# 基于时序一致性约束的视频异常报警优化方法

张师源
陕西警察学院 信息技术系，陕西 西安 710021
DOI:10.61369/ETI.2026060024

**摘　　要：** 针对弱监督视频异常检测中分数波动引起的报警频繁切换和短时误报，提出一种时序一致性报警后处理方法。该方法在16帧片段尺度上依次进行指数移动平均、双阈值判定、短间隔合并和最短持续时间约束，并利用校准子集确定工作点。基于DSANet公开预训练模型，在UCF-Crime和XD-Violence上进行5次重复分层留出实验。F1工作点下，两个数据集的每百视频纯误报事件数分别由59.14降至24.91和10.81降至5.97，报警切换次数分别下降41.1%和19.9%，帧级F1基本保持。该方法无需重新训练，平均处理时间不超过1.94 ms/视频，但事件召回率略有下降，报警延迟有所增加。

**关 键 词：** 视频异常检测；时序一致性；事件级误报

## Video Anomaly Alarm Optimization Based on Temporal-Consistency Constraints

Zhang Shiyuan
Department of Information Technology, Shaanxi Police College, Xi'an , Shaanxi 710021

**Abstract:** A temporal-consistency-based post-processing method is proposed to reduce frequent alarm transitions and short false alarms caused by fluctuating anomaly scores. At the 16-frame segment level, it combines exponential moving average, dual-threshold state determination, short-gap merging, and minimum-duration filtering; operating points are selected on calibration videos. Five repeated stratified holdout experiments are conducted on UCF-Crime and XD-Violence using scores from the publicly available pretrained DSANet. At the F1 operating point, false-alarm events per 100 videos decrease from 59.14 to 24.91 and from 10.81 to 5.97, while alarm transitions decrease by 41.1% and 19.9%, respectively, with frame-level F1 essentially unchanged. The method requires no retraining and takes no more than 1.94 ms per video on average, although event recall decreases slightly and alarm delay increases.

**Keywords:** video anomaly detection; temporal consistency; event-level false alarm

## 引言

视频异常检测旨在从长时视频中定位偏离正常模式的事件，在公共安全等领域具有应用价值[1]。弱监督方法仅需视频级标签，可降低帧级标注成本。Sultani等构建UCF-Crime并引入多示例排序学习[2]，XD-Violence进一步扩展到多场景暴力视频[3]；RTFM利用特征幅值区分正常与异常片段[4]，长短期时间序列关联方法则强化片段间联系[5]。

随着视觉语言模型的发展，CLIP提供了图文语义特征[6]，VadCLIP[7]和DSANet[8]进一步将视觉语言对齐用于弱监督视频异常检测。上述方法主要改进异常表征与分数排序，连续分数通常仍需通过阈值转换为报警状态。

现有研究多采用AUC和AP评价异常分数排序能力，但实际系统还需要将连续分数转换为离散报警。分数在阈值附近波动时，单阈值判定容易产生频繁切换和短时误报，模型部署还需考虑报警周期与响应时间[9]。本文不改变基础检测器，而是研究异常分数的报警稳定化，主要工作为：构建由EMA、双阈值、最短持续时间和间隔合并组成的后处理流程；在相同F1或F0.5目标下比较处理前后结果；在两个公开数据集上从帧级、事件级、报警稳定性和计算开销等方面评价收益及代价。

项目课题：2025年度陕西省教育厅重点科学研究计划项目（高校工程研究中心项目）——面向智慧警务的轻量化视频异常行为识别模型及特定场景定制方法研究（项目编号:25JR032）。
作者简介：张师源（1995-），男，陕西西安人，硕士研究生，助教，研究方向为公安大数据，智慧公安。

 Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.

## 一、相关工作

弱监督方法将视频视为包、片段视为实例，通过排序或分类约束学习异常程度，研究重点多为特征表示和语义对齐，最终报警仍常由连续分数与单阈值获得。本文采用 DSANet 预训练模型生成分数，仅处理其输出，不依赖检测器内部结构。时序平滑虽能抑制噪声，但仍需转换为报警状态；双阈值滞回已用于信号处理与边缘检测 [10]。本文进一步加入持续时间约束和间隔合并。

## 二、时序一致性报警优化方法

### （一）问题定义与总体流程

设基础检测器输出片段异常分数 $s=\{s_1,s_2,\cdots,s_T\}$，$s_t \in [0,1]$。实验以 16 帧为一个片段，平滑、状态判定和形态处理均在片段尺度完成，再映射至视频帧，总体流程如图 1 所示。

