Video Anomaly Detection Using Open Data Filter and Domain Adaptation

Chen Zhang[1], Guorong Li[1, 2], Li Su[1, 2], Weigang Zhang[3], Qingming Huang[1, 2, 4]
[1]*School of Computer Science and Technology, UCAS, Beijing, China*
[2]*Key Lab of Big Data Mining and Knowledge Management, UCAS, Beijing, China*
[3]*Harbin Institute of Technology, Weihai, China*
[4]*Key Lab of Intelligent Information Processing, ICT, CAS, Beijing, China*
zhangchen181@mails.ucas.ac.cn, {liguorong, suli}@ucas.ac.cn, wgzhang@hit.edu.cn, qmhuang@ucas.ac.cn

***Abstract*—Video anomaly detection is a very challenging task because of the rarity, openness, and the definition of the anomalies. Researchers pay more attention to the characteristics of anomalies and have proposed a variety of anomaly detection models. However, most existing methods only use normal events to construct anomaly detection models and ignore the diversity and openness of normal events. Actually, because real-world video data often have an open-ended distribution, some normal patterns hardly ever appeared in the training data. In addition, analogous to human experience in identifying anomalies, rare abnormal events can play a certain role in the detection of similar abnormal events in the dataset. Therefore, assuming that a small number of abnormal events are known, we propose a novel supervised anomaly detection model which explicitly detects open normal events and open abnormal events in the dataset and treats open data and seen data with different classifiers. First, we use the training video to train an imbalanced classifier as the seen data classifier. Then, during the testing phase, an open data filter module is used to divide the test data into seen data and open data. Finally, we directly use the seen data classifier to generate anomaly scores for the seen test data. For the open test data, we adopt a domain adaptation method to reduce the distribution difference between it and the training data and train a new classifier to score for it. Extensive experimental results prove the effectiveness of our model.**

***Index Terms*—anomaly detection, imbalanced classifier, open data, domain adaptation**

I. Introduction

Anomaly detection is a key technology for realizing intelligent video monitoring, which can automatically detect abnormal events from the video. One of the challenges of this task is the rarity of abnormal events. Given training data containing only normal events, most previous methods [1], [2], [3], [4] used semi-supervised abnormal event detection models, and events that do not meet the characteristics of normal data are determined as abnormal events. However, due to the diversity and openness of events, normal events that have not been seen may not meet the characteristics of known normal events in the training set, will be misjudged as abnormal events, leading to a higher false positive rate.

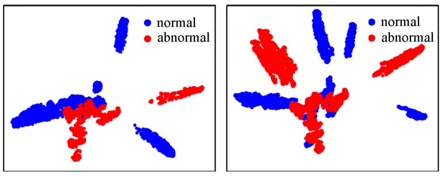

Fig. 1. The ShanghaiTech dataset features: training set features (left), test set features (right).

For the difficulty of defining abnormal events, unsupervised models [5], [6] didn't use known normal data and only detected anomalies by estimating the resolvability of frames with reference to the context of the video. If a frame can be easily distinguished from other frames in the same video, it is labeled as an abnormal frame; otherwise it is considered as normal. Unsupervised models dont't use the supervision information in the training data, but the abnormal events in the test set are annotated with reference to the training set. Therefore, unsupervised models are easy to judge the originally normal events as abnormal, causing some false positives.

Considering the openness of abnormal events, a supervised open set anomaly detection method [7] using large amounts of normal data and small amounts of abnormal data is proposed. This method used triplet loss based on a prediction model to effectively solve the open anomaly detection problem. Nevertheless, it ignores the unpredictability of certain specific normal events and open normal events. As shown in Fig. 1, the training set contains both normal and abnormal data, but the test set contains both open abnormal data and open normal data that have never been seen in the training set. Therefore, although [7] can effectively detect abnormal events that have not been seen by using a margin learning embedded prediction framework, it may fail for some open normal events and unpredictable normal events because of the unpredictability of these events.

