EUROPEAN CONFERENCE ON COMPUTER VISION
MILANO 2024

Learning Anomalies with Normality Prior for Unsupervised Video Anomaly Detection

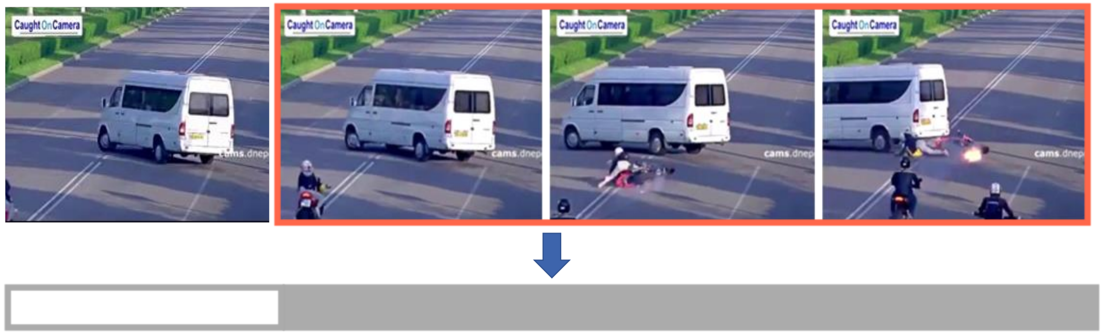
Haoyue Shi[1][3], Le Wang*[1], Sanping Zhou[1], Gang Hua[2], Wei Tang[3]

[1]Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University

[2]Multimodal Experiences Research Lab, Dolby Laboratories [3]University of Illinois Chicago

Introduction

- **Video Anomaly Detection (VAD)**
  - Detect abnormal events in video sequences along the temporal dimension.
  - Unsupervised VAD(UAVD) aims to detect abnormal events without any annotations.



Normal frame　　Abnormal frame

- **Application Scenario**

Intelligent surveillance

Road accident warning

Crime detecting

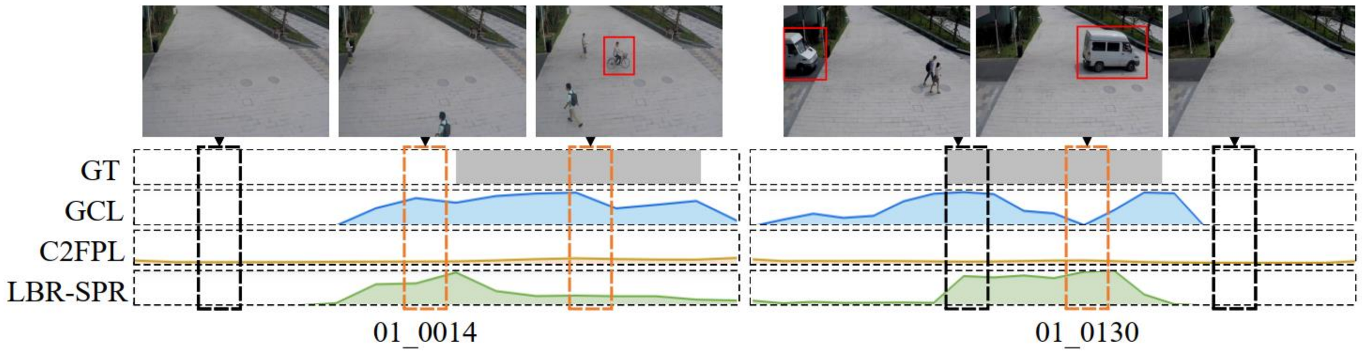
Introduction

➢ **Current UVAD methods**

- Recent methods are data-driven, performing unsupervised learning by identifying abnormal patterns in videos.

➢ **Limitation**

- These methods heavily rely on the feature representation and data distribution, often struggling to capture less salient anomalies.



