

Exploiting Completeness and Uncertainty of Pseudo Labels for Weakly Supervised Video Anomaly Detection - Supplementary Material

1. Results for Each Category of Anomalies

Since the UCF-Crime [2] dataset provides anomaly category annotations, we further evaluate the performance on each category of anomalies. As shown in Figure 1, we compare the performance of our method with our baseline model and the state-of-the-art two-stage self-training method MIST [1]. Our method achieves the best results on 9 out of 13 types of anomalies, especially on Arson and Burglary achieving 9.62% and 14.36% improvement respectively compared to MIST.

In Table 1, for each type of abnormal video, we summarize the total number of frames, the average duration of abnormal events (i.e., the average number of frames in an abnormal event), the number of videos containing multi-segment abnormal events (multiple discontinuous abnormal intervals) and the performance gain compared to baseline and MIST [1]. For the multi-segment or long-term abnormal events in Table 1, our method can achieve significant performance improvement compared to baseline and MIST [1], such as types Arson, Burglary, Explosion, Robbery, Shooting, Shoplifting and Stealing. However, our method suffers performance degradation regarding types of Assault and Vandalism. We found that our method has a high false positive rate on these two types of anomalous events, and the reason may be that the model over-improves the completeness of pseudo labels for them. We will further improve our method in future work to address this issue.

2. Additional Ablation Results

In Figure 2, we show the ROC curves of several variants of our model. The baseline model (blue curve) directly uses the MIL-based method as a pseudo label generator and then uses all the pseudo labels to train a clip-level classifier, which means that the completeness and uncertainty properties of pseudo labels are not considered. “With Completeness” (orange curve) denotes that we take into account the completeness via the multi-head classifier constrained by a diversity loss as the pseudo label generator. We observe that “With Completeness” has a higher true positive rate (TPR) than the baseline when false positive rate (FPR) is lower than 0.4. This indicates that the completeness property of

pseudo labels can effectively reduce the miss alarm rate of anomalies. “With Uncertainty” (green curve) denotes that we exploit the uncertainty property and adopt an iterative uncertainty aware pseudo label refinement strategy. The results show that uncertainty property can also bring performance improvements over the baseline. Finally, our full model (red curve) combines the advantages of completeness and uncertainty properties, leading to the best results among these variants.

3. Additional Qualitative Results

As shown in Figure 3, we compare the frame-level anomaly scores predicted by our method and MIST [1] for different types of anomalous events on the UCF-Crime dataset. On the video Burglary, our method can accurately predict two abnormal intervals (Figure 3(b)), and predict high anomaly scores for abnormal frames. However, MIST [1] only predicts high anomaly scores for part of abnormal frames (Figure 3(a)). On the video Robbery, MIST [1] predicts low anomaly scores for all frames, missing all anomalous frames (Figure 3(c)). In contrast, our method can completely detect all abnormal frames with high anomaly scores (Figure 3(d)).

References

- [1] Jia-Chang Feng, Fa-Ting Hong, and Wei-Shi Zheng. MIST: Multiple instance self-training framework for video anomaly detection. In *CVPR*, pages 14009–14018, 2021. 1
- [2] Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos. In *CVPR*, pages 6479–6488, 2018. 1

