NEURAL INFORMATION PROCESSING SYSTEMS

Advancing Video Anomaly Detection: A Concise Review and a New Dataset

Liyun Zhu, Lei Wang, Arjun Raj, Tom Gedeon, Chen Chen

Paper: https://arxiv.org/abs/2402.04857

Project page: https://msad-dataset.github.io

Australian National University
DATA 61
CSIRO
Curtin University
UCF
TIME LAB

1.1 Challenge & Motivation

Video Anomaly Detection (VAD) presents a challenge in real-world scenarios, particularly in security and surveillance applications.

Characteristic: unknown, diverse, and infrequent

Motivation

- Current methods struggle with diverse anomalies and complex environments.
- Although current surveys are comprehensive, they are not portable and lightweight.

1.2 Contribution

We are motivated to offer a concise review that highlights current challenges, research trends, and future directions, providing valuable insights and guidance for researchers.

We propose a new Multi-Scenario Anomaly Detection (MSAD) dataset, a high-resolution, real-world anomaly detection benchmark encompassing diverse scenarios and anomalies.

We propose a novel Scenario Adaptive Anomaly Detection (SA^2D) model, using few-shot learning for efficient adaptation to new scenarios.

2.1 Related Works – Existing Datasets

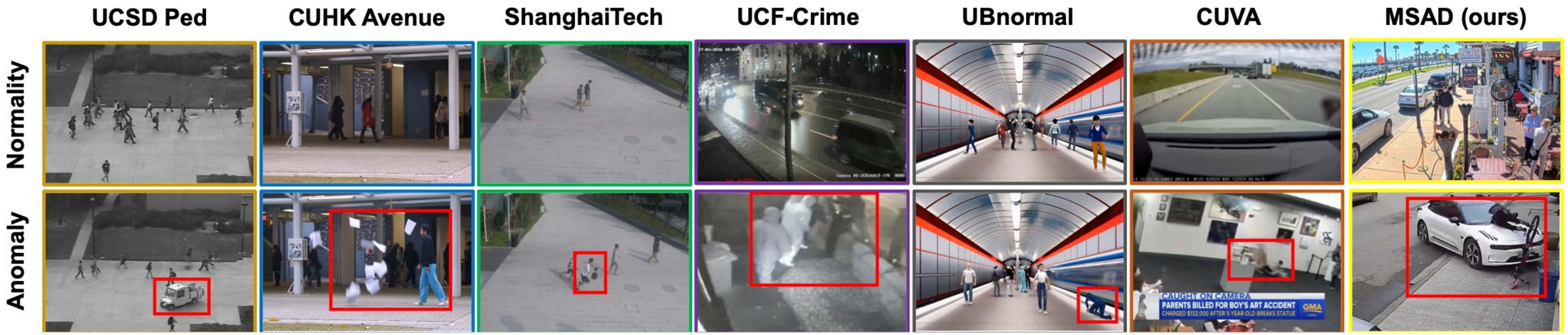


Figure 1: A comparison of existing datasets such as UCSD Ped, CUHK Avenue, ShanghaiTech, UCF-Crime, UBnormal and CUVA *vs*. our Multi-Scenario Anomaly Detection (MSAD) dataset.

Drawbacks of existing datasets: Poor quality, Limited scenarios, Unreasonable anomaly types, Lack of non-human-related anomalies…

3.1 Our MSAD Dataset

Dataset	Year	Source	Domain	#Video	#HRA	#NHRA	#View	#Scenario	Modality	Resolution	Variations
Subway Entrance [3]	2008	Surveillance	Pedestrian	1	5	-	1	1	RGB	512×384	✗
Subway Exit [3]	2008	Surveillance	Pedestrian	1	3	-	1	1	RGB	512×384	✗
UMN [45]	2009	Surveillance	Behavior	5	1	-	3	1	RGB	320×240	✗
UCSD Ped1 [79]	2010	Surveillance	Pedestrian	70	5	-	1	1	RGB	238×158	✗
UCSD Ped2 [79]	2010	Surveillance	Pedestrian	28	5	-	1	1	RGB	238×158	✗
CUHK Avenue [35]	2013	Surveillance	Pedestrian	35	5	-	1	1	RGB	640×360	✗
ShanghaiTech [37]	2017	Surveillance	Pedestrian	437	13	-	13	1	RGB	856×480	✗
UCF-Crime [72]	2018	Online Surv.	Crime	1900	12	1	NA	NA	RGB	Multiple	✓
Street Scene [55]	2020	Surveillance	Traffic	81	17	-	1	1	RGB	1280×720	✗
IITB Corridor [63]	2020	Surveillance	Pedestrian	358	10	-	1	1	RGB	1920×1080	✗
XD-Violence [92]	2020	Films/Online	Violence	4754	5	1	NA	NA	RGB+Audio	640×360	✓
UBnormal [2]	2022	3D modeling	Pedestrian	543	20	2	29	8	RGB	1080×720	✓
NWPU Campus [6]	2023	Surveillance	Pedestrian	547	27	1	43	1	RGB	Multiple	✗
CUVA [14]	2024	News/Online	Multiple	1000	27	15	NA	NA	RGB+Text	Multiple	✓
MSAD (ours)	2024	Online Surv.	Multiple	720	**35**	**20**	**∼500**	**14**	RGB	Multiple	✓

