



# Follow the Rules: Reasoning for Video Anomaly Detection with Large Language Models

Yuchen Yang, Kwonjoon Lee, Behzad Dariush, Yinzhi Cao, Shao-Yuan Lo

Johns Hopkins University, Honda Research Institute USA

Presenter: Sunghyun Ahn

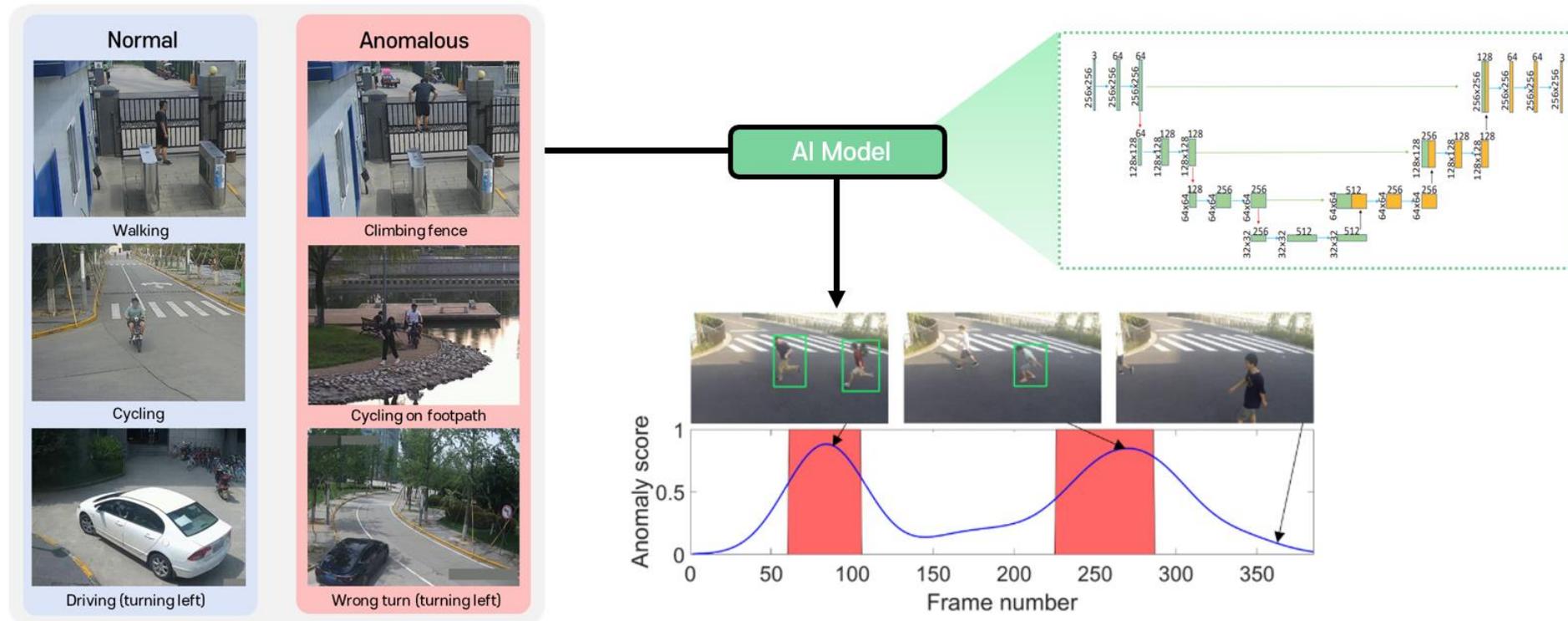
[sunghyun.ahn@pyler.tech](mailto:sunghyun.ahn@pyler.tech)



## Background

# Video Anomaly Detection

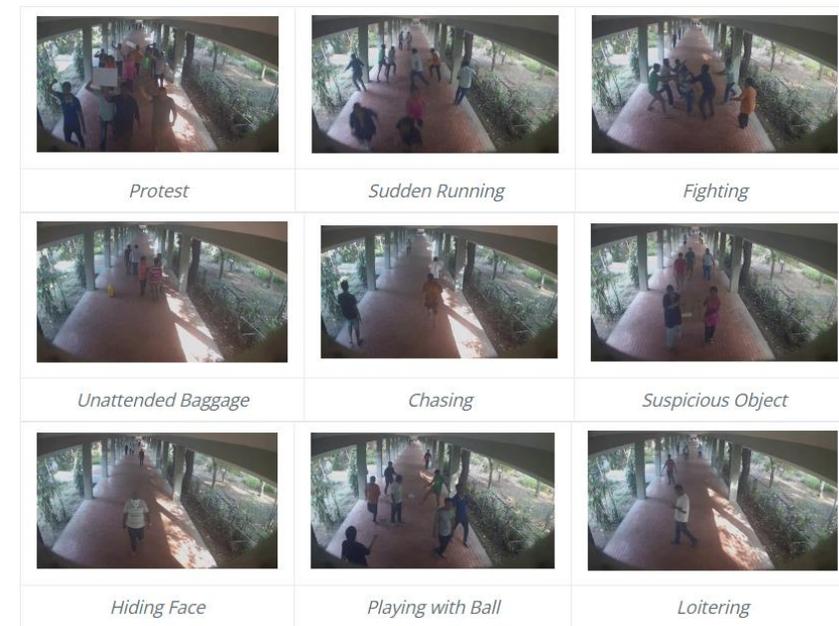
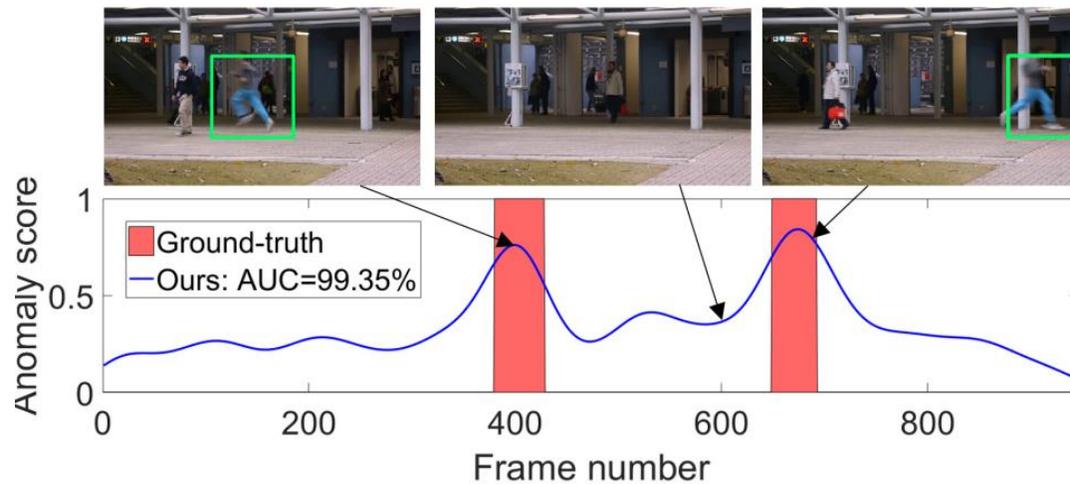
- Video Anomaly Detection (VAD) aims to determine whether abnormal events occur within video streams
- Abnormal events include the appearance or action of objects that are not suitable for the situation
- The goal is to do **Binary Classification** on each frame



## Background

# One Class Classification

- Class imbalance problem  $|\{x_i | y_i=0\}| \gg |\{x_i | y_i=1\}|$
- Diverse anomaly
- **One-Class Classification (OCC)** is utilized that learns exclusively from normal data and classifies anything not resembling the patterns of normal data as abnormal.