➢ **Learning Anomalies with Prior**

- Our idea is to leverage ***data-irrelevant prior*** knowledge about normal and abnormal events to aid in identifying anomalies.

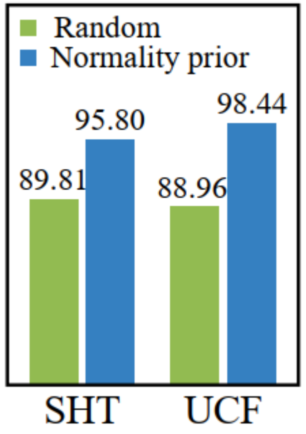
Introduction

EUROPEAN CONFERENCE ON COMPUTER VISION

MILANO 2024

➢ *What is good prior knowledge for UVAD?*

- It should be both **informative** and of **high quality**.
- We introduce **a normality prior**: the start and end of a video are mostly normal.
- Using **our normality prior** to select normal snippets is **significantly more accurate** than random selection.



➢ *How should the prior knowledge be used?*

- Direct comparison: **normal frames can be easily mislabled**. ✗
- We propose **Normality Propagation** to propagate normal information based on temporal and semantic relations of snippets. ✓

Introduction

Contributions of our method

Contribution1:

Unlike previous data-driven methods, we propose to use the **data-irrelevant normality prior** to identify abnormal events. To the best of our knowledge, **such prior has never been studied before in the area of UVAD.**

Contribution2:

We introduce **normality propagation** to effectively propagate the normality prior to unlabeled snippets for pseudo label generation.

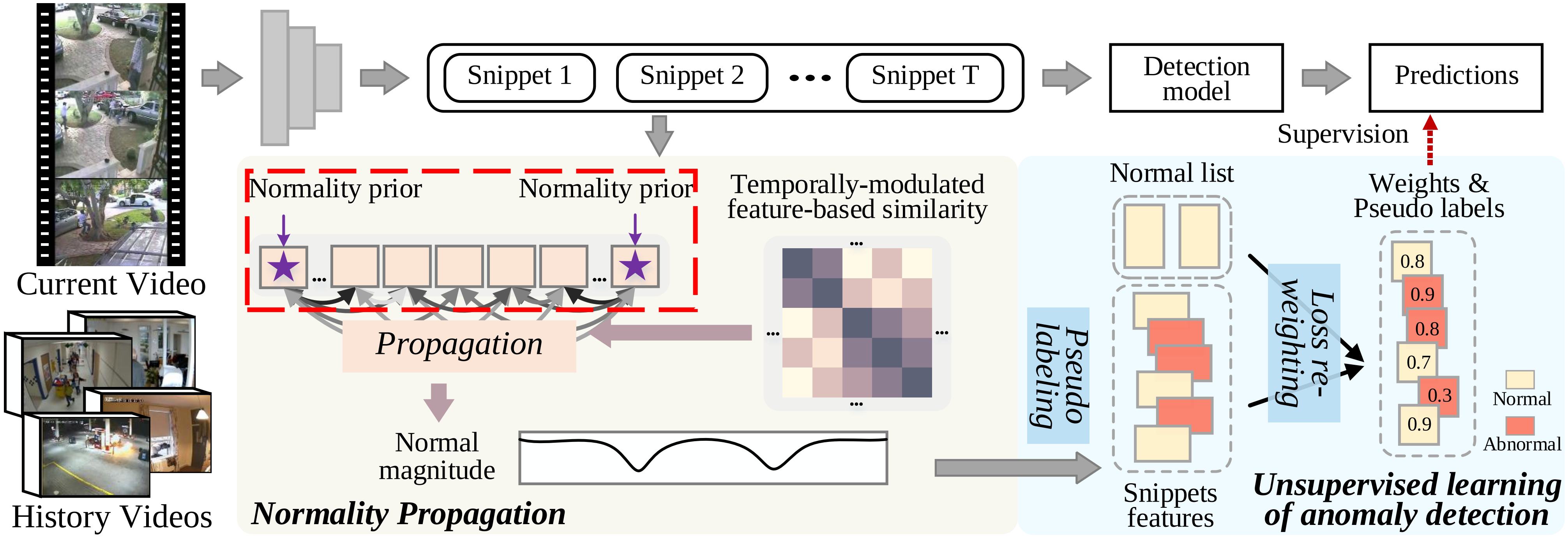
Contribution3:

We perform **unsupervised learning of abnormal detection** based on **propagated labels** and **a new loss re-weighting method**. They are complementary to normality propagation and mitigate the negative impact of incorrectly labels.