3.1 MSAD Dataset

Diverse scenarios, objects, weather and lighting conditions

14 scenarios

3.1 Our MSAD Dataset

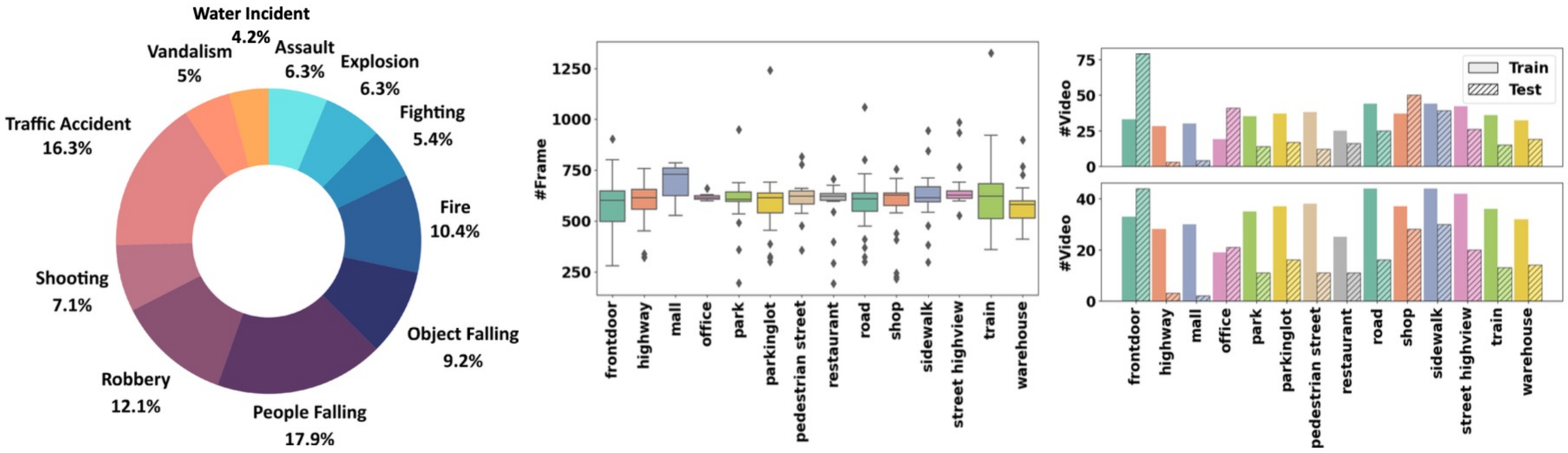
11 Anomaly types:

Assault, Explosion, Fighting, Fire, Object falling, People falling, Robbery, Shooting, Traffic accident, Vandalism, Water incident

