IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134115

Improving Anamoly Detection in Live Streams Using Deep Multiple Instance Learning And Weak Labels

Sravan Kumar Reddy M[1], Yaswanth Kumar Reddy Bussa[2],

Mohammad Peera Thondaladinne[3], Gowthami Nagappagari[4], Divya Byreddy[5]

Associate Professor, Department of Computer Science and Engineering, Rajeev Gandhi Memorial College of Engineering and Technology, Nandyal, Andhra Pradesh, India[1]

Students, Department of Computer Science and Engineering, Rajeev Gandhi Memorial College of Engineering and Technology, Nandyal, Andhra Pradesh, India[2-5]

Abstract: The aim of the project is to develop an effective method for anomaly detection and recognition in live streams. Anomaly detection refers to the task of identifying instances or events that deviate from the expected or normal patterns. In the context of live streams, anomalies can include activities such as fighting, accidents, robbery, and other unusual events. Rather than requiring continuous human monitoring of surveillance feeds, automated anomaly detection can relieve the burden on security staff. The system can alert operators only when potentially significant anomalies are detected. Effective anomaly detection can help reduce false alarms in live streams. By accurately distinguishing between the normal and anomalous activities, security personnel can avoid unnecessary responses to non-threatening situations. The proposed method works better by recognizing the abnormal behaviors in long-distance captured images and reducing the false alarm rate by effectively distinguishing between the normal and abnormal activities.

Keywords: Anomaly detection, live streams, false alarms, unusual events, abnormal behaviors, normal activities, effective recognition

I. INTRODUCTION

Live streams are progressively being utilized in broad daylight places for example roads, convergences, banks, shopping centres, and so forth to increment public wellbeing. Be that as it may, the checking ability of policing has not kept pace. The outcome is that there is a lack of glaring in the usage of observation cameras and an impossible proportion of cameras to human screens. One basic undertaking in video observation is identifying irregular occasions like car crashes, wrongdoings or criminal operations.

For the most part, abnormal occasions seldom happen when contrasted with typical exercises. Consequently, to forestall the misuse of work and time, creating wise PC vision calculations for programmed video oddity identification is a squeezing need. The objective of a pragmatic inconsistency discovery framework is to convenient sign an action that strays typical examples and distinguish the time window of the happening oddity. Thusly, irregularity recognition can be considered as coarse level video understanding, which channels out peculiarities from ordinary examples. When an abnormality is recognized, it can additionally be arranged into one of the particular exercises utilizing order strategies.

Certifiable abnormal occasions are confounded and various. Posting the conceivable irregular events is all troublesome. Thusly, it is advantageous that the oddity recognition calculation depends on no earlier data about the occasions. All in all, irregularity discovery ought to be finished with least oversight. Scanty coding-based approaches are considered as agent strategies that accomplish cutting edge inconsistency recognition results.

These strategies expect that main a little starting piece of a video contains typical occasions, and in this manner the underlying part is utilized to construct the ordinary occasion word reference. Then, the primary thought for oddity location is that abnormal occasions are not precisely reconstructable from the ordinary occasion. Be that as it may, since the climate caught by observation cameras can change definitely throughout the time (for example at various seasons of a day), these methodologies produce high phony problem rates for various ordinary ways of behaving.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134115

It is very important for academics to study the fusion rules used by fusion centres in Wireless Sensor Networks [1]. Nowadays, the homeland security field faces more difficulties in identifying suspicious or abnormal entities in huge datasets [2]. At present, the field of homeland security faces many obstacles while determining abnormal or suspicious entities within the huge set of data [3].

The procedures outlined above may seem attractive, but they operate under the premise that any pattern that differs from the taught usual patterns should be regarded as abnormal. This assumption might not hold true, though, as it might be exceedingly challenging or impossible to describe a normal event that accounts for all plausible typical patterns and behaviours. More significantly, it's not always clear where abnormal behavior ends and normal behavior begins. Furthermore, in practical settings, an identical conduct may exhibit normalcy or abnormality depending on the circumstances. It is therefore claimed that an anomaly detection system can learn more effectively by using training data of both normal and abnormal occurrences. In this research, we present an algorithm for anomaly identification based on training videos with weak labels.