Rodrigues, Royston, et al. "Multi-timescale trajectory prediction for abnormal human activity detection." *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2020.

## Background

# One Class Classification

- **Classification-based:** Learning cluster that group normal feature vectors within a specific range.

$$\mathcal{A}_\theta(x) = \text{Diff}(\phi_\theta(x), c)$$

- Key idea
- **normal data:** clustered in a 'normal' cluster
  - **abnormal data:** away from the 'normal' cluster
- $\phi_\theta$ : kernel function  
 $c$ : center of hypersphere

- **Distance-based:** Storing features of normal data in memory and using the Nearest Neighbor search for classification.

$$\mathcal{A}_\theta(x) = \min_{x_{tr} \in X_{tr}} \text{Diff}(\phi_\theta(x), \phi_\theta(x_{tr}))$$

- Key idea
- **normal data:** similar to training data in a feature space
  - **abnormal data:** NOT similar to training data in a feature space

- **Reconstruction-based:** Learning to reconstruct normal samples using a generative model.

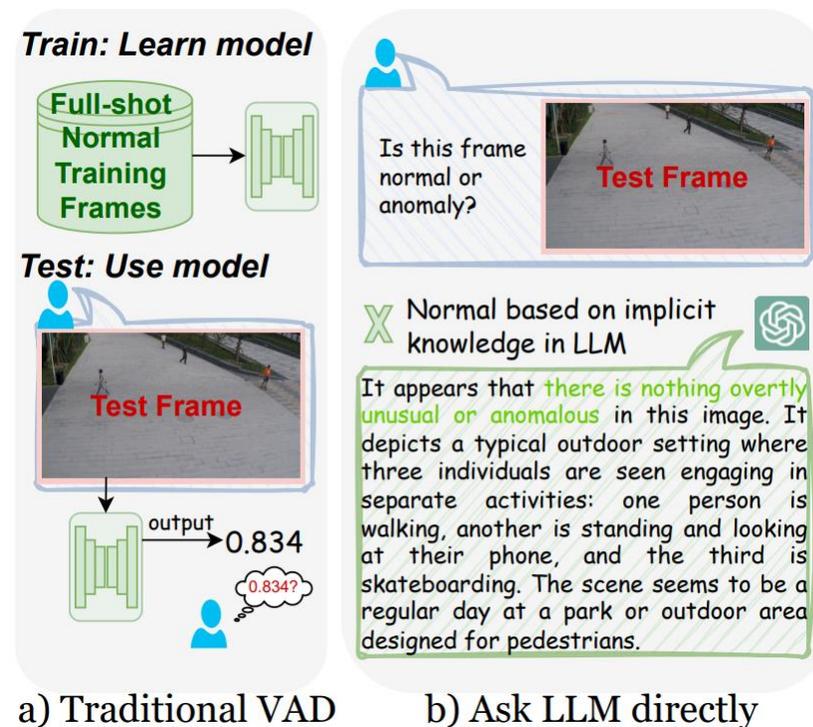
$$\mathcal{A}_\theta(x) = \text{Diff}(x, \text{Recon}_\theta(x))$$

- Key idea
- **normal data:** reconstructed well
  - **abnormal data:** NOT reconstructed well

## Introduction

# Limitations of Existing Methods

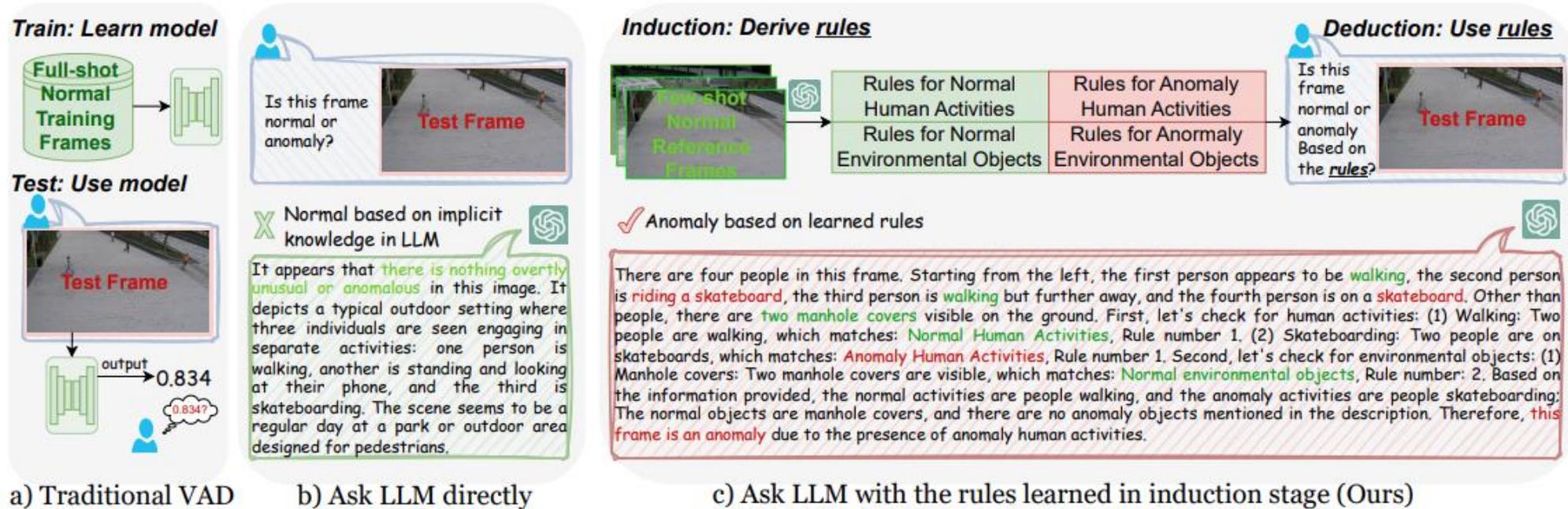
- **Traditional VAD:** Output only anomaly scores without providing clear rationale
  - **Hinder earning public trust in real-world deployment**
- **Ask LLM:** Rely on implicit knowledge from pre-trained LLMs that focuses on general context
  - **Misalign with specific real-world VAD applications**