Non-human-related Anomalies

Human-related Anomalies

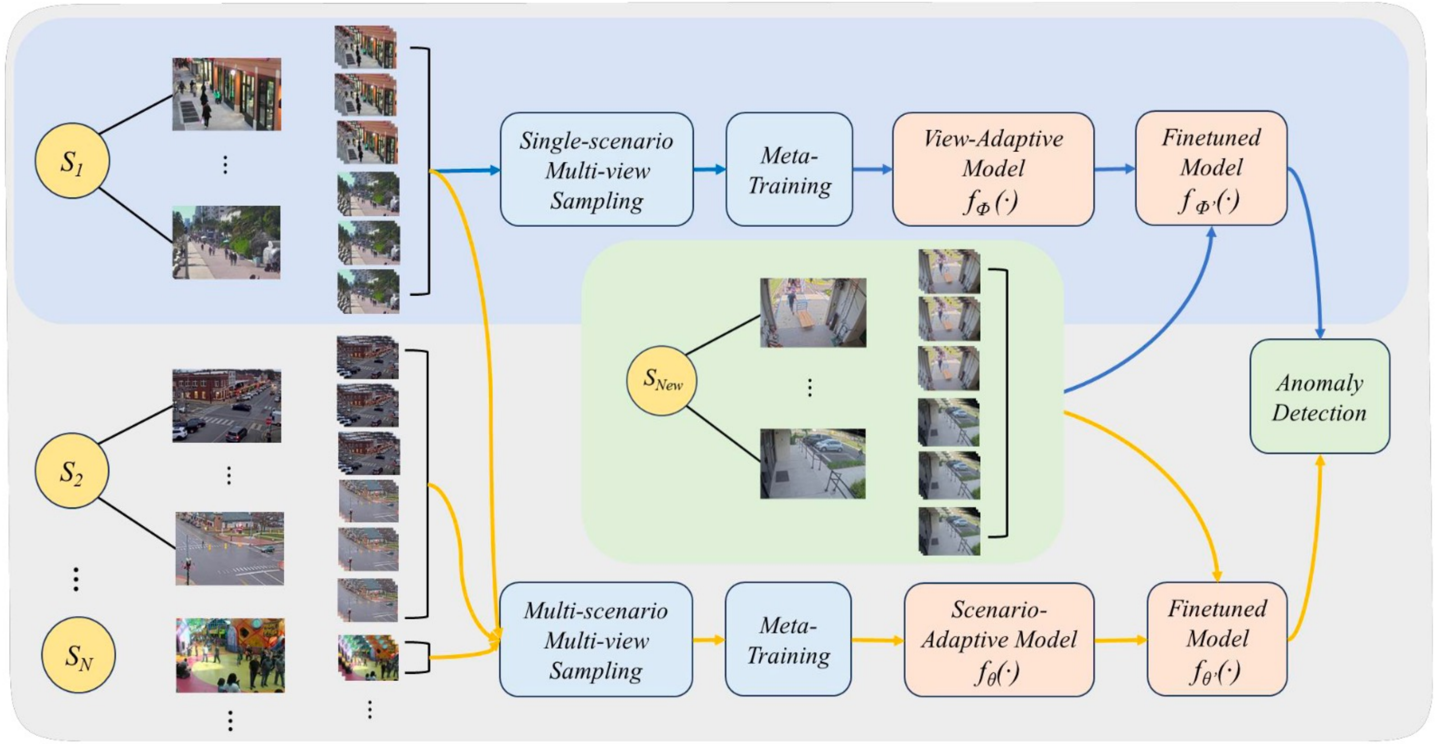
3.1 Our MSAD Dataset



(a) The proportions of anomalies. (b) Variations in frame numbers. (c) Distributions of train/test splits.

Figure 3: The statistics of our MSAD dataset include: (a) a breakdown of main anomaly types and their respective percentages, (b) a boxplot illustrating frame number variations across scenarios in MSAD training set, and (c) the distributions of train/test splits across scenarios for two evaluation protocols (see Sec. 3 evaluation protocols): (*top* plot) generalizability and adaptability, and (*bottom* plot) practical applicability and effectiveness.

