S2N: Suppression-Strengthen Network for Event-based Recognition under Variant Illuminations

Zengyu Wan, Yang Wang, Ganchao Tan, Yang Cao, and Zheng-Jun Zha

University of Science and Technology of China, Hefei, China

In this supplementary material, we first present the results of ablation studies on each loss term proposed in the Noise Suppression Network (NSN). Secondly, we show the detailed recognition results under various illumination conditions, which differ significantly in terms of contrast and noise levels. Then, we visualize more feature strengthen results to demonstrate the superiority of our method. Next, we compare the feature representation before and after introducing the motion evolution map feature to verify its guiding role. Finally, we report more details of DAVIS346Gait and DAVIS346Character datasets.

1 Contribution of each loss term in NSN

We show the results of the NSN module trained by various combinations of loss functions in Fig. 1. Without considering the noise suppression loss L_N, the results are severely corrupted by the noise, especially in low illumination scenes. Removing the enhancement loss L_E will result in the inability to boost the contrast of the event frames and information loss of effective signal. And removing the consistency loss L_C leads to severe channel imbalances and corrupts the relevant information of each channel.

2 Cross-illumination generalization

In order to verify the generalization of our proposed Suppression-Strengthen Network (S2N), we conduct extensive cross-illumination validation experiments on three datasets, the DVS128Gesture [2], DAVIS346Gait and DAVIS346Character datasets, as show in Table 1 - Table 5, Table 6 - Table 9 and Table 10 - Table 13. The corresponding content of Table 2, Table 3 and Table 4 in main text are **the mean accuracy** of Table 1 - Table 5, Table 6 - Table 9, Table 10 - Table 13, respectively. In addition, we show the results of testing models on the entire dataset, represented as ***all*** in the tables. We can observe that our proposed S2N model outperforms the SOTA methods and the methods combing event denoising and recognition models, which demonstrates that our proposed S2N can effectively suppress the influence of event degradation and strengthen the robustness of feature representation.

2.1 N-ImageNet dataset

We also evaluate our S2N on the N-ImageNet dataset [7], which contains 1,000 classes of objects captured in diverse illumination conditions, to demonstrate the robustness of our method. Without loss of generality, we randomly select 30 categories with 500 samples in each category and test them under four illumination conditions. We compare our S2N model with Sorted Time Surface [1], Event Histogram [8], and DiST [7] representations with ResNet34 recognition model, results as shown in Table 14. We can observe that our S2N still achieves better robustness results in different scenarios. It implies that our proposed S2N can generalize to various recognition tasks.

3 More feature strengthen results

In this part, we evaluate our S2N on different illumination conditions, including fluorescent, led, and natural scenes sampled from the DVS128Gesture dataset. As shown in Fig. 2 - Fig. 4, our method can stably extract the structural features of objects, even in low contrast and noisy conditions. It demonstrates that our proposed S2N model is robust and applicable to variant illumination conditions.

4 The guidance of the motion evolution map

To verify the role of the motion evolution map, we compare the feature representation before and after introducing the motion evolution map, as shown in Fig. 5 - Fig. 6. We can observe that under the guidance of the motion evolution map, the structural features of objects are more complete. Besides, the motion evolution map can stably strengthen the feature representation under variant illumination conditions, illustrating our method's robustness.

5 Dataset description

This section will describe the capture process of the DAVIS346Gait and DAVIS346Character datasets in more details. During the capturing of the DAVIS346Gait dataset, volunteers stand approximately 5m in front of the camera, at 90 degrees to the capturing direction (**no facial data are recorded**), and walk naturally along a straight line. Each sample is captured for a duration of approximately 300 ms. 36 volunteers are involved in the shooting process, and 4,512 samples are obtained, of which half is used as the training set and the rest as the test set. And there is no overlap between the training and test sets. In the process of capturing the DAVIS346Character dataset, we scan the handwritten character from 8 directions, including top to bottom, left to right, right to left, bottom to top, top left to bottom right, top right to bottom left, bottom left to top right, and bottom right to top left. To simulate different illumination scenes, we place three different neutral filters in front of the lens and obtain four scenes with the illumination of 6lux, 15lux, 75lux, and 300lux (no filter). We show some samples from the DAVIS346 character and DAVIS346Gait datasets in Fig. 7 - Fig. 8.

