Autonomous Driving Anomaly Detection: A Weakly Supervised Horizon

Utkarsh Tiwari

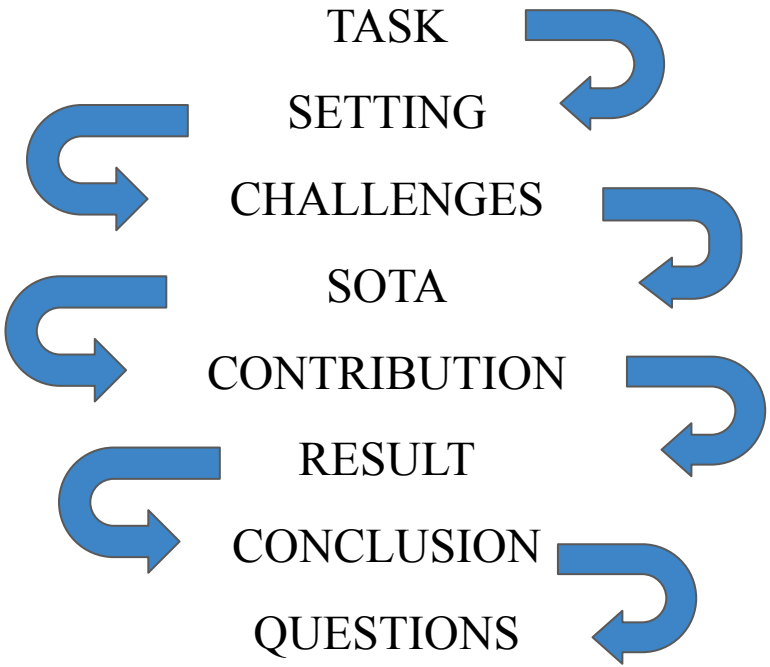
Snehashis Majhi

Michal Balazia

François Brémond

INRIA Sophia Antipolis, France

Presentation Flow



The Task: Video Anomaly Detection

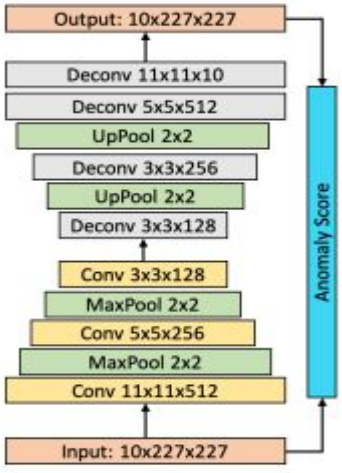
How to identify, localise, and classify anomalies in videos

How to learn a discriminative representation for real-world anomalies?

The Setting: Unsupervised Vs. Supervised Vs. Weakly-supervised

Unsupervised : Pixel Reconstruction Based Approach

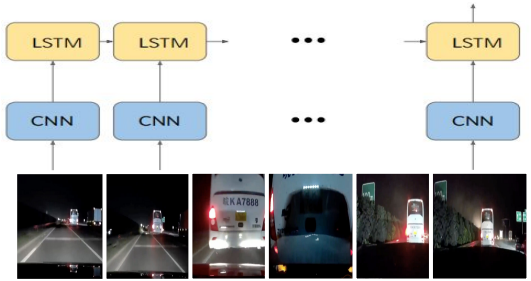
- Requires NO Annotation



- Low Generalization Ability to diverse scenarios
- High False positive in unseen training sample

Supervised : Frame-level Classification Approach

- Requires Full Annotations



Weakly-supervised : Multiple-instance Learning Approach

- Only Video-Level Annotations required
- Higher generalization ability w.r.t unsupervised methods

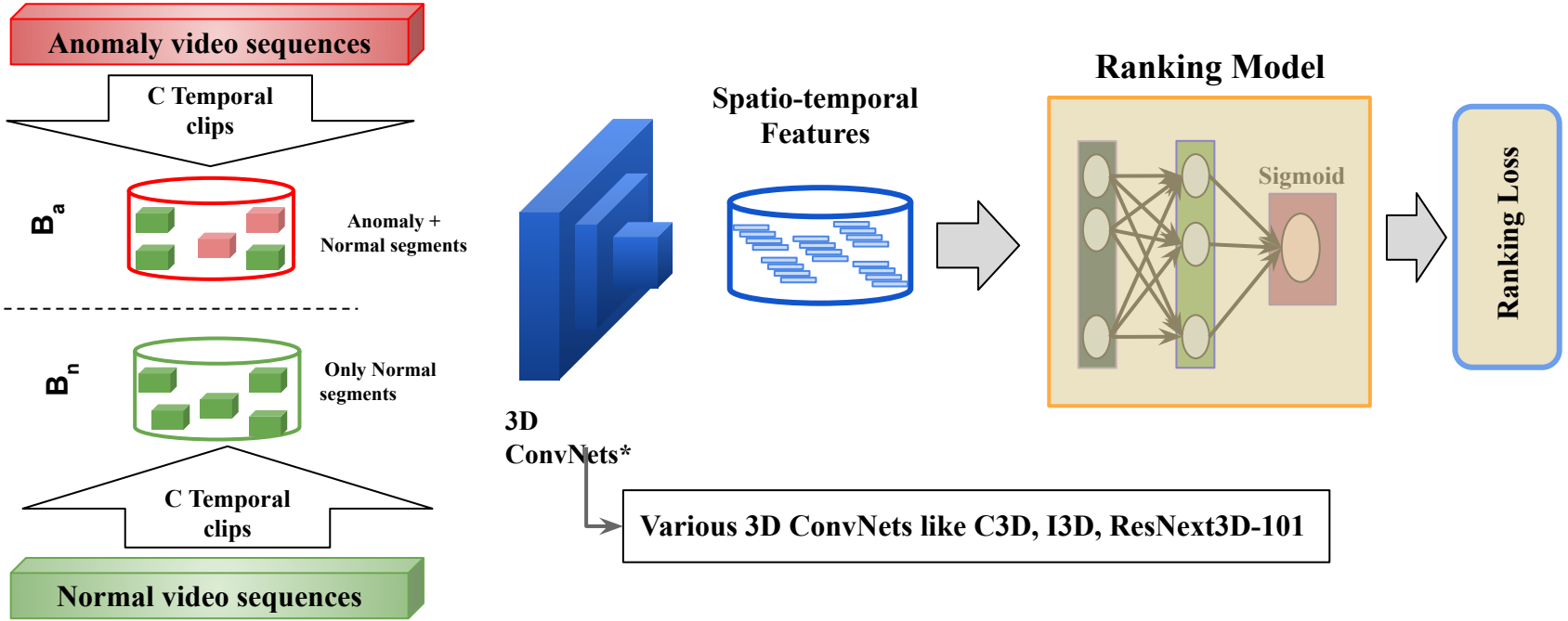
Frame-level Annotations [Supervised]

- Laborious and Time consuming
- Prone to Error

Video-level Annotations [Weakly Supervised]

- Easy to Obtain
- Less Error

Basic Multiple Instance Learning Training Framework



The (Modified) Task: Weakly Supervised Video Anomaly Detection for Traffic Scenarios

How to identify anomaly instances (frames) with video-level labels

How to learn a discriminative representation for real-world anomalies?