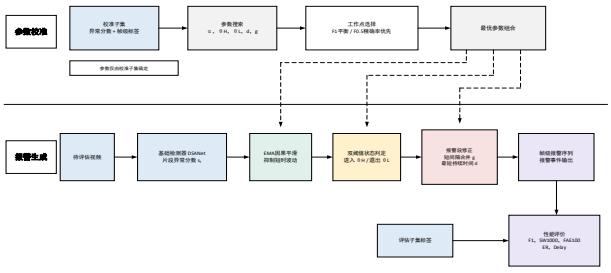



图 1 视频异常分数时序一致性优化流程

### （二）指数移动平均

采用指数移动平均（exponential moving average，EMA）对片段异常分数进行因果平滑：

$$\hat{s}_1=s_1,\ \hat{s}_t=\alpha s_t+(1-\alpha)\hat{s}_{t-1},\ t=2,\ldots,T \quad (1)$$

式中，$\alpha \in (0,1]$ 为当前片段权重；$\alpha$ 越小，平滑越强。EMA 仅使用当前及历史分数。

### （三）滞回双阈值状态判定

分别设置报警进入阈值 $\theta H$ 和报警维持阈值 $\theta L$，且 $\theta L \leqslant \theta H$。令二值报警状态 $y_t \in \{0,1\}$，递推过程为：

$$y_t=1,\ 若\ (y_{t-1}=0\ 且\ \hat{s}_t \geq \theta H)\ 或\ (y_{t-1}=1\ 且\ \hat{s}_t \geq \theta L);$$
$$否则\ y_t=0 \quad (2)$$

未报警时，分数达到 $\theta H$ 后进入报警；已报警时，分数低于 $\theta L$ 后退出。$\theta H > \theta L$ 可减少阈值附近波动造成的状态切换。

### （四）持续时间约束与间隔合并

若相邻报警段间连续为 0 的片段数不超过 g，则合并两段；随后删除长度小于 d 的报警段。先合并后删除可避免同一事件因短时降分被切分，d 和 g 的单位均为片段。

### （五）两类工作点与校准子集选参

F1 工作点取校准子集 F1 最大值，F0.5 工作点侧重精确率；原始分数按相同目标确定阈值。搜索范围为 $\alpha \in \{0.3,0.5,0.7\}$、$d \in \{1,3,5,9,15\}$、$g \in \{0,1,3\}$，$\theta H$ 与 $\theta L$ 差值取 $\{0,0.05,0.10,0.15\}$，全部参数仅由校准子集确定。

## 三、实验设计

### （一）数据集与基础检测器

实验采用 UCF-Crime（1900 个视频，官方测试视频 290 个）和 XD-Violence（4754 个视频，官方测试视频 800 个）。本文仅使用视觉 CLIP 特征和二分类异常分数，不使用 XD-Violence 音频模态。

基础检测器采用 DSANet 公开代码和预训练权重 [8]，不再训练或修改模型。完整测试集上，UCF-Crime 的 AUC/AP 为 0.8944/0.3742，XD-Violence 为 0.9540/0.8695。

### （二）划分协议与评价指标

公开训练集缺少帧级标签，本文将官方测试集按正常、异常视频分层，每次随机抽取 20% 作为校准子集、其余 80% 作为评估子集，共重复 5 次，随机种子为 234、42、7、123 和 2026。同一划分中各方法共享数据，评估集不参与选参。不同划分可能包含重复视频，因此结果仅用于考察处理前后的配对变化，不与官方完整测试结果进行排名。

连续分数采用 AUC 和 AP 评价，二值报警报告精确率 P、召回率 R 和 F1。另定义每千负样本帧误报数 FP1000、每千帧报警切换次数 SW1000 和每百视频纯误报事件数 FAE100：

$$FP_{1000}=1000\cdot FP/Nneg, SW_{1000}=1000\cdot \Sigma_i \Sigma_t \cdot (y_{i,t} \neq y_{i,t-1})/N \quad (3)$$

$$FAE_{100}=100\cdot NFA/Nvideo \quad (4)$$

式中，Nneg、N 和 NFA 分别为正常帧数、总帧数和未与真实异常重叠的预测报警段数。事件召回率 ER 表示被报警覆盖的真实事件比例；Delay 为成功检出事件起点至首次报警的片段数。各指标报告 5 次实验均值和样本标准差。

### （三）对比设置

对比原始 F1/F0.5 阈值、3/9 片段中值滤波、EMA、滞回、EMA+滞回及完整时序方法。实验平台为 Ryzen 9 8945H、RTX 4060 Laptop GPU 和 PyTorch 2.1.0；后处理在 CPU 完成。