References

1. Alzugaray, I., Chli, M.: Asynchronous Corner Detection and Tracking for Event Cameras in Real Time. IEEE Robot. Autom. Lett. **3** (2018) 2, 8
2. Amir, A., Taba, B., Berg, D., Melano, T., McKinstry, J., Di Nolfo, C., Nayak, T., Andreopoulos, A., Garreau, G., Mendoza, M., Kusnitz, J., Debole, M., Esser, S., Delbruck, T., Flickner, M., Modha, D.: A Low Power, Fully Event-Based Gesture Recognition System. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Honolulu, HI (2017) 1
3. Bi, Y., Chadha, A., Abbas, A., Bourtsoulatze, E., Andreopoulos, Y.: Graph-Based Object Classification for Neuromorphic Vision Sensing. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, Seoul, Korea (South) (2019) 7
4. Carreira, J., Zisserman, A.: Quo vadis, action recognition? a new model and the kinetics dataset. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6299–6308 (2017) 4
5. Gehrig, D., Loquercio, A., Derpanis, K., Scaramuzza, D.: End-to-End Learning of Representations for Asynchronous Event-Based Data. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, Seoul, Korea (South) (2019) 4, 5
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. arXiv:1512.03385 [cs] (2015) 4
7. Kim, J., Bae, J., Park, G., Zhang, D., Kim, Y.M.: N-imagenet: Towards robust, fine-grained object recognition with event cameras. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 2146–2156 (October 2021) 2, 8
8. Maqueda, A.I., Loquercio, A., Gallego, G., Garcia, N., Scaramuzza, D.: Event-Based Vision Meets Deep Learning on Steering Prediction for Self-Driving Cars. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, Salt Lake City, UT (2018) 2, 8
9. Wang, Q., Zhang, Y., Yuan, J., Lu, Y.: Space-Time Event Clouds for Gesture Recognition: From RGB Cameras to Event Cameras. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, Waikoloa Village, HI, USA (2019) 4
10. Wang, Y., Du, B., Shen, Y., Wu, K., Zhao, G., Sun, J., Wen, H.: EV-Gait: Event-Based Robust Gait Recognition Using Dynamic Vision Sensors. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Long Beach, CA, USA (2019) 5
11. Wang, Y., Zhang, X., Shen, Y., Du, B., Zhao, G., Cui Lizhen, L.C., Wen, H.: Event-Stream Representation for Human Gaits Identification Using Deep Neural Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence (2021) 5

Table 1: The comparison results on the **DVS128Gesture dataset**. The model is trained on **S0 (fluorescent led)** and tested on S1 (fluorescent), S2 (natural), S3 (led), and S4 (lab), respectively.

	Training Set	Test Sets				
	S0	S1	S2	S3	S4	all
I3D [4]	**97.00%**	94.80%	89.20%	**93.80%**	**95.30%**	**94.10%**
ResNet34 [6]	90.60%	87.20%	70.80%	78.60%	85.40%	82.60%
PointNet++ [9]	87.40%	81.00%	85.10%	85.20%	85.20%	86.70%
EST [5]	96.60%	**95.10%**	91.00%	92.20%	93.70%	93.80%
VN + I3D	96.40%	93.50%	**93.80%**	93.20%	92.20%	93.90%
YN + I3D	96.10%	94.70%	84.90%	91.00%	94.80%	92.30%
VN + EST	94.70%	94.40%	89.90%	87.40%	90.30%	91.30%
YN + EST	93.90%	91.50%	82.60%	88.10%	92.80%	89.70%
S2N	**98.40%**	**96.40%**	**96.70%**	**97.40%**	**97.50%**	**97.20%**

Table 2: The comparison results on the **DVS128Gesture dataset**. The model is trained on **S1 (fluorescent)** and tested on S0, S2, S3, and S4, respectively.

	Training Set	Test Sets				
	S1	S0	S2	S3	S4	all
I3D	94.70%	96.90%	83.60%	**92.60%**	95.10%	92.70%
ResNet34	87.10%	92.80%	71.80%	80.00%	87.70%	83.90%
PointNet++	89.20%	92.30%	84.40%	84.30%	87.50%	87.60%
EST	95.20%	**97.50%**	90.60%	92.50%	93.20%	**94.00%**
VN +I3D	94.90%	96.20%	**91.80%**	92.20%	92.20%	93.60%
YN +I3D	**95.40%**	96.40%	89.20%	91.30%	**96.60%**	93.70%
VN +EST	90.90%	93.40%	82.90%	80.90%	86.80%	87.10%
YN +EST	94.70%	95.90%	81.90%	88.30%	92.60%	90.80%
S2N	**96.30%**	**97.80%**	**94.30%**	**94.10%**	**96.60%**	**95.80%**

Table 3: The comparison results on the **DVS128Gesture dataset**. The model is trained on **S2 (natural)** and tested on S0, S1, S3, and S4, respectively.

	Training Set	Test Sets				
	S2	S0	S1	S3	S4	all
I3D	94.00%	93.10%	90.20%	91.20%	88.10%	91.40%
ResNet34	92.70%	88.80%	81.40%	86.00%	79.20%	85.70%
PointNet++	87.40%	86.30%	81.40%	83.40%	81.90%	85.00%
EST	93.90%	89.50%	83.20%	88.60%	86.60%	88.10%
VN +I3D	**94.70%**	**93.40%**	89.80%	**93.40%**	88.60%	**92.10%**
YN +I3D	92.50%	**93.40%**	**90.70%**	93.20%	**89.20%**	91.90%
VN +EST	92.20%	90.60%	83.30%	86.80%	86.30%	87.60%
YN +EST	88.70%	90.00%	85.50%	84.40%	84.80%	86.60%
S2N	**95.70%**	**96.40%**	**95.50%**	**96.00%**	**91.00%**	**95.30%**

Table 4: The comparison results on the **DVS128Gesture dataset**. The model is trained on **S3 (led)** and tested on S0, S1, S2, and S4, respectively.