The Challenges

1) Complex dynamic scenarios due to moving cameras
2) Low camera field of view
3) Little to no prior cues before anomaly occurrence
4) Obstructions/reflections due to the camera on the dashboard

State-of-the-art Weakly-supervised VAD methods

Methods:

1. MGFN (AAAI 2023 Oral)
2. OE-CTST (WACV 2023)
3. RTFM (ICCV 2021)
4. UR-DMU (AAAI 2023)

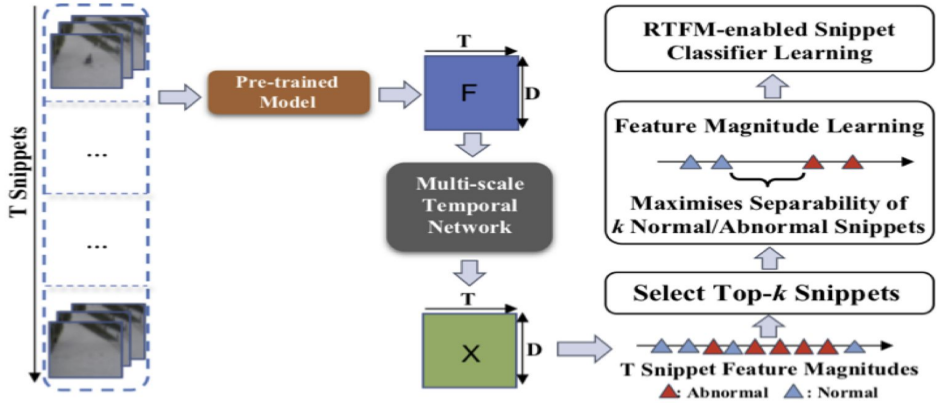


Fig. 5: RTFM [19] Framework

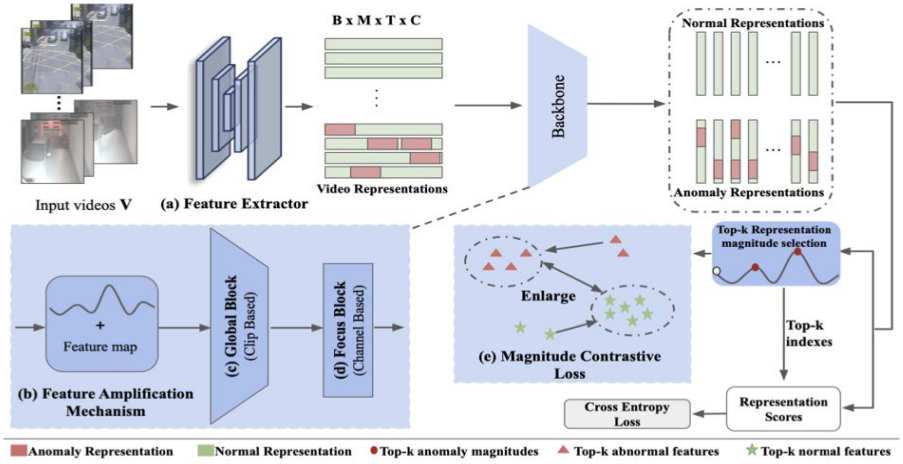


Fig. 3: MGFN. [3] Framework

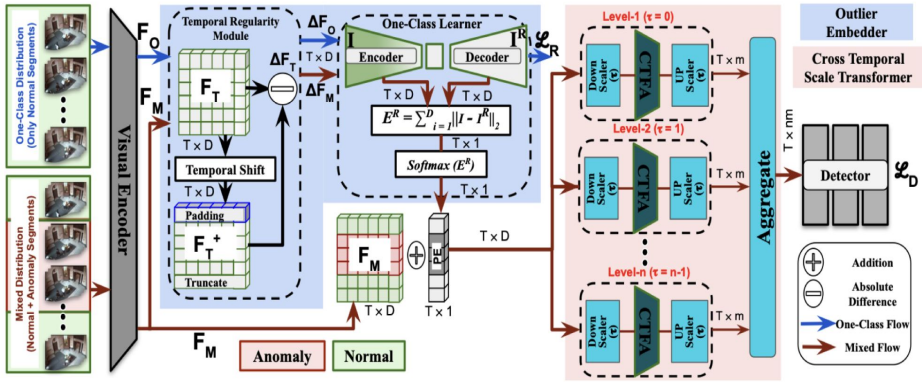


Fig. 4: OECTST [14] Framework

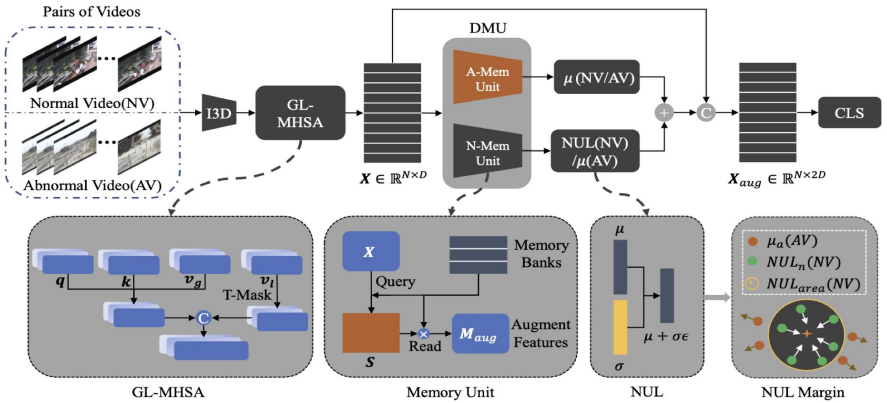


Fig. 6: UR-DMU [28] Framework

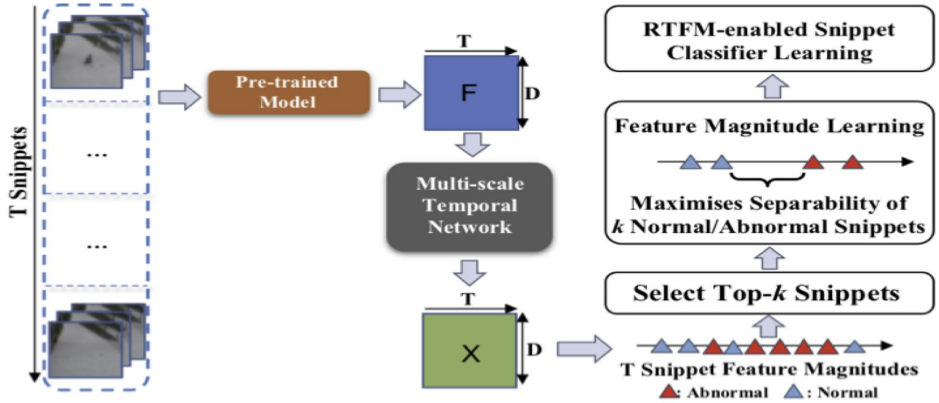


Fig. 5: RTFM [19] Framework

