COMMUNICATION UNIVERSITY OF CHINA

MCC 媒体融合与传播国家重点实验室
State Key Laboratory of Media Convergence and Communication

Locality-aware Attention Network with Discriminative Dynamics Learning For Weakly Supervised Anomaly Detection

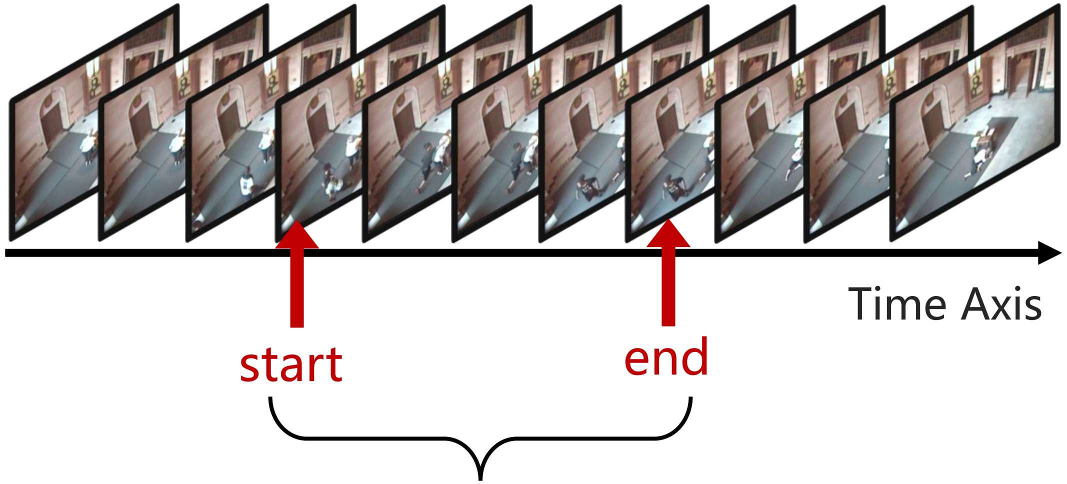
Yujiang Pu[1], Xiaoyu Wu[1]*

ICME 2022

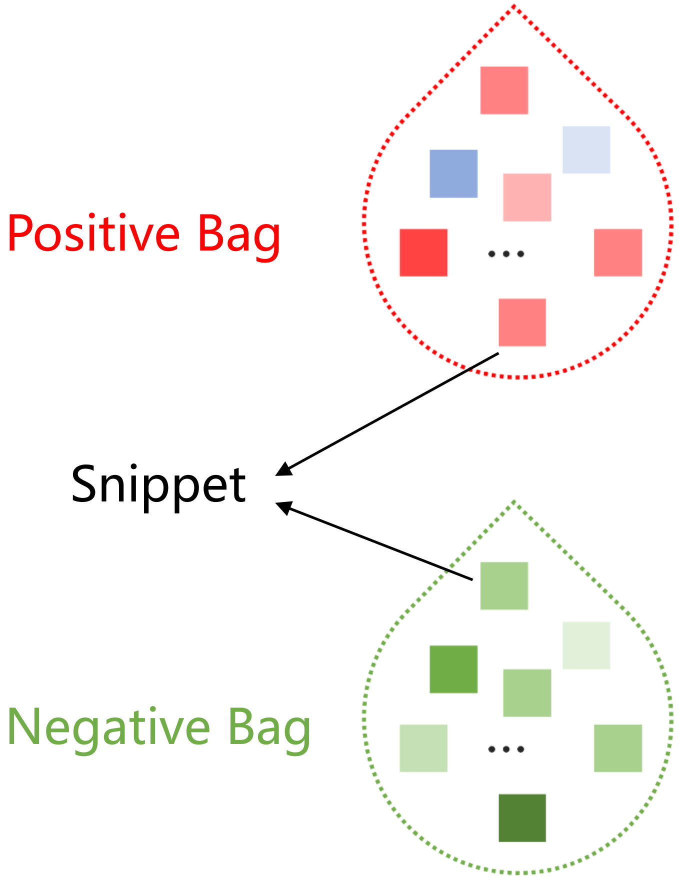
[1]State Key Laboratory of Media Convergence and Communication,
Communication University of China, Beijing, China