# Introduction

## Contributions

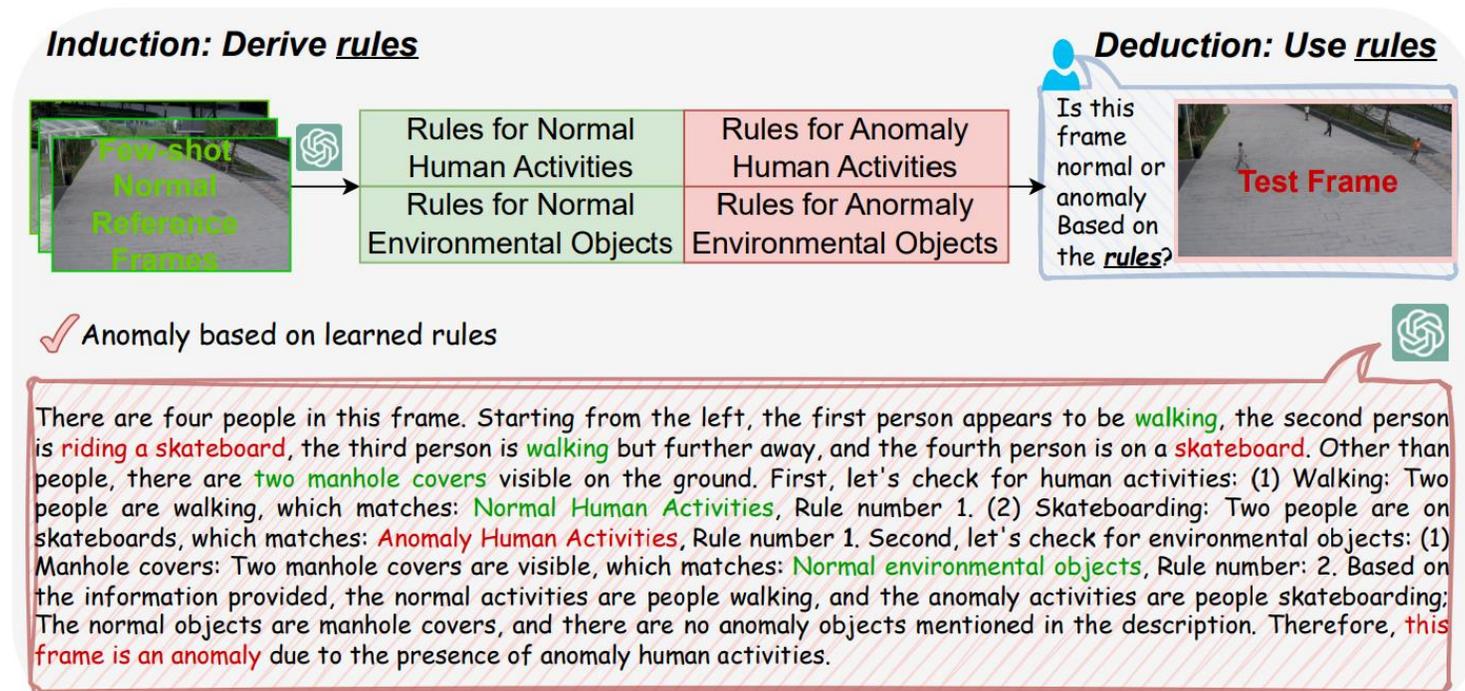
- Propose AnomalyRuler, a rule-based reasoning framework for VAD
- Adopt few-normal-shot prompting for fast and cost-effective adaptation
- Apply rule aggregation, perception smoothing, and robust reasoning to enhance robustness



## Introduction

# AnomalyRuler

- AnomalyRuler comprises two main stages:
- **Induction:** Derive rules from few-shot normal reference frames using VLM + LLM
- **Deduction:** Detect anomalies based on derived rules

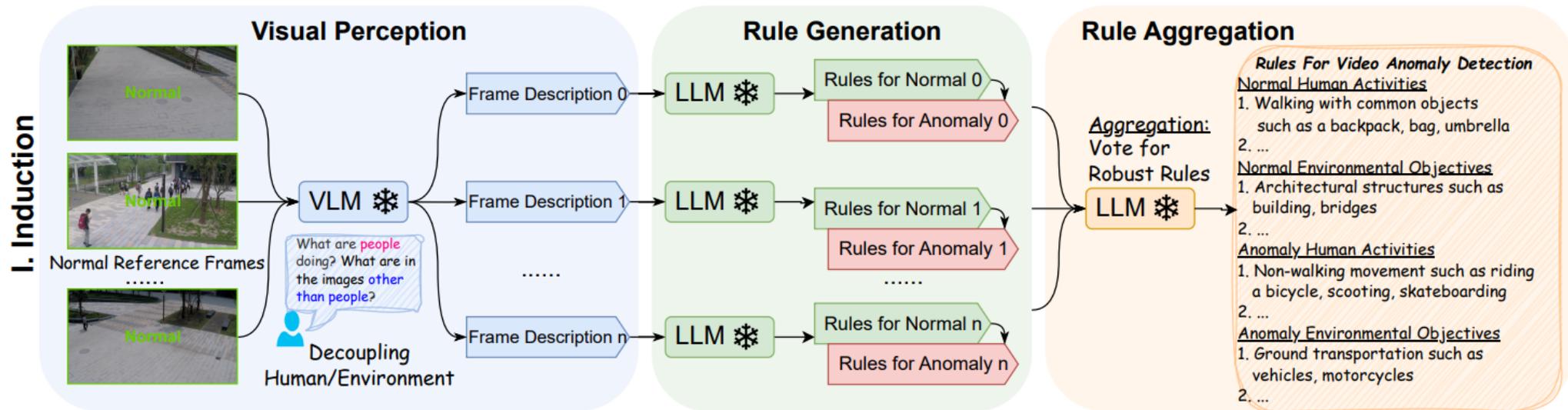


c) Ask LLM with the rules learned in induction stage (Ours)

# Methods

## Induction

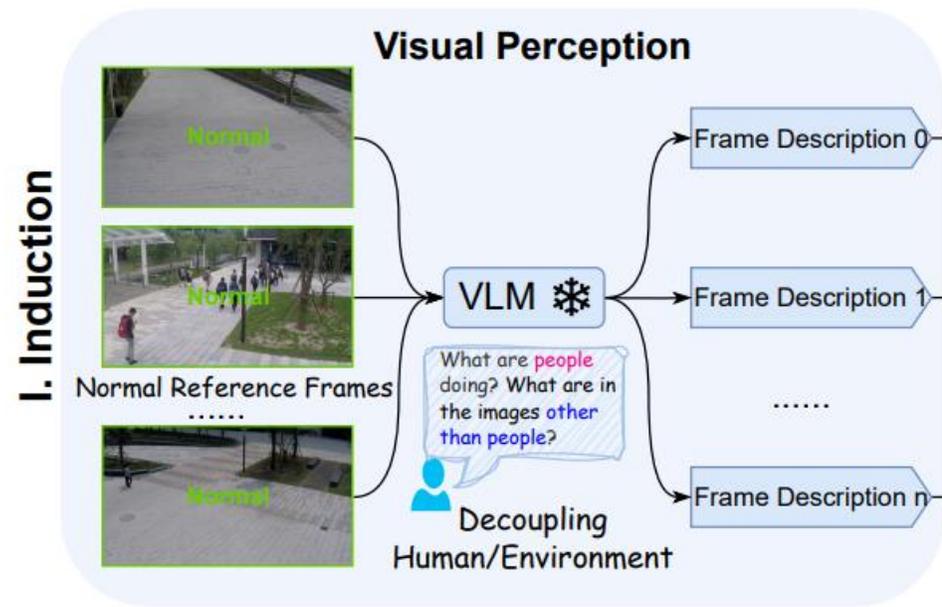
- **Visual Perception:** utilizes a VLM which takes a few normal reference frames as inputs and **outputs frame descriptions**
- **Rule Generation:** use an LLM to **generate rules based on these descriptions**
- **Rule Aggregation:** employ a voting mechanism to **mitigate the errors from rule generation**