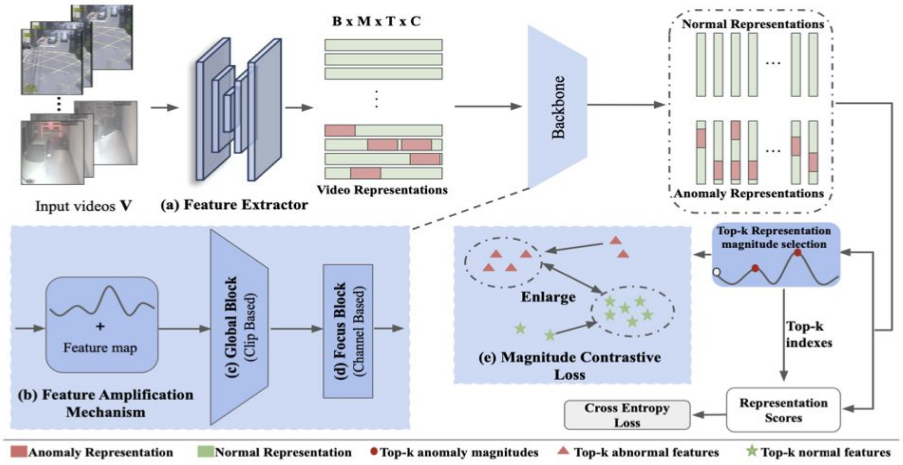


Fig. 3: MGFN. [3] Framework

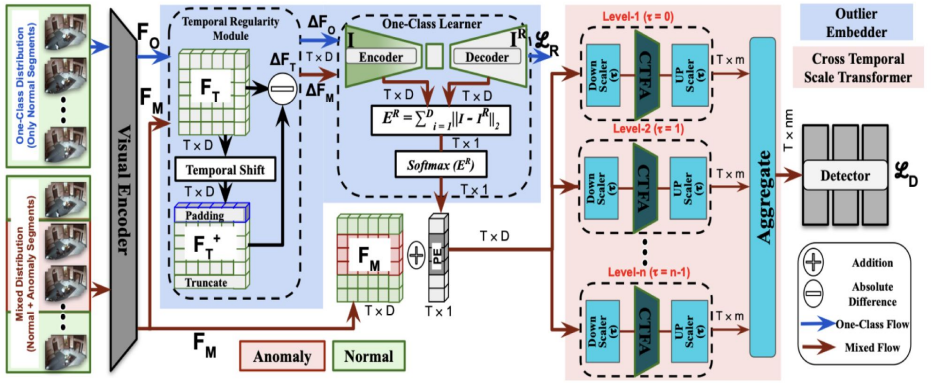


Fig. 4: OECTST [14] Framework

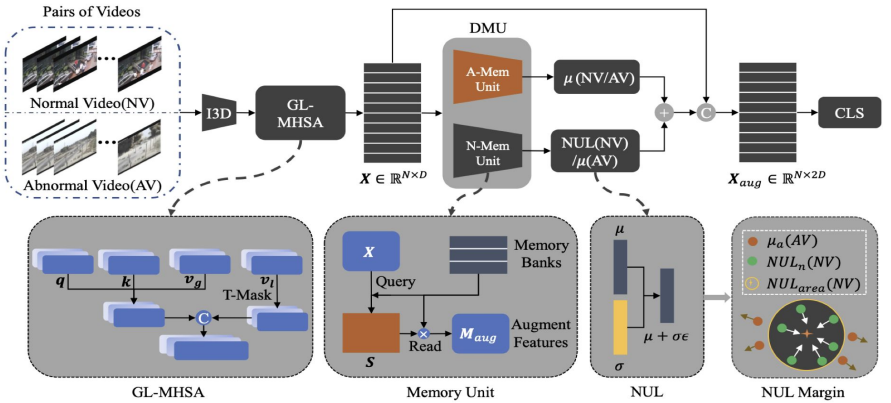


Fig. 6: UR-DMU [28] Framework

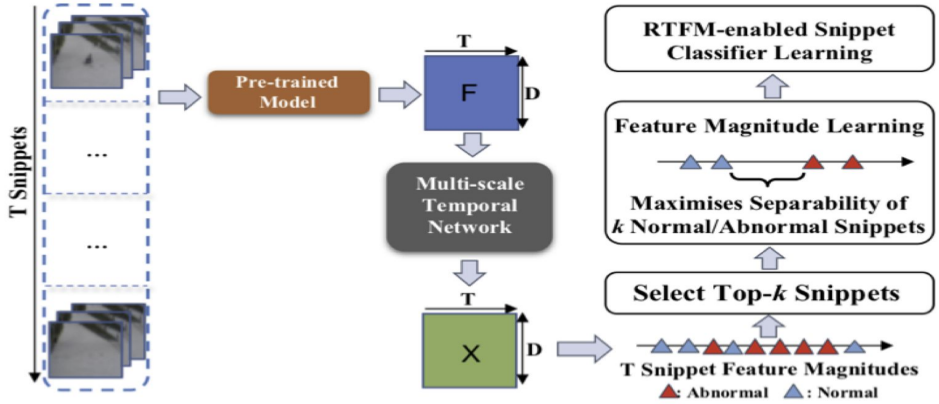


Fig. 5: RTFM [19] Framework

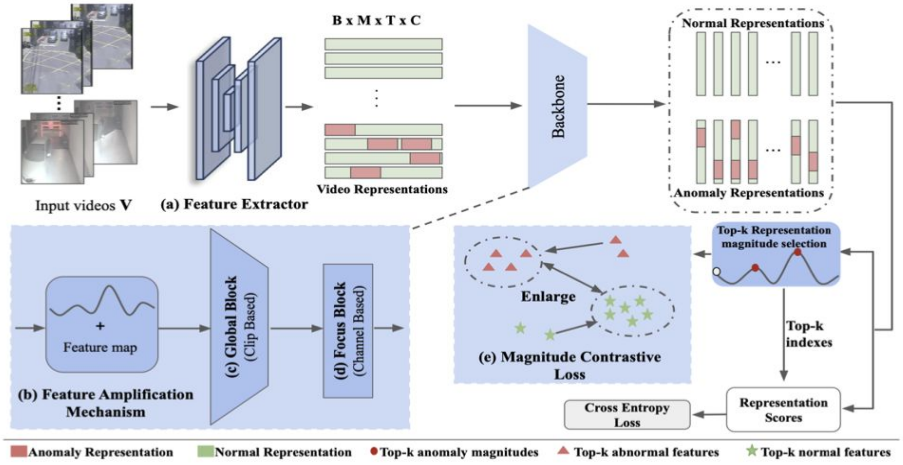


Fig. 3: MGFN. [3] Framework

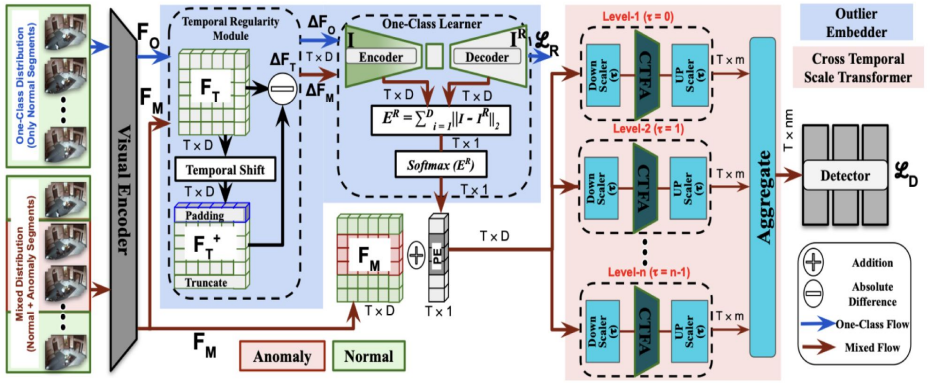


Fig. 4: OECTST [14] Framework

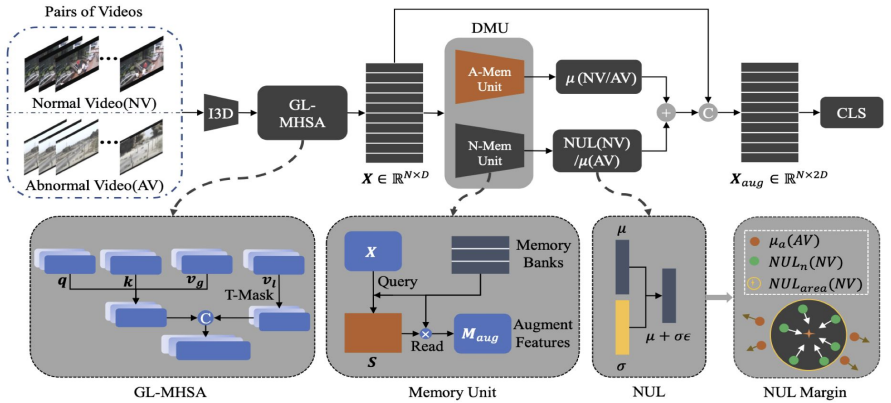


Fig. 6: UR-DMU [28] Framework

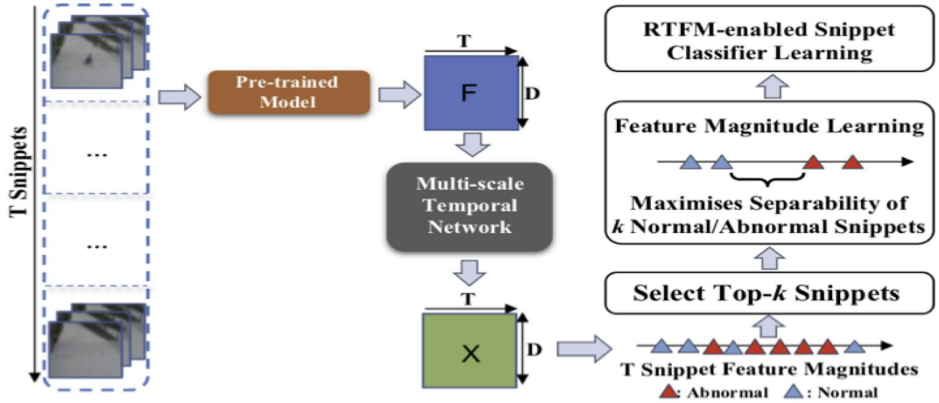


Fig. 5: RTFM [19] Framework

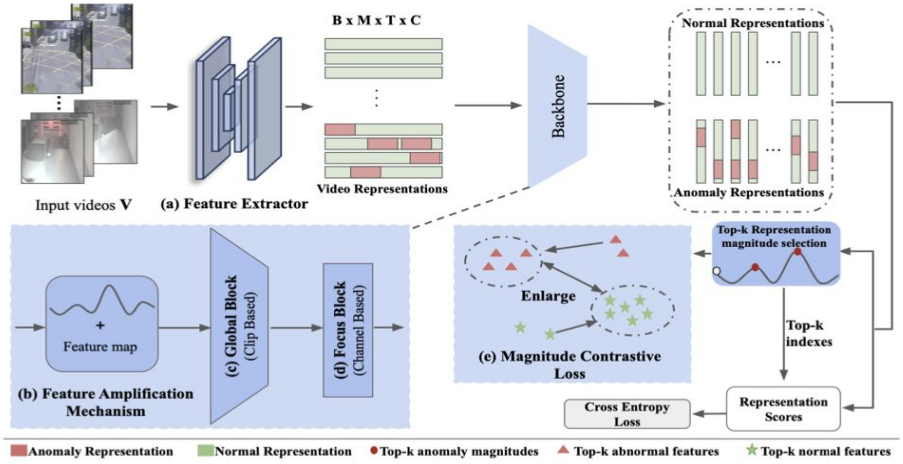


Fig. 3: MGFN. [3] Framework

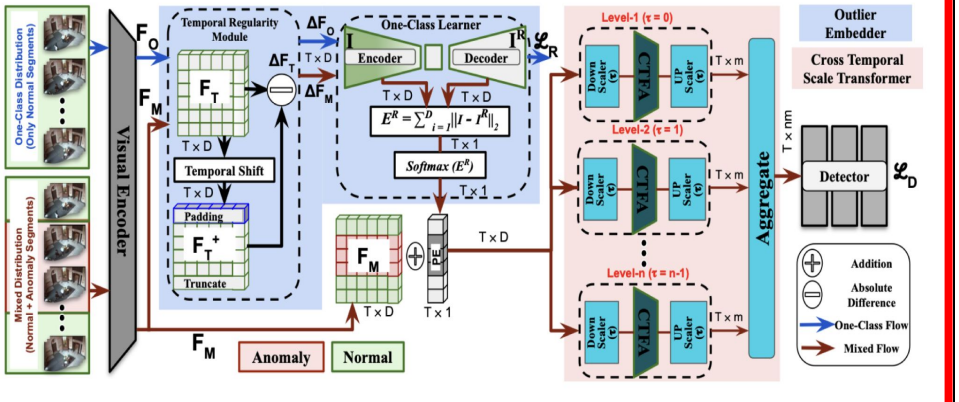


Fig. 4: OECTST [14] Framework

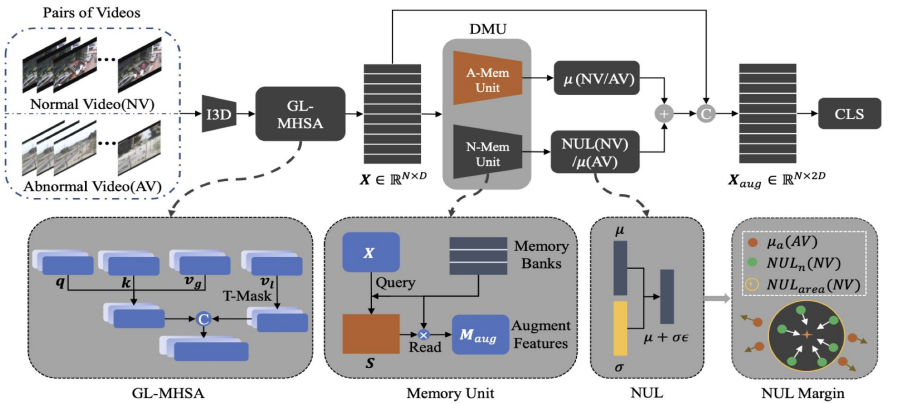