To solve the above problem, we propose a novel supervised model that takes the openness of normal data and abnormal data in the test set into account. In the training phase, we use the training video to train an imbalanced classifier for the seen data. During the inference phase, we design a filter module which can filter out the open data (open normal data and open

This work was supported in part by the Italy-China collaboration project TALENT:2018YFE0118400, in part by National Natural Science Foundation of China: 61620106009, 61772494, 61931008, U1636214, 61836002 and 61976069, in part by Key Research Program of Frontier Sciences, CAS: QYZDJ-SSW-SYS013, in part by Youth Innovation Promotion Association CAS.

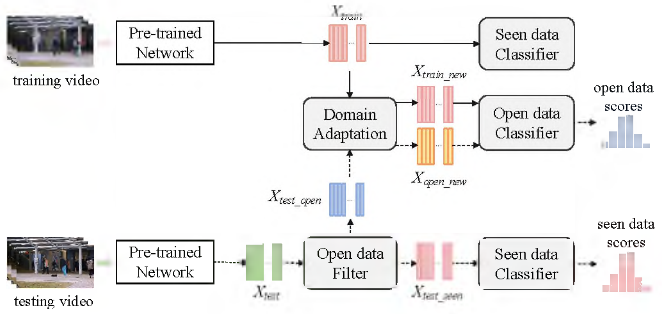

Fig. 2. Overview of our supervised anomaly detection model. The solid arrows represent the training phase, and the dotted arrows represent the testing phase.

abnormal data) in the test set. Then for the remained seen data in the test set, we use the trained imbalanced classifier to obtain anomaly scores. For the open data, because its data distribution is different from that of the training set, we regard the training data and open data as two domains. Using a domain adaptation method named Joint Distribution Adaptation [8] to minimize the difference in both the marginal distribution and conditional distribution between domains at the same time, new feature representation that is effective for substantial distribution difference, is constructed. Then we learn a classifier on the new source data features to generate anomaly scores of the open data in the test set. Finally, we integrate anomaly scores of the open data and seen data to get the anomaly detection results of the entire test set.

The main contributions of this article are summarized as follows. (1) We propose an effective open data filter module that can filter out open data, i.e. open normal data and open abnormal data. Therefore, we can efficiently use the trained classifier for the seen data; (2) For the open data, we adopt a domain adaptation method to minimize the joint distribution difference between it and the training data and obtain a new feature space. After that, a new classfier is trained in the new feature space; (3) Experimental results on challenging anomaly detection datasets prove the effectiveness of our method.

II. Proposed Model

The proposed anomaly detection model consists of four parts: feature extraction, open data filter, domain adaptation of open data and anomaly score prediction. Fig. 2 illustrates the framework of our method. In the training stage, we use a pre-trained neural network to extract the appearance features of all the training video frames X_{train}, and then use X_{train} to train an imbalanced classifier. In the testing phase, the appearance features X_{test} extracted by the pre-trained neural network are put into the open data filter, and X_{test} are divided into open data X_{test_open} and seen data X_{test_seen}. For the seen data, we directly use the seen data classifier to calculate its anomaly scores. For the open data, a domain adaptation module is used to reduce the distribution difference between it and the training data, and construct new feature representations denoted as X_{train_new} and X_{open_new}. Then we use X_{train_new} to retrain a new imbalanced classifier, and use it to generate anomaly scores for the open data. Finally, anomaly scores of the open data and seen data are integrated to get the anomaly detection results of the entire testing videos.

A. Feature Extraction

Similar to [5], we use [9] to extract appearance features $X = \{X_1 = \{x_1^1, x_1^2, \dots, x_1^l\}, \dots, X_i = \{x_i^1, x_i^2, \dots, x_i^m\}\}$ of input videos, where x_i^j is the jth frame-level feature representation of the ith video. We use X_{train}, X_{test} to denote the features of all the training videos and testing videos, respectively. Some approaches [1], [2], [3], [4] extracted motion features to improve anomaly detection performance, but it is time consuming. So we only use appearance features.

