Towards Multi-Domain Learning for Generalizable Video Anomaly Detection

MyeongAh Cho*
Kyung Hee University
maycho@khu.ac.kr

Taeoh Kim
NAVER Cloud
taeoh.kim@navercorp.com

Minho Shim
NAVER Cloud
minho.shim@navercorp.com

Dongyoon Wee
NAVER Cloud
dongyoon.wee@navercorp.com

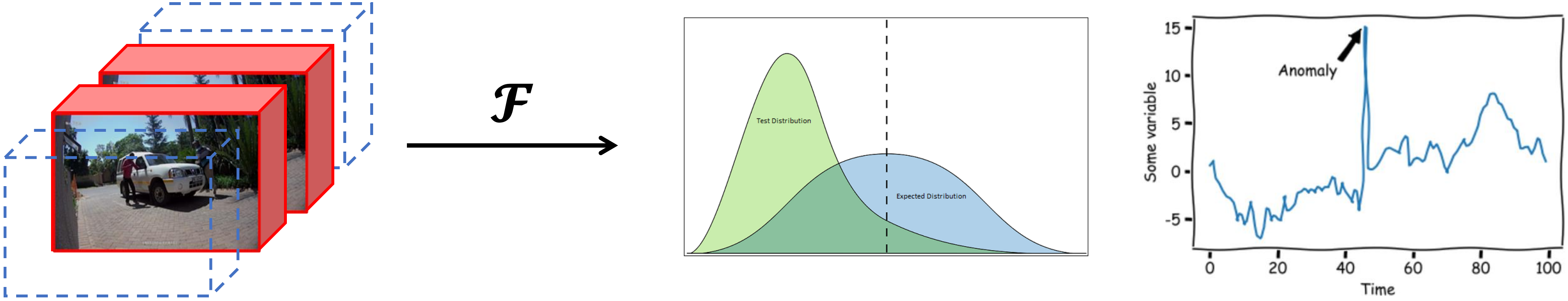
Sangyoun Lee†
Yonsei University
syleee@yonsei.ac.kr

NEURAL INFORMATION PROCESSING SYSTEMS

MVP LAB
Yonsei Image and Video Pattern Recognition Lab

NAVER Cloud

Video Anomaly Detection



Method	Training data	Test data
Unsupervised Learning	Normal videos **(Unlabeled)**	Unseen Abnormal videos **(frame-level detection)**
Weakly-supervised Learning	Normal videos + Abnormal videos **(video-level labeled)**	

Introduction

What is the problem with the existing VAD model?

Table 2: Anomaly detection performances (Area under curve, AUC) of single-domain models. Diagonal elements are in-domain results and off-diagonal elements are cross-domain results.

	Target					
Source	UCFC	XD	LAD	UBIF	TAD	ST
UCFC	82.32	68.06	75.75	71.12	73.75	59.24
XD	68.38	90.87	77.60	67.23	71.10	46.87
LAD	59.60	75.26	86.97	59.27	73.80	47.29
UBIF	74.79	75.22	70.29	93.63	68.16	54.21
TAD	50.83	45.38	52.02	61.95	90.71	41.58
ST	55.75	52.87	48.96	59.00	48.57	91.88