To sum up, the following contributions are made by this paper:

• Our proposal for anomaly identification with only weakly labelled training films is a MIL solution. In order to train a deep learning network to identify anomaly scores for video segments, we suggest a MIL ranking loss with smoothness and sparsity restrictions.

• We present a comprehensive dataset for video anomaly detection, encompassing 190 real-world surveillance videos featuring 13 distinct abnormal events and regular activities that were recorded by security cameras. With a total of 128 hours of footage, it is by far the largest dataset, having more than 15 times as many videos as current anomaly datasets.

• Our suggested strategy outperforms the most advanced anomaly detection techniques, according to experimental results on our newly created dataset.

• Because of the substantial intra-class variability and complexity of the activities, our dataset also provides a demanding benchmark for activity detection on untrimmed videos. We provide the outcomes of the baseline techniques, C3D and TCNN.

Accordingly, detecting anomalies among data is a very important task, and that has several high-impact appliances in fields like security, health care, law enforcement, and finance [4]. Dynamic Rumor Influence Minimization with User Experience Model in Social Network [5]. The Ensemble Training set consists of labelled dataset described in various formats and its instances are defined in attributes based vectors [6].

II. RELATED WORK

P. Kuppusamy et al. [9] concentrated on two particular aberrant behaviors: falling and loitering. The authors cover a number of techniques for loitering, such as the Generalized Sequential Patterns (GSP) algorithm, 3DCNN with slow fusion and Brox's optical flow approach, GMM and fuzzy C-means clustering, GMM and pedestrian activity area categorization using MeanShift, and Two-stream 3D CNN. Based on the CAVIAR and UMN datasets, the authors found that the two-stream 3D CNN model performs the best, with an accuracy of 97%.

The authors address the following techniques for falling: Shape feature encoding via Fisher Vector for effective fall detection, Background subtraction and rule-based classifier, CNN is used for patient monitoring. Fall detection systems for homes with cameras, fall detection systems based on vision utilizing VGG16, fall detection systems for the elderly based on multi-stream CNN, and two-stream 3D CNN. The authors discover that, with an accuracy of 99.72% on the FDD dataset with data augmentation, the geriatric fall detection based on multi-stream CNN model performs the best.

A 3D convolutional neural network (CNN) is the suggested approach by J. Li et al. [10] for identifying and identifying fight scenes in videos. The model's architecture aims to achieve both computing efficiency and accuracy. The suggested method's salient characteristics are: Dense blocks: Dense blocks help the model become more generalizable by allowing features to be reused. Bottleneck architecture: This architecture reduces the size of feature maps and boosts computer performance. Global average pooling: To connect convolutional networks to a fully connected layer for classification, global average pooling is utilized. The suggested approach produced state-of-the-art outcomes on both of the battle video datasets that were used for evaluation. As you can see, instead of just displaying one column on the last page, the formatting makes sure that the text ends in two equal-sized columns.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.134115

A system for real-time human subject discrimination and anomalous behavior detection was proposed by K.E. Ko et al. [11]. There are three primary components to the system:

- A real-time object detector based on deep learning (YOLO v2) that can identify human beings.
- A deep convolutional framework for the detection of anomalous behavior.
- An object entity estimator based on the Kalman filter to distinguish between human individuals in the same scene.

The YOLO v2 network is initially used by the algorithm to identify every human subject present in the scene. Then, to distinguish between the identified human individuals, an object entity estimator based on the Kalman filter is employed. Lastly, each distinguished human subject's current behavior is identified using the deep convolutional framework for anomalous behavior identification. The system demonstrated efficacy in detecting deviant behavior and discriminating between human subjects using a benchmark video dataset.

The suggested framework by W. Song et al. [12] is a machine learning-based method for detecting anomalous human behavior. This implies that a dataset of labeled video footage must be used to train the framework. The caliber of the training data will determine how well the suggested framework performs. The framework might not be able to generalize well to new data if the training set is not representative of the real-world data. It is possible to expand the suggested framework to detect several kinds of anomalous behavior, like object tracking, crowd monitoring, and traffic monitoring. All things considered, the suggested paradigm offers a promising method for identifying aberrant human behavior. It is simple to use, efficient in terms of computing, and effective. It is crucial to remember that the caliber of the training data will determine how well the framework performs.