B. Open Data Filter

We design the open data filter module to analyze test data to get the seen data X_{test_seen} and open data X_{test_open}. We can analyze from Fig. 1 that the data far from the training data can actually be regarded as open data, so we use Euclidean distance between the test data and the training data to measure the openness of the test data. Due to the diversity of events, we apply k-means clustering to divide the training data into multiple categories, and obtain clusters representing various types of the training data. After that we calculate the distance between each frame x_i^j in the test video and the cluster centroid $C = \{C_1, C_2, \dots, C_k\}$ of each type of the training data. The greater the distance, the more likely a frame is to be open data. Therefore, we take the minimum distance of the cluster centroid as the minimum open probability of a frame in the test set. Formally, for each test frame x_i^j in the test set, the open probability relative to the training set $p_{open}(x_i^j, X_{train})$ is defined as follow:

$$p_{open}(x_i^j, X_{train}) = \min_n ||x_i^j - C_n||_2, \forall n \in \{1, 2, \dots, k\} \quad (1)$$

After getting the open probability of all the test video frames, we sort it and filter out a certain percentage of test frames with large open probability as the open data. In detail, we use the Mann-Whitney U test [10] to determine the optimal proportion of open data. Assume that the ratio of open data to test data is $[0.1, 0.5]$, we can filter out open data X_{test_open} according to the ratio range to obtain a series of seen data X_{test_seen}. Then the mean values of X_{test_seen}, X_{train_nor} and X_{train_abnor} are calculated and denoted as M_{test_seen}, M_{nor} and M_{abnor}, where X_{train_nor} and X_{train_abnor} correspond to the normal data and abnormal data in the training set, respectively. Finally, the U-test method [10] is used to obtain two P values denoted as P_{nor} and P_{abnor}, respectively, where P_{nor} is the P value of M_{test_seen} and M_{nor} and P_{abnor} is that of M_{test_seen} and M_{abnor}. The difference between the training data and the seen data is measured by the sum of P_{nor} and P_{abnor}. And the larger the sum of these two P values, the smaller the difference between the train data and seen data. We know that when the filtering ratio is the best, the difference between the seen data and the training data is least. So the best open ratio is the ratio corresponding to the maximum sum of P values:

$$\operatorname{argmax} P_{nor}(r) + P_{abnor}(r), \forall r \in [0.1, 0.5] \tag{2}$$

C. Domain Adaptation of Open Data

The domain adaptation module is used to reduce the distribution difference between the open data X_{test_open} and the training data X_{train}. As shown in Fig. 1, the original training data and test data are slightly different in distribution. However, the difference between the open data and the training data is obvious. At this time, the training data and the open data can actually be regarded as two different domains. From the perspective of transfer learning, we can regard the training data as the source domain X_s and open data as the target domain X_t. Considering that there may be difference in both the marginal distribution and conditional distribution between domains, we use the Joint Distribution Adaptation (JDA) [8] to simultaneously minimize these two distributions between the source and target domains. In detail, the JDA method [8] minimizes the distribution difference between domains through Maximum Mean Discrepancy distance and get a transformation matrix A. Then we can use A to obtain new feature representations of the training data and open data:

$$X_{train\_new} = A^T X_s \tag{3}$$

$$X_{open\_new} = A^T X_t \tag{4}$$

Finally, the new training data representations are used to learn a new imbalanced classifier for the open data.

D. Anomaly Score Prediction

Since there are a large amount of normal data and a little abnormal data in the training set, the classes are imbalanced. Therefore, we use an imbalanced classifier to obtain the probability that a test frame is abnormal, and regard it as the anomaly score.

In detail, we first use the Sequential model in Keras to build a neural network. Since anomaly detection can actually be regarded as a binary classification, where 1 indicates the abnormal class and 0 indicates the normal class, the sigmoid function is used in the output layer. And we use focal loss [11] to optimize our model. The sigmoid function can smoothly map the real number field to a number between $[0, 1]$, so the function value of sigmoid can be used as the probability that a test frame is abnormal.