3.2 Our Proposed Method - SA²D



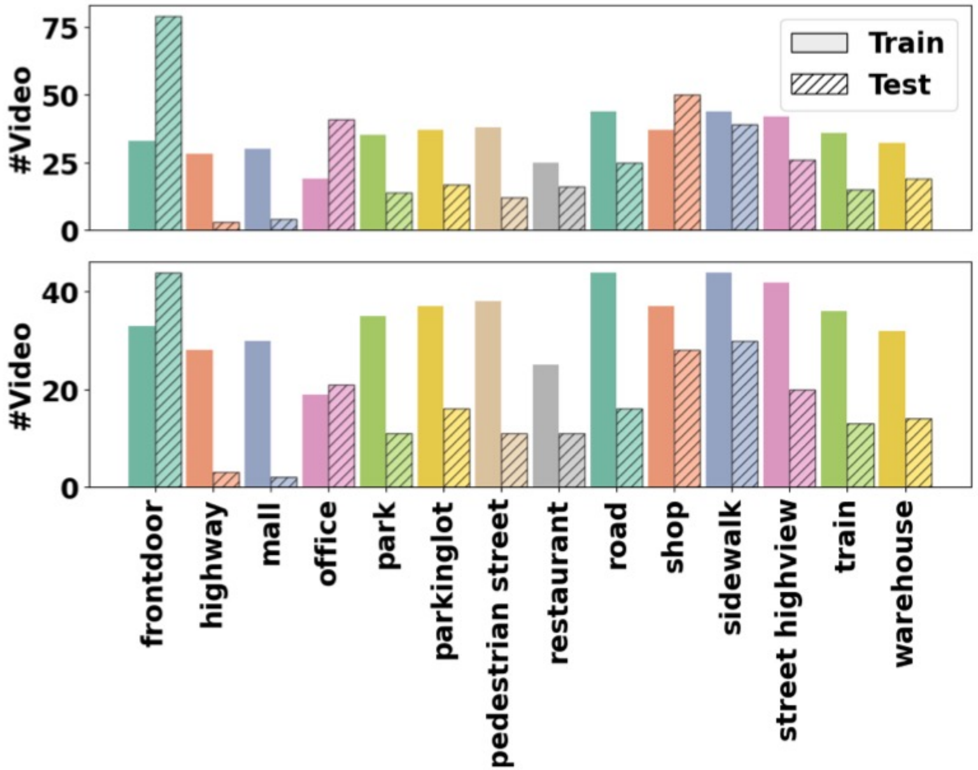
Current method only train the model with different viewpoints under single scenario.

Scenario-Adaptive: Apply the few-shot learning in various scenarios and finetune the model in a new scenario.

3.3 Evaluation - Two Protocols

(i) Train on 360 normal videos from 14 scenarios and test on the remaining 120 normal videos and 240 abnormal videos. This protocol is suitable for evaluating self-supervised methods.

(ii) Train on 360 normal and 120 abnormal videos, and test on 120 normal and 120 abnormal videos. During training, we only provide video-level annotations. This protocol is suitable for evaluating weakly-supervised methods trained with our video-level annotations.



3.4 Evaluation - Self-supervised Methods

Table 2: Experimental results on single-scenario evaluation. On ShanghaiTech (ShT), only 7 views are used during training and the rest views are individually used for testing. The notation ShT-$v*$ denotes the use of different camera views.

Test view	Train	FSAD [36]		Train	FSAD [36]		**SA²D (ours)**	
		Micro	Macro		Micro	Macro	Micro	Macro
ShT-$v1$	ShT (7 views)	61.36	55.34	**MSAD**	63.74	62.92	**68.96**	**77.89**
ShT-$v3$		26.51	26.58		64.39	62.56	**67.59**	**73.43**
ShT-$v5$		53.40	53.32		55.04	54.63	**55.74**	**54.02**
ShT-$v6$		78.36	78.27		70.26	71.02	**75.47**	**72.35**
ShT-$v8$		50.02	52.54		59.97	57.45	**60.85**	**61.52**

Table 3: Evaluations on cross-scenario setups. We use FSAD [36] and SA^{2}D (ours) for training on ShanghaiTech (ShT) and MSAD, respectively.

Train	Test	AUC	
		Micro	Macro
ShT	UCSD Ped2	57.38	58.36
	CUHK Avenue	69.98	78.32
	MSAD	**63.92**	**64.92**
MSAD	UCSD Ped2	**70.35**	**65.74**
	CUHK Avenue	**79.57**	**84.49**
	MSAD	**69.96**	**69.60**

3.4 Evaluation - Weakly-supervised Methods

Table 4: Comparison of six methods with varying backbones on UCF-Crime, ShanghaiTech, and our MSAD dataset using three popular backbones: C3D, I3D, and SwinTransformer (SwinT).

