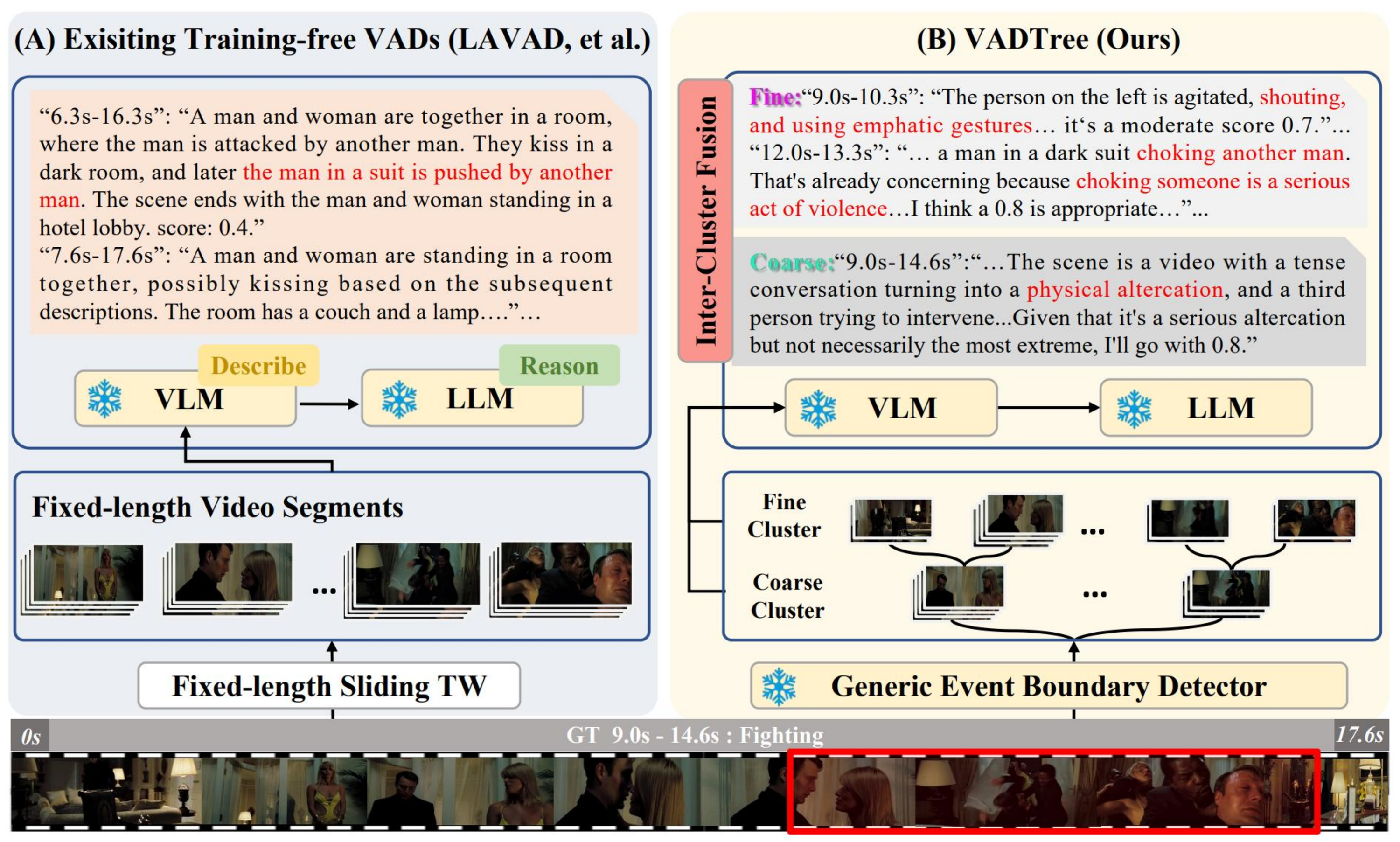
VADTree: Explainable Training-Free Video Anomaly Detection via Hierarchical Granularity-Aware Tree