The suggested approach by C. Direkoglu et al. [13] operates as follows: In the video, optical flow is calculated for every frame. The suggested MII generation scheme is used to generate the MII for every frame. The CNN receives the MIIs in order to classify them. For every MII, the CNN generates a prediction indicating whether the MII is normal or aberrant. The suggested approach for detecting atypical crowd events has a number of benefits over current approaches, including: It can withstand noise and occlusion better. Computing is more effective. It can be applied to a wider range of anomalous occurrences. The suggested approach has been assessed using UMN and PETS2009, two publically accessible datasets. The findings demonstrate that, when applied to both datasets, the suggested strategy outperforms the current approaches and yields the best results. All things considered, the suggested technique is a viable one for detecting anomalous crowded events. It is efficient, accurate, and broadly applicable.

In the research, deep learning-based automatic food classification methods were studied, using the Squeeze Net and VGG-16 convolutional neural networks [7]. In terms of humanity, paddy agriculture is quite important, particularly in the Asian subcontinent. Convolutional Neural Network was chosen because of how well it handles images [8].

III. PROPOSED ANOMALY DETECTION METHOD

During training, the suggested method starts by segmenting surveillance footage into a predetermined number of segments. These sections create situations within a bag. We use the suggested deep MIL ranking loss to train the anomaly detection model utilizing both positive (anomalous) and negative (normal) bags.

3.1. Multiple Instance Learning

The labels of all positive and negative samples are supplied in standard supervised classification problems with support vector machines, and the classifier is trained using the optimization function that follows:

$$\min_{\mathbf{w}} \left[\frac{1}{k} \sum_{i=1}^{k} \overbrace{max(0, 1 - y_i(\mathbf{w}.\phi(x) - b))}^{①} \right] + \frac{1}{2} \|\mathbf{w}\|^2, \quad (1)$$

where w is the classifier to be taught, k is the total number of training examples, b is a bias, y_i is the label of each example, φ(x) is the feature representation of an image patch or a video segment, and 1 is the hinge loss. Accurate annotations of positive and negative instances are required in order to train a robust classifier. For supervised anomaly identification, a classifier requires temporal annotations of every video segment. However, it takes a lot of effort and time to collect temporal annotations for videos.

The requirement of having these precise temporal annotations is loosened by MIL. The exact time locations of anomalous incidents in videos are unknown in MIL. Rather, all that is required are video-level labels that indicate if an anomaly is present across the entire film.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134115

Videos with abnormalities are classified as positive, while those without any anomalies are classified as negative. Afterwards, we depict a positive video as a positive bag Ba, in which distinct temporal segments constitute separate instances (p^1, p^2, ..., p^m), with m representing the total number of instances in the bag. We presume that the abnormality is present in at least one of these cases. Similarly, temporal segments in a negative bag, Bn, generate negative instances (n^1, n^2, ..., n^m), which represent the negative video. There isn't an oddity in any of the examples in the negative bag. One can optimize the objective function with respect to the maximum scored instance in each bag while the precise information (i.e., instance-level label) of the positive instances is unknown.

$$\min_{\mathbf{w}} \left[\frac{1}{z} \sum_{j=1}^{z} max(0, 1 - Y_{\mathcal{B}_j} (\max_{i \in \mathcal{B}_j} (\mathbf{w}.\phi(x_i)) - b)) \right] + \|\mathbf{w}\|^2 , \quad (2)$$

where z is the total number of bags, Y_{Bj} stands for the bag-level label, and all other variables are the same as in Eq. 1.

3.2. Deep MIL Ranking Model

Since abnormal conduct is highly subjective and varies greatly from person to person, it is challenging to characterize with precision. Furthermore, it's not clear how anomalies should be given 1/0 labels. Furthermore, anomaly detection is typically approached as a low likelihood pattern detection problem rather than a classification problem because there aren't enough examples of anomalies [10, 5, 20, 26, 28, 42, 18, 26].

We formulate anomaly identification as a regression problem in our suggested methodology. Higher anomaly scores than normal segments are what we are aiming for in the anomalous video parts. The simplest method would be to employ a ranking loss that promotes high scores for unusual video portions relative to regular segments, like:

$$f(\mathcal{V}_a) > f(\mathcal{V}_n), \quad (3)$$

where the associated predicted scores are denoted by $f(V_a)$ and $f(V_n)$, respectively, and V_a and V_n stand for anomalous and normal video segments. If the segment-level annotations are known during training, the above ranking function ought to function properly.