Introduction

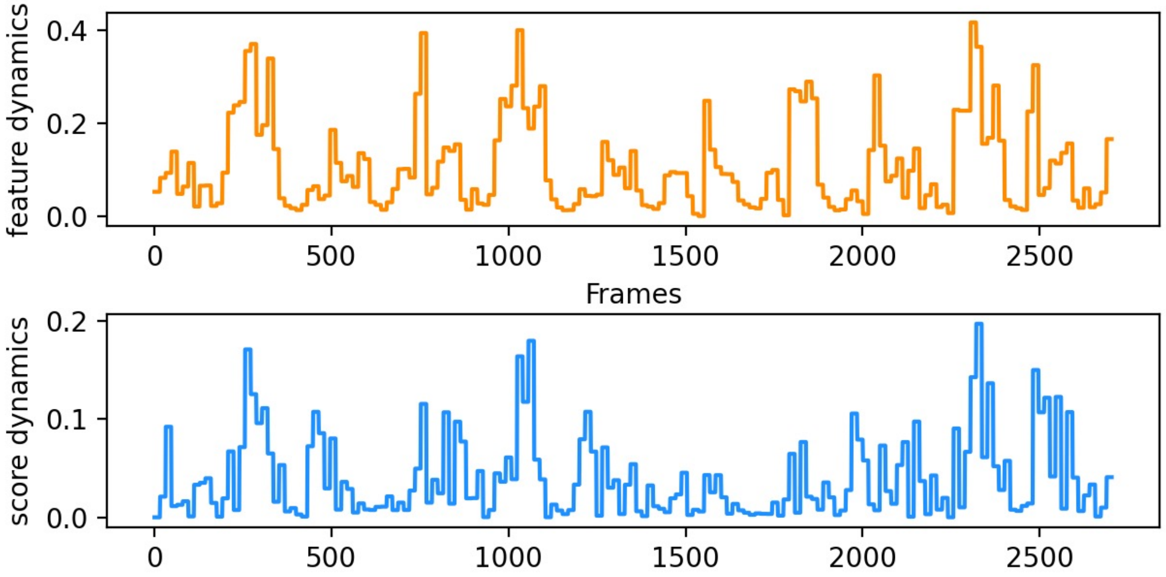
Video Anomaly Detection



locating the start and end of the event at the frame level



Motivation



Temporal Dynamics

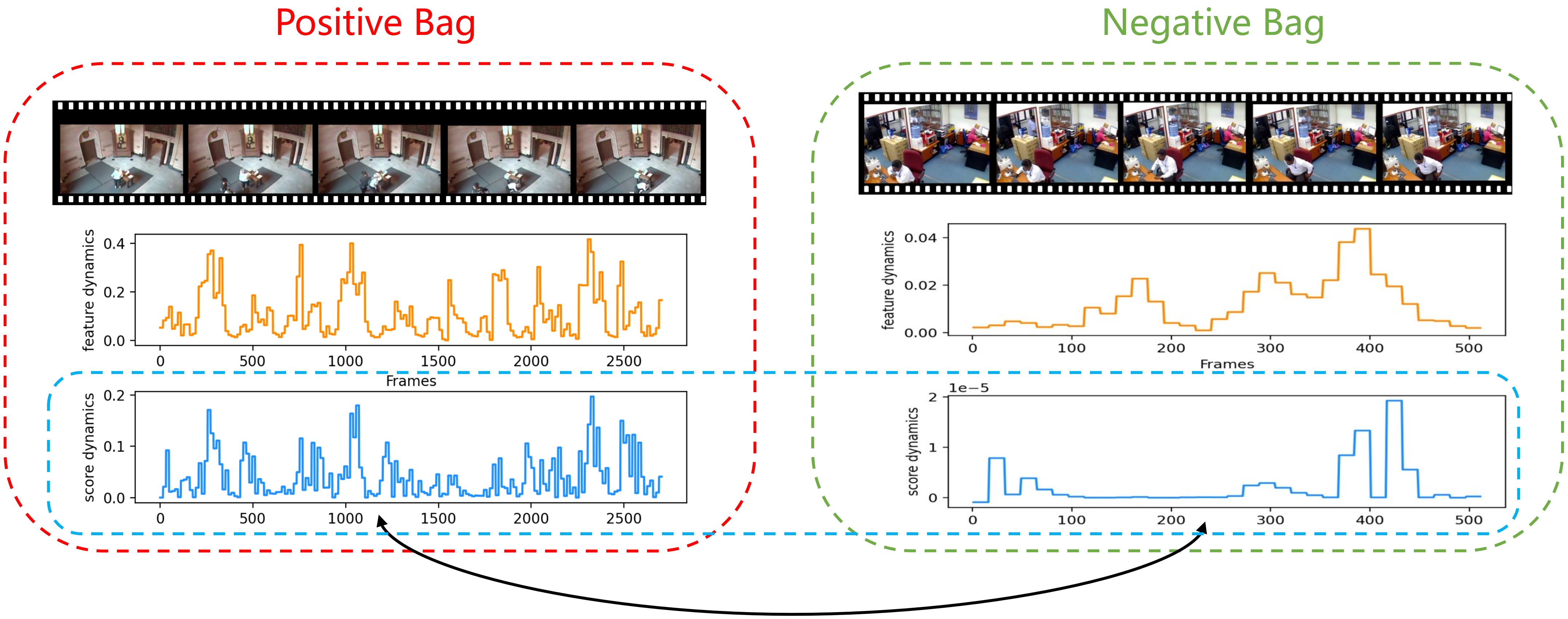
- **Feature Dynamics (FD)**

Feature Difference between adjacent snippet