	Training Set	Test Sets				
	S3	S0	S1	S2	S4	all
I3D	**94.10%**	93.60%	90.80%	94.90%	89.50%	93.00%
ResNet34	85.10%	88.70%	84.10%	86.50%	81.70%	85.30%
PointNet++	86.00%	90.80%	85.80%	88.00%	90.80%	87.20%
EST	92.60%	**95.40%**	90.80%	**95.60%**	**92.60%**	**93.30%**
VN +I3D	93.40%	91.70%	89.60%	94.70%	91.50%	92.10%
YN +I3D	93.90%	94.70%	**91.10%**	92.60%	89.90%	92.60%
VN +EST	87.20%	89.00%	85.70%	93.40%	87.90%	88.50%
YN +EST	89.90%	93.40%	88.10%	88.00%	90.80%	90.00%
S2N	**94.70%**	**96.70%**	**93.40%**	**95.50%**	**93.70%**	**94.80%**

Table 5: The comparison results on the **DVS128Gesture dataset**. The model is trained on **S4 (lab)** and tested on S0, S1, S2, and S3, respectively.

	Training Set	Test Sets				
	S4	S0	S1	S2	S3	all
I3D	**97.50%**	90.30%	**87.00%**	75.00%	78.40%	84.50%
ResNet34	87.00%	79.40%	79.90%	56.80%	61.20%	72.20%
PointNet++	91.40%	85.20%	85.80%	**77.10%**	78.70%	83.30%
EST	95.30%	**92.00%**	86.60%	76.40%	**79.50%**	**85.30%**
VN +I3D	94.20%	86.50%	86.80%	74.10%	76.20%	83.00%
YN +I3D	97.30%	83.70%	84.10%	70.50%	77.90%	81.90%
VN +EST	92.80%	80.90%	82.00%	69.60%	74.10%	79.10%
YN +EST	97.30%	86.90%	82.50%	61.50%	69.10%	78.30%
S2N	**97.30%**	**97.20%**	**96.10%**	**92.10%**	**91.50%**	**94.70%**

Table 6: The comparison results on the **DAVIS346Gait dataset**. The model is trained on **L0 (300lux)** and tested on L1 (120lux), L2 (15lux), and L3 (6lux), respectively.

	Training Set	Test Sets			
	L0	L1	L2	L3	all
EV-Gait-3DGraph [11]	86.75%	48.24%	33.39%	**25.81%**	25.89%
EST [5]	99.00%	12.80%	5.40%	3.10%	29.70%
VN + EV-Gait-IMG [10]	97.80%	90.60%	65.90%	21.00%	68.60%
YN + EV-Gait-IMG	**99.10%**	**95.90%**	**84.40%**	21.30%	**75.00%**
VN + EST	97.80%	82.80%	22.90%	3.20%	51.80%
YN + EST	99.20%	43.60%	14.10%	2.00%	39.50%
S2N	**99.70%**	**95.70%**	**89.60%**	**71.10%**	**88.40%**

Table 7: The comparison results on the **DAVIS346Gait dataset**. The model is trained on **L1 (120lux)** and tested on L0, L2, and L3, respectively.

	Training Set	Test Sets			
	L1	L0	L2	L3	all
EV-Gait-3DGraph	43.60%	9.29%	36.97%	29.08%	29.19%
EST	98.60%	52.70%	41.00%	2.90%	48.60%
VN + EV-Gait-IMG	99.10%	82.40%	**97.70%**	30.50%	56.80%
YN + EV-Gait-IMG	**99.80%**	**92.80%**	**97.70%**	**39.10%**	**82.30%**
VN + EST	97.00%	65.50%	29.60%	2.90%	49.10%
YN + EST	98.10%	80.40%	73.20%	3.90%	63.60%
S2N	**100.00%**	**95.20%**	**98.70%**	**85.60%**	**94.30%**

Table 8: The comparison results on the **DAVIS346Gait dataset**. The model is trained on **L2 (15lux)** and tested on L0, L1 and L3, respectively.

	Training Set	Test Sets			
	L2	L0	L1	L3	all
EV-Gait-3DGraph	26.29%	5.50%	10.30%	21.64%	21.66%
EST	**98.70%**	19.00%	51.90%	19.10%	47.50%
VN + EV-Gait-IMG	99.20%	72.80%	**97.90%**	**68.20%**	**84.60%**
YN + EV-Gait-IMG	99.70%	**78.70%**	94.30%	27.40%	75.00%
VN + EST	91.30%	23.00%	50.80%	19.10%	47.30%
YN + EST	97.70%	44.90%	81.20%	8.00%	58.20%
S2N	**99.80%**	**92.30%**	**99.10%**	**93.70%**	**96.30%**

Table 9: The comparison results on the **DAVIS346Gait dataset**. The model is trained on **L3 (6lux)** and tested on L0, L1, and L2, respectively.

	Training Set	Test Sets			
	L3	L0	L1	L2	all
EV-Gait-3DGraph	19.86%	4.99%	5.92%	10.22%	19.64%
EST	97.80%	2.50%	3.20%	16.80%	29.80%
VN + EV-Gait-IMG	**99.80%**	15.30%	33.60%	56.70%	51.60%
YN + EV-Gait-IMG	97.30%	**26.20%**	**38.40%**	**59.70%**	**55.60%**
VN + EST	88.10%	2.50%	2.90%	13.70%	26.50%
YN + EST	93.20%	2.50%	4.10%	5.60%	26.20%
S2N	**99.80%**	**76.80%**	**89.60%**	**95.00%**	**90.30%**

Table 10: The comparison results on the **DAVIS346Character dataset**. The model is trained on **L0 (300lux)** and tested on L1 (120lux), L2 (15lux), and L3 (6lux), respectively.