We use $g_s(\theta_s)$ and $g_n(\theta_n)$ to represent the classifier trained with the original training data X_s and new source training data X_{train_new}, where θ_s and θ_n are the parameters of the classifiers. For the seen data $X_{test\_seen} = \{x_1^1, x_1^2, ..., x_i^j\}$ in the test set, we directly use $g_s(\theta_s)$ to generate anomaly score of the jth frame of the ith video:

$$s(x_i^j) = g_s(\theta_s, x_i^j) \tag{5}$$

For the open data, we use its new feature representation $X_{open\_new} = \{x_{1\_new}^1, x_{1\_new}^2, ..., x_{i\_new}^j\}$ to generate anomaly score with the classifier $g_n(\theta_n)$:

$$s(x_{i\_new}^j) = g_n(\theta_n, x_{i\_new}^j) \tag{6}$$

Finally, anomaly scores of the seen data and open data are integrated to evaluate the anomaly detection results of the entire testing videos.

III. Experiments

We evaluate the performance on the two most challenging anomaly datasets, Avenue [12] and ShanghaiTech [2]. And the results of experiments are showed to prove that our method can significantly improve the performance of anomaly detection.

A. Datasets

Avenue. The Avenue dataset [12] contains 16 training video clips with 15328 frames and 21 test video clips with a total of 15324. For each test frame, the ground truth is given in pixel level. This dataset is challenging due to the variety of abnormal events it contains.

ShanghaiTech. The ShanghaiTech Campus dataset [2] is more challenging. First of all, the scenes of the video are diverse, including 13 different scenes with complex light conditions and camera angles. Second, the dataset size is very large, consisting of 330 training videos with over 270, 000 frames and 107 test videos with 130 abnormal events. The ground truth of abnormal events is also given in pixel level.

B. Implementation Details and Evaluation Metric

Training/Testing Split. Since there are only normal events in the training set of the above two datasets, similar to [7], we select a small number of abnormal events from the test set and add them into the training set to verify the effectiveness of our supervised model in processing more realistic datasets. In detail, we divide the abnormal events in the test set into K types $\{A_1, A_2, \ldots, A_k\}$ by k-means clustering. Then we select one cluster at a time as the open anomaly data. A certain proportion of data is sampled from the other $K - 1$ clusters by stratified sampling and added into the training set. The remaining abnormal data is still left in the test set. In other words, there are K ways to divide training data and test data, which is similar to K-fold cross-validation. This division not only ensures that there is both normal and abnormal data in the training set, but also that the test set contains the types of abnormal events that have never been seen in the training set, which can verify the effectiveness of our method when facing open data. Besides, this division method ensures that all frames of the original test set participate in the performance evaluation.

Ratio Setting. For the fairness of comparison, we adopt the same setting as [7] for the ratio between the normal frames and abnormal frames in the training set. The ratios of normal frames to abnormal frames in the Avenue and ShanghaiTech training set are around 50:1 and 85:1, respectively. Since [7] does not specify the method of defining anomaly types, we cannot guarantee that the proportion of open abnormal data in our experiments is the same as it. In our experiments, open abnormal data accounts for 20% of the total abnormal data in the test set in terms of the types of anomalies. Since we use the clustering method to roughly divide the anomaly types

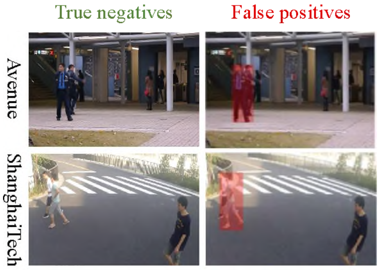

Fig. 3. Detection results on the Avenue and ShanghaiTech datasets. The detection results in the first column correspond to our method, and the results in the second column are obtained by [4].

into K categories, the proportion of types of open anomaly data may actually be higher.

Evaluation Metric. We use the frame-level area under curve (AUC) to evaluate the performance of our proposed model. At the frame-level, the model only needs to detect whether video frames contain abnormal events.

C. Experimental Results

After obtaining anomaly scores of the seen data and open data through different imbalanced classifier, we use these scores to calculate the final AUC.

Quantitative Results Analysis. The frame-level AUC metrics computed on the Avenue dataset [12] and ShanghaiTech dataset [2] are presented in Tabel I. From the results, we can see that compared with these semi-supervised methods [1], [2], [3], [4] and the unsupervised method [6], by adding a small number of abnormal frames into the training set, our supervised model significantly improves the performance of anomaly detection. The results prove that a small number of known abnormal events in the training set can be very helpful for detecting anomalies in the test set. Compared with the supervised model [7], our method can obtain better results on the Shanghaitech dataset [2] since it takes into account the existence of open normal data. This proves that our method is more efficient for anomaly detection in large multi-scene videos, which may contain a higher percentage of open data.