Extensive experiments on ShanghaiTech and UCF-Crime demonstrate the effectiveness of the proposed method.

Method

- **Overview**



- We first use the **normality prior** to specify that the start and end of a video are normal.
- Then our **normality propagation** propagates the normal information to estimate normal magnitudes.
- After that, We perform **unsupervised learning of abnormal detection** based on the **propagated labels** and a **new loss re-weighting method**.

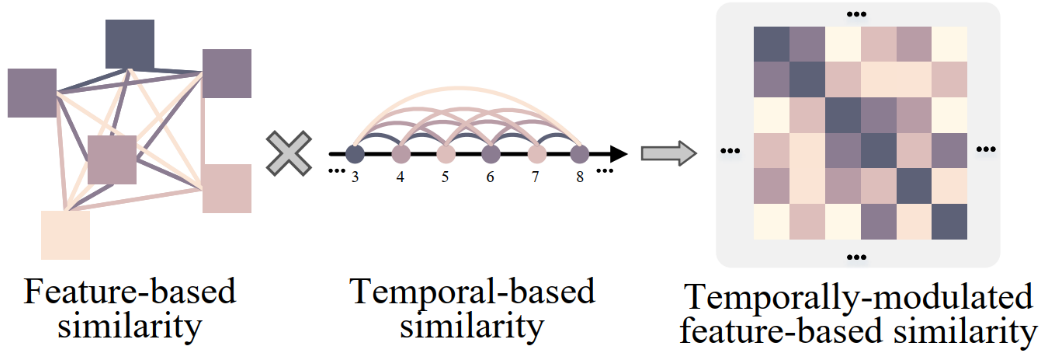
Method

➢ Normality Propagation

- It iterates the propagation below until convergence.

$$\mathbf{z}(n+1) = \alpha\mathbf{S}\mathbf{z}(n) + (1-\alpha)\mathbf{y}$$

y: *label vector*, **S**: *similarity matrix*, **z**: normal magnitudes.



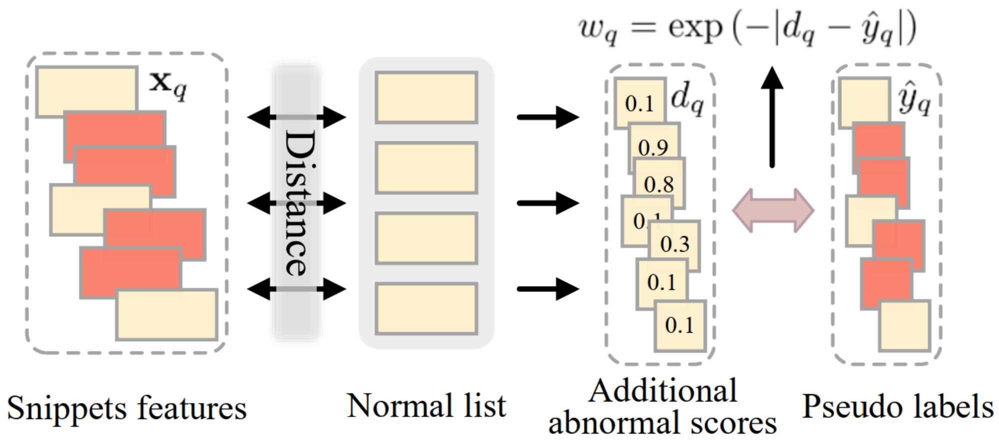
- We contrast **temporally-modulated feature-based similarity matrix** by computing the snippets' **feature space similarities** and then modulating them by their respective **temporal positions**.