	Training Set	Test Sets			
	L0	L1	L2	L3	all
ResNet34	**97.70%**	4.80%	2.80%	2.80%	27.20%
GCNN [3]	76.47%	45.04%	**31.25%**	24.40%	24.45%
EST	97.00%	14.00%	2.80%	3.80%	29.70%
VN + ResNet34	96.50%	29.10%	4.50%	9.10%	35.10%
YN + ResNet34	95.10%	22.20%	3.40%	2.80%	31.30%
VN + EST	96.60%	**59.30%**	4.60%	**30.40%**	**47.90%**
YN + EST	96.00%	35.40%	4.10%	3.20%	35.00%
S2N	**98.00%**	**97.60%**	**73.20%**	**46.70%**	**78.80%**

Table 11: The comparison results on the **DAVIS346Character dataset**. The model is trained on **L1 (120lux)** and tested on L0, L2, and L3, respectively.

	Training Set	Test Sets			
	L1	L0	L2	L3	all
ResNet34	**97.80%**	79.30%	82.80%	20.00%	69.90%
GCNN	35.71%	29.93%	27.64%	22.03%	22.21%
EST	97.50%	94.30%	**84.40%**	**86.40%**	**90.70%**
VN + ResNet34	95.40%	93.10%	64.30%	48.50%	75.40%
YN + ResNet34	91.50%	77.00%	53.90%	44.30%	66.80%
VN + EST	96.10%	**95.70%**	74.30%	66.90%	83.50%
YN + EST	94.50%	53.60%	47.60%	9.90%	51.30%
S2N	**97.60%**	**97.80%**	**85.30%**	**84.90%**	**91.50%**

Table 12: The comparison results on the **DAVIS346Character dataset**. The model is trained on **L2 (15lux)** and tested on L0, L1, and L3, respectively.

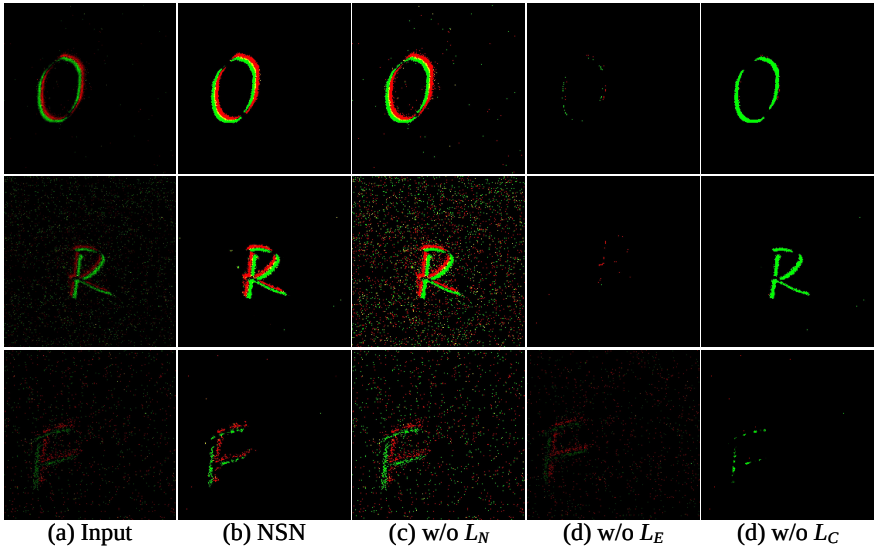
	Training Set	Test Sets			
	L2	L0	L1	L3	all
ResNet34	**97.20%**	82.90%	97.20%	84.80%	90.50%
GCNN	5.48%	3.85%	4.85%	4.99%	5.03%
EST	95.20%	80.40%	**95.70%**	**88.60%**	**90.00%**
VN + ResNet34	82.70%	73.50%	85.20%	64.20%	76.40%
YN + ResNet34	77.00%	40.20%	72.20%	63.80%	63.20%
VN + EST	89.50%	**89.20%**	93.00%	72.10%	86.20%
YN + EST	76.90%	40.80%	77.40%	40.60%	58.50%
S2N	**97.20%**	**95.90%**	**97.00%**	**92.50%**	**95.60%**

Table 13: The comparison results on the **DAVIS346Character dataset**. The model is trained on **L3 (6lux)** and tested on L0, L1, and L2, respectively.

	Training Set	Test Sets			
	L3	L0	L1	L2	all
ResNet34	94.10%	72.00%	75.80%	72.00%	78.60%
GCNN	4.40%	3.22%	3.84%	3.83%	4.50%
EST	**94.80%**	**86.00%**	**90.60%**	**83.40%**	**89.00%**
VN + ResNet34	76.30%	67.90%	73.20%	57.10%	68.70%
YN + ResNet34	82.80%	38.90%	40.60%	46.20%	52.20%
VN + EST	82.30%	71.80%	79.50%	75.30%	77.50%
YN + EST	86.00%	13.20%	28.60%	48.80%	43.90%
S2N	**97.50%**	**84.90%**	**92.80%**	**93.40%**	**92.10%**

Table 14: The comparison results on the **N-ImageNet dataset**. The model is tested on brightness 6, 7, 8 and 9, which corresponding to 12.75 lux, 23.38 lux, 95.50 lux and 111.00 lux, respectively.