- **Score Dynamics (SD)**

Score Difference between adjacent snippet

Motivation

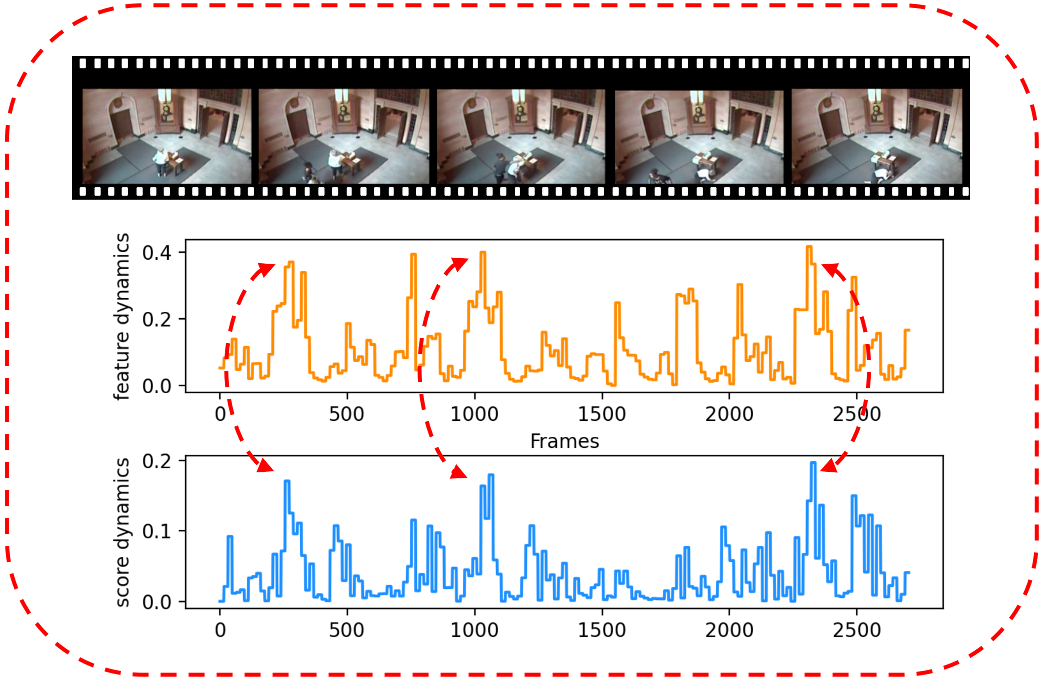


SD in Positive Bag ≫ **SD** in Negative Bag

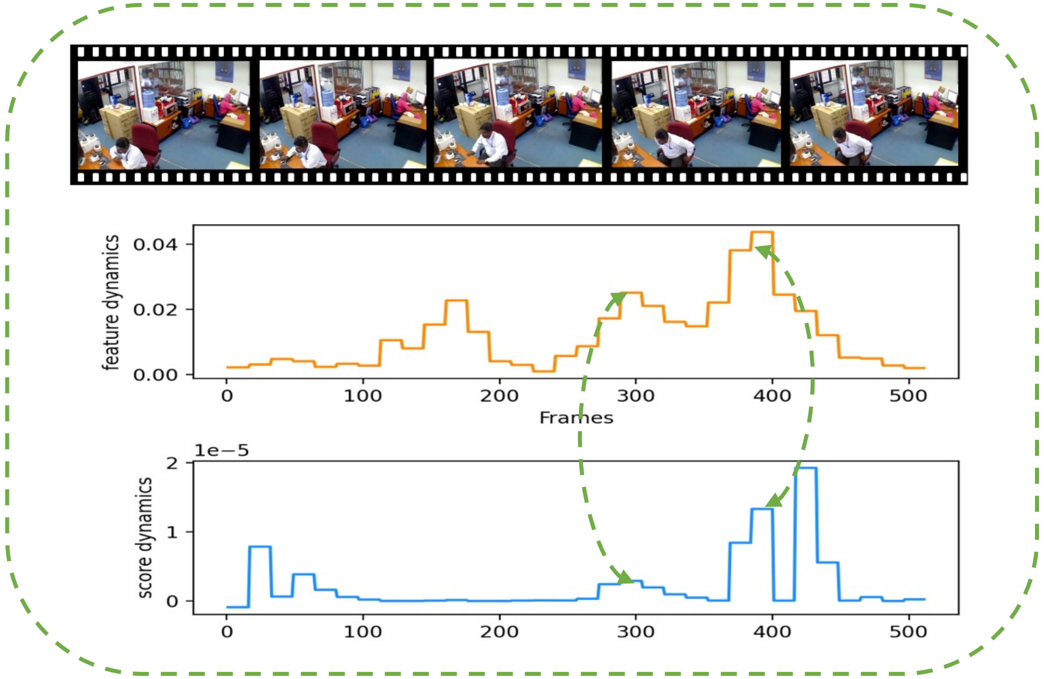
Score Dynamics Ranking $\mathcal{L}_{DR}$