Method

➢ Unsupervised Learning of Abnormal Detection

- Pseudo Labeling:

  Video-level pseudo labels are generated first, and then snippet-level pseudo labels are generated.



- Loss re-weighting strategy:

  The loss weight of a snippet is measured by the **discrepancy** between its **pseudo label** and its **additional normal-based anomaly score**.

Experiments

- **Dataset**

ShanghaiTech

UCF-Crime

- **Evaluation Metric**

  - We use the **frame-level area under the ROC curve** (AUC) for evaluation and comparisons.

Experiments

EUROPEAN CONFERENCE ON COMPUTER VISION

MILANO 2024

➢ Comparison with the State-of-the-Art

- We establish a new **state-of-the-art** on two datasets with ResNext as the backbone.
- Our method achieves better performance than one-class methods.
- As no annotations are provided, we still perform inferior to weakly-supervised methods.

	Method	Features	ShanghaiTech	UCF-Crime
Weakly-supervised	Sultani *et al.* [33]	C3D	-	75.41
	CLAWS [49]	C3D	89.67	83.03
	CLAWS Net+ [50]	C3D	90.12	83.37
	MIST [5]	C3D	93.13	81.40
	RTFM [36]	C3D	91.51	83.28
	MIST [5]	I3D	94.83	82.30
	RTFM [36]	I3D	97.21	84.30
	Zhang *et al.* [51]	I3D	-	86.22
	CLAWS [49]	ResNext	-	82.61
	CLAWS Net+ [50]	ResNext	91.46	84.16
	Zaheer *et al.* [47]	ResNext	86.21	79.84
One-class	Lu *et al.* [23]	-	68.00	65.51
	BODS [40]	I3D	-	68.26
	GODS [40]	I3D	-	70.46
	OGNet [48]	ResNext	69.90	69.47
	Zaheer *et al.* [47]	ResNext	79.62	74.20
Unsupervised	DyAnNet [35]	I3D	-	79.76
	C2FPL [1]	I3D	-	**80.65**
	Ours	I3D	**88.32**	80.02
	Kim *et al.* [15]	ResNext	56.47	52.00
	LBR-SPR [46]	ResNext	77.12	57.18
	GCL [47]	ResNext	78.93	71.04
	Tur *et al.* [37]	ResNext	68.88	62.91
	Tur *et al.* [38]	ResNext	66.36	63.52
	C2FPL [1]	ResNext	67.36	74.71
	Ours	ResNext	**86.46**	**76.64**

