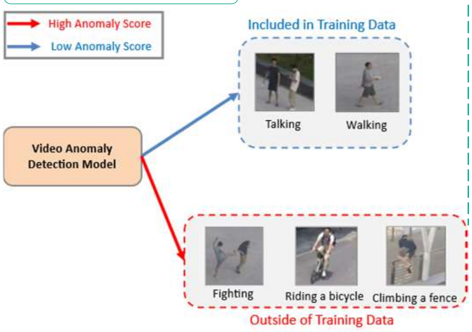
Language-Guided Decision Override for Adaptive and Retraining-Free Video Anomaly Detection

Ryo Moriyama[1,3], Shin Suzuki[1], Naoshi Kaneko[2], Kazuhiko Sumi[1]
[1]Aoyama Gakuin University, [2]Tokyo Denki University, [3]Fujitsu Limited

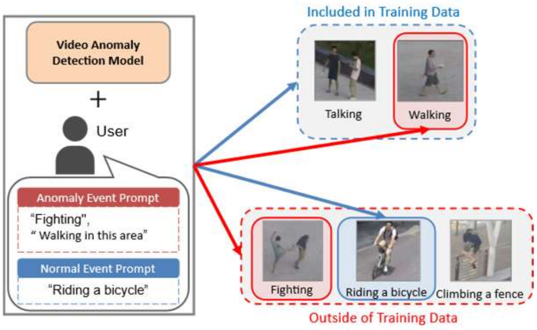
BMVC 2025

Introduction

Previous OC-VAD



LinGuard with Event Prompt



➢One-class Video Anomaly Detection (OC-VAD)

OC-VAD is trained only on normal videos and detects anomalies based on deviations from learned patterns.

➢ Limitations of OC-VAD ?

- All events outside the learned distribution as anomalies
  ➡ Low flexibility
- Costly model retraining is required to redefine anomalous
  ➡ High retraining costs

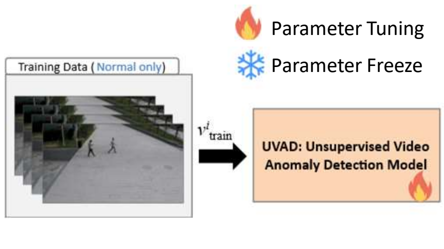
➢ Our approach

We propose **LinGuard**, a framework that **allows users to redefine anomalies via language Event Prompts**. It fuses scores from a VAD model and a Large Vision Language Model (LVLM) to update the detection result **without retraining**.

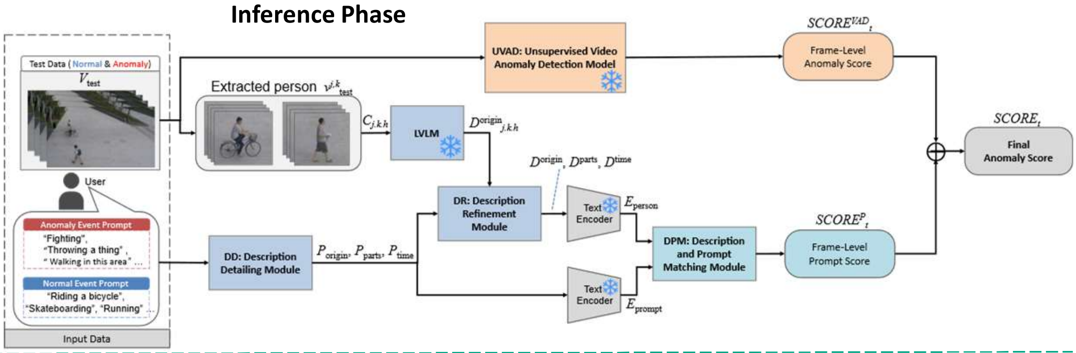
Method

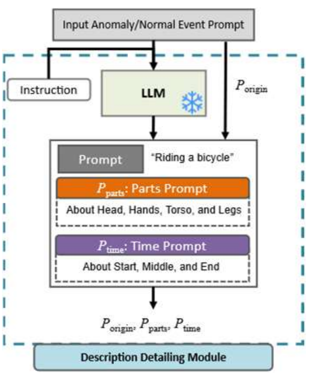
This method updates pre-trained model scores using **Prompt Scores** without retraining. Prompt Scores are calculated from cosine similarity between user **Event Prompt** and **LVLM-generated action descriptions**.