Motivation

Positive Bag



Negative Bag

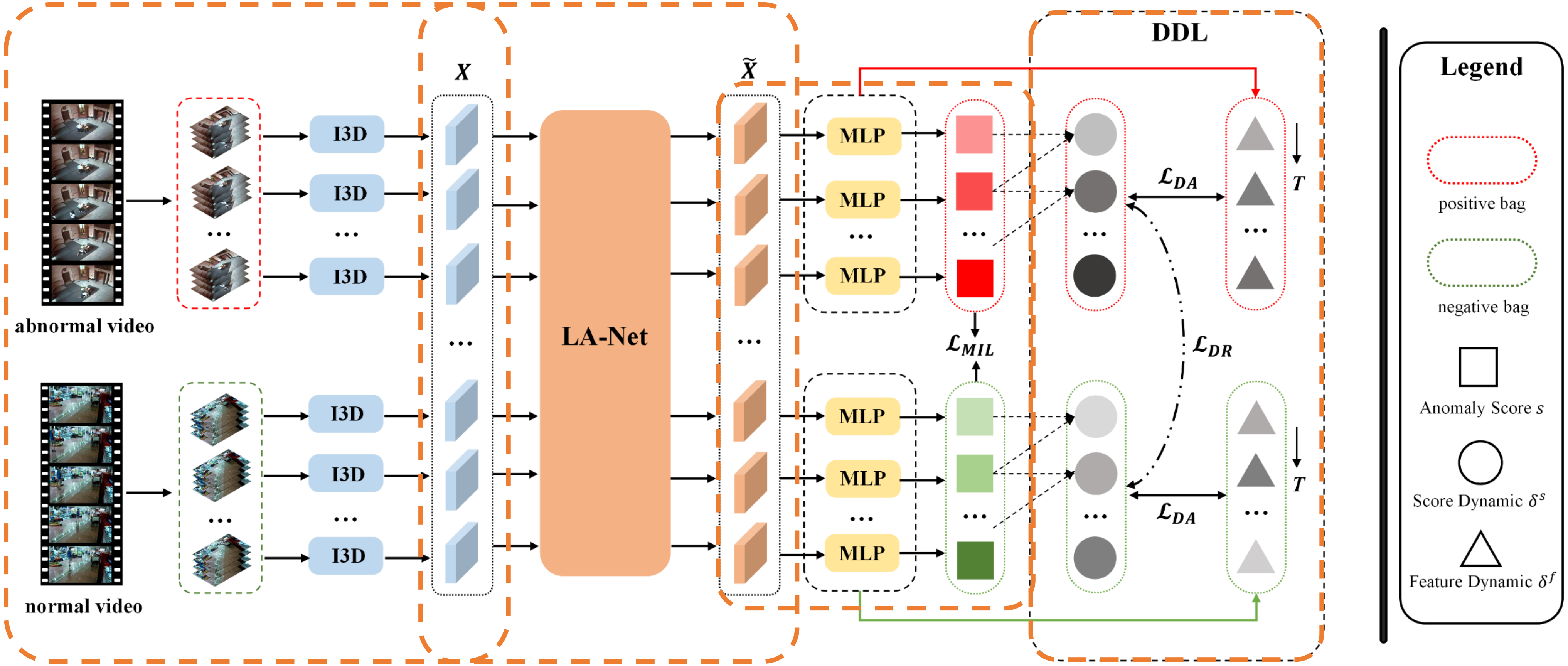


The **FD** and **SD** within a bag show a certain temporal consistency

Feature Dynamics Alignment $\mathcal{L}_{DA}$

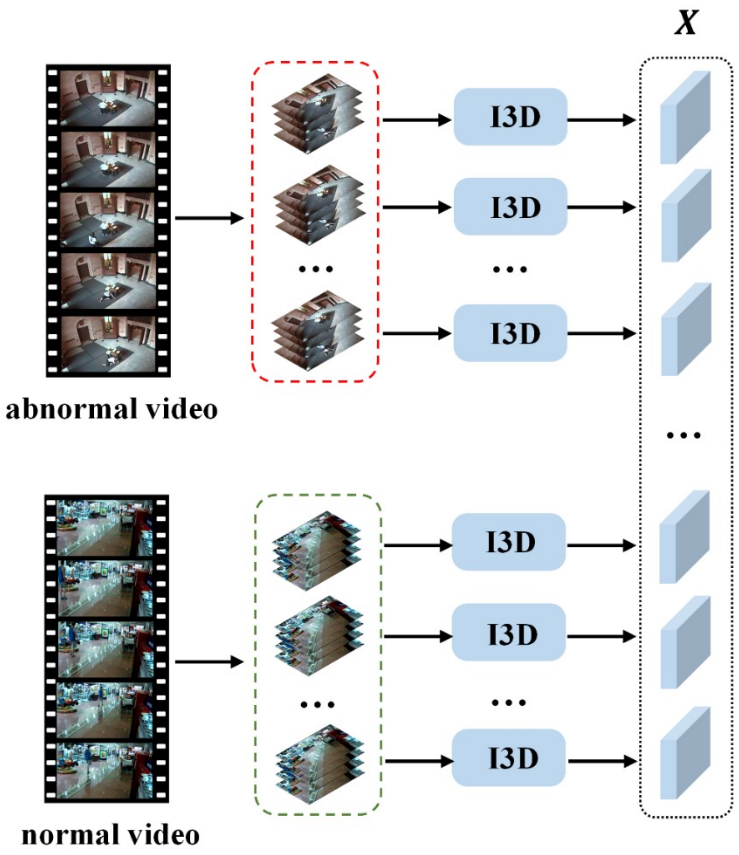
Methodology

Overall Structure of LA-Net with DDL method



Methodology

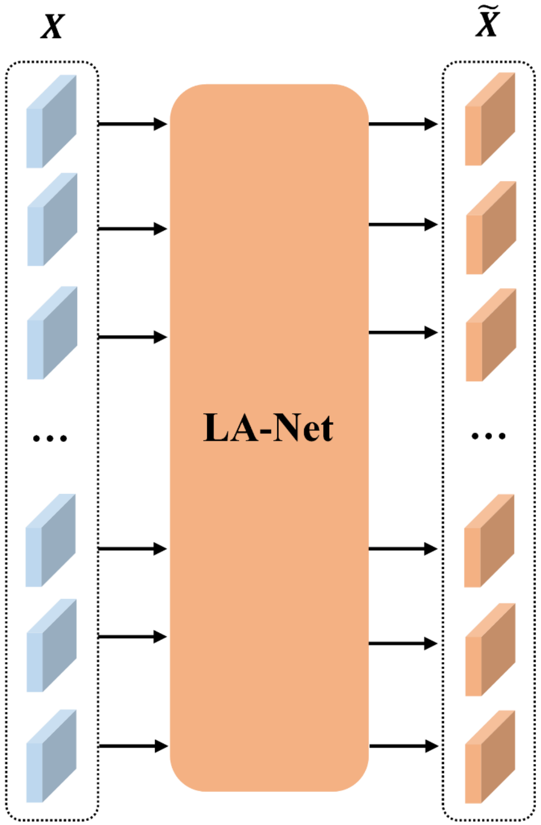
Feature Extraction

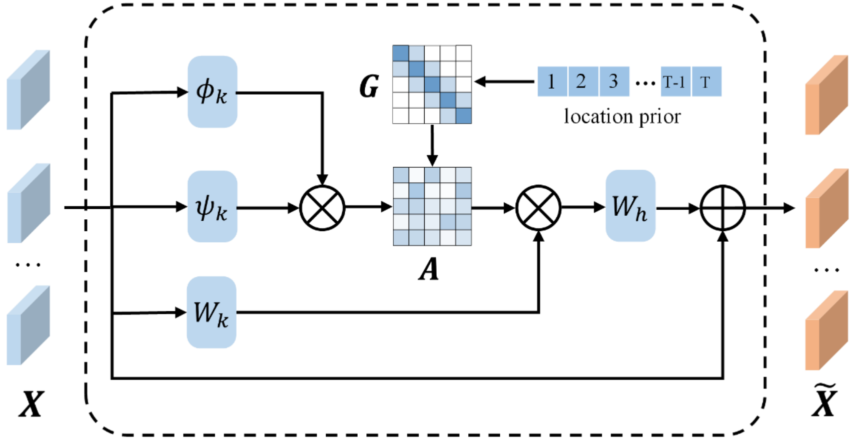


- Untrimmed Video $\mathcal{V}$
- Sliding Window $\chi = \{x_i\}_{i=1}^{T}$
- Snippet Feature $X = I3D(\chi) \in \mathbb{R}^{T \times D}$

Methodology

Locality-aware Attention Network





$$\mathcal{A}_{ij} = \frac{exp\{\mathcal{R}_k(x_i, x_j)\}}{\sum_{n=1}^{T} exp\{\mathcal{R}_k(x_i, x_n)\}}$$