## Methods

# Visual Perception

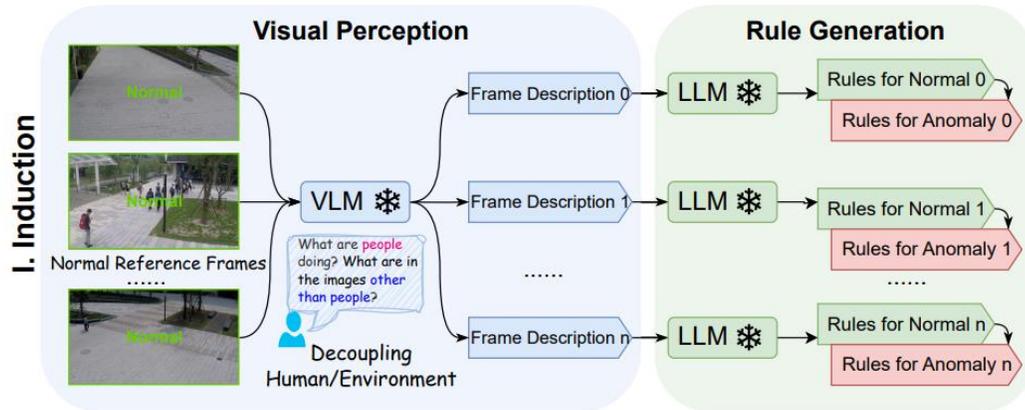
- Define  $F_{normal} = \{f_{normal_0}, \dots, f_{normal_n}\}$  as the few-normal-shot reference frames, with each frame  $f_{normal_i} \in F_{normal}$  randomly chosen from the training set
- Output the text description of each normal reference frame:  $D_{normal} = \{\text{VLM}(f_{normal_i}, p_v) \mid f_{normal_i} \in F_{normal}\}$ , with  $p_v$  as the prompt “What are people doing? What are in the images other than people?”



## Methods

# Rule Generation

- Design a Rule Generation module that uses a frozen LLM to generate rules (denoted as  $R$ )
- In formal terms,  $R = \{\text{LLM}(d_{normal_i}, p_g) \mid \in D_{normal}\}$ , where  $p_g$  is the prompt



$n$ : number of batches (rule sets)  
 $m$ : number of frames used per batch  
 $n = 2, m = 10$ : generate 2 rule sets, each from 10 descriptions

```
- system: You are a city surveillance monitor. Your task is to derive rules for normal and abnormal behaviors.

- user (1): Derive normal rules starting from abstract concepts and then move to concrete examples.
- assistant (1):
Rules for Normal Human Activities
1.
Rules for Normal Environmental Objects
1.

- user (2): Now, based on these, derive abnormal rules.
- assistant (2):
Rules for Anomaly Human Activities
1.
Rules for Anomaly Environmental Objects
1.

- user (3): Now, given the frame description {}, apply the Normal/Anomaly rules and provide the result.
```

## Methods

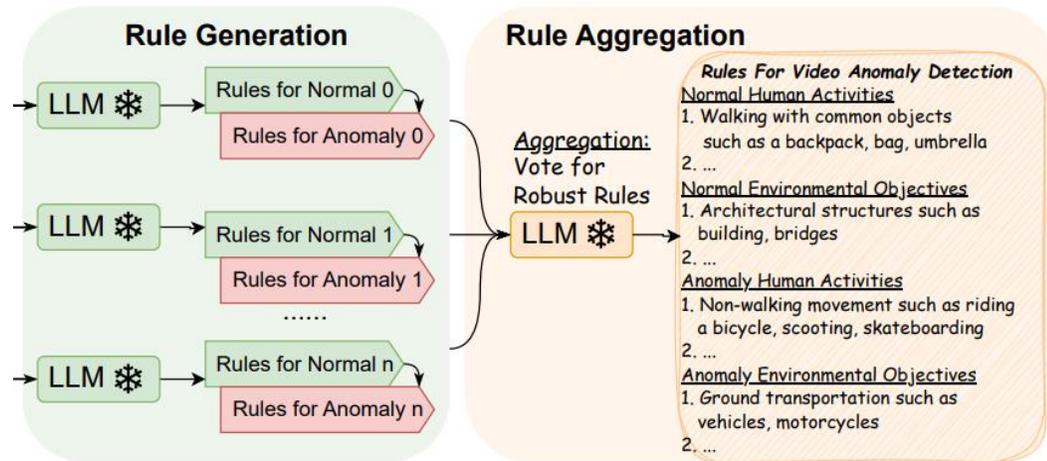
# Rule Generation

- **Contrast:** rules based on  $D_{normal}$  are assumed to be ground-truth normal, generates rules for anomalies by contrasting them with the rules for normal  
if “walking” is a common pattern in  $D_{normal}$ , it becomes a normal rule, and then “non-walking movement” will be included in the rules for anomaly
- **Analogy:** start from an abstract concept and then effectively generalizes to more concrete examples  
“walking” → “walking, whether alone or with others.”  
consequently, the anomaly rule evolves to include specific non-walking movements such as “riding a bicycle”, “scooting”, or “skateboarding.”
- **Separation:** lead the LLM to pay attention separately to environmental elements (e.g., vehicles or scene factors) and human activities, separately  
This enriches the rule set for VAD tasks, where anomalies often arise from interactions between humans and their environment.

## Methods

# Rule Aggregation

- Use LLM as an aggregator with voting mechanism, combine  $n$  rule sets  $R$  from normal reference frames  $\rightarrow R_{robust}$
- Assume **randomize smoothing: single errors unlikely to persist across multiple random samples**
- Mitigate errors from earlier stages (e.g., misinterpreting walking as skateboarding)



- System: You are a city surveillance monitor. Your job is to organize rules for detecting abnormal activities and objects.

- User (1): You are given  $\{n\}$  independent sets of rules for Normal and Anomaly. For the organized Normal Rules, list the given normal rules with high-frequency elements. For the organized Anomaly Rules, list all the given anomaly rules.