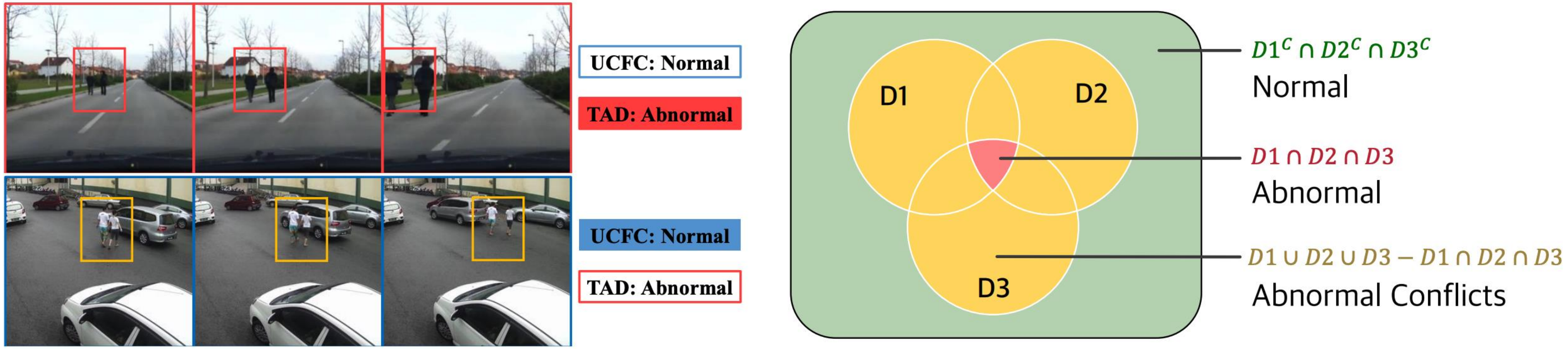
Introduction

Why do we need a general VAD model?

1) A single generalized model removes the need for multiple specific models for different domains, analogous to **multi-task learning**
2) Proper pre-training on multiple domains embodies generalized representation, leading to **better performance in unseen target domains**
3) A general VAD model will be highly beneficial for **practical scenarios**

Introduction

Is it possible to create a general VAD model?



(a) Examples of Abnormal Conflict between datasets

(b) Venn diagrams of events

Figure 1: **(a)** An example of abnormal conflict: *Pedestrian on the road* is normal in UCFC dataset but is abnormal in TAD. **(b)** Each circle represents each domain. MDVAD aims to design a general model that effectively considers abnormal conflicts to separate general normal and abnormal events.

Goal

“Construct a general VAD model by conducting multi-domain learning
while recognizing abnormal conflicts
and exploring representations of general normality and abnormality”

1) Multiple Domain VAD, along with a benchmark and new evaluation protocols
2) Domain-specific multiple heads to mitigate abnormal conflicts
3) Abnormal Conflict (AC) Classifier to explore general features while being aware of abnormal conflicts