Fig. 6: UR-DMU [28] Framework

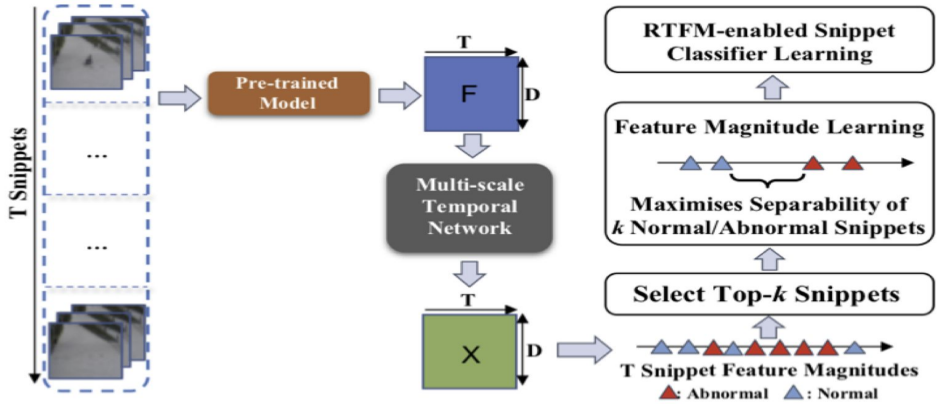


Fig. 5: RTFM [19] Framework

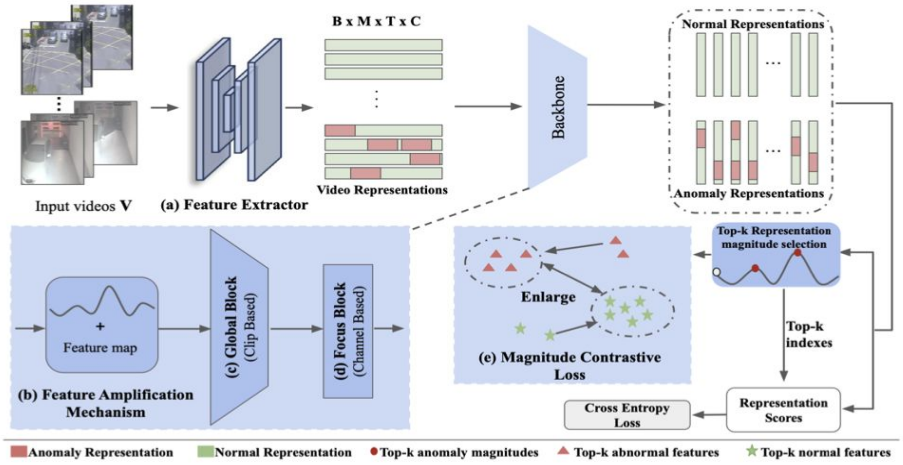


Fig. 3: MGFN. [3] Framework

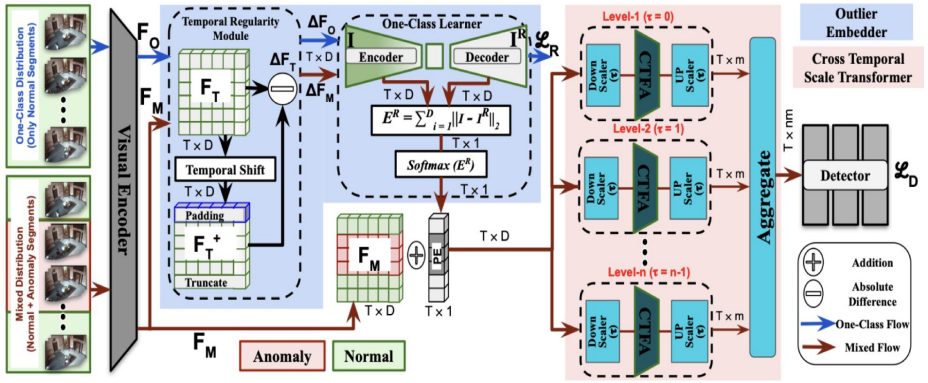


Fig. 4: OECTST [14] Framework

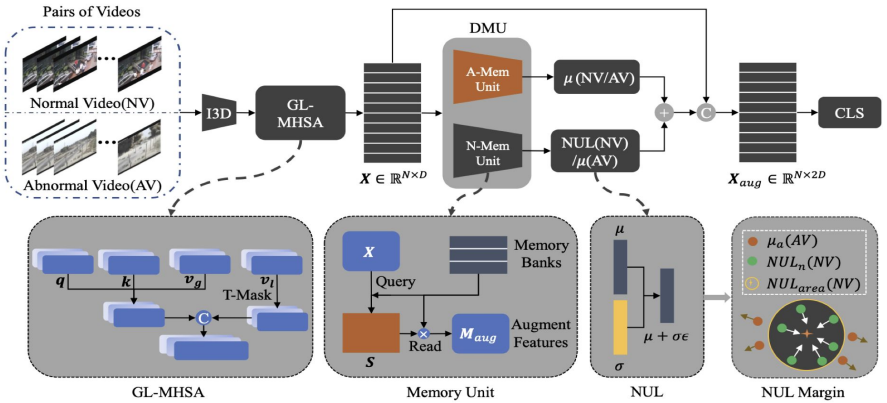


Fig. 6: UR-DMU [28] Framework

Our Contributions

1) **A reorganized dataset, named WS-DoTA:** To promote Weakly Supervised methods exploration on Traffic Anomaly scenarios.

2) **Analysis and Evaluation of SoTA methods:** To benchmark on our reorganized dataset.

3) **Proposed a Feature Transformation Block:** To improve the salient feature learning of SoTA methods.

WS-DoTA: DoTA+D^2-City (training set) + DoTA (test set)

Frame Count	Train Split		Test Split							
	Normal	Anomaly	ST	AH	LA	OC	TC	VP	VO	OO
Average	737.8	104.6	25.5	32.6	36.7	28.4	29.1	30.1	30.4	49.2
Minimum	287	30	9	7	4	5	1	10	12	9
Maximum	750	299	50	84	158	203	135	71	75	143
Total Videos	3592	2689	24	164	168	115	390	35	29	106

ST : Collision with another vehicle which starts, stops, or is stationary
AH : Collision with another vehicle moving ahead or waiting
LA : Collision with another vehicle moving laterally in the same direction
OC : Collision with another oncoming vehicle
TC : Collision with another vehicle which turns into or crosses a road
VP : Collision between vehicle and pedestrian
VO : Collision with an obstacle in the roadway
OO : Out-of control and leaving the roadway to the left or right.

