A Proof of Theorem 1

Proof. Denote the random variable T for the test statistics $\hat{t}$, then

$$T = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m} V_{f_j}^{(i)}\right),$$

where $V_{f_j}^{(i)}$ follows the same distribution but independent to each other for all i. Next, we discuss the property of $V_{f_j}^{(i)}$. To simplify the notation, we use V_{f_j} interchangeably.

Under H_1, V_{f_j} is a Bernoulli random variable with $p = 0.5 + \varepsilon_{f_j}$. Therefore, the mean and variance for $\frac{1}{m}\sum_j V_{f_j}$ are $0.5 + e_1$ and $\frac{0.25 - e_2 + c_1}{m}$, where $e_1 = \frac{1}{m}\sum_j \varepsilon_{f_j}$, $e_2 = \frac{1}{m}\sum_j \varepsilon_{f_j}^2$, and $c_0 = \frac{1}{m}\sum_{j \neq k} \mathrm{cov}_{H_0}(V_{f_j}, V_{f_k})$. Similarly, under H_0, V_{f_j} is with $p = 0.5$. The mean and variance for $\frac{1}{m}\sum_j V_{f_j}$ Each V_{f_j} is a Bernoulli random variable, with $p = 0.5$ and $p = 0.5 + \varepsilon_{f_j}$ under H_0 and H_1. Therefore, the mean and variance of $\frac{1}{m}\sum_j V_{f_j}$ are 0.5 and $\frac{0.25 + c_0}{m}$

By central limit theorem, the test statistics is $T \sim \mathcal{N}(0.5, \frac{0.25 + c_0}{mn})$ under H_0 and $T \sim \mathcal{N}(0.5 + e_1, \frac{0.25 - e_2 + c_1}{mn})$ under H_1. Therefore, given the significance level α, the threshold for accepting or rejecting H_0 is $z_\alpha = 0.5 + \Phi^{-1}(1-\alpha)\sqrt{\frac{0.25 + c_0}{mn}}$. We then derive the type-II error as

$$\begin{aligned}
\mathbb{P}_{T \sim \mathcal{N}\left(0.5 + e_1, \frac{0.25 - e_2 + c_1}{mn}\right)}(T < z_\alpha) &= \mathbb{P}_{T \sim \mathcal{N}\left(0, \frac{0.25 - e_2 + c_1}{mn}\right)}\left(T < \Phi^{-1}(1-\alpha)\sqrt{\frac{0.25 + c_0}{mn}} - e_1\right) \\
&= \Phi\left(\sqrt{\frac{mn}{0.25 - e_2 + c_1}}\left(\Phi^{-1}(1-\alpha)\sqrt{\frac{0.25 + c_0}{mn}} - e_1\right)\right) \\
&= \Phi\left(\frac{\Phi^{-1}(1-\alpha)\sqrt{0.25 + c_0} - e_1\sqrt{mn}}{\sqrt{0.25 - e_2 + c_1}}\right)
\end{aligned}$$

Therefor, the power is

$$1 - \Phi\left(\frac{\Phi^{-1}(1-\alpha)\sqrt{0.25 + c_0} - e_1\sqrt{mn}}{\sqrt{0.25 - e_2 + c_1}}\right) = \Phi\left(\frac{e_1\sqrt{mn} - \Phi^{-1}(1-\alpha)\sqrt{0.25 + c_0}}{\sqrt{0.25 - e_2 + c_1}}\right)$$

□

B Implementation details

B.1 Implementation on Toy dataset in Section 3.2

For C2ST in the toy dataset, we train a 2-layer MLP with hidden layer size 20 for 50 epochs. We use RELU as activation function, with the batch size of 16. For optimizer we use the default Adam Optimizer. For $m = 5$ cases, we simply use different random seeds with early stopping for training the classifier.

B.2 Implementation on Benchmark datasets in Section 4

Our frame level motion features have size 120 × 160. We choose different motion thresholds for different datasets to reproduce results in Ionescu et al. [12], and use the same threshold to report results of the proposed algorithm.

For filtering out the noise of the features, we follow Lu et al. [21]. We calculate the motion responses of each region across 5 consecutive frames to obtain a corresponding value that records motion intensity inside each position. Using a threshold, we can find and eliminate features in regions with small motion responses.