Baselines

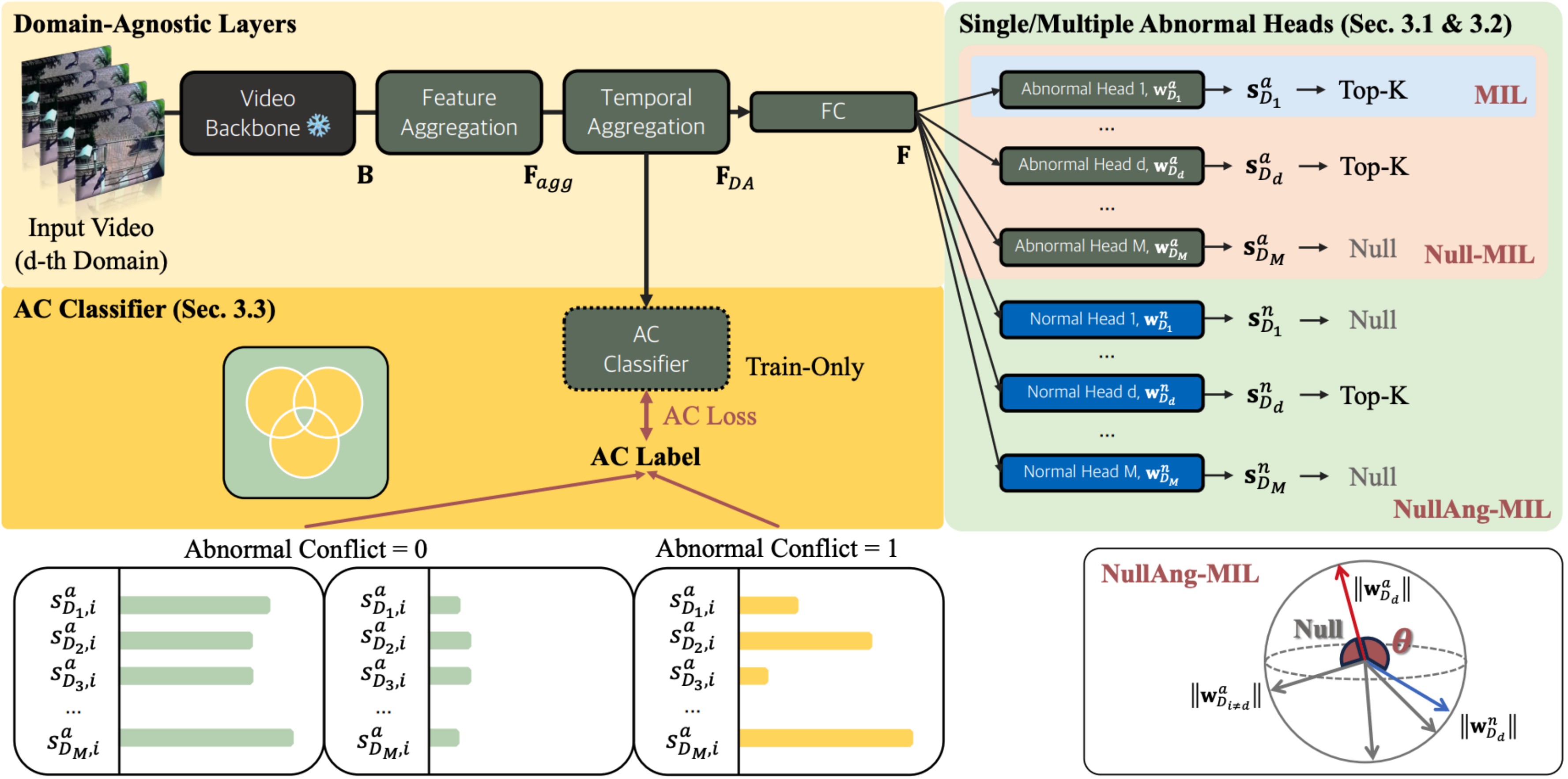
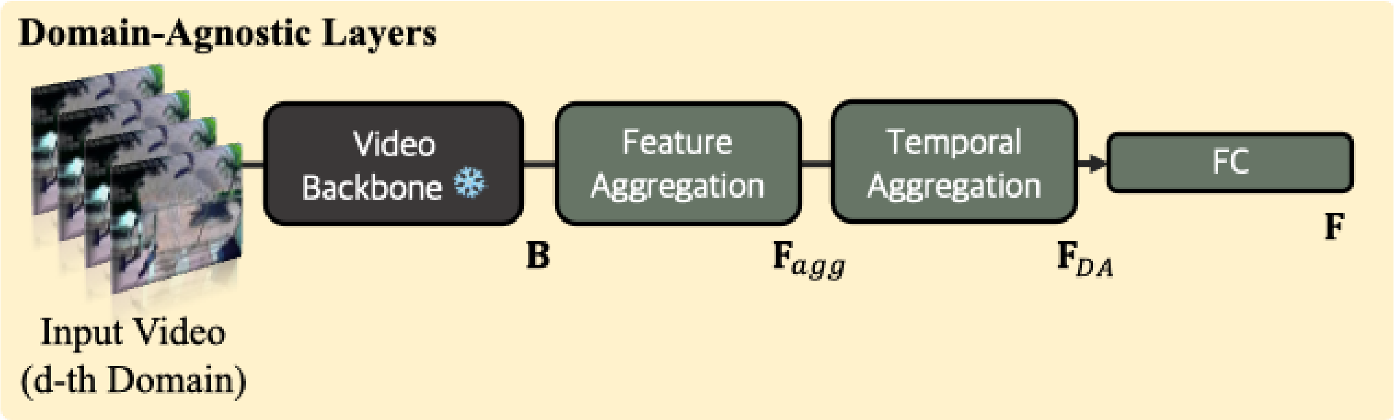