Wenlong Li, Yifei Xu✉, Yuan Rao, Zhenhua Wang, Shuiguang Deng

Introduction

1. Key Challenges in Video Anomaly Detection

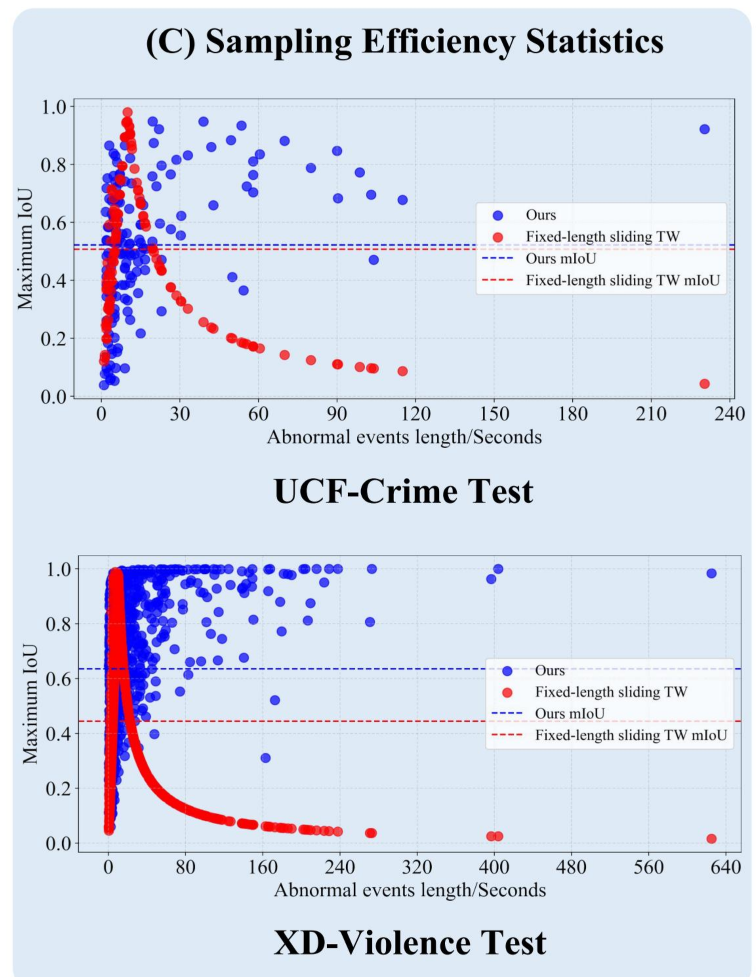
- **Fixed-Length Sampling Limitation**: Clash with variable anomaly durations, causing misalignment. and discontinuity.
- **Complex Event Modeling Gap:** Conflicts between temporal scales hinder contextual reasoning for extended anomalies.