	Orig.	Brightness			
		6	7	8	9
Sorted Time Surfece [1]	40.70%	32.20%	31.20%	32.70%	32.90%
Event Histogram [8]	41.40%	32.60%	32.00%	33.70%	33.90%
DiST [7]	**46.50%**	**35.70%**	**34.80%**	**36.30%**	**36.20%**
S2N	**46.70%**	**37.70%**	**37.20%**	**38.20%**	**38.40%**



(a) Input (b) NSN (c) w/o L_N (d) w/o L_E (d) w/o L_C

Fig. 1: Ablation study of the contribution of each loss term (L_N: noise suppression loss, L_E: enhancement loss, L_C: consistency loss). 'w/o' means without.

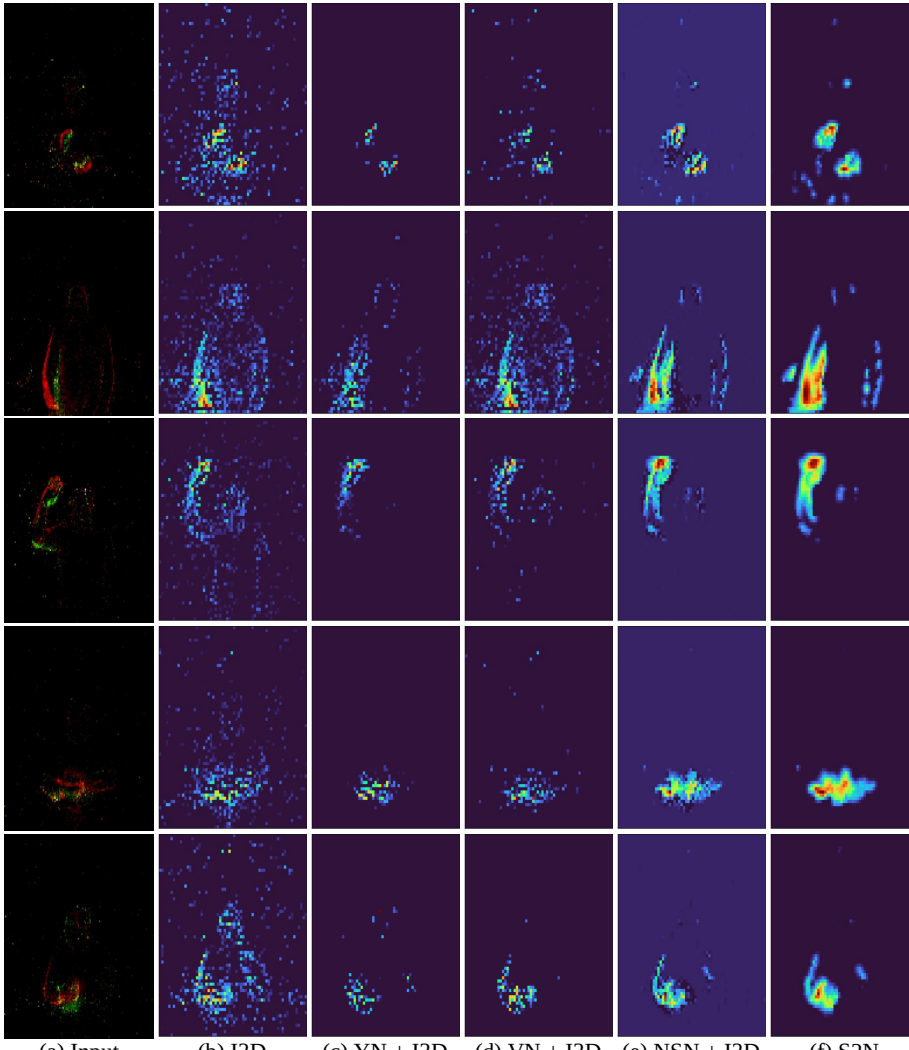


Fig. 2: The feature map visualization of different methods, including I3D, YN + I3D, VN + I3D, NSN + I3D and our model S2N (NSN + FSN) on **fluorescent** scene of DVS128Gesture dataset.