$$\mathcal{R}_k = \phi(x_i)^T \psi(x_j)$$

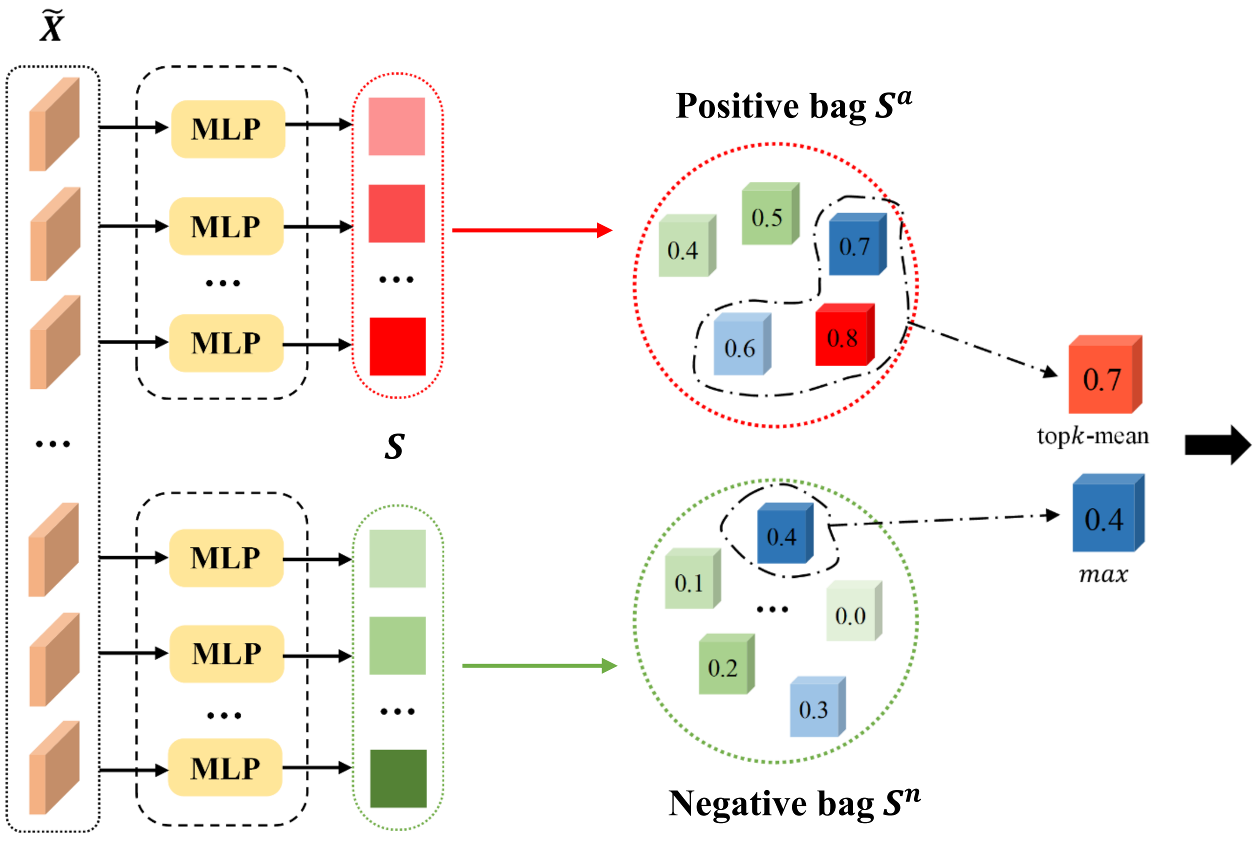
$$\mathcal{G}_{ij} = exp(-\frac{|i-j|^2}{2\sigma})$$

$$\widetilde{\mathcal{A}} = \mathcal{A} + \mathcal{G}$$

$$\widetilde{X} = Norm(stack(\widetilde{\mathcal{A}} X W_k) W_h + X)$$

Methodology

Multiple Instance Learning



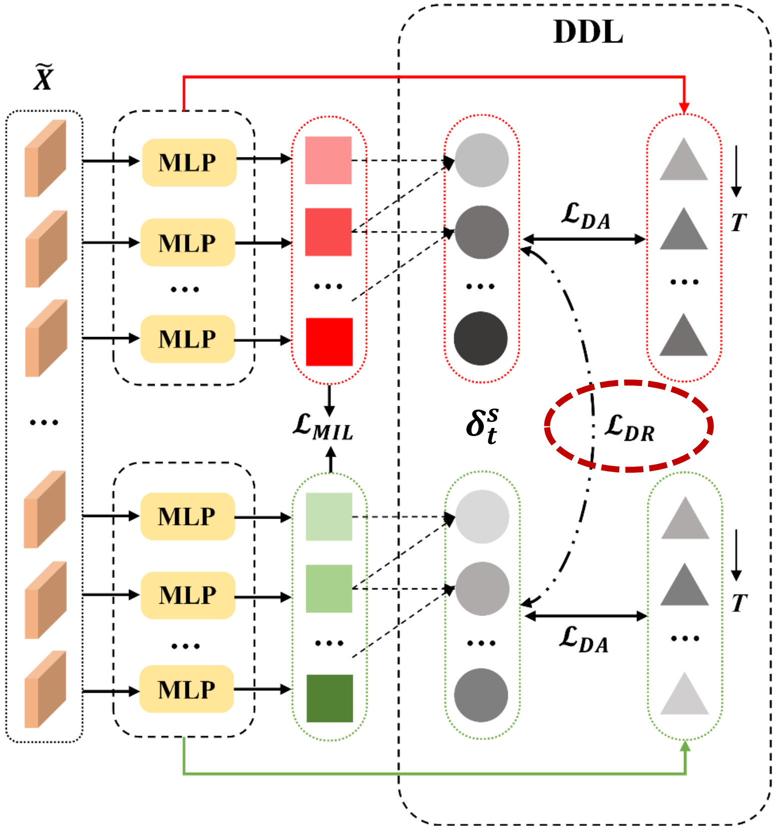
$$\mathcal{L}_{MIL} = \frac{1}{N}\sum_{i=1}^{N} -y_i \log(p_i)$$

