Supplemental Material

Nuo Chen[1], Chao Xiao[1], Yimian Dai[2], Shiman He[1], Miao Li[1], Wei An[1]✉
[1]National University of Defense Technology, [2]Nankai University
{chennuo97,anwei}@nudt.edu.cn

1. Evaluation Metrics

In this work, we evaluate the event-by-event segmentation performance of the algorithm using Intersection over Union (IoU) and accuracy (ACC), and assess the algorithm's localization performance using probability of detection (P_d) and false alarm rate (F_a).

Intersection over Union. Intersection over Union (IoU) is a pixel-level evaluation metric. It's calculated by the ratio of the intersection and union events between the prediction and the label:

$$IoU = \frac{E_{inter}}{E_{union}}, \tag{1}$$

where E_{inter} and E_{union} represent the interaction events and union events, respectively.

Accuracy. Accuracy (ACC) is another pixel-level evaluation metric that measures the proportion of correctly classified events in a segmentation model:

$$ACC = \frac{E_{right}}{E_{target}}, \tag{2}$$

where E_{right} represents the accurately segmented target events by the model and E_{target} stands for all target events.

Probability of Detection. Probability of detection (P_d) is a target-level evaluation metric. It measures the ratio between the number of correctly predicted targets $T_{correct}$, and the total number of targets T_{all}. P_d is defined as follows:

$$P_d = \frac{T_{correct}}{T_{all}} \times 100\%, \tag{3}$$

False-Alarm Rate. False-Alarm Rate (F_a) is another target-level evaluation metric. It measures the ratio of falsely predicted pixels P_{false} to all image pixels P_{all}, as follows:

$$F_a = \frac{P_{false}}{P_{all}}. \tag{4}$$

Note that for target-level evaluation metrics (i.e., P_d and F_a), it is necessary to convert the event-by-event segmentation results into boundingbox results before conducting the evaluation.

2. Results on Other Datasets

Quantitative results in Table 1 show that our method achieves the best performance on the NeRDD and EvDET200K datasets.

Table 1. Performance comparison on other datasets.

Methods	EvDET200K-UAV				NeRDD			
	IoU	ACC	P_d	F_a	IoU	ACC	P_d	F_a
YOLOV10-S	<u>59.16</u>	<u>89.01</u>	<u>90.21</u>	4.21	56.86	86.33	89.32	5.87
RVT	56.89	88.78	88.39	6.32	<u>57.21</u>	<u>86.63</u>	<u>90.26</u>	3.82
EMS-YOLO	52.12	84.32	86.72	7.23	54.19	83.98	87.39	8.37
COSeg	54.83	85.86	87.38	<u>0.98</u>	55.82	84.38	83.78	<u>1.56</u>
Ours	**60.32**	**89.33**	**92.43**	**0.51**	**59.12**	**88.35**	**93.68**	**0.75**

The **bold** and the <u>underline</u> represent the best and second-best performance, respectively. EvDET200K-UAV is the UAV-only subset of EvDET200K.

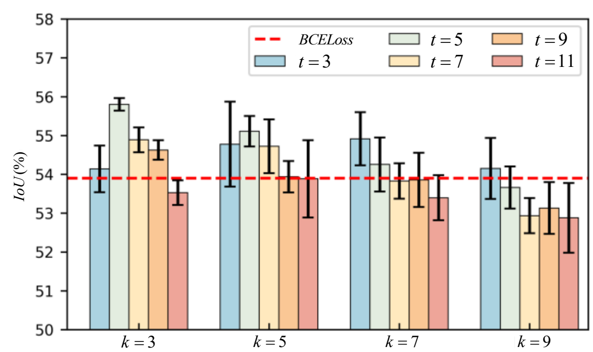

Figure 1. Results of different hyperparameters for the STC loss. The height of each colored bar corresponds to the average IoU, while the range of the black bar represents the variation between the maximum and minimum IoU values.

3. Optimal Parameters of STC Loss

To select the optimal hyperparameters for the loss function, we compare the performance of different hyperparameter combinations. We conduct five experiments with different random seeds and report the average performance. As shown in Fig. 1, the model achieves the best performance when k and t are set to 3 and 5, respectively. As k and

t increase, the model performance decreases and the variance grows. This is because a larger neighborhood range may incorrectly include many irrelevant events as supporting events, leading to inaccurate spatiotemporal correlation calculations.

4. Visualization of GDSCA Module.

As shown in Fig. 2, the GDSCA module enables the network to more precisely distinguish the target from complex interferences, such as background clutter and continuous curves formed by fixed noise.

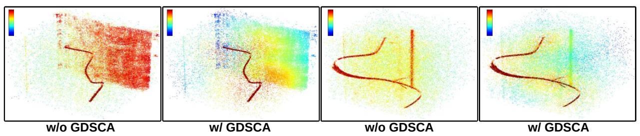

Figure 2. Comparison of feature maps with and without GDSCA.