However, using Eq. 3 is not feasible in the lack of video segment level annotations. Alternatively, we suggest the multiple instance ranking objective function that follows:

$$\max_{i \in \mathcal{B}_a} f(\mathcal{V}_a^i) > \max_{i \in \mathcal{B}_n} f(\mathcal{V}_n^i), \quad (4)$$

where the maximum is applied to every video segment in every bag. We only enforce ranking on the two occurrences of the bag that have the highest anomaly score in the positive and negative bags, as opposed to applying ranking to every instance of the bag. The real positive instance (anomalous segment) is most likely the segment in the positive bag that has the greatest anomaly score. The segment that appears to be the most anomalous but is actually a typical instance is the one that corresponds to the greatest anomaly score in the negative bag. This negative case is regarded as a hard instance that could lead to an anomaly detection false alarm.

$$l(\mathcal{B}_a, \mathcal{B}_n) = \max(0, 1 - \max_{i \in \mathcal{B}_a} f(\mathcal{V}_a^i) + \max_{i \in \mathcal{B}_n} f(\mathcal{V}_n^i)). \quad (5)$$

The fact that the previously mentioned loss ignores the abnormal video's underlying temporal structure is one of its limitations. Initially, anomalies in real-world situations typically last for a brief period of time. The scores of the instances (segments) in the anomalous bag in this scenario should be sparse, meaning that the anomaly may only be present in a small number of segments. Secondly, because the video is composed of multiple segments, there should be a smooth transition between each section and the anomaly score. Thus, by minimizing the difference in scores for neighboring video segments, we enforce temporal smoothness between anomaly scores of temporally adjacent video segments. The loss function is transformed by incorporating the smoothness and sparsity restrictions on the instance scores.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134115

$$l(\mathcal{B}_a, \mathcal{B}_n) = \max(0, 1 - \max_{i \in \mathcal{B}_a} f(\mathcal{V}_a^i) + \max_{i \in \mathcal{B}_n} f(\mathcal{V}_n^i))$$

$$+\lambda_1 \overbrace{\sum_{i}^{(n-1)} (f(\mathcal{V}_a^i) - f(\mathcal{V}_a^{i+1}))^2}^{①} + \lambda_2 \overbrace{\sum_{i}^{n} f(\mathcal{V}_a^i)}^{②}, \quad (6)$$

where the sparsity term is represented by 2 and the temporal smoothness term by 1. The fault is back-propagated from the maximum scored video segments in both positive and negative bags in this MIL ranking loss. We anticipate that the network will acquire a generalized model to forecast high scores for anomalous regions in positive bags after training on a sizable number of positive and negative bags (refer to Figure 8). Lastly, the whole goal function we have is provided by

$$\mathcal{L}(\mathcal{W}) = l(\mathcal{B}_a, \mathcal{B}_n) + \|\mathcal{W}\|_F, \quad (7)$$

where W represents model weights.

Bags Formations. Each video is divided into an equal number of temporally separated segments that do not overlap, and these segments are used as bag instances. We extract the 3D convolution features for every video segment. This feature representation is what we employ for video action identification because it is computationally efficient and clearly capable of capturing appearance and motion dynamics.

IV. DATASET

4.1. Previous Datasets

We give a quick overview of the current datasets for video anomaly detection. Five distinct staged videos that show people walking around before they eventually start jogging in different directions make up the UMN dataset. The sole action that distinguishes the oddity is running. There are 70 and 28 surveillance videos in the UCSD Ped1 and Ped2 collections respectively. There is only one location where those videos were taken. The minor anomalies in the videos—such as persons crossing the path and non-pedestrian objects like wheelchairs, skateboarders, and bikes—do not accurately depict real-world oddities in video surveillance. The Avenue dataset includes 37 different videos. Even though there are more anomalies, they are all posed and photographed in one place.

This dataset's films are comparable in length, and some of the anomalies—like throwing paper—are implausible. Each of the datasets for the subway entrance and exit has a single, lengthy surveillance video. Simple anomalies like walking in the opposite direction and omitting payment are captured in the two recordings. The BOSS dataset is gathered from a train-mounted surveillance camera. It includes both regular footage and anomalies like harassment, a sick individual, and a panic attack. Every anomaly is acted out by actors. In general, the quantity and duration of the videos in the prior datasets for video anomaly detection are low. There are also few variations in anomalies. Furthermore, a few oddities appear implausible.