Sample Frames of WS-DoTA

NEXT Slides->

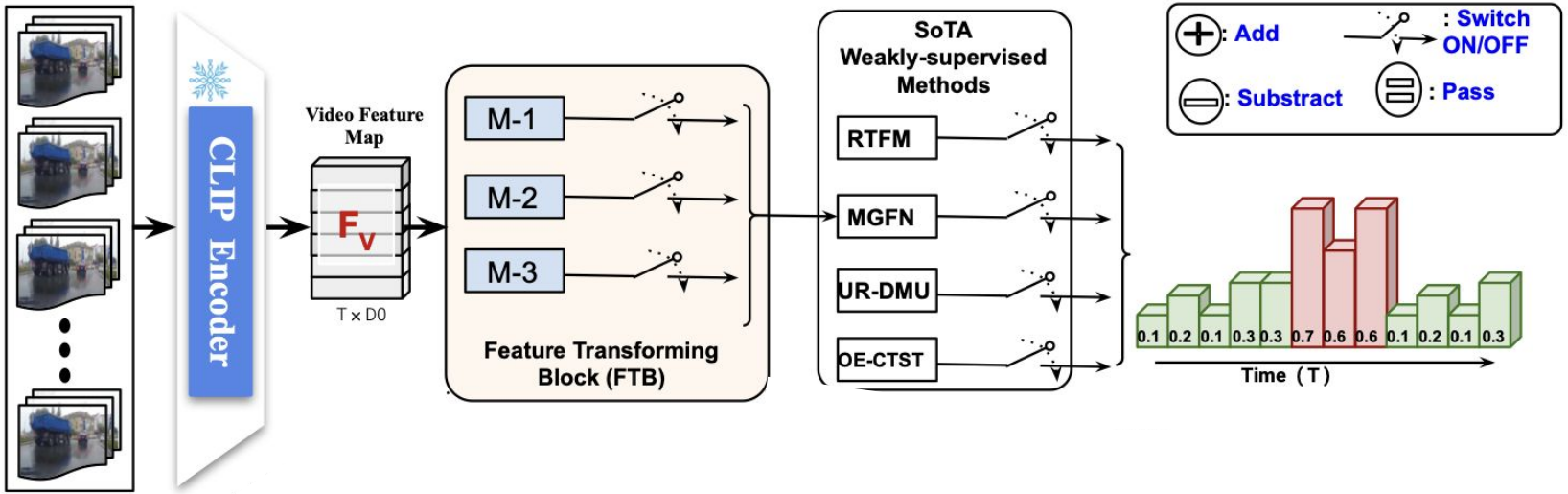

65KM/H 2019/04/10 16:19:13


112KM/H 2019/03/19 20:17:08

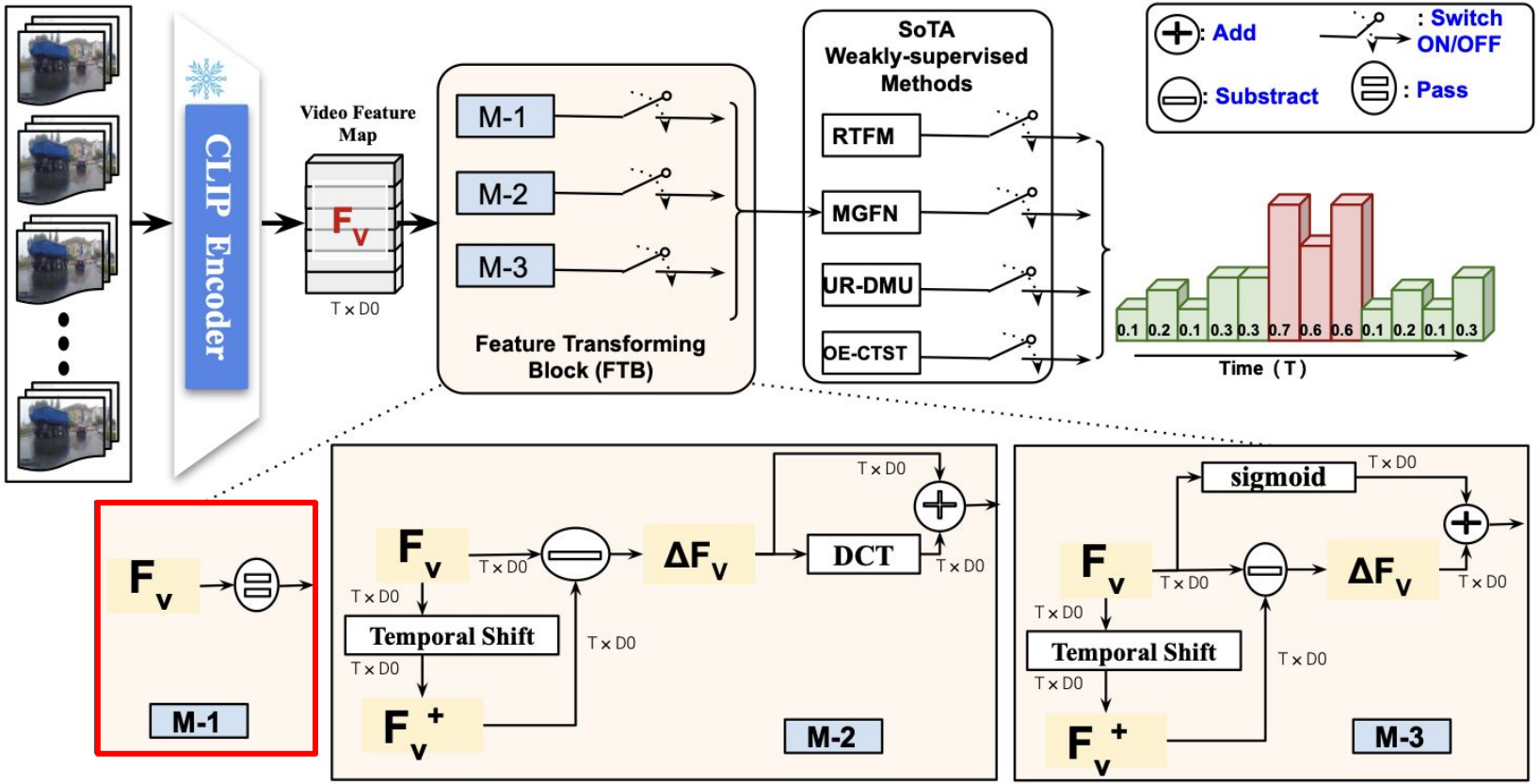

CAR CRASHES TIME
YOUTUBE CHANNEL
18+
2017/02/14
