SL	I_1,I_2,E	C2	sparse	0.57	0.30	0.70	0.30	0.58	0.15	**0.74**	**0.29**	0.82	**2.34**
	DCEIFlow (Ours)	SL	I_1,I_2,E	M	dense	0.64	0.87	0.74	1.16	0.70	1.08	0.20	0.00	×	×
	DCEIFlow (Ours)	SL	I_1,I_2,E	M	sparse	0.75	1.55	0.90	2.10	0.80	1.77	0.22	0.00	×	×
I_1+I_2 +E	Bl-RAFTE2I	SL	I_1,I_2,E	C2	dense	0.39	0.11	0.49	0.04	0.44	0.00	0.75	0.08	0.58	2.65
	Bl-RAFTE2I	SL	I_1,I_2,E	C2	sparse	0.41	0.11	0.55	0.10	0.47	0.01	0.65	0.09	0.58	2.84
	Bl-RAFTE2I	SL	I_1,I_2,E	M	dense	0.48	0.09	0.57	0.32	0.49	0.06	0.22	0.00	×	×
	Bl-RAFTE2I	SL	I_1,I_2,E	M	sparse	0.52	0.13	0.68	0.81	0.53	0.12	0.24	0.00	×	×
	Bl-RAFTE2IFusion	SL	I_1,I_2,E	C2	dense	0.39	0.12	0.47	0.03	0.43	0.00	0.78	0.12	0.63	3.21
	Bl-RAFTE2IFusion	SL	I_1,I_2,E	C2	sparse	0.41	0.13	0.52	0.08	0.44	0.00	0.65	0.11	0.60	2.82
	Bl-RAFTE2IFusion	SL	I_1,I_2,E	M	dense	0.43	0.01	0.56	0.11	0.48	0.02	0.26	0.00	×	×
	Bl-RAFTE2IFusion	SL	I_1,I_2,E	M	sparse	0.46	0.01	0.66	0.21	0.52	0.03	0.25	0.00	×	×

Input $dt=4$	Method	Train Mann.	Train D.Type	Train D.Set	Eval. Metric	*indoor_flying1* EPE	*indoor_flying1* %Out	*indoor_flying2* EPE	*indoor_flying2* %Out	*indoor_flying3* EPE	*indoor_flying3* %Out	*outdoor_day1* EPE	*outdoor_day1* %Out	*outdoor_day2* EPE	*outdoor_day2* %Out
E	Baseline-EV	SL	E	C2	dense	1.76	13.21	2.05	19.09	1.99	18.73	2.63	30.84	2.10	20.83
	Baseline-EV	SL	E	C2	sparse	1.72	12.37	2.24	22.98	1.91	16.57	1.96	17.91	2.01	18.86
	Baseline-EV	SL	E	M	dense	2.59	29.67	3.53	39.82	2.87	31.21	1.55	12.24	×	×
	Baseline-EV	SL	E	M	sparse	3.04	36.87	4.55	52.58	3.38	36.72	1.64	13.64	×	×
I_1+E	Baseline-EI	SL	I_1,E	C2	dense	1.66	11.17	2.13	20.17	1.78	14.56	2.25	25.19	1.92	19.39
	Baseline-EI	SL	I_1,E	C2	sparse	1.65	11.21	2.37	25.70	1.79	14.57	1.66	13.20	1.83	17.55
	Baseline-EI	SL	I_1,E	M	dense	2.08	21.13	2.77	29.20	2.32	24.59	1.16	5.98	×	×
	Baseline-EI	SL	I_1,E	M	sparse	2.21	23.61	3.37	36.99	2.47	26.50	1.11	5.01	×	×
	Baseline-EIFusion	SL	I_1,I_2,E	C2	dense	1.58	9.93	2.08	19.08	1.71	12.92	2.10	22.42	1.89	18.97
	Baseline-EIFusion	SL	I_1,I_2,E	C2	sparse	1.60	10.74	2.39	25.62	1.75	13.74	1.57	11.48	1.84	17.76
	Baseline-EIFusion	SL	I_1,I_2,E	M	dense	1.97	18.29	2.86	31.65	2.31	24.18	1.27	7.64	×	×
	Baseline-EIFusion	SL	I_1,I_2,E	M	sparse	2.10	21.38	3.45	39.52	2.52	27.67	1.18	6.51	×	×
	DCEIFlow (Ours)	SL	I_1,I_2,E	C2	dense	**1.49**	**8.14**	**1.97**	**17.37**	**1.69**	**12.34**	1.87	19.13	1.62	14.73