4.2. Our Dataset

We create a new large-scale dataset to assess our approach because the old ones had limitations. Long, unedited surveillance footage covering thirteen actual anomalies—abuse, arrest, arson, assault, accident, burglary, explosion, fighting, robbery, shooting, stealing, shoplifting, and vandalism—make up the collection. The reason these anomalies were chosen is that they significantly affect public safety.

Gathering of Videos. We train ten annotators with varying degrees of computer vision experience to gather the dataset in order to guarantee its quality. For each abnormality, we use text search phrases (with minor variations, such as "car crash," "road accident," etc.) to look for videos on YouTube1 and LiveLeak2. Thanks to Google Translator, we can also perform text queries for each anomaly in many languages (e.g., French, Russian, Chinese, etc.) to obtain the maximum number of videos. Videos that fit any of the following categories are eliminated: videos that have been hand-picked, edited manually, are pranks, weren't recorded by CCTV, were taken from the news, were taken with a hand-held camera, or contain compilations. Videos in which the anomaly is unclear are also discarded. A total of 950 unedited real-world surveillance recordings with glaring anomalies are gathered using the video pruning restrictions mentioned above. With the same restrictions, 950 regular videos are collected, bringing the total number of videos in our sample to 1900.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134115

Annotation. All that is needed for training our anomaly detection system is video-level labels. However, we must be aware of the temporal annotations, or the beginning and ending frames of the anomalous event in each testing anomalous video, in order to assess its performance on testing videos. In order to identify the temporal span of each abnormality, we assign the identical videos to a number of different annotators. By averaging the annotations from several annotators, the final temporal annotations are produced. The entire dataset is finished after months of rigorous work.

Training and Testing sets. Our dataset is split into two categories: the testing set, which consists of the remaining 150 normal and 140 anomalous videos, and the training set, which consists of 800 normal and 810 anomalous videos (details presented in Table 2). All 13 anomalies are present in the training and testing sets at different points in time throughout the videos. Moreover, there are several irregularities in some of the videos.

V. EXPERIMENTAL RESULTS

5.1. Comparison with the State-of-the-art

We contrast our approach with two cutting-edge anomaly detection techniques. According to Lu et al., [15] a dictionary-based method was suggested for learning typical behaviors and using reconstruction errors to identify abnormalities. We compute gradient-based features in each volume and extract 7000 cuboids from each regular training video by following their method. Sparse representation is used to learn the dictionary after PCA is used to reduce the feature dimension. A completely convolutional feedforward deep auto-encoder based method for learning local features and classifiers was proposed by Hasan et al. [14].

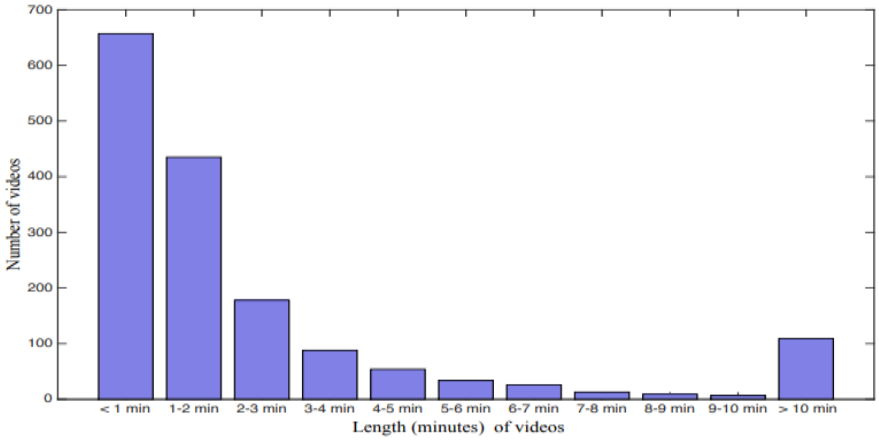

Figure 1. Distribution of videos according to length (minutes) in the training set.

We train the network on regular videos with a temporal window of 40 frames by using their implementation. Reconstruction error is akin to anomaly measurement. As a baseline, we also employ a binary SVM classifier. To be more precise, we treat regular videos as a different class and all anomalous videos as one. For every video, C3D characteristics are calculated, and a linear kernel is used to train a binary classifier. This classifier gives the likelihood that every video clip is abnormal for testing purposes.