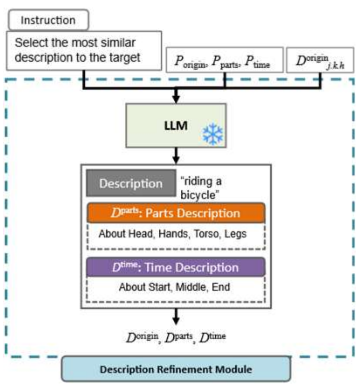
Training Phase

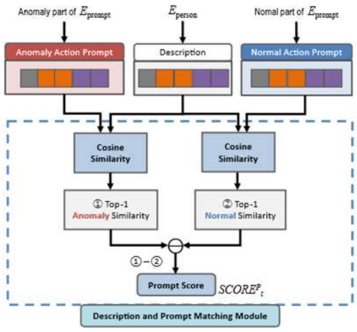


Inference Phase









Original Description / Embedding
Parts Description / Embedding
Time Description / Embedding

➢ **DD: Description Detailing Module**

DD Module expands Event Prompts into detailed descriptions of body part-level movements and temporal transitions.

➢ **DR: Description Refinement Module**

DR Module refines event description from LVLM by using LLM to select the most similar option from a list of detailed Event Prompts.

➢ **DPM: Description and Prompt Matching Module**

DPM Module calculates cosine similarity between the event description and both Anomaly and Normal Event Prompts, producing a Prompt Score.

Results

Anomaly/Normal Event Prompt

Dataset : HR-ShanghaiTech Campus[1,2]

Setting	Anomaly Event Prompt	Normal Event Prompt
1	General anomaly description with an LLM	None
2	General anomalies not in Normal Event Prompt	riding a bicycle/ motorcycle/segway, skateboarding
3	As above	riding a bicycle/segway, skateboarding
4	As above	running, pushing/carrying cart, jumping
5	As above	riding bicycle/motorcycle/Segway, skateboarding, running, pushing/carrying cart, jumping

Quantitative Results

Result on HR-ShanghaiTech (AUC(%))

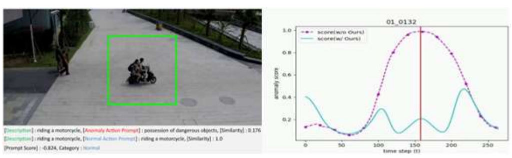
Parentheses show the difference from the base AUC in Setting1. Score degradation suppression is the difference between these values for w/o LinGuard and w/LinGuard.

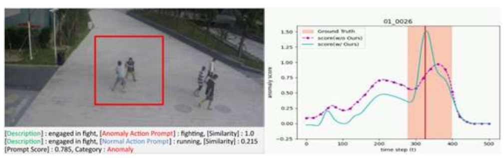
	Setting	w/o LinGuard	w/ LinGuard	Suppression of Score Degradation (pt)
STG-NF [3]	1	87.4(base)	89.5(base)	-
	2	71.9(-15.5)	82.9(-6.6)	8.9pt
	3	76.8(-10.6)	85.8(-3.7)	6.9pt
	4	84.1(-3.3)	89.5(±0.0)	3.3pt
	5	68.1(-19.3)	83.9(-5.6)	**13.7pt**

VAD backbone + LinGuard vs. LinGuard (Prompt Score only) vs. LLM/LVLM based method

Setting	STG-NF[3] +LinGuard	STJP[4] +LinGuard	LinGuard (Prompt Score only)	Anomaly Ruler[5] (LLM/LVLM based approach)
1	**89.5**	87.4	80.4	85.3
5	**83.9**	84.7	78.6	65.2

Qualitative Results





Magenta=VAD alone
Cyan=LinGuard

(Left): Reduce anomaly score by Normal Prompt.
(Right): Increase anomaly score by Anomaly Prompt.

[1] W.Liu, W.Luo, D.Lian, S.Gao: Future frame prediction for anomaly detection - A new baseline. In CVPR, 2018.
[2] R.Morais, et al.: Learning Regularity in Skeleton Trajectories for Anomaly Detection in Videos. In CVPR, 2019.
[3] O.Hirschorn, S.Avidan: Normalizing Flows for Human Pose Anomaly Detection, In ICCV, 2023.
[4] G.Wang, et al.: Video Anomaly Detection by Solving Decoupled Spatio-Temporal Jigsaw Puzzles. In ECCV, 2022.
[5] Y.Yang, et al.: Follow the Rules: Reasoning for Video Anomaly Detection with Large Language Models