Supplementary Material for Spike-FlowNet: Event-based Optical Flow Estimation with Energy-Efficient Hybrid Neural Networks

Chankyu Lee[1][0000−0002−6933−7459], Adarsh Kumar Kosta[1][0000−0001−6377−6701], Alex Zihao Zhu[2][0000−0002−2195−014X], Kenneth Chaney[2][0000−0003−1768−6136], Kostas Daniilidis[2][0000−0003−0498−0758], and Kaushik Roy[1][0000−0002−0735−9695]

[1] Purdue University, West Lafayette IN, 47907, USA
{lee2216,akosta,kaushik}@purdue.edu
[2] University of Pennsylvania, Philadelphia PA 19104, USA
{alexzhu,chaneyk,kostas}@seas.upenn.edu

In this supplementary material, we present the ablation studies to explore the optimal design choices of hybrid networks, input data representation and weight factor (λ) of the smoothness loss in the loss function.

1 Hybrid Network

In addition to the described architecture (denoted Spike-FlowNet), we train additional network topologies to test different hybrid design options. We use two more networks in which residual blocks are composed of SNN layers: one where only first residual block is converted to SNN (Spike-FlowNet_1R), and second where both residual blocks are converted to SNN (Spike-FlowNet_2R). Note, results for a fully ANN architecture are given in EV-FlowNet [1]. We do not consider converting the decoder layers to construct a fully SNN architecture, as they use analog inputs from intermediate optical flows and output accumulators.

Rows 1-3 in table 1 show the AEE results for the different network topologies. We find that AEE results degrade as more layers are transferred to SNNs for both $dt = 1$ and $dt = 4$. This is because the spike vanishing phenomenon aggravates with the network depth, leading to the degradation in the quality of predicted optical flow. The best AEE results are achieved by Spike-FlowNet case which is advocated throughout the manuscript.

2 Input representation

We validate the influence of the number of groups (N) in input representation. In the case of $N = 3$ and $N = 4$, AEE results are provided in rows 4-5 in table 1. Note, Spike-FlowNet represents $N = 2$ case. With the increase in the number of input groups (N), the results show that $dt = 1$ case achieves worse AEE while $dt = 4$ converges to a reasonably accurate flow estimate. This is because each input group requires to have a certain number of events for proper training, and we find that $N = 2$ case provides optimal results for both $dt = 1$ and $dt = 4$.

3 Loss function

To find the optimal ratio between photometric and smoothness losses, we train networks with a variety of weight factors (λ) over the range $[1, 100]$. Rows 6-8 in table 1 highlight AEE results for $\lambda = 1$, 10, 100. We observe that $\lambda = 10$, 100 cases converge to more accurate flow estimate for $dt = 1$ while $\lambda = 1$ case works better for $dt = 4$. This is because inputs are greatly sparse in $dt = 1$, hence its corresponding flow outputs have more scarce and discontinuous structures, requiring a higher degree of smoothness.

Table 1. Average Endpoint Error (AEE) for ablation studies with different design choices

	dt=1 frame				dt=4 frame			
	indoor1	indoor2	indoor3	outdoor1	indoor1	indoor2	indoor3	outdoor1
Spike-FlowNet	0.84	1.28	1.11	0.49	2.24	3.83	3.18	1.09
Spike-FlowNet_1R	0.88	1.55	1.31	0.51	2.73	4.46	3.66	1.15
Spike-FlowNet_2R	0.90	1.56	1.29	0.56	2.75	4.61	3.76	1.19
N=3	0.92	1.34	1.18	0.50	2.34	4.05	3.29	1.12
N=4	1.07	1.76	1.57	0.60	2.27	3.81	3.10	1.15
λ=1	0.91	1.38	1.23	0.50	2.24	3.83	3.18	1.09
λ=10	0.84	1.28	1.11	0.49	2.42	4.22	3.44	1.18
λ=100	0.84	1.30	1.14	0.49	2.50	4.01	3.28	1.19

References

1. Zhu, A.Z., Yuan, L., Chaney, K., Daniilidis, K.: Ev-flownet: Self-supervised optical flow estimation for event-based cameras. arXiv preprint arXiv:1802.06898 (2018)