$$\mathcal{S} = \sigma(W\mathcal{F}^{\Theta}(\widetilde{X}) + b)$$

$$p_i = \frac{1}{k}\sum_{i=1}^{k} Rank(S_i)$$

Methodology

Discriminative Dynamics Learning

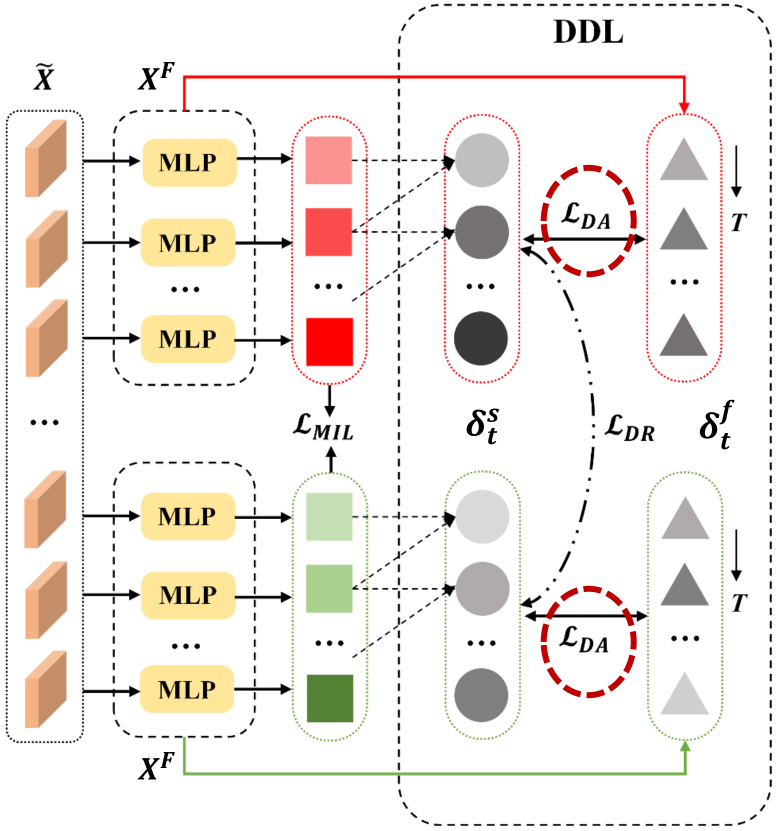


Score Dynamics Ranking --> **Outer Bag**

$$\left.\begin{aligned} S &= \{s_1, s_2, \dots, s_t\} \\ \delta_t^s &= |s_t - s_{t+1}| \\ \varepsilon_{\Delta S} &= \frac{1}{k}\sum_{t=1}^{k} |\delta_t^s|^2 \end{aligned}\right\} \quad \mathcal{L}_{DR} = max(0, \zeta - \mathcal{E}_{\Delta S}^{a} + \mathcal{E}_{\Delta S}^{n})$$

Methodology

Discriminative Dynamics Learning



Feature Dynamics **Alignment** --> Inner Bag

$$X^F = \{x_1^F, x_2^F, \dots, x_t^F\}$$

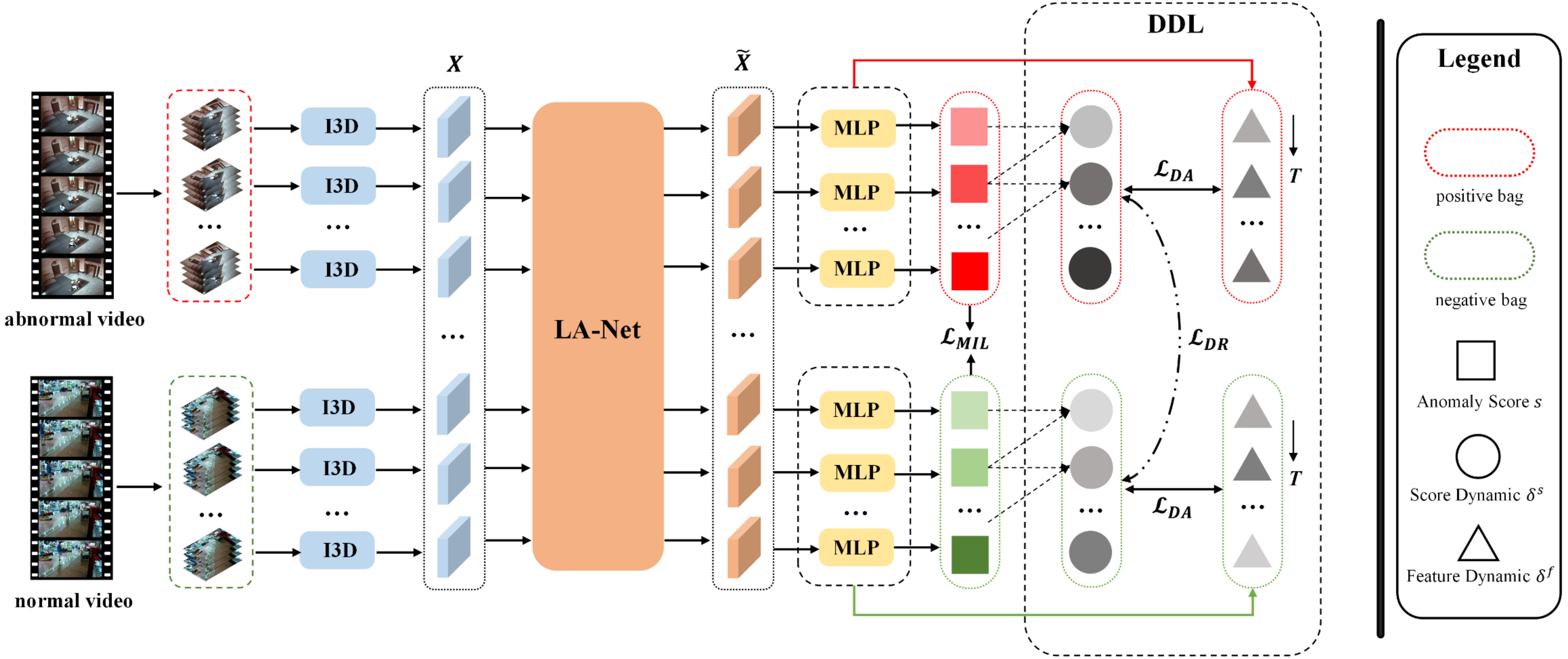
$$\delta_t^f = 1 - \frac{x_t^F x_{t+1}^F}{\|x_t^F\|\|x_{t+1}^F\|}$$

$$\delta_t^s = |s_t - s_{t+1}|$$

$$\mathcal{L}_{DA} = \frac{1}{N \times (T-1)} \sum_{i=1}^{N} \Big(\sum_{t=1}^{T-1} -\delta_t^s log(\delta_t^f + \epsilon) \Big)_i$$

Methodology

Overall Objective Function



$$\mathcal{L} = \mathcal{L}_{MIL} + \lambda_1 \mathcal{L}_{DR} + \lambda_2 \mathcal{L}_{DA}$$

Experimental Results

State-Of-The-Art Performance

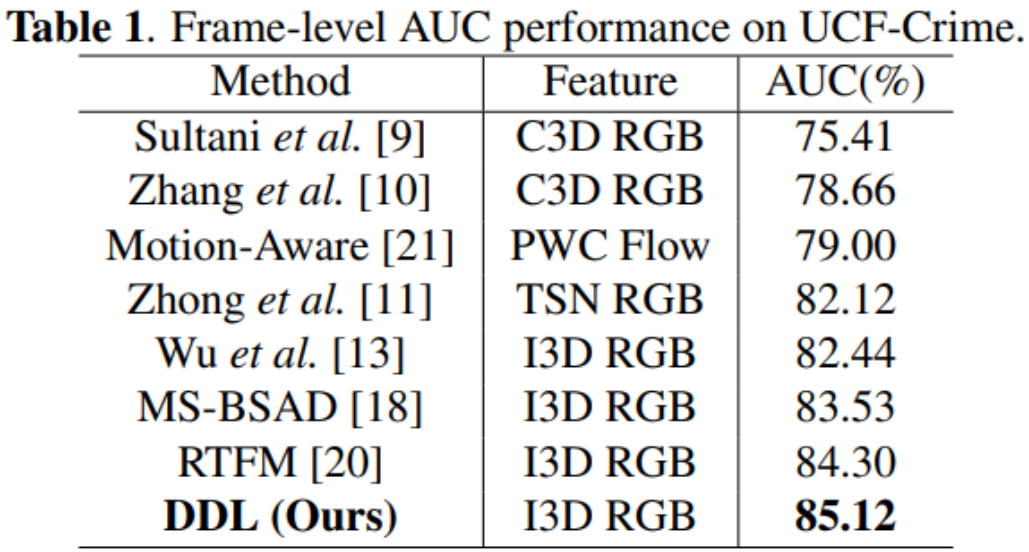


Table 1. Frame-level AUC performance on UCF-Crime.