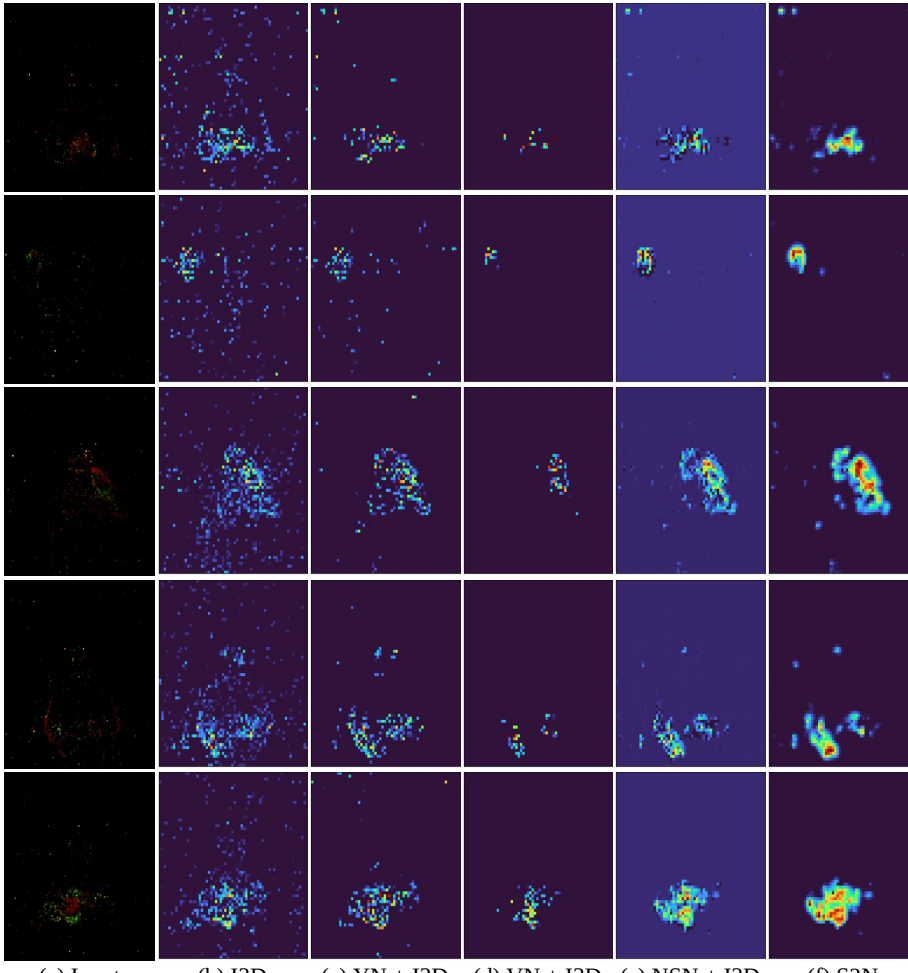


Fig. 3: The feature map visualization of different methods, including I3D, YN + I3D, VN + I3D, NSN + I3D and our model S2N (NSN + FSN) on **led** scene of DVS128Gesture dataset.

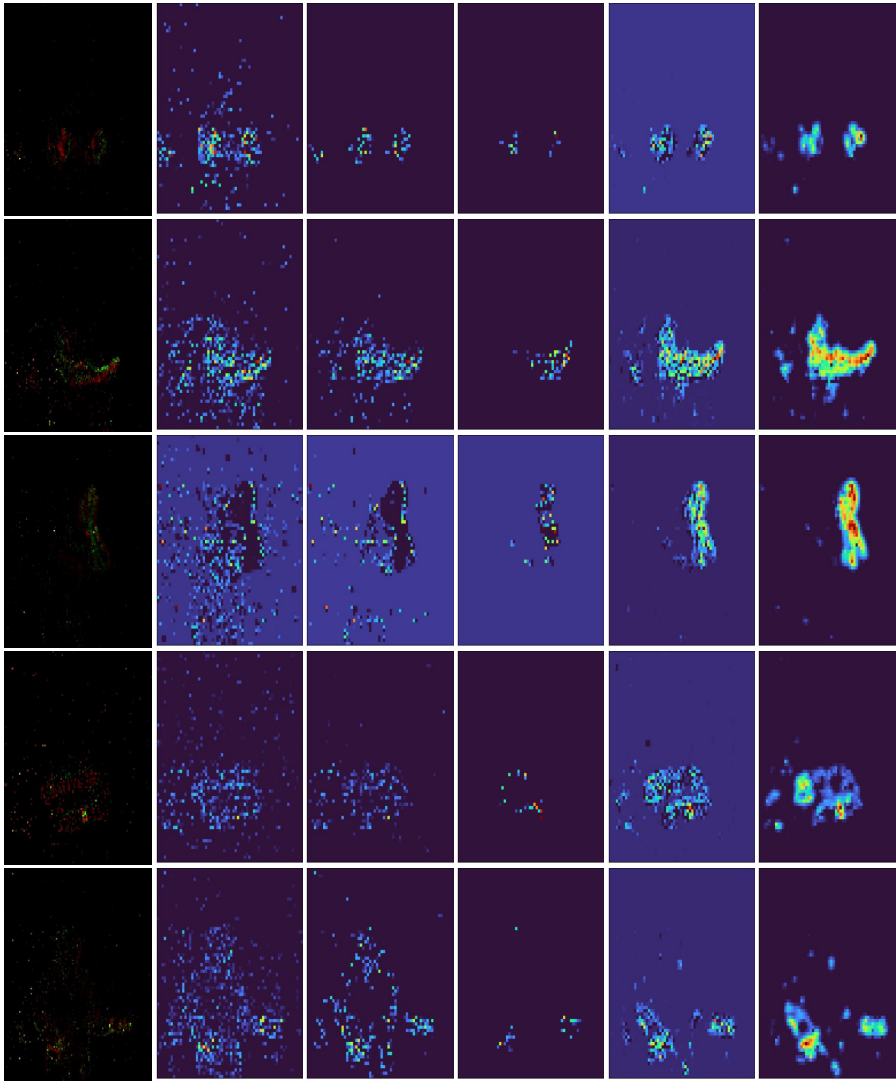
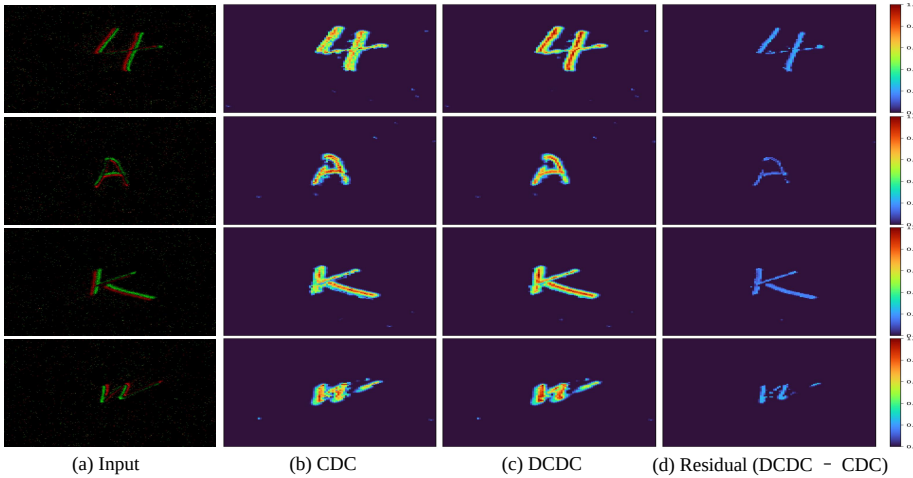