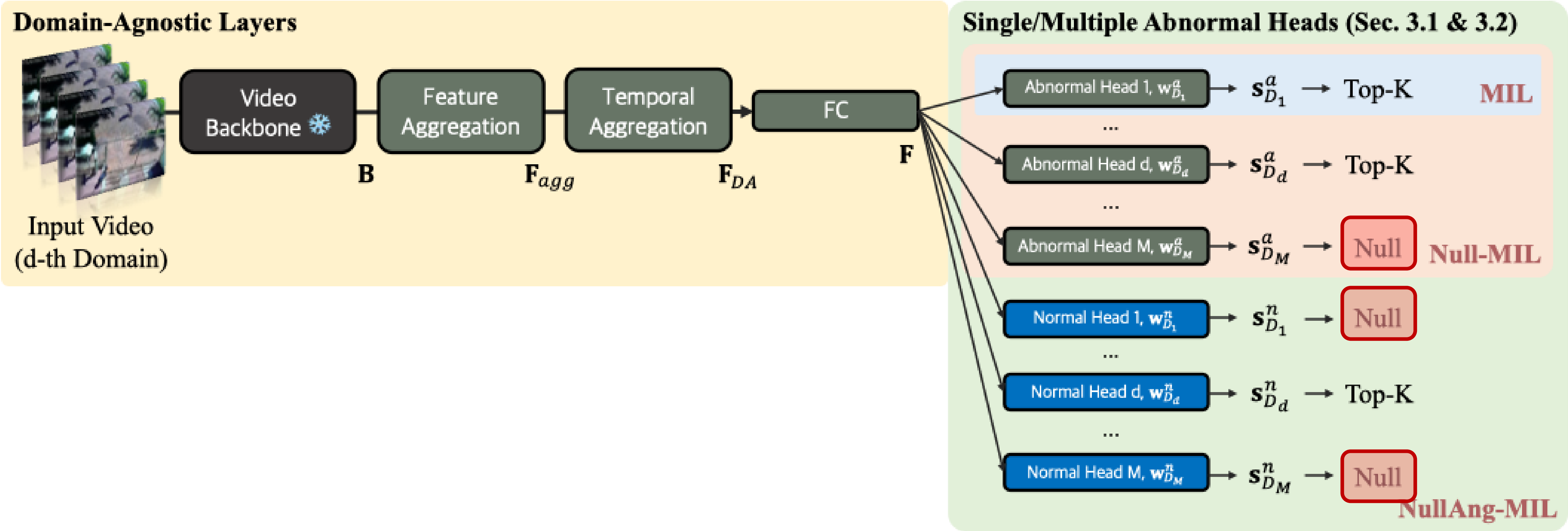
Figure 2: The overall framework of our MDVAD baselines that consists of domain-agnostic layers, single abnormal head (Sec. 3.1), multiple abnormal heads (Sec. 3.2), and AC classifier (Sec. 3.3).