- Assistant(1):

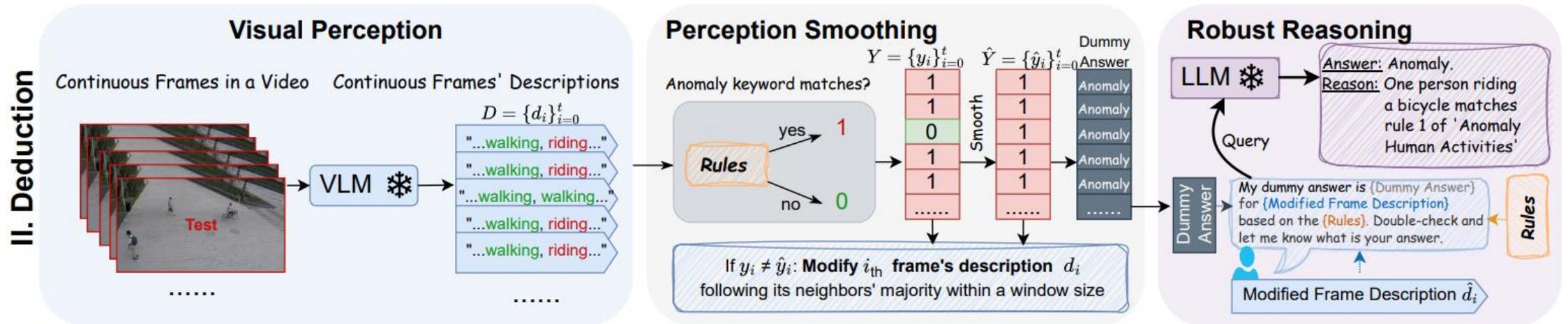
```
Rules for Anomaly Human Activities
1.
Rules for Anomaly Environmental Objects
1.
Rules for Normal Human Activities
1.
Rules for Normal Environmental Objects
1.
```

- User (2): Now you are given  $\{n\}$  independent sets of rules as the sublists of  $\{R\}$ . What rules for Anomaly and Normal do you get? Think step by step, and reply following the above format.

## Methods

# Deduction

- **Visual Perception:** Process continuous test video frames  $\rightarrow$  output frame descriptions  $D = \{d_0, \dots, d_t\}$
- **Perception Smoothing:** Reduce errors with Exponential Majority Smoothing
- **Rule Aggregation:** Recheck detection results with LLM and apply reasoning based on rules

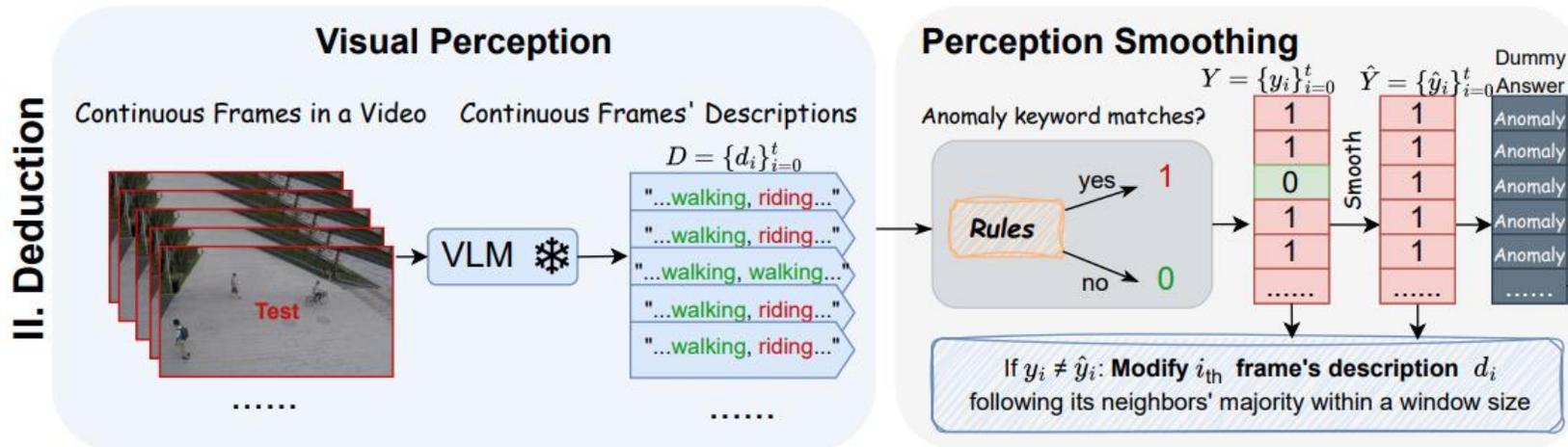


## Methods

# Perception Smoothing

- Visual perception errors can occur in deduction too
  - Propose Exponential Majority Smoothing to address this issue
  - Use smoothed results to correct frame descriptions and enhance robustness
- (1) Initial Anomaly Matching → (2) Exponential Majority Smoothing → (3) Description Modification
- Match anomaly keywords  $K$  from induction rules → assign each  $d_i$  a predicted label  $y_i$  → initial predictions as  $Y = \{y_0, \dots, y_t\}$

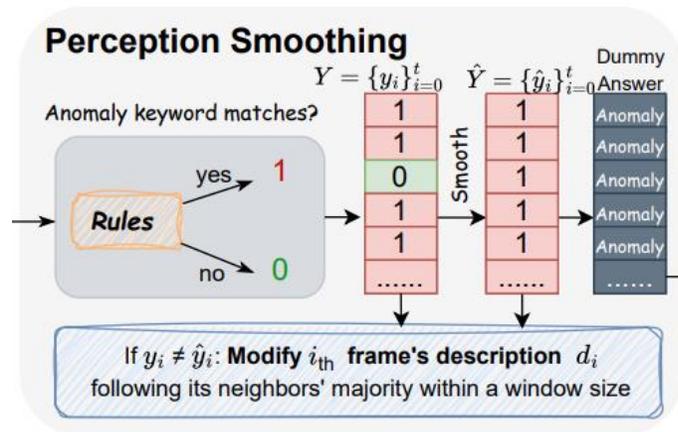
```
You will be given a set of rules for detecting abnormal activities and objects ; please extract the anomaly keywords , activities using 'ing ' verbs , and anomaly objects using nouns , and provide a combined Python list with each represented by a single word . The output should be in the format : [ " object1 " , ... , " activity1 " , " activity2 " , ...]. Now you are given {R_robust} :
```



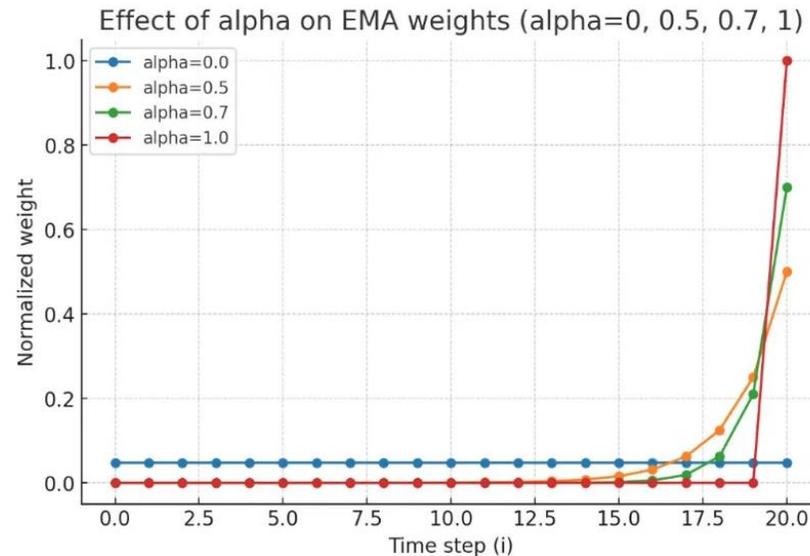
## Methods

# Perception Smoothing

- **Exponential Majority Smoothing**: approach that combines **Exponential Moving Average (EMA)** and **Majority Vote**
- **EMA**: method of calculating the current value by considering past values
- **Small  $\alpha$**  : consider past values more evenly, **Large  $\alpha$**  : focus on recent values ( $\alpha = 1 \rightarrow s_t = y_t$ )