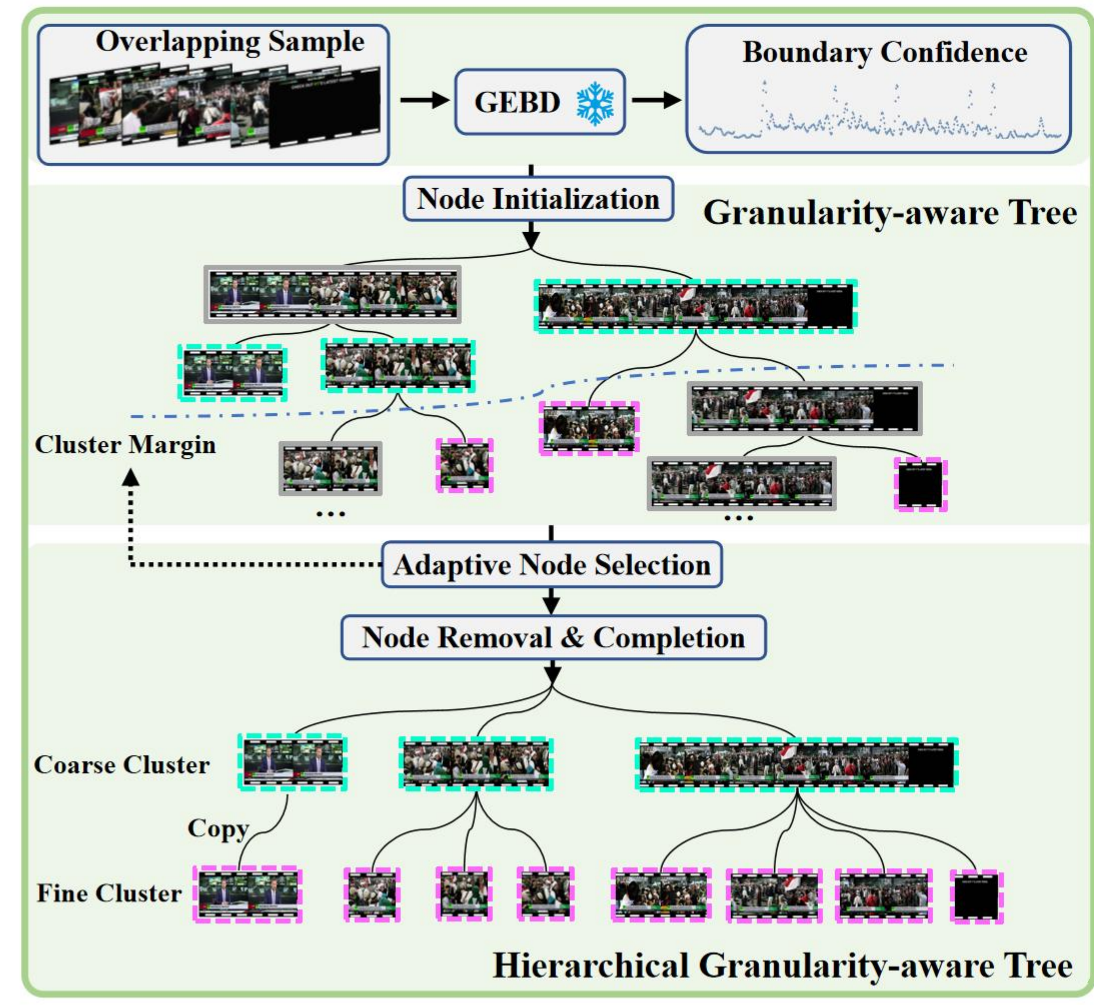


2. Core Contributions

- **VADTree:** Tree structure driven video anomaly detection.
- **Hierarchical Granularity-Aware Tree (HGTree):** Constructs multi-granularity event nodes from pre-trained knowledge for robust representation.