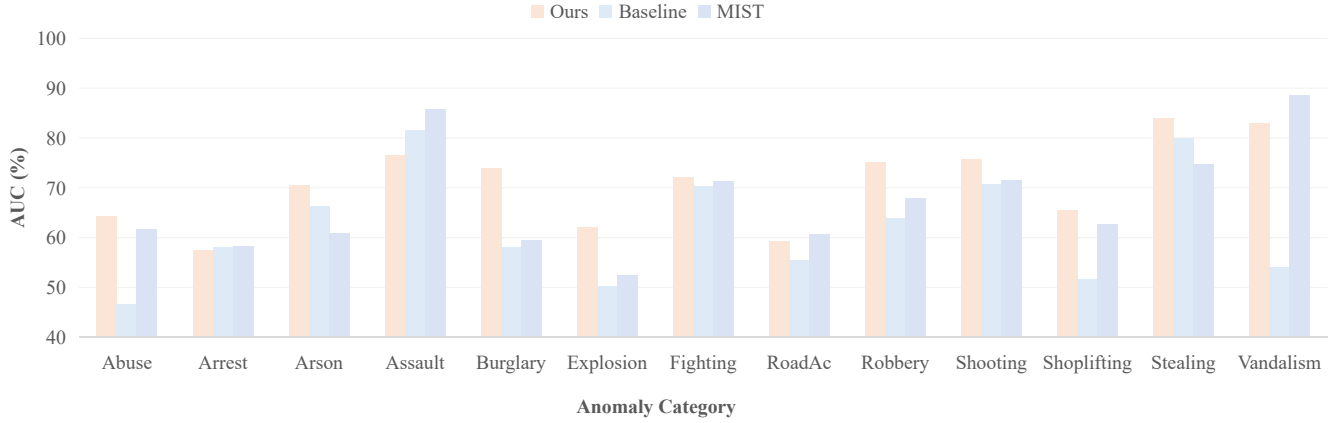


Figure 1. Results on each type of anomalies on the UCF-Crime dataset.

Category	Abuse	Arrest	Arson	Assault	Burglary	Explosion	Fight	RoadAc	Robbery	Shoot	Shoplift	Steal	Vandalism
Frame Number	2956	33624	27888	26561	76524	64997	12292	26525	8327	76187	76134	19811	11077
Anomaly Duration	81	1564	907	2840	1238	363	887	97	763	427	358	1201	415
Multi Anomaly	0	0	1	1	2	1	0	0	0	2	4	2	3
AUC Gain (B)	17:67	0:74	4:08	4:92	15:9	11:83	1:89	3:77	11:16	5:0	13:83	3:99	28:88
AUC Gain (M)	2:55	0:83	9:62	9:22	14:36	9:63	0:81	1:45	7:24	4:06	2:67	9:29	5:58

Table 1. Detailed statistics for each type of anomalies on the UCF-Crime dataset. “Anomaly Duration” refers to the average number of abnormal frames in each type of abnormal video. “Multi Anomaly” denotes the number of videos containing multiple discontinuous abnormal events. “AUC Gain (B)” and “AUC Gain (M)” denote the AUC performance gain of our method on each class of anomalies compared with Baseline and MIST, respectively.

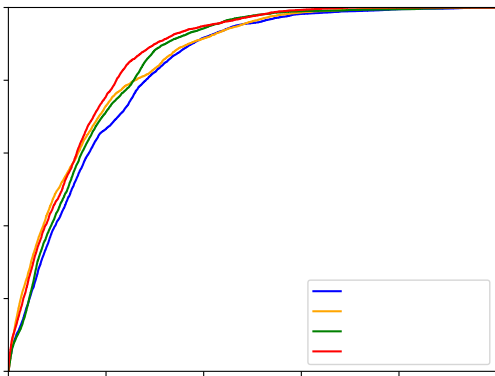


Figure 2. ROC curves of the variants of the proposed model. “With Completeness” and “With Uncertainty” denote adding completeness and uncertainty properties to the baseline model, respectively. “Ours” represents our full model that exploits both completeness and uncertainty properties.

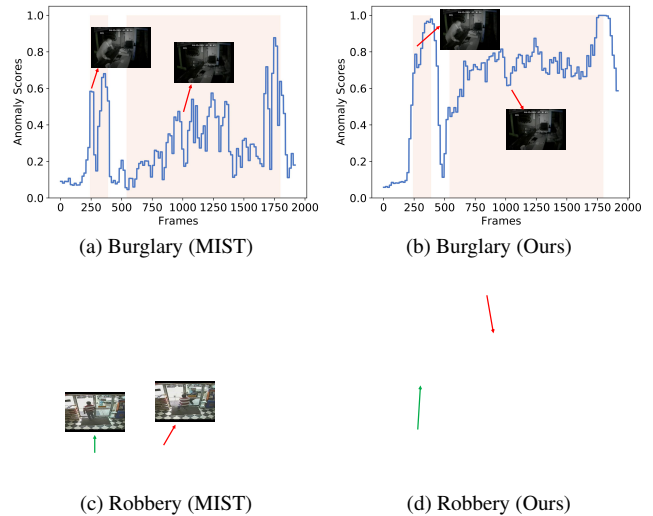


Figure 3. Qualitative results on two UCF-Crime (Burglary and Robbery) test videos. The pink background denotes frames of abnormal events. The blue curve indicates the predicted anomaly scores of each video frame. The left column is the result obtained by the MIST method, and the right column shows the result of our method.