## 四、结果与分析

### （一）主要结果

表 1 两个数据集 F1 工作点的主要报警指标（均值 ± 标准差）

| 数据集 | 方法 | F1 | SW1000 | FAE100 | ER | Delay/片段 |
|---|---|---|---|---|---|---|
| UCF | 原始 -F1 | 0.441 ± 0.018 | 0.508 ± 0.028 | 59.14 ± 1.44 | 0.724 ± 0.040 | 2.317 ± 0.337 |
| UCF | 时序 -F1 | 0.454 ± 0.021 | 0.299 ± 0.027 | 24.91 ± 2.70 | 0.624 ± 0.040 | 2.783 ± 0.215 |
| XD | 原始 -F1 | 0.770 ± 0.014 | 0.507 ± 0.030 | 10.81 ± 1.80 | 0.694 ± 0.096 | 1.065 ± 0.511 |
| XD | 时序 -F1 | 0.768 ± 0.014 | 0.406 ± 0.034 | 5.97 ± 2.77 | 0.673 ± 0.110 | 1.231 ± 0.602 |

Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License. 

由表1可见，在F1工作点下，UCF-Crime的F1由0.441升至0.454，SW1000由0.508降 至0.299，FAE100由59.14降 至24.91；XD-Violence的F1由0.770变为0.768，SW1000由0.507降至0.406，FAE100由10.81降至5.97。F0.5工作点能够进一步减少误报，但会降低异常召回率，因此更适合人工复核成本较高的场景。

**（二）报警稳定性与检出代价**

处理前后的AUC和AP变化不超过0.004，说明该方法主要改变离散报警序列，对异常分数排序影响较小。误报下降同时带来一定检出代价：UCF-Crime的事件召回率由0.724降至0.624，报警延迟由2.317增至2.783个片段；XD-Violence的事件召回率由0.694降至0.673，延迟由1.065增至1.231个片段。

**（三）参数敏感性、稳健性与计算开销**

不同划分选择的参数并不完全一致；最短持续时间d增大时，SW1000和FAE100总体下降，事件召回率也随之降低。分层配对Bootstrap的2000次重采样结果显示，两个数据集在F1和F0.5工作点下的 ΔSW1000与 ΔFAE100的95%置信区间均不包含0。由于划分间存在视频重叠，该结果不代表独立外部场景的统计保证。

后处理在AMD Ryzen 9 8945H CPU上完成，UCF-Crime和XD-Violence的平均耗时分别为1.93 ms/视频和1.65 ms/视频。计时不含文件读取、视频解码、特征提取和模型推理。

## 五、结论

本文提出一种由EMA、双阈值、持续时间约束和间隔合并组成的报警后处理方法。两个数据集上的实验表明，该方法在帧级F1基本保持的情况下能够减少报警切换和纯误报事件，平均处理时间不超过1.94 ms/视频，可作为异常检测模型输出端的轻量级处理模块。

误报减少会伴随事件召回率下降和报警延迟增加，实际应用中应结合漏报损失、响应时效和人工复核成本选择工作点。本文的校准数据由官方测试集重新划分得到，实验结果不能直接等同于官方完整测试结果；方法在不同帧率、音频、多摄像头及流式视频中的适用性仍需验证。

*参考文献*

[1] 汪洋，周脚根，严俊，等.基于深度学习的监控视频异常检测方法综述[J].中国图象图形学报，2025,30(3):615-640.

[2] SULTANI W, CHEN C, SHAH M. Real-world anomaly detection in surveillance videos[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 6479-6488.

[3] WU P, LIU J, SHI Y, et al. Not only look, but also listen: learning multimodal violence detection under weak supervision[C]//European Conference on Computer Vision. Cham: Springer, 2020: 322-339.

[4] TIAN Y, PANG G, CHEN Y, et al. Weakly-supervised video anomaly detection with robust temporal feature magnitude learning[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 4975-4986.

[5] 朱新瑞，钱小燕，施俞洲，等.长短期时间序列关联的视频异常事件检测[J].中国图象图形学报，2024,29(7):1998-2010.

[6] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//International Conference on Machine Learning. PMLR, 2021: 8748-8763.

[7] WU P, ZHOU X, PANG G, et al. VadCLIP: adapting vision-language models for weakly supervised video anomaly detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38(6): 6074-6082.

[8] YIN W, ZHANG H, WANG X, et al. Learning to tell apart: weakly supervised video anomaly detection via disentangled semantic alignment[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2026, 40(14): 12027-12035.

[9] KARIM H, DOSHI K, YILMAZ Y. Real-time weakly supervised video anomaly detection[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024: 6848-6856.

[10] CANNY J. A computational approach to edge detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8(6): 679-698.

 | Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.