Hierarchical Granularity-Aware Tree

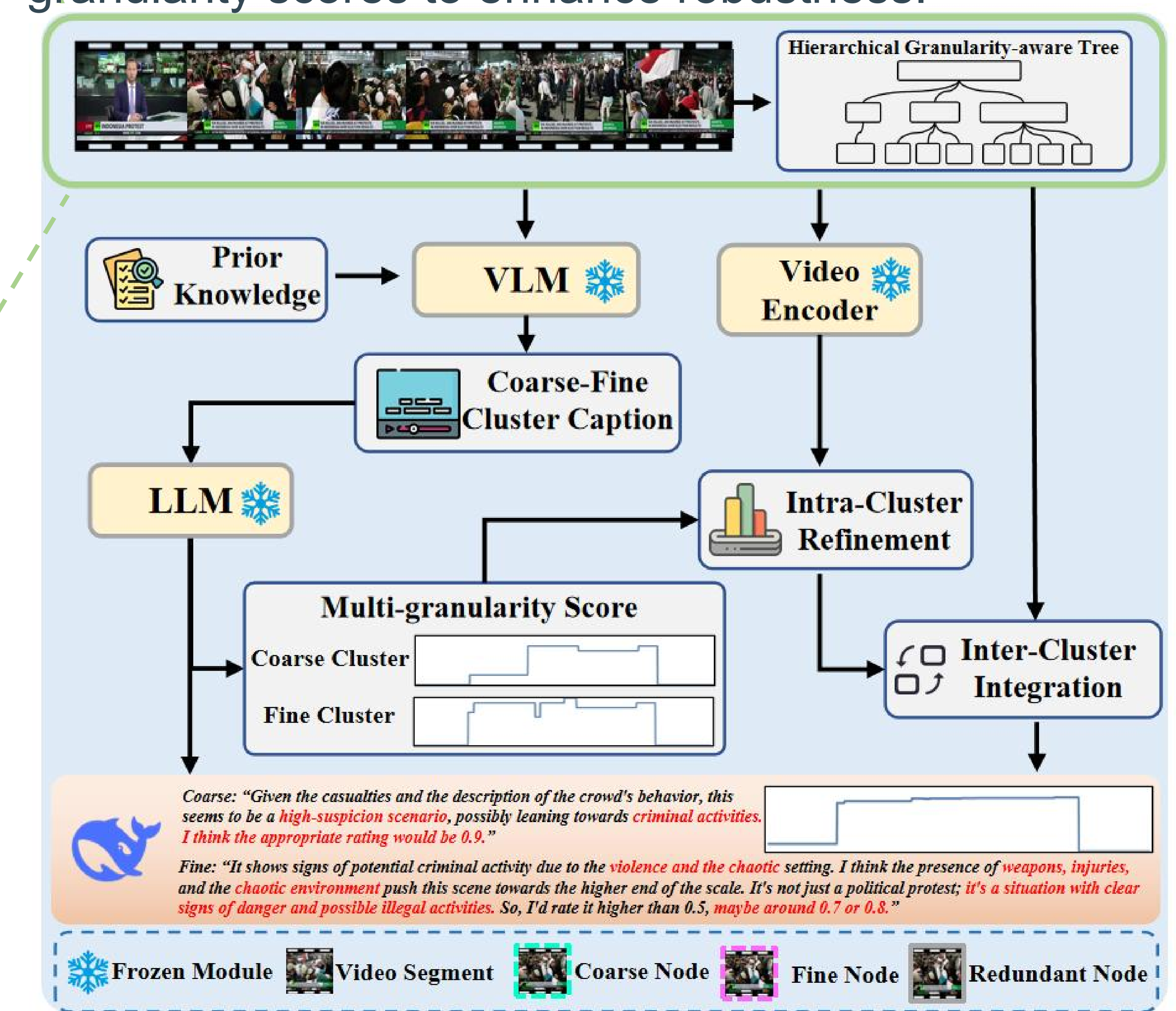




- **Segmentation Confidence Sequence:** Based on pre-trained GEBD model and overlapping sampling.
- **Generic Event Node Initialization:** Depth first traversal for representation initialization.
- **Adaptive Node Stratification:** Maximization of inter-hierarchical event confidence divergence via K-Means, RemoveDup, and Complete.
- **Sampling Efficiency Statistics:** Our sampling method more effectively captures anomalous events of varying lengths and generates fewer segments.

VADTree

- **Prior-infused Node Scoring:** Injecting prior knowledge to enhance anomaly perception.
- **Intra-cluster Node Refinement:** Refine scoring through intra cluster similarity.
- **Inter-cluster Node Correlation:** Integrating multi-granularity scores to enhance robustness.



