Weakly Supervised Video Anomaly Detection with Anomaly-Connected Components and Intention Reasoning

Supplementary Material

6. Ablation Study on UCF-Crime Dataset

To validate the utility of the proposed LAS-VAD, we further conduct ablation studies on UCF-Crime dataset. We analyze the impact of our model's components on detection performance, with results presented in Tab. 7. Specifically, we examine the utility of our intention awareness mechanism (IAM), anomaly-connected components (ACC), and anomaly attribute clues (AAT) to elucidate their individual contributions to the model's overall performance. As shown in Tab. 7, incorporating each component yields performance gains across all metrics, and the full model achieves superior performance compared to its ablated variants.

ATT	ACC	IAM	mAP@IoU					
			0.1	0.2	0.3	0.4	0.5	AVG
✘	✘	✘	11.67	7.70	6.24	4.48	2.85	6.59
✔	✘	✘	14.08	10.92	9.55	6.99	4.96	9.30
✘	✔	✘	17.95	13.77	12.81	9.63	6.98	12.23
✘	✘	✔	18.36	13.97	12.94	9.61	7.15	12.41
✔	✔	✘	20.81	16.85	14.68	10.67	7.90	14.18
✔	✘	✔	20.74	16.68	14.55	10.62	7.86	14.09
✘	✔	✔	21.34	18.99	15.36	10.91	8.05	14.93
✔	✔	✔	**22.07**	**19.96**	**16.18**	**11.24**	**8.64**	**15.62**

Table 7. Ablation Studies on UCF-Crime. "IAM" denotes the intention awareness module, "ACC" denotes the anomaly-connected components, and "AAT" denotes anomaly attribute clues.

7. Ablation Study on Coarse-grained VAD

In this section, we further investigate the impact of our proposed components on coarse-grained VAD performance, with results reported in Tab. 8. We observe similar results both on the XD-Violence and UCF-Crime datasets where our proposed three modules all yield positive gains and the full model achieves optimal performance.

ATT	ACC	IAM	AP(%)
✘	✘	✘	84.06
✔	✘	✘	84.72
✘	✔	✘	85.94
✘	✘	✔	85.85
✔	✔	✘	86.64
✔	✘	✔	86.93
✘	✔	✔	87.50
✔	✔	✔	**87.92**

(a) XD-Violence

ATT	ACC	IAM	AUC(%)
✘	✘	✘	87.75
✔	✘	✘	88.39
✘	✔	✘	89.41
✘	✘	✔	89.37
✔	✔	✘	89.76
✔	✘	✔	89.85
✘	✔	✔	90.32
✔	✔	✔	**90.86**

(b) UCF-Crime

Table 8. Ablation Studies about coarse-grained prediction with CLIP features on XD-Violence.

8. Inference Time Analysis

In this section, we further analyze the average inference time (in seconds) of each video for fine-grained predictions on XD-Violence and UCF-Crime, and compare results with representative methods with a single Nvidia 3090 GPU, with results presented in Tab. 9. We observe that our method outperforms the SOTA LEC-VAD and ReFLIP in terms of speed. When compared to VadCLIP, it incurs only a marginal increase in processing time, yet achieves substantial improvements in performance.

Methods	XD-Violence	UCF-Crime
VadCLIP	0.26	0.49
ReFLIP	0.45	0.71
LEC-VAD	0.29	0.54
LAS-VAD (Ours)	0.28	0.51

Table 9. Runtime comparisions of each video with a single Nvidia 3090 GPU during inference (in seconds).

9. Attribute Information Generated by LLMs

Tab. 10 presents detailed information regarding the attribute descriptions of anomaly categories used in this paper. These descriptive attributes are generated by prompting large language models (LLMs) with the instruction: "Please describe the attribute characteristics of the following anomalies within 50 characters: [X]" to LLMs, where [X] is replaced by the specific name of each anomaly category. The character constraint is deliberately set to balance information richness and conciseness, ensuring that the generated content remains focused on core attributes without excessive redundancy.

Notably, from the detailed entries in Tab. 10, we can clearly observe that the generated anomaly attribute information exhibits a high degree of comprehensiveness. Specifically, these descriptions not only capture the intrinsic behavioral features of each anomaly but also subtly reflect contextual cues and potential consequences associated with their occurrence. This multi-dimensional coverage of attributes provides the proposed LAS-VAD model with more nuanced semantic clues. As a result, the model is better equipped to distinguish between anomalous and normal behaviors, effectively reducing ambiguities in recognition and thereby enabling more accurate and efficient identification of anomalous actions.

Dataset	Attribute Description
XD-Violence	"normal": "regular surroundings, no chaotic elements, peaceful scene" "fighting": "people scuffling, physical clashes, aggressive gestures, pushing/hitting actions, angry facial expressions" "shooting": "visible gun, muzzle flash, people fleeing in panic, possible bullet traces, tense postures" "riot": "chaotic crowd, broken property/windows, aggressive mob actions, scattered debris, yelling people" "abuse": "aggressive physical contact (grabbing/hitting), victim's distressed face, tense body language, intimidating gestures" "car accident": "crushed vehicle bodies, scattered debris, bent metal, possible injured people nearby, damaged parts" "explosion": "flames, thick smoke, flying debris, damaged structures, bright flash at blast site"
UCF-Crime	"normal": "regular surroundings, no chaotic elements, peaceful scene" "abuse": "aggressive physical contact (grabbing/hitting), victim's distressed face, tense body language, intimidating gestures" "arrest": "legal detention of suspects by law enforcement authorities for violating criminal or civil laws" "arson": "intentional act of setting fire to buildings, property, or vegetation, causing destruction or danger" "assault": "unlawful physical attack or threat of attack on an individual, resulting in potential harm" "burglary": "illegal entry into a premises (residence/business) to steal property or commit other crimes" "explosion": "flames, thick smoke, flying debris, damaged structures, bright flash at blast site" "fighting": "people scuffling, physical clashes, aggressive gestures, pushing/hitting actions, angry facial expressions." "roadAccidents": "unexpected incidents involving vehicles on roads, leading to injuries, deaths, or property damage" "robbery": "forcible seizure of others' property through violence, threats, or intimidation in public or private spaces" "shooting": "visible gun, muzzle flash, people fleeing in panic, possible bullet traces, tense postures", "shoplifting": "secret theft of merchandise from retail stores without payment, often done discreetly" "stealing": "illegal taking of another person's belongings without permission or consent, violating property rights" "vandalism": "malicious and intentional damage or defacement of public or private property, including graffiti or destruction"

Table 10. Generated attribution information of different anomaly categories by LLM.