Method	Feature	AUC(%)
Sultani *et al.* [9]	C3D RGB	75.41
Zhang *et al.* [10]	C3D RGB	78.66
Motion-Aware [21]	PWC Flow	79.00
Zhong *et al.* [11]	TSN RGB	82.12
Wu *et al.* [13]	I3D RGB	82.44
MS-BSAD [18]	I3D RGB	83.53
RTFM [20]	I3D RGB	84.30
DDL (Ours)	I3D RGB	**85.12**

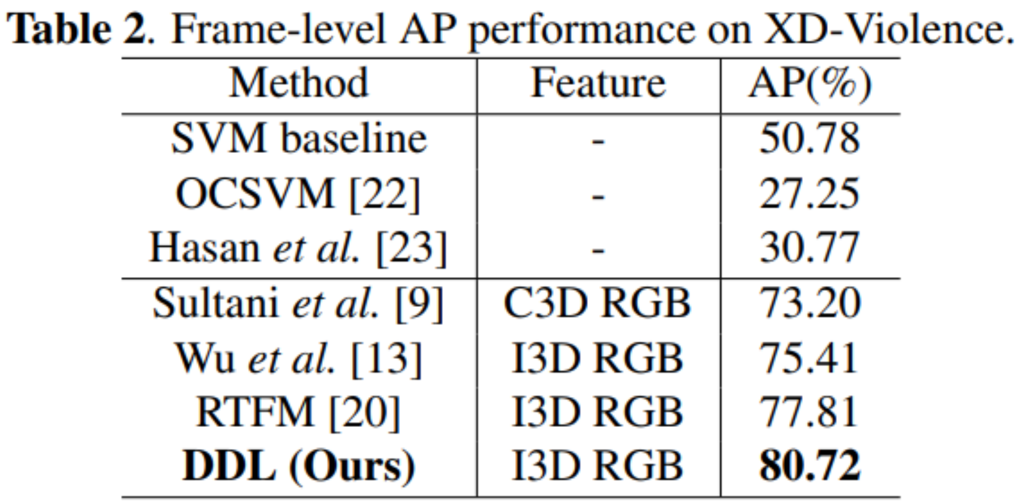


Table 2. Frame-level AP performance on XD-Violence.

Method	Feature	AP(%)
SVM baseline	-	50.78
OCSVM [22]	-	27.25
Hasan *et al.* [23]	-	30.77
Sultani *et al.* [9]	C3D RGB	73.20
Wu *et al.* [13]	I3D RGB	75.41
RTFM [20]	I3D RGB	77.81
DDL (Ours)	I3D RGB	**80.72**

Experimental Results

Ablation Study

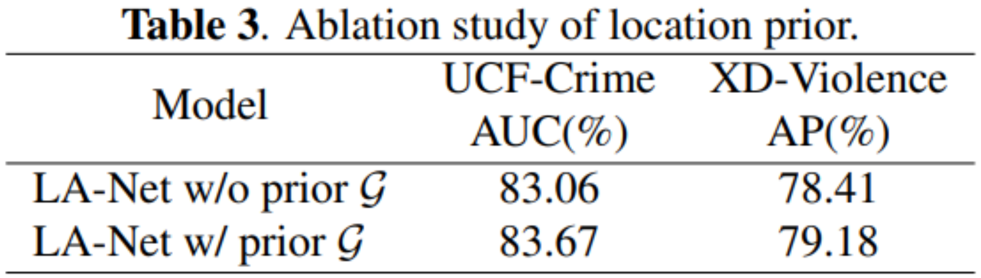


Table 3. Ablation study of location prior.

Model	UCF-Crime AUC(%)	XD-Violence AP(%)
LA-Net w/o prior $\mathcal{G}$	83.06	78.41
LA-Net w/ prior $\mathcal{G}$	83.67	79.18