Code

Paper

WeChat

Experiments Results

UCF-Crime dataset (AUC)

Explainable VAD Methods		
VADor [24]	Fine-tuning	88.13
Holmes-VAD [59]	Fine-tuning	89.51
Holmes-VAU [60]	Fine-tuning	88.96
VERA [50]	Verbalized Learning	86.55
Blip2 [18]	Training-free	46.42
ZS CLIP [27]	Training-free	53.16
ZS ImageBind (Image) [12]	Training-free	53.65
ZS ImageBind (Video) [12]	Training-free	55.78
LLaVA-1.5 [20]	Training-free	72.84
Video-Llama2 [58]	Training-free	74.42
LAVAD [55]	Training-free	80.28
SUVAD [11]	Training-free	83.90
MCANet [9]	Training-free	82.47
EventVAD [30]	Training-free	82.03
VADTree(Ours)	Training-free	**84.74**

XD-Violence dataset (AP & AUC)

Explainable VAD Methods			
Holmes-VAD [59]	Fine-tuning	90.67	-
Holmes-VAU [60]	Fine-tuning	87.68	-
VERA [50]	Verbalized Learning	70.54	88.26
Blip2 [18]	Training-free	10.89	29.43
ZS CLIP [27]	Training-free	17.83	38.21
ZS ImageBind (Image) [12]	Training-free	27.25	58.81
ZS ImageBind (Video) [12]	Training-free	25.36	55.06
LLaVA-1.5 [20]	Training-free	50.26	79.62
Video-Llama2 [58]	Training-free	53.57	80.21
LAVAD [55]	Training-free	62.01	85.36
SUVAD [11]	Training-free	**70.10**	-
MCANet* [9]	Training-free	69.72	87.43
EventVAD [30]	Training-free	64.04	87.51
VADTree (Ours)	Training-free	67.82	**90.44**
VADTree* (Ours)	Training-free	68.85	**90.55**

MSAD dataset

Method	Supervision	AUC (%)	AUC_a (%)	AP (%)	AP_a (%)
RTFM [37]	Weakly Supervised	86.65	-	-	-
MGFN [4]	Weakly Supervised	84.96	-	-	-
TEVAD [3]	Weakly Supervised	86.82	-	-	-
UR-DMU [26]	Weakly Supervised	85.78	67.95	67.35	75.30
GS-MoE [7]	Weakly Supervised	87.72	69.54	68.26	76.68
π-VAD [26]	Weakly Supervised	88.68	**71.25**	71.26	**77.86**
VADTree (Ours)	Training-free	**89.32**	67.85	**71.41**	75.49

HGTree Ablation

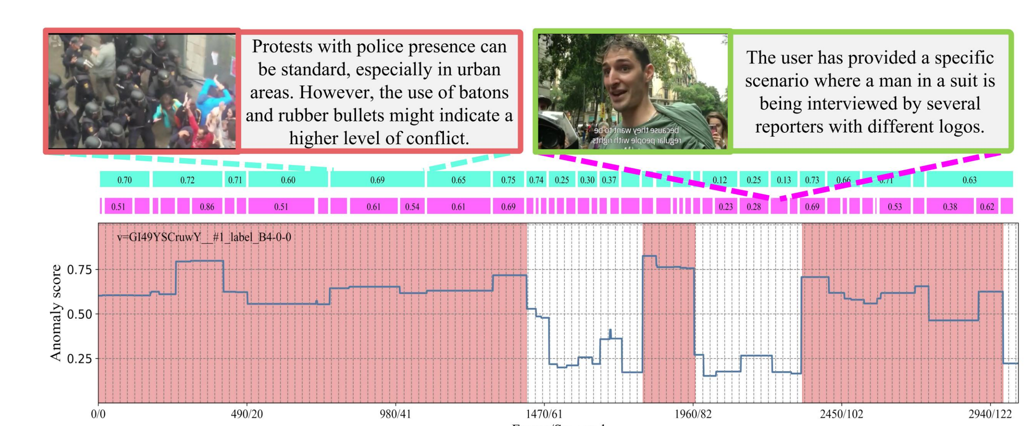
γ_{min}	Cluster Tool	Clusters	AUC (%)
0.3	-	Fine	80.89
0.4	-	Fine	82.81
0.5	-	Fine	80.85
0.3	K-Means	Coarse + Fine	83.74
0.4	K-Means	Coarse + Fine	84.74
0.5	K-Means	Coarse + Fine	82.40
0.4	K-Medoids	Coarse + Fine	**85.24**

- **Comparison with SOTA:** Surpassing the best training-free methods. Significantly outperforms weakly supervised methods on MSAD.
- **Ablation Study:** Hierarchical structure brings robust performance gains. Prior information effectively enhances anomaly perception.
- **Qualitative Results:** VADTree exhibits a superior ability to capture multi-granularity anomalies.