Baselines

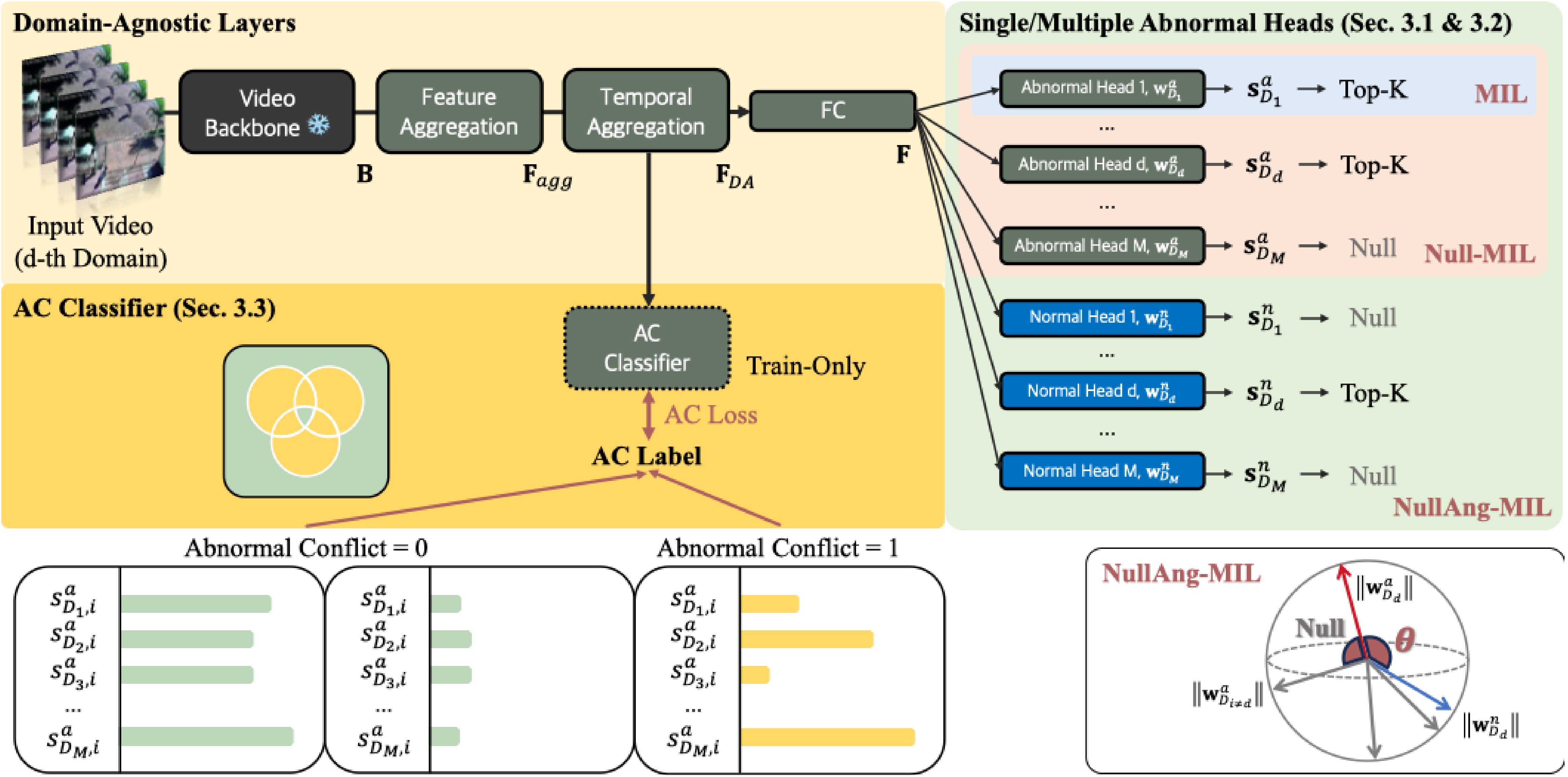


Domain-Agnostic Layers
Video Backbone
Feature Aggregation
Temporal Aggregation
FC
B
$\mathbf{F}_{agg}$
$\mathbf{F}_{DA}$
F
Input Video
(d-th Domain)