$$s_t = \frac{\sum_{i=0}^t (1-\alpha)^{t-i} y_i}{\sum_{i=0}^t (1-\alpha)^i}$$



$$w_{t,i} = \frac{(1-\alpha)^{t-i}}{\sum_{j=0}^t (1-\alpha)^j}$$

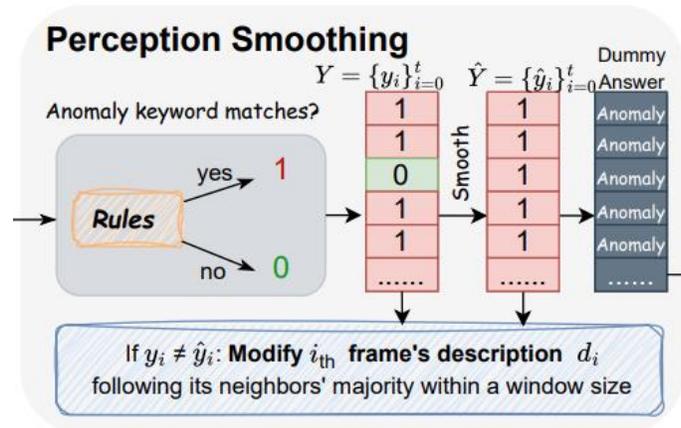
$$w_{t,i} \approx \frac{1}{t+1}$$

$$w_{t,t} \rightarrow 1, \quad w_{t,i < t} \rightarrow 0$$

## Methods

# Perception Smoothing

- **Majority Vote:** smooth the prediction within a window centered at each EMA value  $s_i$  with a padding size  $p$
- For each  $s_i$ , check neighboring EMA values within the window  $\rightarrow$  Decide  $\hat{y}_i$  by majority vote (above/below threshold  $\tau$ )
- Define threshold  $\tau$  as the mean of all EMA values



$$\hat{y}_i = \begin{cases} 1 & \text{if } \sum_{j=\max(1, i-p)}^{\min(i+p, t)} \mathbb{1}(s_j > \tau) > \frac{\min(i+p, t) - \max(1, i-p) + 1}{2} \\ 0 & \text{otherwise} \end{cases}$$

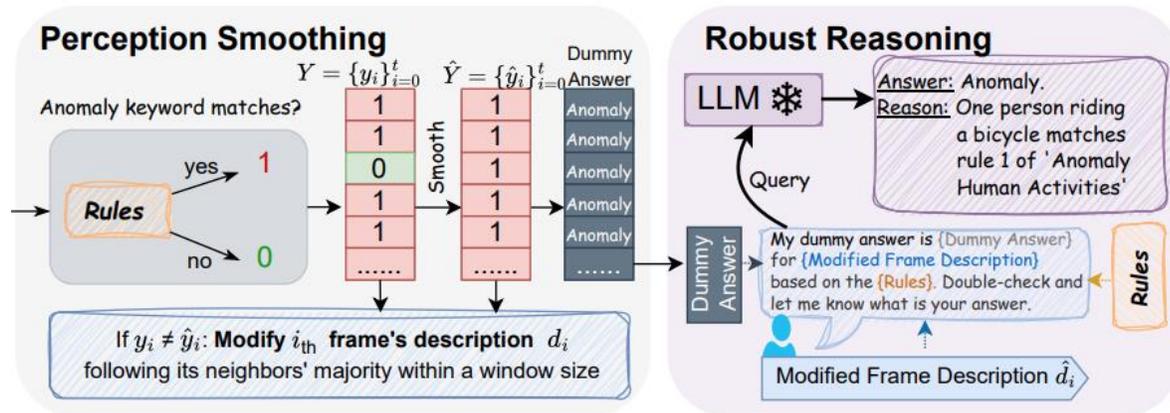
$$\hat{y}_i = \begin{cases} 1 & \text{if more than half of the values inside the window are larger than } \tau \\ 0 & \text{otherwise} \end{cases}$$

- **Description Modification:** modify the description  $D$  comparing  $Y$  and  $\hat{Y}$  and outputs the modified  $\hat{D}$
- $y_i = 0$  but  $\hat{y}_i = 1$  (false negative): Add "There is a person  $\{k\}$ ", where  $k \in K$  is the most frequent keyword in the window
- $y_i = 1$  but  $\hat{y}_i = 0$  (false positive): Remove parts of  $d_i$  containing the anomaly keyword  $k$

## Methods

# Robust Reasoning

- Recheck whether  $\hat{y}_i$  matches description  $\hat{d}_i$  according to  $R_{robust}$
- Use prompt  $p_r$  to guide the LLM
- Reduce missed anomalies (false negatives) and align reasoning more closely with rules



You will be given an description of frame , your task is to double check my initial anomaly detection result based on the rules .  
The rules are : {R\_robust}  
My initial result is {y^i}

First, if human activity is present , which rule is matching ?  
List the rule category , e.g. , normal or anomaly , with the rule number.

Second, if an environmental object is present, which rule is matching? List the rule category , e.g ., normal or anomaly , with the rule number .

Third , are the human activities or environmental objects anomaly?  
Answer : anomaly , if ANY anomaly rule ( even if only one , no matter human activities or environmental objects ) matches , otherwise answer : normal .

Now you are given the frame {d^i}, think step by step .

# Experiments

## Datasets

- Use campus datasets (Ped2, Avenue, ShanghaiTech) and UBnormal with diverse scenes
- Campus anomalies include **throwing objects, jumping, pushing, riding a bike, running, driving a car, etc**
- UBnormal covers 29 diverse scenes and 22 abnormal event types

Data set	#frames						#anomalies	#scenes	#anomaly types	open set
	total	training	validation	test	normal	abnormal				
CUHK Avenue [33]	30,652	15,328	-	15,324	26,832	3,820	77 <sup>‡</sup>	1	5	✓
ShanghaiTech [35]	317,398	274,515	-	42,883	<b>300,308</b>	17,090	158 <sup>†</sup>	13	11	✓
Street Scene [43]	203,257	56,847	-	146,410	159,341	43,916	205	1	17	✓
Subway Entrance [1]	144,250	76,453	-	67,797	132,138 <sup>†</sup>	12,112 <sup>†</sup>	51 <sup>†</sup>	1	5	✓
Subway Exit [1]	64,901	22,500	-	42,401	60,410 <sup>†</sup>	4,491 <sup>†</sup>	14 <sup>†</sup>	1	3	✓
UCF-Crime [53]	<b>13,741,393</b>	<b>12,631,211</b>	-	<b>1,110,182</b>	NA	NA	NA	NA	13	✗
UCSD Ped1 [37]	14,000	6,800	-	7,200	9,995	4,005	61 <sup>‡</sup>	1	5	✓
UCSD Ped2 [37]	4,560	2,550	-	2,010	2,924	1,636	21 <sup>‡</sup>	1	5	✓
UMN [38]	7,741	NA	-	NA	6,165	1,576	11	3	1	✓
<b>UBnormal (ours)</b>	236,902	116,087	<b>28,175</b>	92,640	147,887	<b>89,015</b>	<b>660</b>	<b>29</b>	<b>22</b>	✓



CUHK Avenue (Ave)



ShanghaiTech (ShT)

Table 1. Statistics about our novel benchmark versus existing anomaly detection data sets. Compared with other open-set benchmarks, we have a higher number of abnormal events from a broader set of action categories (anomaly types), occurring across a larger set of scenes. Top number in each column is in bold text. Legend: † – computed based on tracks from [18]; ‡ – computed based on tracks from [43].