Qualitative Results Analysis. As shown in Fig. 3, for some false abnormal event detections in [4], our model can filter them as open data by using the open data filter and correctly detect them as true normal events. These examples can also prove the effectiveness of our model.

Ablation Study. The baseline of our method is an imbalanced classifier using focal loss [11]. In order to process open data, we add the open filter and domain adaptation modules into our model. In addition, we calculate the frame-level AUC when removing the open data filter module and directly using the domain adaptation module, which can prove the importance and effectiveness of the open filter module in the model. The ablation results on Avenue dataset [12] in Table II indicate that adding open data filter and domain adaptation modules into our model is indeed important and useful.

TABLE I
FRAME-LEVEL AUC (%) COMPARISON ON AVENUE AND SHANGHAITECH

Method	**Used Features**	**Frame AUC**	
		Avenue	**Shanghai Tech**
Con-AE [1]	Appearance + Motion	70.2	60.9
Unmasking [6]	Appearance + Motion	80.6	N/A
Stacked RNN [2]	Appearance + Motion	81.7	68.0
Frame Prediction [3]	Appearance + Motion	84.9	72.8
Margin Learning [7]	Appearance + Motion	92.8	76.8
Object-centric AE [4]	Appearance + Motion	90.4	84.9
Our method	Appearance	88.2	**91.3**

TABLE II
ABLATION RESULTS (%) ON AVENUE

Method	**Frame AUC**
Without Filter and Domain Adaptation	70.5
Without Filter but With Domain Adaptation	75.8
With Filter and Domain Adaptation (Our method)	**88.2**

IV. CONCLUSION

In this paper, we propose a novel framework for abnormal event detection in video containing open normal data and open abnormal data. Specifically, we design the open data filter and domain adaptation modules to reduce the impact of open data on the performance of abnormal event detection. Experimental results show that our method can effectively deal with open data, and performs well in large-scale surveillance videos with multiple scenes. Since most of the surveillance videos in the real world are large and multi-scenes, and may contain open data, our method is more in line with real applications.

REFERENCES

[1] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, “Learning temporal regularity in video sequences,” in *CVPR*, 2016, pp. 733–742.

[2] W. Luo, W. Liu, and S. Gao, “A revisit of sparse coding based anomaly detection in stacked rnn framework,” in *ICCV*, 2017, pp. 341–349.

[3] W. Liu, W. Luo, D. Lian, and S. Gao, “Future frame prediction for anomaly detection–a new baseline,” in *CVPR*, 2018, pp. 6536–6545.

[4] R. T. Ionescu, F. S. Khan, M.-I. Georgescu, and L. Shao, “Object-centric auto-encoders and dummy anomalies for abnormal event detection in video,” in *CVPR*, 2019, pp. 7842–7851.

[5] Y. Liu, C.-L. Li, and B. Póczos, “Classifier two sample test for video anomaly detections,” in *BMVC*, 2018, p. 71.

[6] R. Tudor Ionescu, S. Smeureanu, B. Alexe, and M. Popescu, “Unmasking the abnormal events in video,” in *ICCV*, 2017, pp. 2895–2903.

[7] W. Liu, W. Luo, Z. Li, P. Zhao, S. Gao *et al.*, “Margin learning embedded prediction for video anomaly detection with a few anomalies,” in *IJCAI*, 2019, pp. 3023–3030.

[8] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, “Transfer feature learning with joint distribution adaptation,” in *ICCV*, 2013, pp. 2200–2207.

[9] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *ICLR*, 2015.

[10] H. B. Mann and D. R. Whitney, “On a test of whether one of two random variables is stochastically larger than the other,” *The annals of mathematical statistics*, pp. 50–60, 1947.

[11] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *ICCV*, 2017, pp. 2980–2988.

[12] C. Lu, J. Shi, and J. Jia, “Abnormal event detection at 150 fps in matlab,” in *ICCV*, 2013, pp. 2720–2727.