	DCEIFlow (Ours)	SL	I_1,I_2,E	C2	sparse	1.52	8.79	2.21	22.13	1.74	13.33	**1.37**	**8.54**	**1.61**	**14.38**
	DCEIFlow (Ours)	SL	I_1,I_2,E	M	dense	1.90	17.43	2.97	34.38	2.32	26.07	0.87	3.12	×	×
	DCEIFlow (Ours)	SL	I_1,I_2,E	M	sparse	2.08	21.47	3.48	42.05	2.51	29.73	0.89	3.19	×	×
I_1+I_2 +E	Bl-RAFTE2I	SL	I_1,I_2,E	C2	dense	1.38	6.99	1.76	12.98	1.57	10.16	2.63	32.00	1.48	13.12
	Bl-RAFTE2I	SL	I_1,I_2,E	C2	sparse	1.41	7.45	1.88	15.25	1.60	10.89	2.04	21.12	1.50	12.45
	Bl-RAFTE2I	SL	I_1,I_2,E	M	dense	1.52	9.98	2.01	16.50	1.69	11.25	0.99	4.61	×	×
	Bl-RAFTE2I	SL	I_1,I_2,E	M	sparse	1.60	11.94	2.34	22.42	1.81	13.10	1.02	5.38	×	×
	Bl-RAFTE2IFusion	SL	I_1,I_2,E	C2	dense	1.37	6.99	1.70	12.06	1.56	9.65	2.74	34.07	1.56	14.45
	Bl-RAFTE2IFusion	SL	I_1,I_2,E	C2	sparse	1.40	7.24	1.82	14.41	1.58	10.37	2.10	22.14	1.43	12.46
	Bl-RAFTE2IFusion	SL	I_1,I_2,E	M	dense	1.33	6.64	1.82	14.65	1.49	8.35	1.01	4.29	×	×
	Bl-RAFTE2IFusion	SL	I_1,I_2,E	M	sparse	1.40	8.35	2.11	19.93	1.57	9.58	0.90	3.45	×	×

TABLE III
DENSE FLOW PREDICTION ANALYSIS. THE RESULTS ARE AVERAGED ON THE MVSEC *indoor_flying1-3* SEQUENCES.

Input	Method	Claim Setting	Train. set (dt)	Eval. dt	Full valid pixels EPE	Full valid pixels %Out	Event Masked EPE	Event Masked %Out	Event Excluded EPE	Event Excluded %Out	Dense Ratio
$\mathbf{I_1}$+$\mathbf{I_2}$	PWC-Net [14]	dense	C+T	1	1.58	3.03	1.62	3.73	1.58	2.96	1.013
	RAFT [9]	dense	C	1	0.49	0.06	0.55	0.07	0.49	0.06	1.060
	ARFlow [15]	dense	SR+S	1	0.43	0.07	0.42	0.05	0.43	0.08	0.996
	SMURF [16]	dense	C	1	0.46	0.19	0.44	0.08	0.46	0.20	0.981
	PWC-Net [14]	dense	C+T	4	2.05	17.47	2.11	18.99	2.03	16.83	1.019
	RAFT [9]	dense	C	4	1.63	10.92	1.69	11.68	1.61	10.49	1.024
	ARFlow [15]	dense	SR+S	4	1.44	7.93	1.50	9.15	1.41	7.20	1.034
	SMURF [16]	dense	C	4	1.53	9.29	1.54	9.72	1.52	8.86	1.007
$\mathbf{E}$	SpikeFlowNet [7]	sparse	M (1)	1	1.14	4.67	1.08	4.03	1.15	5.16	0.969
	Stoffregen *et al.* [6]	dense	ESIM	1	0.75	0.57	0.65	0.41	0.76	0.60	0.927
	E-RAFT [17]	dense	DSEC	1	0.82	1.54	0.97	2.76	0.80	1.33	1.105
	Baseline-EV	dense	M (1)	1	0.83	2.38	0.99	4.60	0.80	1.96	1.114
	Baseline-EV	dense	M (1&4)	1	0.88	1.42	1.00	2.66	0.86	1.15	1.080
	Baseline-EV	dense	C2	1	0.97	1.24	0.99	1.60	0.97	1.18	1.010
	SpikeFlowNet [7]	sparse	M (4)	4	3.65	45.42	3.08	33.45	3.78	49.01	0.906
	Stoffregen *et al.* [6]	dense	ESIM	4	3.08	35.91	2.29	21.03	3.35	40.79	0.835