Experiments

ECCV
EUROPEAN CONFERENCE ON COMPUTER VISION
MILANO 2024

➢ Validation of Normality Prior

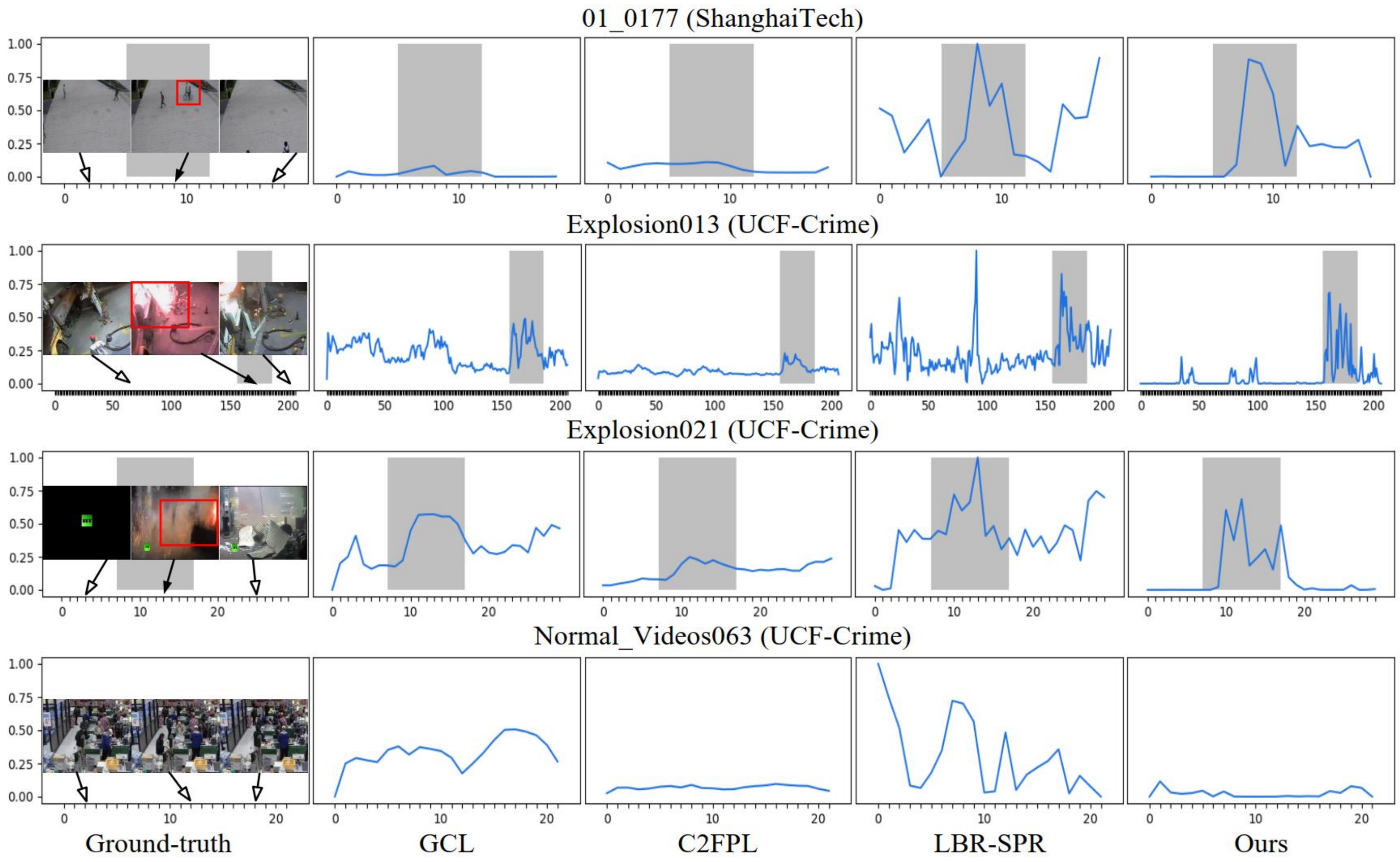
	ShanghaiTech			UCF-Crime		
	Precision	Recall	TestAUC	Precision	Recall	TestAUC
Random	17.90	29.27	81.69	13.28	42.41	60.33
Data-driven Prior	19.74	24.64	83.59	15.19	48.23	68.51
Normality Prior	34.39	42.92	85.96	20.37	54.28	75.99

➢ Ablation Study of Normality Propagation

	Pairwise Similarity	ShanghaiTech	UCF-Crime
Direct Comparison	-	72.33	71.22
Normality Propagation	T	79.62	57.99
	F	79.73	63.01
	T&F	85.96	75.99

Experiments

➢ **Qualitative Analysis**



Conclusion

- **Conclusion**
  - We propose to learn anomalies with **normality prior** for video anomaly detection.
  - We perform unsupervised learning of abnormal detection based on proposed **normality propagation** and a **new loss re-weighting method**.

- **Limitation**
  - As our method is build on the semantic consistency, it performs weak in a video with multiple scenes.
  - In the future, we will explore combining with other methods that focus on **detecting anomalies in multiple scenes**.