	Venue	UCF-Crime			ShanghaiTech			**MSAD**	
		C3D	I3D	SwinT	C3D	I3D	SwinT	I3D	SwinT
MIST [18]	CVPR 2021	81.40	82.30	-	93.13	94.83	-	-	-
RTFM [76]	ICCV 2021	83.28	83.14	83.31	91.51	97.94	96.76	86.65	85.67
MSL [31]	AAAI 2022	82.85	85.30	85.62	94.23	95.45	97.32	-	-
UR-DMU [102]	AAAI 2023	82.65	86.19	83.74	94.67	96.15	95.71	85.02	72.36
MGFN [13]	AAAI 2023	82.37	83.44	84.30	90.82	93.97	93.58	84.96	78.94
TEVAD [12]	CVPRW 2023	83.39	84.54	84.65	92.05	98.10	97.63	86.82	83.60

3.4 Evaluation - Cross dataset

Table 9: Comparison of cross-dataset results using four recent anomaly detection models with the I3D backbone. UCF, ShT, CUHK, and Ped2 denote UCF-Crime, ShanghaiTech, CUHK Avenue, and UCSD Ped2, respectively. Improvements from using models pre-trained on the MSAD dataset are highlighted in red, while performance drops are indicated in blue.

Method	UCF→ShT	UCF→CUHK	UCF→Ped2	**MSAD**→ShT	**MSAD**→CUHK	**MSAD**→Ped2
RTFM [76]	42.62	50.76	60.03	39.59 (↓3.03%)	63.23 (↑12.47%)	57.97 (↑2.06%)
UR-DMU [102]	46.69	45.67	62.90	35.05 (↓11.64%)	58.86 (↑13.19%)	66.84 (↑3.94%)
MGFN [13]	37.58	44.48	51.75	48.10 (↑10.52%)	56.66 (↑12.18%)	62.09 (↑10.34%)
TEVAD [12]	59.34	43.39	36.96	45.27 (↓14.07%)	64.82 (↑21.43%)	62.56 (↑25.60%)

Evaluating the generalization of different weakly-supervised methods.

3.4 Evaluation - Different Anomalies & Scenarios

Table 7: Performance evaluations by anomaly type (a total of 11 main anomaly types) on our MSAD test set are conducted. We use frame-level Micro AUC (%) and Average Precision (AP, in %) as evaluation metrics for models pretrained on either UCF-Crime or our MSAD. We use I3D as the backbone for all methods. The best training scheme for each method is highlighted in bold.

Training set	Method	Assault		Explosion		Fighting		Fire	
		AUC	AP	AUC	AP	AUC	AP	AUC	AP
UCF-Crime	RTFM [76]	60.6	62.2	**69.3**	**79.0**	68.5	80.7	36.0	64.5
	MGFN [13]	**60.5**	**62.0**	65.5	**74.3**	53.6	63.9	21.6	55.0
	UR-DMU [102]	**59.4**	60.5	**69.3**	**82.0**	71.2	85.2	36.2	66.5
MSAD	RTFM [76]	**68.1**	**67.3**	46.8	60.4	**89.6**	**93.0**	**61.3**	**81.2**
	MGFN [13]	59.7	59.0	**64.5**	71.9	**89.4**	**93.5**	**86.0**	**93.0**
	UR-DMU [102]	56.9	**64.5**	67.9	74.5	**83.9**	**90.4**	**61.2**	**82.9**

Training set	Method	Object Falling		People Falling		Robbery		Shooting	
		AUC	AP	AUC	AP	AUC	AP	AUC	AP
UCF-Crime	RTFM [76]	82.0	88.8	**69.5**	**63.0**	**76.8**	**90.6**	59.7	65.7
	MGFN [13]	65.5	73.1	**57.2**	**59.5**	72.0	**89.1**	42.1	57.6
	UR-DMU [102]	72.4	76.5	**69.3**	**57.6**	**69.7**	**81.5**	59.9	73.8
MSAD	RTFM [76]	**94.7**	**96.7**	56.5	50.4	65.7	81.2	**78.2**	**84.7**
	MGFN [13]	**90.9**	**94.8**	52.7	47.8	**73.9**	86.7	**86.8**	**88.5**
	UR-DMU [102]	**92.1**	**95.8**	42.5	43.7	63.5	79.3	**81.4**	**87.8**