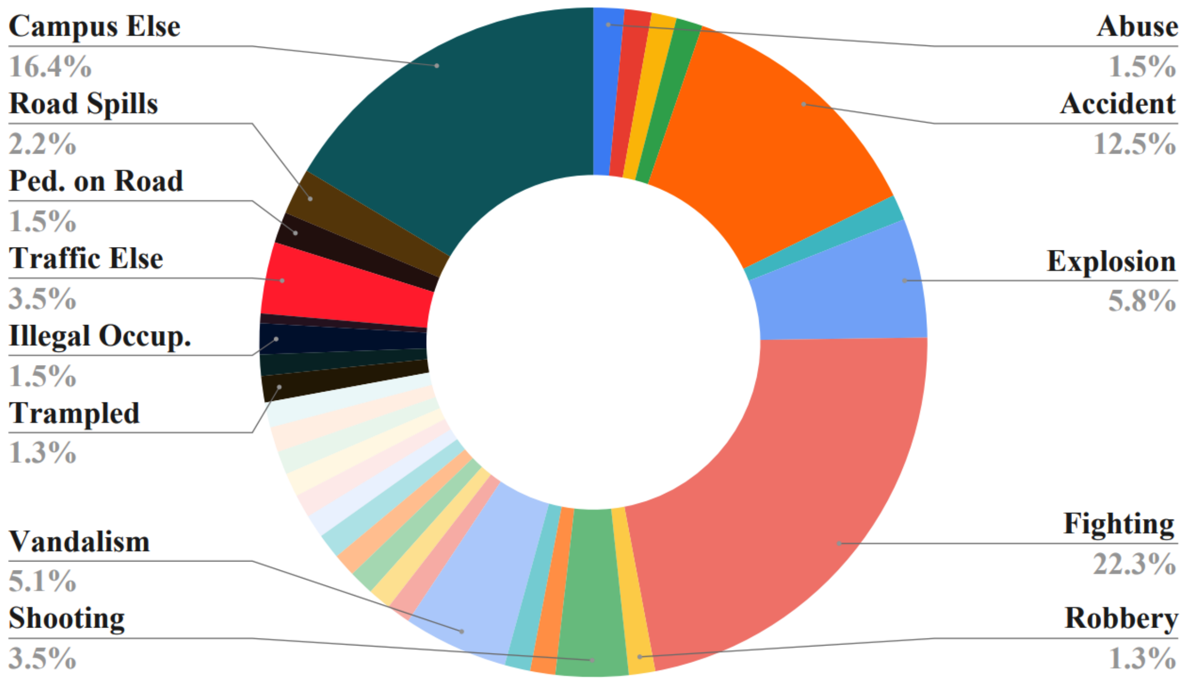
Baselines


Domain-Agnostic Layers
Input Video
(d-th Domain)
Video Backbone
Feature Aggregation
Temporal Aggregation
FC
B
$\mathbf{F}_{agg}$
$\mathbf{F}_{DA}$
F
Single/Multiple Abnormal Heads (Sec. 3.1 & 3.2)
Abnormal Head 1, $\mathbf{w}^a_{D_1}$
$\mathbf{s}^a_{D_1}$
Top-K
MIL
Abnormal Head d, $\mathbf{w}^a_{D_d}$
$\mathbf{s}^a_{D_d}$
Top-K
Abnormal Head M, $\mathbf{w}^a_{D_M}$
$\mathbf{s}^a_{D_M}$
Null
Null-MIL
Normal Head 1, $\mathbf{w}^n_{D_1}$
$\mathbf{s}^n_{D_1}$
Null
Normal Head d, $\mathbf{w}^n_{D_d}$
$\mathbf{s}^n_{D_d}$
Top-K
Normal Head M, $\mathbf{w}^n_{D_M}$
$\mathbf{s}^n_{D_M}$
Null
NullAng-MIL

Baselines


Domain-Agnostic Layers
Input Video
(d-th Domain)
Video Backbone
Feature Aggregation
Temporal Aggregation
FC
B
F_agg
F_DA
F
Single/Multiple Abnormal Heads (Sec. 3.1 & 3.2)
Abnormal Head 1
Abnormal Head d
Abnormal Head M
Normal Head 1
Normal Head d
Normal Head M
Top-K
Null
MIL
Null-MIL
NullAng-MIL
AC Classifier (Sec. 3.3)
AC Classifier
Train-Only
AC Loss
AC Label
Abnormal Conflict = 0
Abnormal Conflict = 1
NullAng-MIL
Null
θ

MDVAD benchmark



Four evaluation protocols for the MDVAD benchmark

- E1: Held-in
- E2: Leave-one-out
- E3: Low-shot adaptation
- E4: Full fine-tuning

Empirical studies

Table 5: **E1:** Multi-domain training: held-in results (AUC).

MDVAD Benchmarks								
		Target						Avg.
		UCFC	XD	LAD	UBIF	TAD	ST	
Single-domain								
Out Avg. (In-domain)		61.39 (77.93)	63.10 (83.23)	66.83 (83.82)	66.17 (92.62)	61.90 (90.75)	49.21 (90.79)	61.43 (86.52)
E1: Held-in								
Head	AC	UCFC	XD	LAD	UBIF	TAD	ST	Avg.
MIL	–	80.05	83.77	86.01	85.76	88.92	88.82	85.56
	✓	80.11	83.91	85.15	87.72	90.05	87.98	85.82
Null-MIL **(Ours)**	–	79.01	81.96	85.08	93.06	90.57	91.04	86.79
	✓	79.15	82.96	85.82	92.41	91.16	89.67	86.86
NullAng MIL(**Ours**)	–	76.32	82.74	82.32	92.30	91.82	91.26	86.13
	✓	77.21	82.09	83.88	91.90	91.36	91.12	86.26

Training: All six datasets / **Testing:** Target dataset
Column-wise coloring with increased intensity for higher values

Table 6: **E2**: Leave-one-out results

MDVAD Benchmarks							
		Target					
		UCFC	XD	LAD	UBIF	TAD	ST
Single-domain							
Out Avg.		61.39	63.10	66.83	66.17	61.90	49.21
E2: Leave-one-out							
Head	AC						
MIL	–	75.98	74.07	76.94	72.01	74.11	49.39
	✓	78.49	76.87	**78.67**	81.81	78.39	65.66
Null-MIL **(Ours)**	–	62.38	59.63	64.91	55.42	66.28	45.60
	✓	68.78	74.65	74.46	55.61	67.72	55.26
NullAng MIL(**Ours**)	–	75.26	73.00	73.91	79.41	77.94	52.98
	✓	**78.55**	**77.68**	77.36	**82.53**	**79.21**	**60.41**