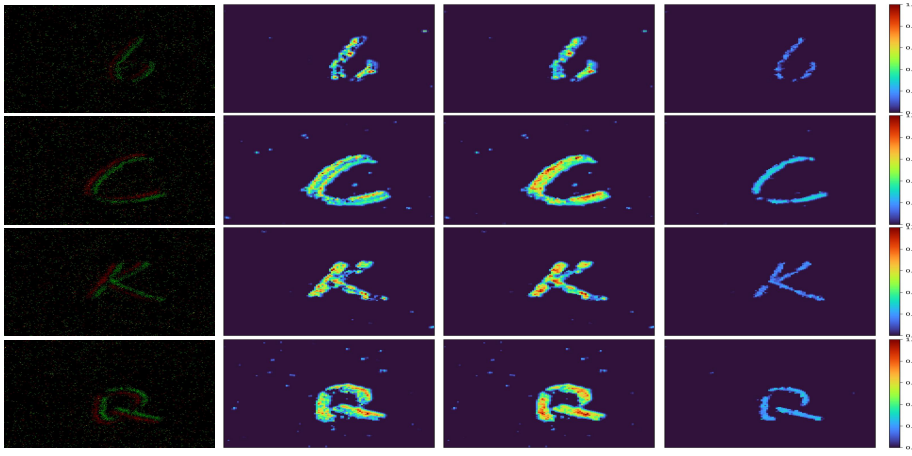
Fig. 4: The feature map visualization of different methods, including I3D, YN + I3D, VN + I3D, NSN + I3D and our model S2N (NSN + FSN) on **natural** scene of DVS128Gesture dataset.

(a) L0 scene

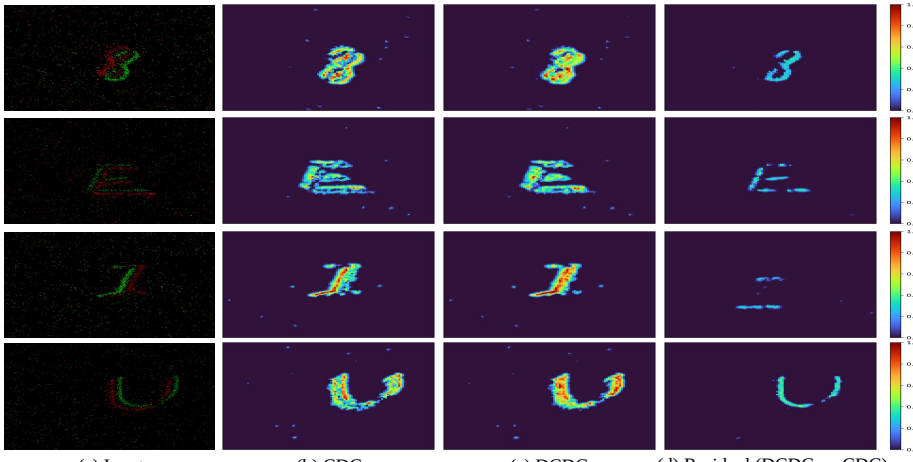


(b) L1 scene

Fig. 5: The guiding role of the motion evolution map under variant illumination scenes. The illumination of L0, L1, L2, and L3 is 300lux, 120lux, 15lux, and 6lux, respectively.