Training set	Method	Traffic Accident		Vandalism		Water Incident		**Overall**	
		AUC	AP	AUC	AP	AUC	AP	AUC	AP
UCF-Crime	RTFM [76]	55.6	45.1	**86.0**	**85.2**	93.5	98.5	71.9	47.4
	MGFN [13]	52.6	45.3	80.7	**81.4**	41.0	81.7	61.8	31.2
	UR-DMU [102]	53.0	47.9	**91.6**	**89.7**	64.6	91.3	74.3	53.4
MSAD	RTFM [76]	**62.2**	**51.8**	85.2	76.1	**96.3**	**99.1**	**86.7**	**66.3**
	MGFN [13]	**68.6**	**54.5**	**82.4**	80.1	**85.5**	**97.0**	**85.0**	**63.5**
	UR-DMU [102]	**62.0**	**55.6**	84.7	77.0	**98.5**	**99.5**	**85.0**	**68.3**

Table 8: Performance evaluations by scenario (a total of 14 scenarios) on our MSAD test set are conducted. We use frame-level Micro AUC (%) and Average Precision (AP, in %) as evaluation metrics for models pretrained on either UCF-Crime or our MSAD. We use I3D as the backbone for all methods. The best training scheme for each method is highlighted in bold.

Training set	Method	Frontdoor		Highway		Mall		Office	
		AUC	AP	AUC	AP	AUC	AP	AUC	AP
UCF-Crime	RTFM [76]	80.8	80.1	37.1	1.4	86.0	**87.1**	68.5	63.2
	MGFN [13]	68.4	70.2	36.3	1.4	**79.6**	**80.4**	64.5	60.2
	UR-DMU [102]	84.7	82.6	18.9	1.1	83.1	80.6	66.6	57.6
MSAD	RTFM [76]	**84.1**	**81.1**	**63.7**	**4.1**	**87.2**	72.2	**78.1**	**68.8**
	MGFN [13]	**86.4**	**85.1**	**79.7**	**4.1**	65.3	56.6	**75.1**	**62.4**
	UR-DMU [102]	**84.8**	**82.8**	**31.5**	**1.3**	**91.0**	**83.8**	**77.8**	**67.3**

Training set	Method	Park		Parkinglot		Pedestrian st.		Restaurant	
		AUC	AP	AUC	AP	AUC	AP	AUC	AP
UCF-Crime	RTFM [76]	**75.3**	23.7	66.7	16.7	84.1	**67.6**	66.5	56.5
	MGFN [13]	55.3	7.9	59.5	12.3	74.4	11.2	47.3	32.4
	UR-DMU [102]	**91.6**	34.8	62.2	17.6	58.5	6.1	75.7	74.4
MSAD	RTFM [76]	69.0	**25.6**	**74.4**	**35.9**	**97.4**	50.6	**96.1**	**91.9**
	MGFN [13]	**77.9**	**38.3**	**68.1**	**14.5**	**88.0**	**20.4**	**95.8**	**91.8**
	UR-DMU [102]	87.8	**36.2**	**91.4**	**53.9**	**81.9**	**11.5**	**93.1**	**87.4**

Training set	Method	Road		Shop		Sidewalk		Street highview	
		AUC	AP	AUC	AP	AUC	AP	AUC	AP
UCF-Crime	RTFM [76]	**82.9**	**47.1**	**85.1**	68.5	**89.1**	**66.1**	**82.6**	**35.9**
	MGFN [13]	54.4	18.3	69.4	60.4	47.4	26.4	37.2	8.3
	UR-DMU [102]	49.5	26.6	78.8	**66.5**	68.0	55.9	62.0	23.0
MSAD	RTFM [76]	54.0	16.8	80.6	**77.3**	52.5	17.1	43.3	12.3
	MGFN [13]	**77.9**	**49.7**	**84.9**	**77.2**	**85.5**	**62.3**	**87.6**	**40.7**
	UR-DMU [102]	**83.0**	**64.4**	**81.3**	64.5	**86.5**	**64.1**	**85.0**	**37.7**

Training set	Method	Train		Warehouse		**Overall**	
		AUC	AP	AUC	AP	AUC	AP
UCF-Crime	RTFM [76]	52.2	**5.0**	**82.3**	**52.8**	71.9	47.4
	MGFN [13]	39.8	2.1	55.4	18.3	61.8	31.2
	UR-DMU [102]	51.3	2.6	**86.9**	54.0	74.3	53.4
MSAD	RTFM [76]	**66.9**	3.9	69.5	37.4	**86.7**	**66.3**
	MGFN [13]	**53.0**	**3.1**	**72.3**	30.9	**85.0**	**63.5**
	UR-DMU [102]	**59.0**	**3.1**	81.2	**59.1**	**85.0**	**68.3**

Thank you!