08:56:24


CAR CRASHES TIME
YOUTUBE CHANNEL
18+
17/12/2016 12:55:20

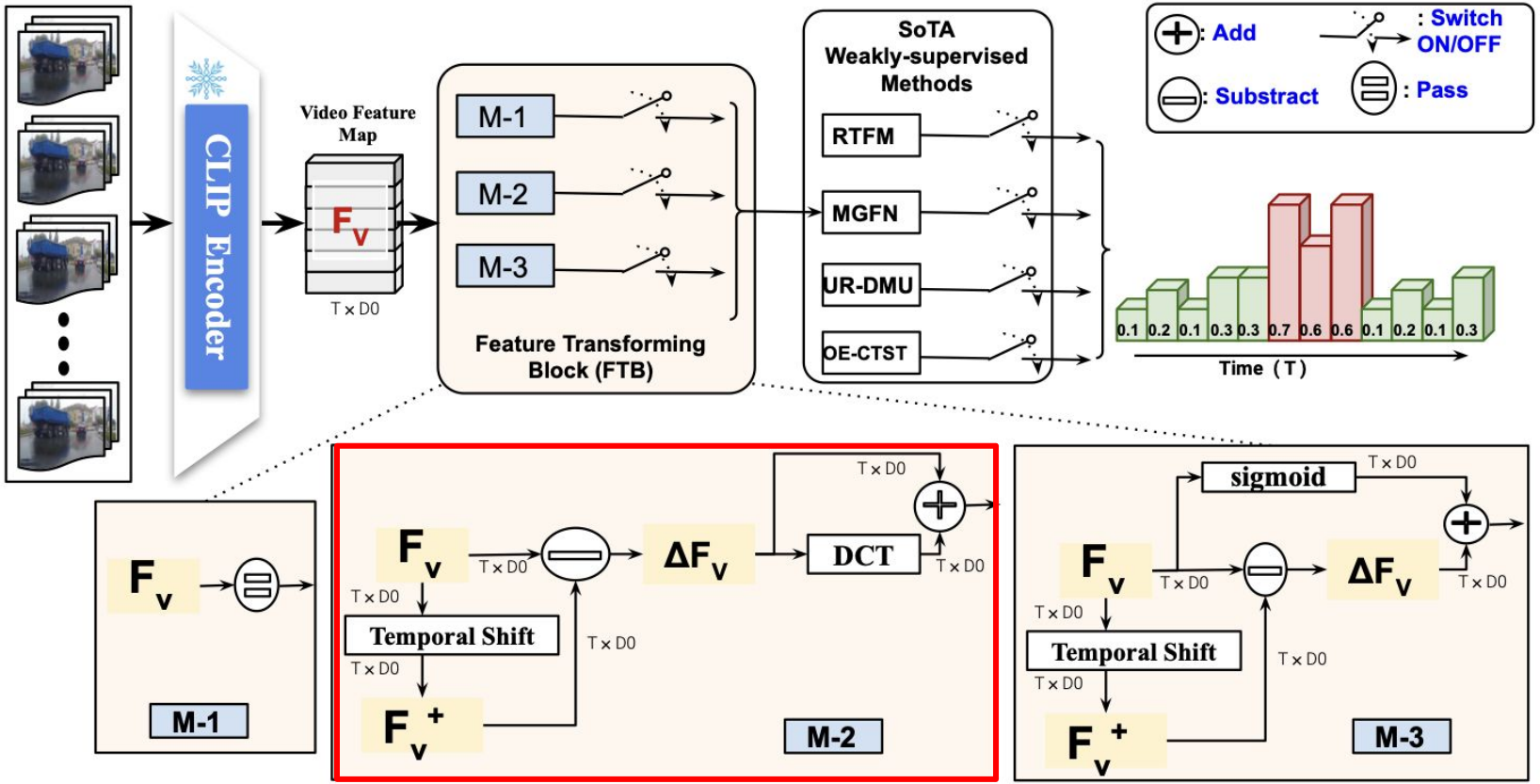
Proposed Experimental Analysis Framework



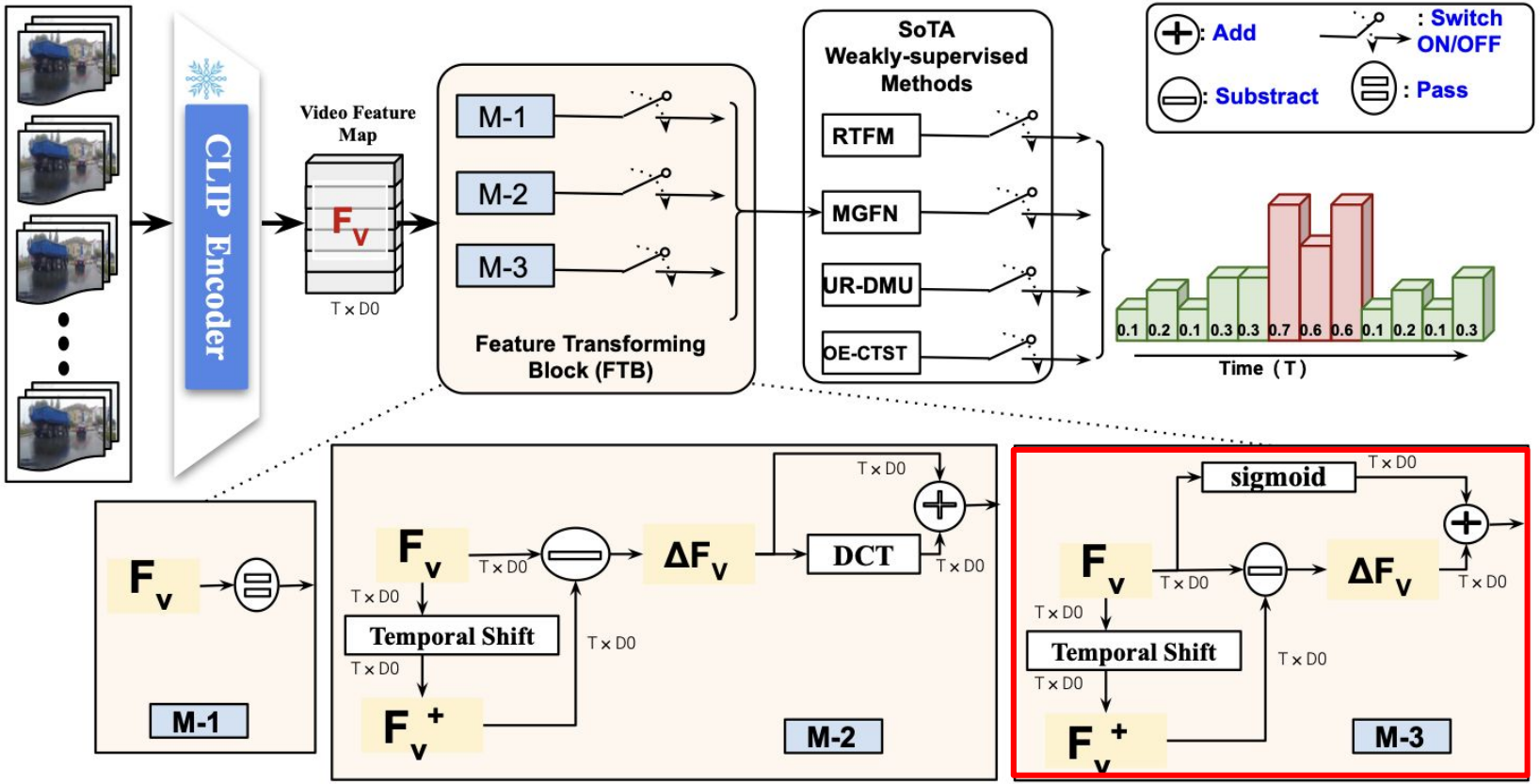
Proposed Experimental Analysis Framework