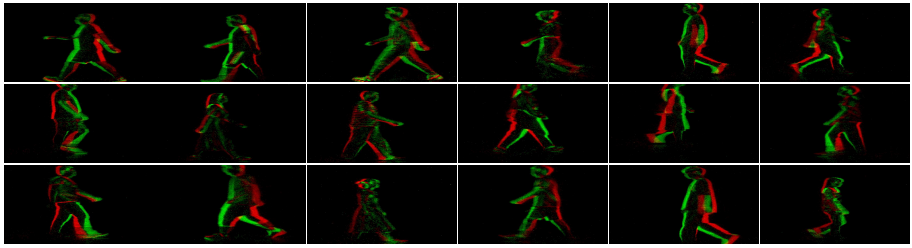


(a) L2 scene

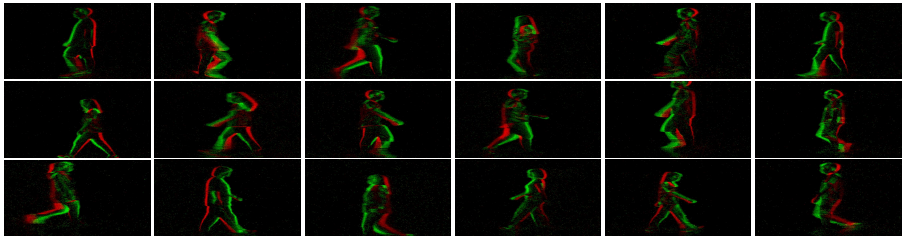


(b) L3 scene

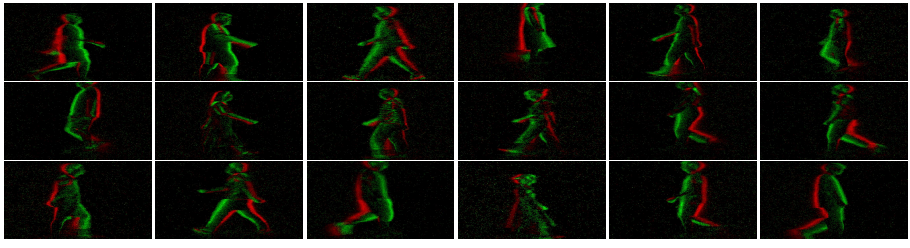
Fig. 6: The guiding role of the motion evolution map under variant illumination scenes. The illumination of L0, L1, L2, and L3 is 300lux, 120lux, 15lux, and 6lux, respectively.



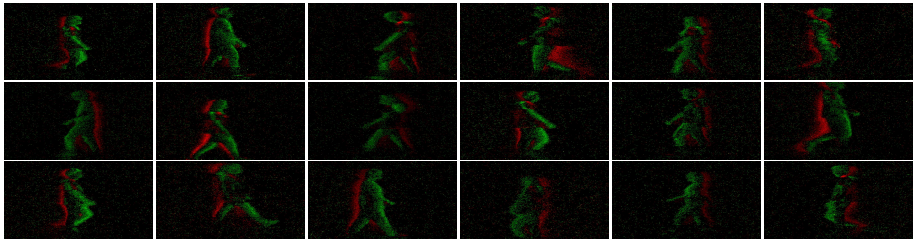
(a) L0 scene



(b) L1 scene

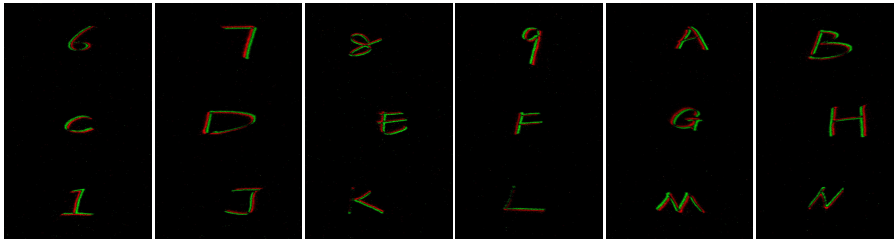


(c) L2 scene

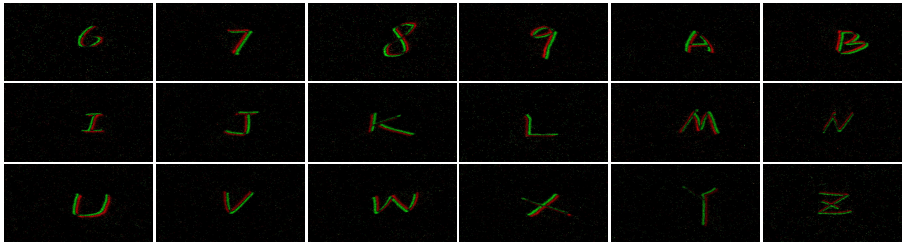


(d) L3 scene

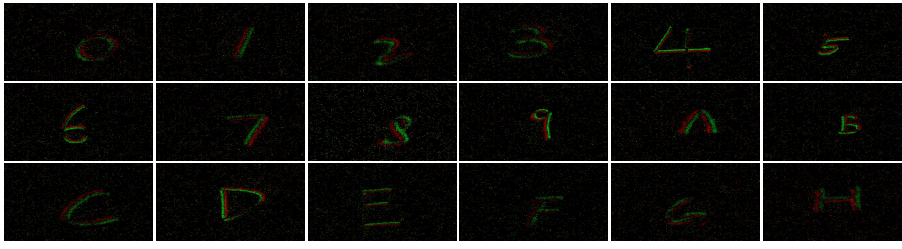
Fig. 7: Samples of **DAVIS346Gait Dataset**. The illumination of L0, L1, L2, and L3 is 300lux, 120lux, 15lux, and 6lux, respectively.



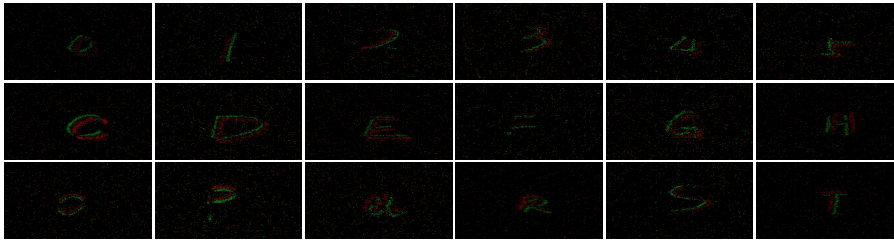
(a) L0 scene



(b) L1 scene



(c) L2 scene



(d) L3 scene

Fig. 8: Samples of **DAVIS346Character Dataset**. The illumination of L0, L1, L2, and L3 is 300lux, 120lux, 15lux, and 6lux, respectively.