	E-RAFT [17]	dense	DSEC	4	2.19	19.55	2.46	23.15	2.05	17.97	1.098
	Baseline-EV	dense	M (4)	4	3.12	35.51	3.60	39.99	2.84	33.36	1.128
	Baseline-EV	dense	M (1&4)	4	3.00	33.57	3.66	42.06	2.66	30.01	1.177
	Baseline-EV	dense	C2	4	1.93	17.01	1.96	17.31	1.90	16.46	1.015
$\mathbf{I_1}$+$\mathbf{E}$	Baseline-EI	dense	M (1)	1	0.77	1.91	0.91	3.61	0.75	1.59	1.106
	Baseline-EI	dense	M (1&4)	1	0.78	0.75	0.84	1.05	0.77	0.68	1.040
	Baseline-EI	dense	C2	1	0.80	0.44	0.84	0.79	0.79	0.40	1.033
	Baseline-EIFusion	dense	M (1)	1	0.69	0.56	0.76	1.11	0.67	0.45	1.060
	Baseline-EIFusion	dense	M (1&4)	1	0.78	0.57	0.84	0.95	0.77	0.50	1.044
	Baseline-EIFusion	dense	C2	1	0.74	0.44	0.78	0.74	0.73	0.41	1.033
	DCEIFlow (Ours)	dense	M (1)	1	0.78	1.60	0.92	2.55	0.76	1.45	1.102
	DCEIFlow (Ours)	dense	M (1&4)	1	0.67	0.48	0.77	0.89	0.66	0.42	1.080
	DCEIFlow (Ours)	dense	C2	1	**0.59**	**0.18**	**0.62**	**0.25**	**0.58**	**0.18**	1.027
	Baseline-EI	dense	M (4)	4	2.73	32.52	3.00	36.79	2.57	30.34	1.081
	Baseline-EI	dense	M (1&4)	4	2.39	24.97	2.69	29.03	2.22	23.02	1.100
	Baseline-EI	dense	C2	4	1.85	15.30	1.94	17.16	1.80	14.18	1.036
	Baseline-EIFusion	dense	M (4)	4	2.41	25.19	2.65	28.61	2.27	23.33	1.081
	Baseline-EIFusion	dense	M (1&4)	4	2.38	24.71	2.69	29.53	2.21	22.42	1.104
	Baseline-EIFusion	dense	C2	4	1.79	13.98	1.91	16.70	1.73	12.54	1.052
	DCEIFlow (Ours)	dense	M (4)	4	2.58	22.00	2.93	25.80	2.41	20.41	1.106
	DCEIFlow (Ours)	dense	M (1&4)	4	2.24	22.42	2.52	27.33	2.10	20.27	1.098
	DCEIFlow (Ours)	dense	C2	4	**1.72**	**12.62**	**1.82**	**14.75**	**1.67**	**11.69**	1.045
$\mathbf{I_1}$+$\mathbf{I_2}$ +$\mathbf{E}$	Bl-RAFTE2I	dense	M (1)	1	0.57	0.31	0.64	0.69	0.56	0.26	1.073
	Bl-RAFTE2I	dense	M (1&4)	1	0.51	0.16	0.58	0.35	0.51	0.12	1.071
	Bl-RAFTE2I	dense	C2	1	0.44	0.05	0.48	0.07	0.44	0.05	1.047
	Bl-RAFTE2IFusion	dense	M (1)	1	0.52	0.08	0.58	0.19	0.51	0.07	1.067
	Bl-RAFTE2IFusion	dense	M (1&4)	1	0.49	0.05	0.55	0.09	0.48	0.04	1.066
	Bl-RAFTE2IFusion	dense	C2	1	0.43	0.05	0.46	0.07	0.43	0.05	1.040
	Bl-RAFTE2I	dense	M (4)	4	1.80	13.45	1.97	16.57	1.72	12.02	1.071
	Bl-RAFTE2I	dense	M (1&4)	4	1.74	12.58	1.92	15.82	1.64	11.07	1.081
	Bl-RAFTE2I	dense	C2	4	1.57	10.04	1.63	11.20	1.55	9.44	1.027
	Bl-RAFTE2IFusion	dense	M (4)	4	1.70	12.29	1.86	15.44	1.63	10.86	1.071
	Bl-RAFTE2IFusion	dense	M (1&4)	4	1.54	9.88	1.69	12.62	1.47	8.60	1.075
	Bl-RAFTE2IFusion	dense	C2	4	1.54	9.57	1.60	10.68	1.52	9.02	1.026