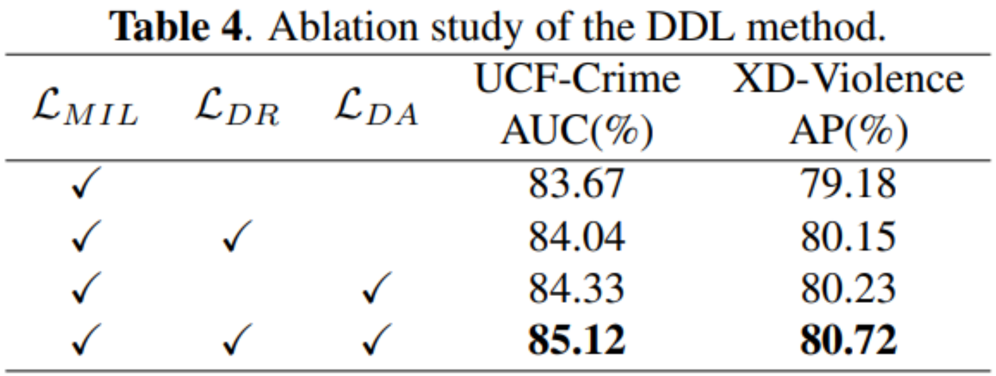
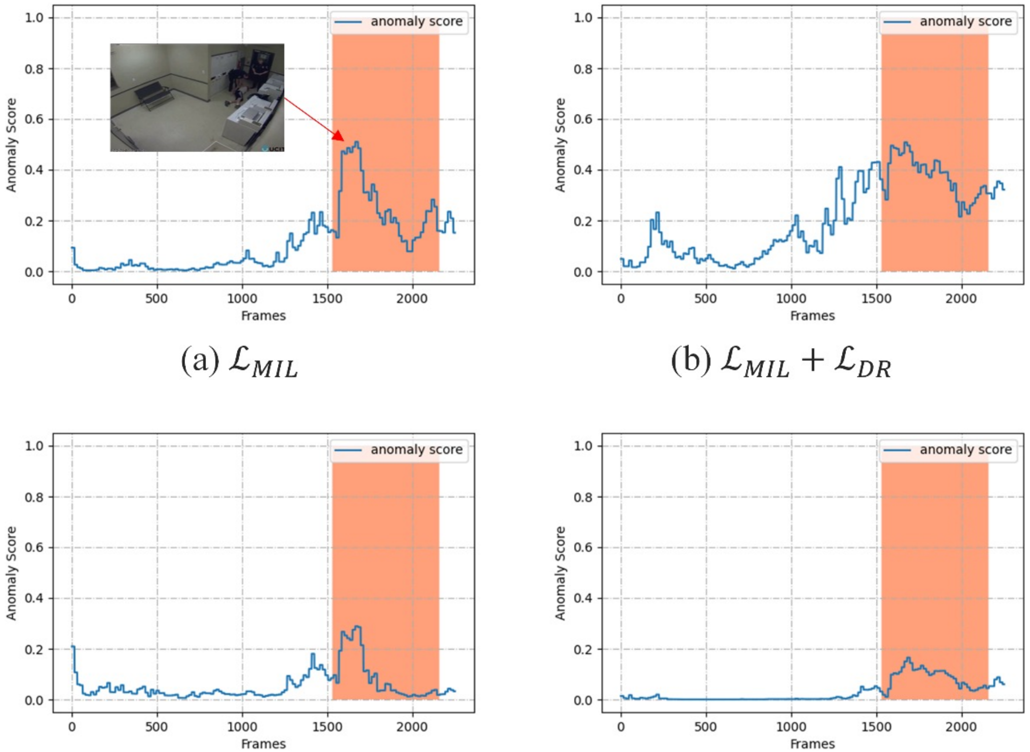


Table 4. Ablation study of the DDL method.

$\mathcal{L}_{MIL}$	$\mathcal{L}_{DR}$	$\mathcal{L}_{DA}$	UCF-Crime AUC(%)	XD-Violence AP(%)
✓			83.67	79.18
✓	✓		84.04	80.15
✓		✓	84.33	80.23
✓	✓	✓	**85.12**	**80.72**



(a) $\mathcal{L}_{MIL}$

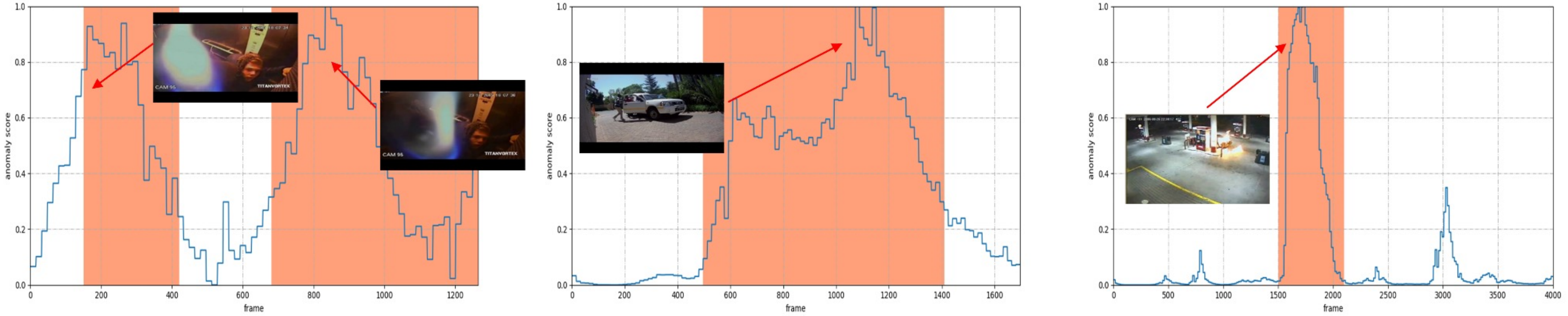
(b) $\mathcal{L}_{MIL} + \mathcal{L}_{DR}$

(c) $\mathcal{L}_{MIL}$+$\mathcal{L}_{DA}$

(d) $\mathcal{L}_{MIL} + \mathcal{L}_{DR} + \mathcal{L}_{DA}$

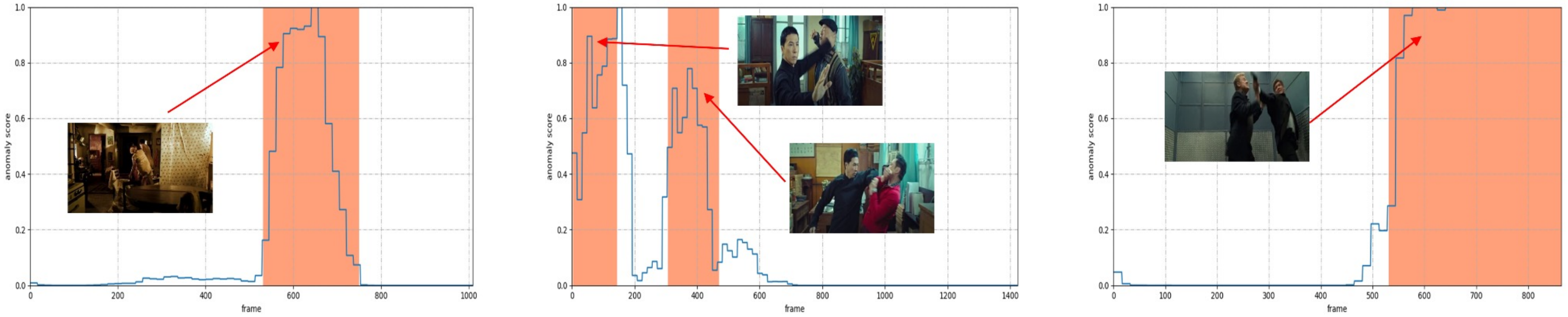
Qualitative Analysis

UCF-Crime



(a) Arson011 (b) Robbery050 (c) Explosion002

XD-Violence



(d) City.of.God.2002 (e) Ip.Man.3.2015 (f) Salt.2010

Thank You!