Acsintoae, Andra, et al. "Ubnormal: New benchmark for supervised open-set video anomaly detection." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.

## Experiments

# Implementation Details

- Employ CogVLM-17B for visual perception, GPT-4-1106-Preview for induction, and Mistral-7B for deduction
- Changing VLM or reasoning LLM in deduction (except GPT-4v) degrades performance
- Show necessity of strong visual understanding and robust reasoning for AnomalyRuler’s effectiveness

**Table 7:** Detection performance with accuracy, precision, and recall (%) using different VLMs/LLMs as backbones in the deduction stage on 100 (limited by GPT-4’s query capacity) randomly selected frames from the ShT test set.

Visual Perception	Robust Reasoning	Accuracy	Precision	Recall	Open Source
CogVLM [51] (default)	Mistral [19] (default)	82.0	88.1	74.0	✓
GPT-4V [1]	GPT-4V	83.0	88.4	76.0	✗
LLaVA [23]	LLaVA	40.0	40.4	42.0	✓
PandaGPT [44]	PandaGPT	37.0	31.4	22.0	✓
OWLViT [33]	Mistral	71.0	82.0	54.0	✓
LLaVA		76.0	79.5	70.0	✓
BLIP-2 [21]		50.0	50.0	94.0	✓
RAM [58]		45.0	47.2	84.0	✓
CogVLM	GPT-3.5 [7]	81.0	86.0	74.0	✗
	LLaMA-2 [47]	60.0	70.8	34.0	✓

## Experiments

# Comparison with LLM-based Baselines

- **Ask LLM Directly:**  $\{\text{LLM}(d_i, p) \mid d_i \in D\}$ , where the LLM is Mistral-7B,  $D$  is frame descriptions generated by CogVLM, and  $p$  is “Is this frame description anomaly or normal?”
- **Ask LLM with Elhafsi et al.:**  $\{\text{LLM}(d_i, p) \mid d_i \in D\}$ , where the LLM is Mistral-7B,  $D$  is frame descriptions generated by CogVLM, and  $p$  is prompts comprising predefined concepts of normality/anomaly.
- **Ask Video-based LLMs Directly:**  $\{\text{VideoLLM}(c_i, p) \mid c_i \in C\}$ , where  $p$  is “Is this clip anomaly or normal?” (Video-LLaMA)

**Table 1:** Detection performance with accuracy, precision, and recall (%) compared with different VAD with LLM methods on the ShT dataset.

Method	Accuracy	Precision	Recall
Ask LLM Directly	52.1	97.1	6.2
Ask LLM with Elhafsi et al. [12]	58.4	<b>97.9</b>	15.2
Ask Video-based LLM Directly	54.7	85.4	8.5
AnomalyRuler	<b>81.8</b>	90.2	<b>64.3</b>

- Achieve **+26.2%** accuracy and **+54.3%** recall by leveraging rule-based reasoning from the induction stage

## Experiments

# Comparison with State-of-the-Art Methods

- AnomalyRuler-base (without robust reasoning) outperforms all Image-Only competitors on ShT and UB
- Performs on par with Image-Only methods on Ped2 and Ave
- Achieves this without tuning → few-normal-shot prompting as effective as full-shot training

**Table 3:** AUC (%) compared with different one-class VAD methods. “Image Only” methods only rely on image features. In contrast, others employ additional features such as bounding boxes from object detectors or 3D features from action recognition networks. “Training” indicates the methods that need a full-shot training process.

Method	Venue	Image Only	Training	Ped2	Ave	ShT	UB
MNAD [36]	CVPR-20	✓	✓	97.0	88.5	70.5	-
rGAN [29]	ECCV-20	✓	✓	96.2	85.8	77.9	-
CDAE [9]	ECCV-20	✓	✓	96.5	86.0	73.3	-
MPN [30]	CVPR-21	✓	✓	96.9	89.5	73.8	-
NGOF [50]	CVPR-21	✗	✓	94.2	88.4	75.3	-
HF2 [25]	ICCV-21	✗	✓	<b>99.2</b>	91.1	76.2	-
BAF [14]	TPAMI-21	✗	✓	98.7	92.3	82.7	59.3
GCL [56]	CVPR-22	✗	✓	-	-	79.6	-
S3R [53]	ECCV-22	✗	✓	-	-	80.5	-
SSL [49]	ECCV-22	✗	✓	99.0	<b>92.2</b>	84.3	-
zxVAD [3]	WACV-23	✗	✓	96.9	-	71.6	-
HSC [45]	CVPR-23	✗	✓	98.1	93.7	83.4	-
FPDM [54]	ICCV-23	✓	✓	-	90.1	78.6	62.7
SLM [43]	ICCV-23	✓	✓	97.6	90.9	78.8	-
STG-NF [18]	ICCV-23	✗	✓	-	-	<b>85.9</b>	71.8
AnomalyRuler-base	-	✓	✗	96.5	82.2	84.6	69.8
AnomalyRuler	-	✓	✗	97.9	89.7	85.2	<b>71.9</b>

**Table 4:** AUC (%) compared with different cross-domain VAD methods. We follow the compared works to use ShT as the source domain dataset for other target datasets.

Method	Venue	Image Only	Training	Ped2	Ave	ShT <sup>1</sup>	UB
rGAN [29]	ECCV-20	✓	✓	81.9	71.4	77.9	-
MPN [30]	CVPR-21	✓	✓	84.7	74.1	73.8	-
zxVAD [3]	WACV-23	✗	✓	95.7	<b>82.2</b>	71.6	-
AnomalyRuler-base	-	✓	✗	<b>97.4</b>	81.6	<b>83.5</b>	<b>65.4</b>

<sup>1</sup> AnomalyRuler employs UB as the source domain when ShT serves as the target domain. The competitors have no cross-domain evaluation on ShT, so we report their same-domain results.

$$a_t = \frac{\sum_{i=0}^t (1 - \tau)^{t-i} \hat{y}_i}{\sum_{i=0}^t (1 - \tau)^i}$$

- Achieve highest AUC across target datasets using ShT as the source domain
- Since Ped2, Ave, and ShT are all campus datasets, similar rules are generated, supporting strong generalization