They have proved that dividing the frames equally to 2x2 bins and processing them separately can slightly improve the results. Thus we follow some of their settings: the same partitions is done on both our features, for every frame the anomaly score is chosen as the maximum score of the four bins.

The classifiers we use are L2-regularized logistic regression classifier from LIBLINEAR [9]. For history sampling, the size of past frames is set to be 5 or 10, for each individual dataset we use the same history size for all sampling methods. For other hyperparameters, such as stride for the sliding window, we follow Ionescu et al. [12].

C Evaluation

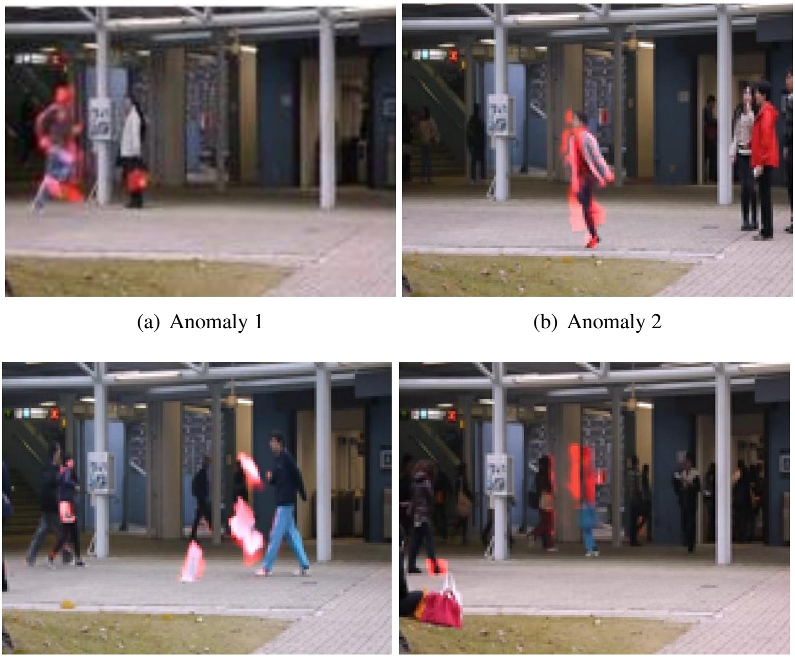
While calculating results on videos, the anomaly scores are smoothed with the same filter as Ionescu et al. [12] before they are used to compute AUC scores with the frame level ground truth. In the setting where pixel level ground truth is provided instead, we consider a frame to be anomalous if it contains abnormal regions.

Also, as the labeling on UMN and Subway datasets are not as precise as Avenue dataset and the UCSD datasets, we manually adjust the labels to be more accurate. For subway dataset, the ground truth is very roughly labeled and only large time intervals are provided for the abnormal activities, so we did the slight adjustment of labeling in reference to the ground truth provided by Adam et al. [1] upon request.

D More Detection Results

The avenue dataset is an interesting case as there are various forms of abnormal activities presented. Our improved motion feature demonstrate a great fit for detecting the anomalies. The visualized results are shown in Figure 5.

The subway dataset has a more complex environment, so as mentioned previously in the experiments, there are some reasonable false positive detections due to the limitation of unsupervised methods. Here we observe a simple case of true positive(wrong direction) and false positive(people jogging to the subway) detections by improved motion feature in Figure 6.



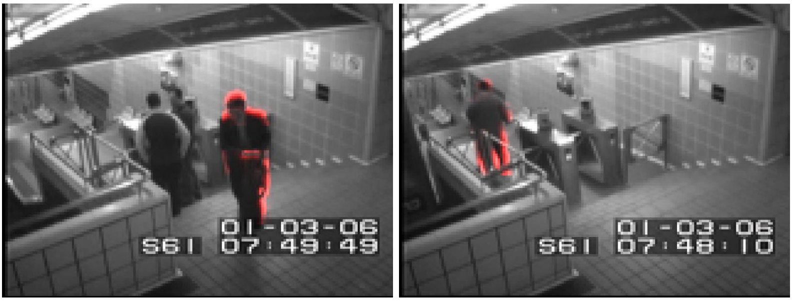
(a) Anomaly 1

(b) Anomaly 2

(c) Anomaly 3

(d) Anomaly 4

Figure 5: More detection results on Avenue dataset



(a) True Positive

(b) False Positive

Figure 6: True positive and false positive detections on Subway dataset