Training: Five datasets except the target dataset
Testing: Target dataset

Empirical studies

VISUAL SCIENCE LABORATORY

Table 7: **E3**: Low-shot adaptation results

MDVAD Benchmarks							
		Target					
		UCFC	XD	LAD	UBIF	TAD	ST
E3: Low-shot Adaptation							
Head	AC						
MIL	–	75.19	68.20	**79.18**	82.13	82.80	71.65
	✓	72.52	71.00	76.69	82.34	78.72	74.88
Null-MIL	–	67.55	60.32	75.11	75.97	62.29	57.72
(Ours)	✓	70.57	66.40	73.58	81.39	71.12	63.02
NullAng	–	77.76	70.67	74.86	83.44	78.57	71.81
MIL**(Ours)**	✓	**78.99**	**75.80**	77.82	**85.75**	**84.06**	**76.23**

Training: E2 + a few target samples
Testing: Target dataset

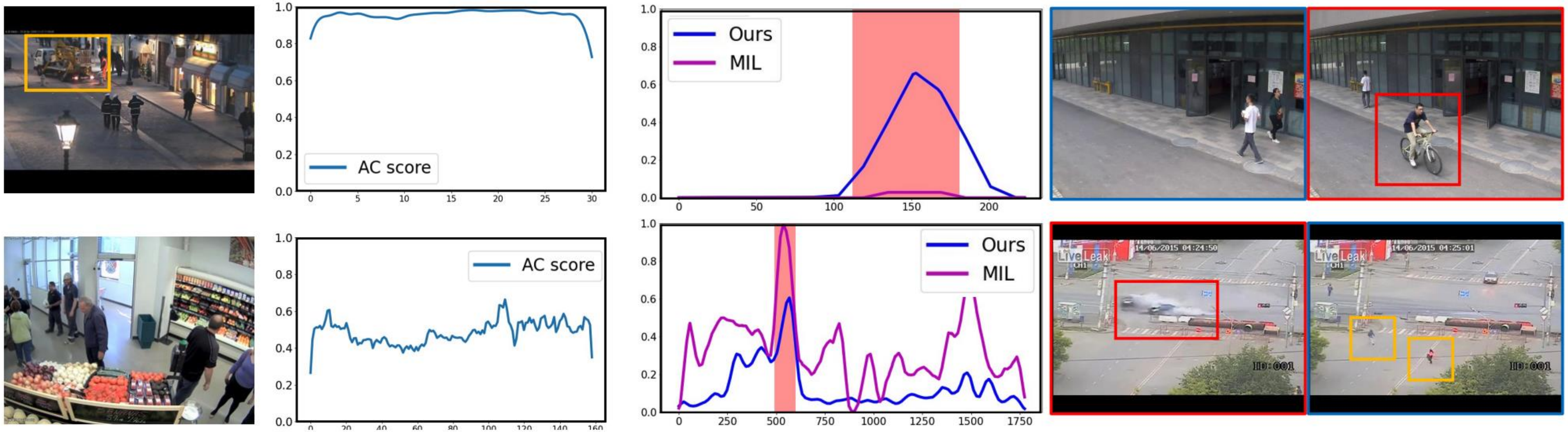
Table 8: **E4**: Comparison between the single-domain model and full fine-tuned models from the E1 and E2.

MDVAD Benchmarks						
	Target					
	UCFC	XD	LAD	UBIF	TAD	ST
Single-domain						
Single	77.93	83.23	83.82	92.62	90.75	90.79
E4: Full fine-tuning						
E1	78.62	82.71	**84.41**	92.42	**92.50**	91.17
E2	**80.24**	**82.77**	83.81	**92.95**	92.07	**91.27**

Finetuning: Target dataset / **Testing:** Target dataset

Analysis

VISUAL SCIENCE LABORATORY



(a) (b)

Figure 3: **(a)** The plot of AC scores. Both scenes are from UCFC and are normal in UCFC. (Top) Yellow box indicates abnormal conflict, which is abnormal in ST. (Bottom) Normal scene. **(b)** Qualitative results. Red box indicates abnormal event in the scene. (Top) *Bicyclist on walkway* abnormal event in ST. (Bottom) *Accident* abnormal event in UCFC and *Pedestrian on Road* abnormal conflict in TAD.

Summary

- Introduced **MDVAD**: A task for generalizable VAD across multiple domains
- Proposed **multi-head framework** with **Null(Ang)-MIL loss** and **AC Classifier**
- Effectively addresses **abnormal conflicts** across domains
- Demonstrates strong results on MDVAD benchmark with diverse protocols
- Focuses on resolving **multi-domain conflicts** rather than single-domain architectures
- Framework compatible with various VAD models; supports future generalization research