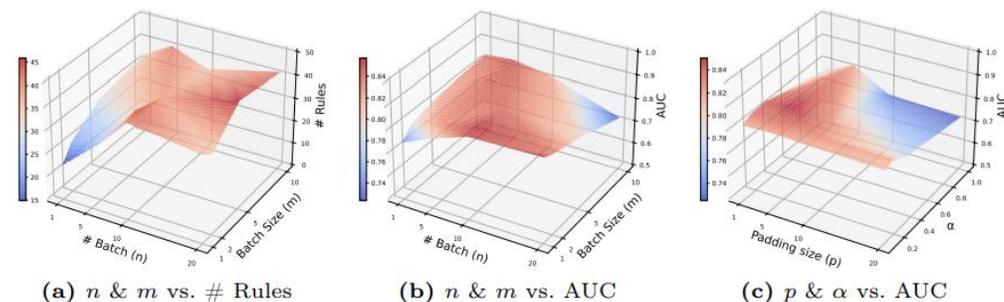
## Experiments

# Ablation Study

- Proposed strategies effectively enhance AnomalyRuler’s performance
- Too few rules cause inadequate coverage, while too many rules create redundancy and errors

**Table 5:** Ablation on strategies. We assess the effects of removing individual strategies in AnomalyRuler. We conduct the experiments five times with different randomly selected normal reference frames for induction and report their mean and standard deviation on the ShT dataset.

Strategy	Stage	# Rules		Accuracy		Precision		Recall		AUC	
		mean	std	mean	std	mean	std	mean	std	mean	std
w. All Below (default)	Both	42.2	4.2	81.6	1.3	90.9	0.8	63.9	2.7	84.5	1.1
w/o. Human and Environment	Both	-20.1	+1.1	-3.3	+0.8	-3.9	+0.8	-1.9	+1.6	-2.4	+2.0
w/o. Normal and Anomaly	Induction	-34.8	-1.3	-20.5	+4.3	-41.2	+7.0	-14.4	+11.6	-18.8	+1.2
w/o. Abstract and Concrete	Induction	+2.3	+2.7	-0.6	-0.2	-0.9	-0.2	-0.3	-0.4	-0.9	+0.1
w/o. Rule Aggregation	Induction	+8.5	+6.1	-9.6	+14.7	+1.1	+2.9	-10.7	+14.1	-15.8	+0.8
w/o. Perception Smoothing	Deduction	NA	NA	-1.7	-0.9	-1.9	+0.1	-3.8	-0.3	-3.3	+0.8



**Fig. 3:** Ablation on hyperparameters of the (a) (b) rule aggregation and (c) perception smoothing modules on the ShT dataset.

- Conduct cross-validation on number of batches  $n = [1, 5, 10, 20]$  and frames per batch  $m = [1, 2, 5, 10]$
- Both rule count and AUC increase with  $n$  and  $m$ , but performance fluctuates when  $n \times m$  is large (e.g., AUC drops from 85.9% to 72.2% at  $n = 20$  with large  $m$  due to redundancy)
- For perception smoothing, test padding size  $p = [1, 5, 10, 20]$  and EMA weight  $\alpha = [0.09, 0.18, 0.33, 1]$
- $p = 5$  found optimal for motion continuity without excessive noise
- Increasing  $\alpha$  from 0.09 to 0.33 generally improves AUC

## Conclusion

# Conclusion

- **Strengths**
- Require only few-normal-shot prompting, avoiding costly full-shot tuning
- Enable fast steering of LLM reasoning for diverse VAD applications
  
- **Limitation**
- Assume strong LLM backbones; performance may degrade with sub-optimal LLMs
  
- **Opinions**
- Frame descriptions may be insufficient for action classification
  - Use video models to generate clip-level descriptions
- Current rules ignore scene-dependent normality/anomaly
  - Separate rules for human activities and environmental objects by scene (e.g., walking → normal in campus, walking → abnormal in road)

# Appendix

## LLM Outputs

–  $R_{\text{robust}} \rightarrow K$  (generate anomaly keywords from anomaly rules, see Section 4).

- Input prompt  $p_k$ :

```
p_k = You will be given a set of rules for detecting abnormal
      activities and objects; please extract the anomaly keywords,
      activities using ‘‘ing’’ verbs, and anomaly objects using
      nouns, and provide a combined Python list with each
      represented by a single word. The output should be in the
      format: ["object1", ..., "activity1", "activity2", ...]. Now
      you are given {R_robust}:
```

- Output  $K$ :

```
anomaly_from_rule = ["trolley", "cart", "luggage", "bicycle", "
                    skateboard", "scooter", "vehicles", "vans", "accident", "
                    running", "jumping", "riding", "skateboarding", "scooting", "
                    lying", "falling", "bending", "fighting", "pushing", "
                    loitering", "climbing", "tampering", "lingering"]
```

$$y_j^* = LLM(\hat{d}_j, \hat{y}_j, R_{\text{robust}}, p_r)$$

```
ŷ_j^* = The human activities present are two people walking. The rules for
      normal human activities include walking alone or with another
      person. So, the rule matching is:
      1. Walking alone or with another person (normal)

      The environmental objects present are two manhole covers on the
      ground, a trash bin, and some trees and plants. So, the rules
      matching are:
      1. Manhole covers on the ground or street (normal)
      2. Trees along the street or Plants present (normal)
      3. Trash bin in vicinity or placed beside the street (normal)

      Answer: normal.
```

– Outputs: Robust rules  $R_{\text{robust}} = LLM(R = \{r_1, \dots, r_n\}, p_a)$

```
**Rules for Anomaly Human Activities:**
1. Using any non-walking movement such as riding a bicycle or scooter,
   skateboarding
2. Running or jumping
3. Moving with the usual speed such as very fast or in a hurry
4. Person lying or bending down on the ground
5. Aggressive or unusual movements such as fighting or pushing
6. Loitering around the trash bin
7. Climbing on trees or into manhole/utility access points
8. Tampering with manhole covers
9. Engaging in physical altercation
10. Lingering without a clear purpose near a sensitive area
11. Falling unexpectedly

**Rules for Anomaly Environmental Objects:**
1. Missing, removed, or ajar manhole covers
2. Ground transportation such as vehicles, vans, bicycles.
3. Unattended bags or objects (security threat)
4. Unusual objects (vehicles, large unattended luggage, etc.)
5. Unauthorized posters or signs
6. Damaged manhole cover or trees
7. Abandoned items in a public area
8. Yellow objects (when not commonly seen)
9. Graffiti on building windows or vandalism to a fence
10. Obstructions on pedestrian crossing
11. Fire hydrant spraying water without presence of emergency services
12. Unidentified objects near the tall structure
13. Smoke or fire coming from a utility access point
14. Objects thrown or falling from a height

**Rules for Normal Human Activities:**
1. Walking alone or with another person
2. Standing and using a mobile device
3. Walking on the sidewalk
4. Walking towards or away from a building
5. Entering a building
6. Standing near a trash bin
7. Waiting at traffic lights
8. Walking on pedestrian crossing
9. Sitting on bench or ground

**Rules for Normal Environmental Objects:**
1. Manhole covers on the ground or street
2. Trees along the street or Plants present
3. Trash bin in vicinity or placed beside the street
4. Posters on glass door
5. Static building with glass windows
6. Fence along the water
7. Pedestrian crossing markings visible
8. Standing fire hydrant
9. Static tall structure in the background
10. Utility access points on the ground
```

**Thank you**