Proposed Experimental Analysis Framework



Proposed Experimental Analysis Framework



Stat-of-the-art Benchmarking - Analysis - Comparison

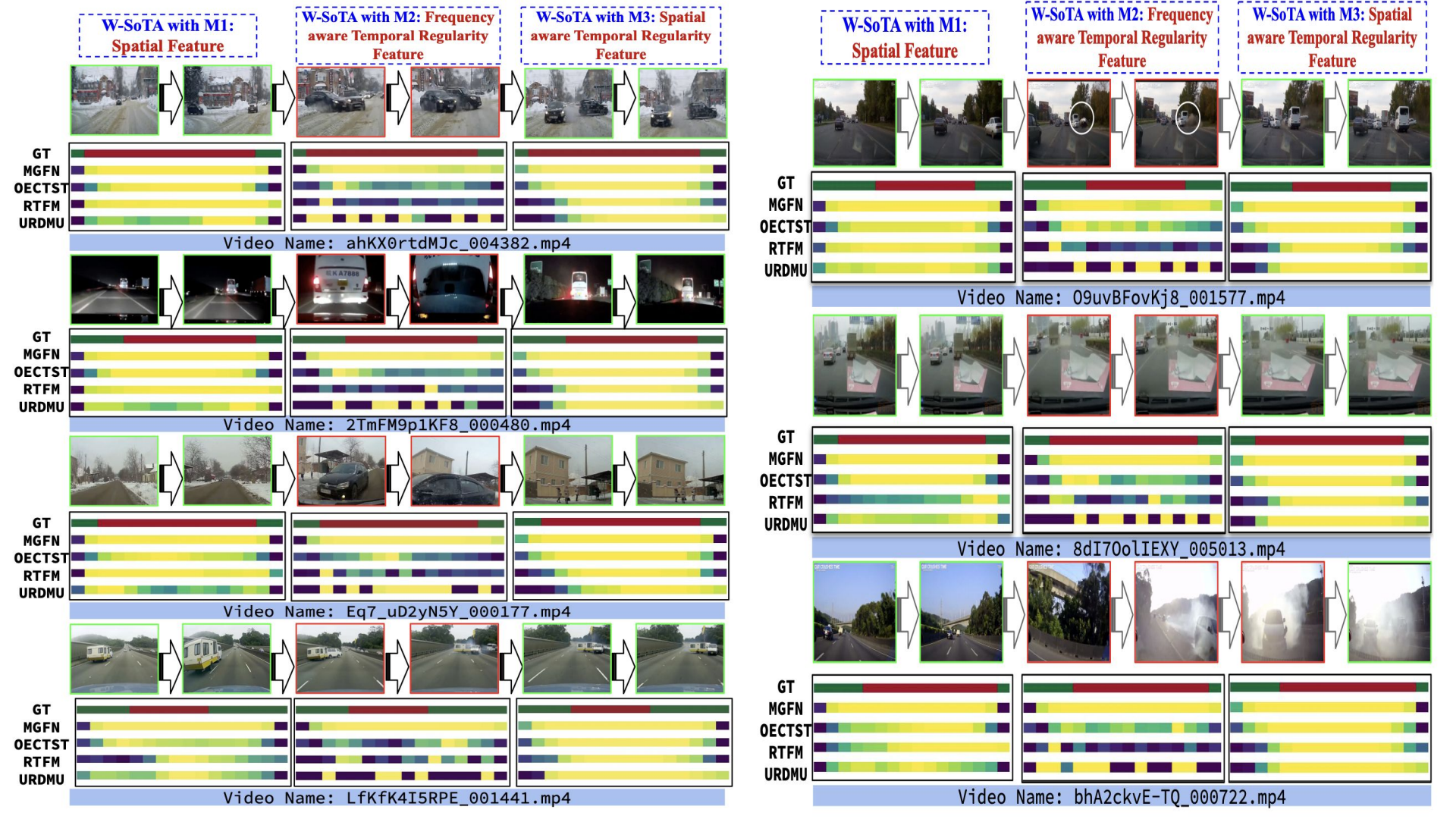
Trained on pre-cursor frames

Trained on WS-DoTA

Methods	Overall AUC (%)	Class-Wise Performance (AUC%) ST	AH	LA	OC	TC	VP	VO	OO
Unsupervised Method with RGB only Feature									
ConvAE (gray) [7]	64.3	-	-	-	-	-	-	-	-
ConvAE (flow) [7]	66.3	-	-	-	-	-	-	-	-
ConvLSTMAE (gray) [5]	53.8	-	-	-	-	-	-	-	-
ConvLSTMAE (flow) [5]	62.5	-	-	-	-	-	-	-	-
AnoPred (RGB) [13]	67.5	**70.4**	68.1	67.6	67.6	69.4	65.6	64.2	57.8
AnoPred (Mask RGB) [13]	64.8	69.6	67.9	62.4	66.1	65.6	65.3	58.8	59.9
TAD (Bbox+ flow) [24]	69.2	-	-	-	-	-	-	-	-
TAD [24] + ML [9] [12] (Bbox+ flow)	69.7	71.2	71.8	68.9	71.3	70.6	67.4	63.8	69.2
Ensemble (RGB + Bbox+ flow)	73.0	75.4	75.5	71.0	75.0	74.5	70.6	65.2	69.6
Supervised method with RGB only Feature									
LSTM [8] (RGB)	63.7	-	-	-	-	-	-	-	-
Encoder-Decoder [4] (RGB)	73.0	-	-	-	-	-	-	-	-
TRN [22] (RGB)	78.0	-	-	-	-	-	-	-	-
*Weakly-Supervised Methods with **M1: Spatial only Feature***									
RTFM [19]	57.9	59.8	58.6	57.6	56.5	56.2	55.2	51.6	60.6
MGFN [3]	66.6	57.1	66.2	64.6	69.6	67.0	63.0	64.3	69.3
URDMU [28]	57.5	50.8	58.8	60.0	57.4	56.7	55.3	53.2	56.2
OE-CTST [14]	70.9	64.2	71.4	71.5	68.2	71.2	66.2	69.6	75.2
*Weakly-Supervised Methods with **M2: Frequency aware Temporal Regularity Feature***									
RTFM [19]	56.0	57.1	56.1	55.7	53.4	56.2	57.9	53.9	58.1
MGFN [3]	67.4	67.1	70.0	66.8	67.9	67.6	67.6	73.7	69.0
URDMU [28]	54.8	58.4	56.3	54.3	53.0	54.7	52.8	54.5	55.1
OE-CTST [14]	71.9	66.3	70.6	72.0	72.1	71.1	67.1	76.4	75.9
*Weakly-Supervised Methods with **M3: Spatial aware Temporal Regularity Feature***									
RTFM [19]	**78.2**	62.7	**79.2**	**78.7**	**76.5**	**77.5**	**74.7**	**79.8**	**83.1**
MGFN [3]	67.4	60.8	68.9	66.5	66.8	67.3	61.2	66.1	68.0
URDMU [28]	73.0	63.8	71.1	72.4	72.9	74.9	65.4	79.5	75.9
OE-CTST [14]	75.6	63.6	77.4	76.0	73.8	74.9	73.3	76.2	78.1

Qualitative Analysis

NEXT
Slide->


W-SoTA with M1: Spatial Feature
W-SoTA with M2: Frequency aware Temporal Regularity Feature
W-SoTA with M3: Spatial aware Temporal Regularity Feature
GT
MGFN
OECTST
RTFM
URDMU
Video Name: ahKX0rtdMJc_004382.mp4
Video Name: 2TmFM9p1KF8_000480.mp4
Video Name: Eq7_uD2yN5Y_000177.mp4
Video Name: LfKfK4I5RPE_001441.mp4
Video Name: O9uvBFovKj8_001577.mp4
Video Name: 8dI7OolIEXY_005013.mp4
Video Name: bhA2ckvE-TQ_000722.mp4

Observations and Future Work

1) Weakly Supervised methods are better suited to VAD task due to dataset constraints, and with the correct feature transformations achieve better or competitive performance to supervised models with the same modalities

2) The proposed FTB (M3) helps improve the SoTA WS methods for autonomous driving AD

3) In future: Develop a deeper semantic understanding of anomalies using VLMs, Unsupervised+FTB, Experiments on the normal videos (to see false +ve rate), analyse STAUC performance

Thanks :D
Questions?

Pre-print of the paper