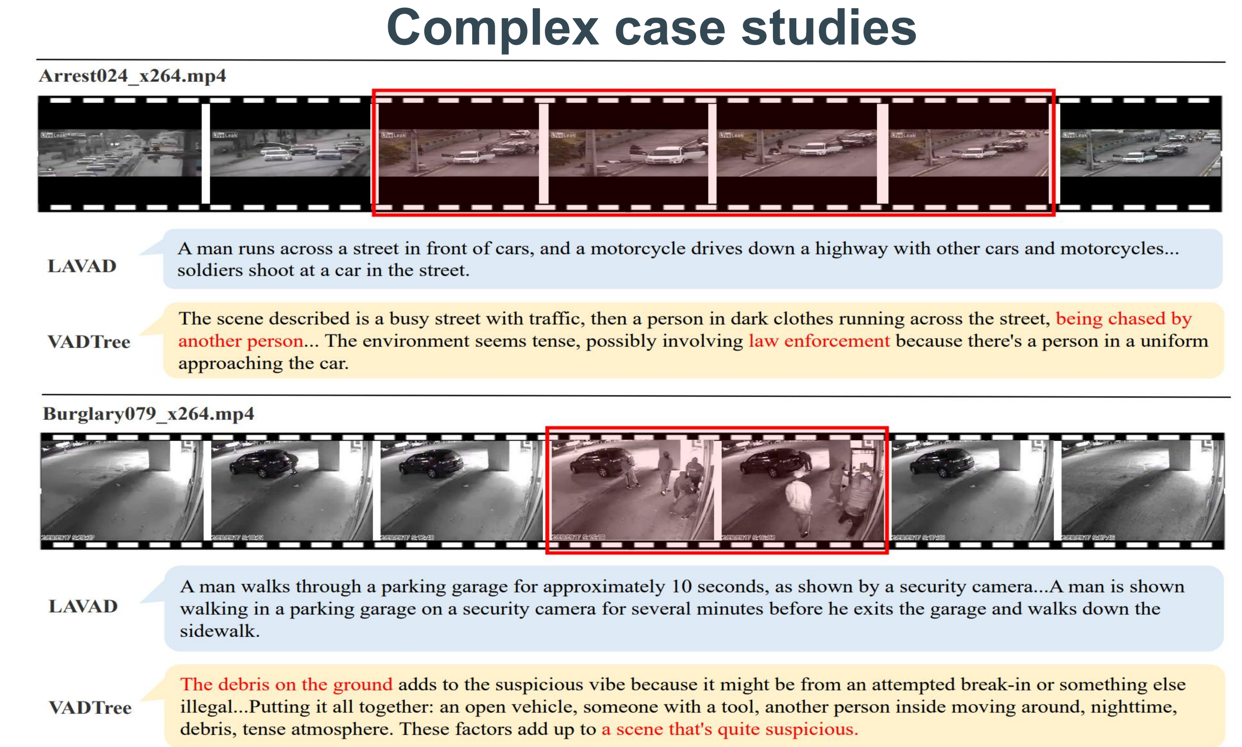
VADTree Ablation

Module	AUC (%)
HGTree Fine Cluster	71.57
+ Prior-infused Node Scoring	75.67
+ Intra-cluster Node Refinement	83.05
+ Inter-cluster Node Correlation	**84.74**

Qualitative Results



Complex case studies



β parameter experiment of correlation