Supplementary to Event-based Temporally Dense Optical Flow Estimation with Sequential Learning

1. Additional Results on Optical Flow Estimation Accuracy

1.1. Results from using a different training-testing set splitting strategy

Given that we partitioned the recordings into training and testing sets using an 80/20 ratio for each recording, our method for dataset generation may raise a concern as it is different from the approach used in [9, 33]. As a result, we include additional results (see Table 1) that demonstrate how the partitioning strategy does not significantly impact the reported trends in the optical flow estimation accuracy. This can be attributed to the fact that, despite variations in scene names within the DSEC dataset, numerous recordings share similar settings. For instance, there are visual similarities between scenes denoted as *interlake_00_f* and *interlake_00_g*, despite their distinct names. For results in Table 1, we specifically exclude some sequences to be used for testing and make them unseen to the models during training. More information about the two dataset generation strategies is available in the accompanying source code.

Table 1. Comparison between the existing and proposed models in terms of the average end-point error (AEE) when using a different method to split training and testing set. Unlike the results reported in the manuscript, each recording on the DSEC dataset is either used for training or testing, but not both.

Architecture	Prediction Rate	AEE
EV-FlowNet [33]	10 Hz	0.81
Spike-FlowNet [19]	10 Hz	1.27
Adaptive-FlowNet [17]	10 Hz	1.86
LSTM-FlowNet	**100 Hz**	**0.65**
EfficientSpike-FlowNet	**100 Hz**	**3.96**

1.2. Results from evaluating optical flow predictions using original optical flow ground truths

Considering that the constant velocity assumption may not universally apply across all motion scenarios, our approach that generates additional optical flow ground truths using linear interpolation may raise a concern. However, due to the lack of a dataset with frequent optical flow ground truths, it is impractical to construct a similar dataset ourselves. Given that the DSEC dataset records cars moving forward, linear interpolation provides a reasonable estimation of the actual optical flows based on visualization as illustrated in the supplementary video. To further substantiate our approach, we offer a result of both existing and proposed models' prediction accuracy when we tested them against actual flow ground truths (see Table 2). Synthetic optical flow ground truths generated from linear interpolation are excluded from the evaluation. The accuracy trend in Table 2 remains consistent with the previously reported results. With the availability of suitable datasets in the future, we expect the proposed framework to still be able to capture non-linear motions as it is end-to-end training.

Table 2. Comparison between the existing and proposed models in terms of AEE when using only actual optical flow ground truths from the DSEC dataset (i.e., synthetic ground truths excluded).

Architecture	Prediction Rate	AEE
EV-FlowNet [33]	10 Hz	0.70
Spike-FlowNet [19]	10 Hz	1.16
Adaptive-FlowNet [17]	10 Hz	1.31
LSTM-FlowNet	**100 Hz**	**0.66**
EfficientSpike-FlowNet	**100 Hz**	**2.74**

1.3. Results from evaluating optical flow predictions using both different dataset splitting strategy and original optical flow ground truths

We repeat the experiment by employing both different dataset generation strategies and utilizing the original optical flow ground truths. This is done to validate a trend in prediction accuracy from the existing and proposed models (see Table 3). The result exhibits a trend in the accuracy similar to Table 1, showing about 13% increase in prediction accuracy when using our proposed training methodology with LSTM-FlowNet. Due to the inherent simplicity in the dynamics of SNNs, EfficientSpike-FlowNet has higher AEE than LSTM-FlowNet and the value increases considerably from the results reported in Table 2. This increase

Table 3. Comparison between the existing and proposed models in terms of AEE when using both a different training-testing set splitting strategy and actual optical flow ground truths from the DSEC dataset.

Architecture	Prediction Rate	AEE
EV-FlowNet [33]	10 Hz	0.82
Spike-FlowNet [19]	10 Hz	1.28
Adaptive-FlowNet [17]	10 Hz	1.89
LSTM-FlowNet	**100 Hz**	**0.71**
EfficientSpike-FlowNet	**100 Hz**	**3.94**

can be attributed to changes in the training and testing sets. While a uniform elevation in AEE is seen across all models, the increase in AEE for EfficientSpike-FlowNet is significantly larger than the other models. In-depth analysis is required and will be an interesting topic for future studies to improve lightweight models for temporally dense optical flow prediction in the future.