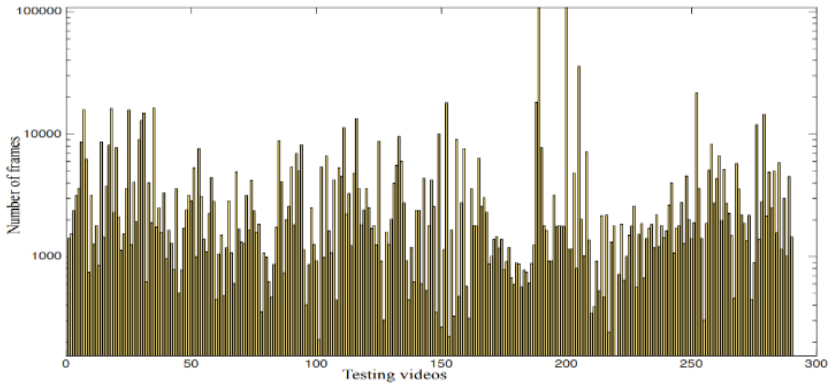

Figure 2. Distribution of video frames in the testing set.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134115

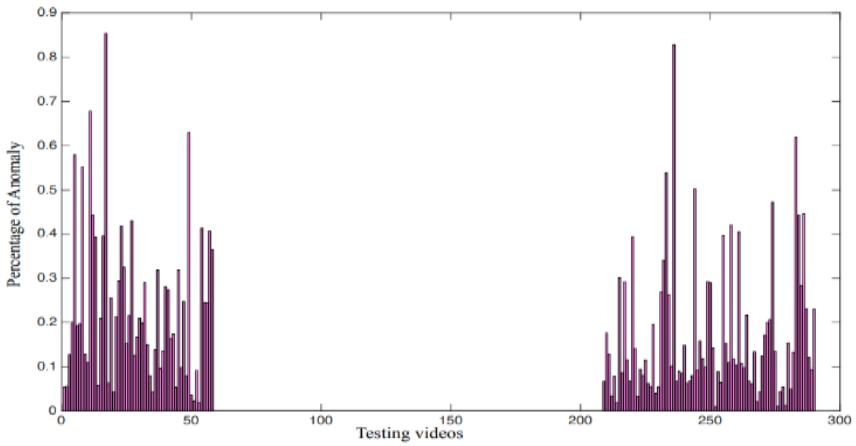

Figure 3. Percentage of anomaly in each video of the testing set. Normal videos (59 to 208) do not contain any anomaly.

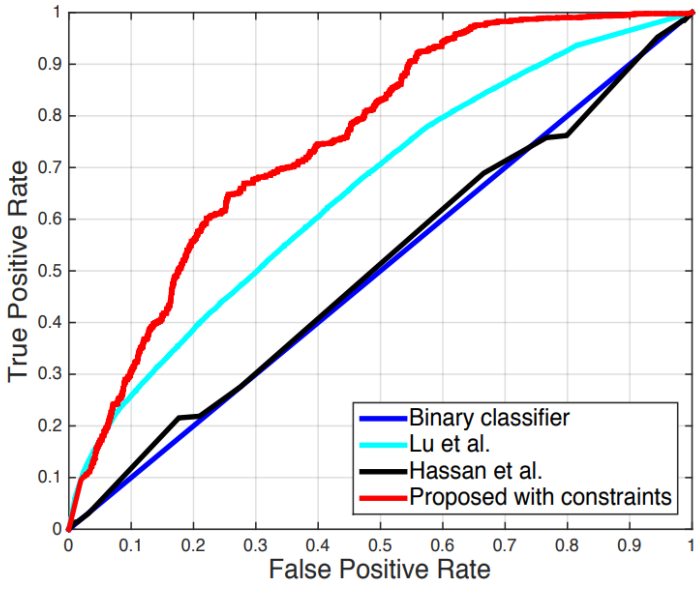

Figure 4. ROC comparison of binary classifier (blue), Lu et al. [15] (cyan), Hasan et al. [14] (black), proposed method without constraints (magenta) and with constraints (red).

Figure 4 and Table 1 displays the quantitative comparisons in terms of AUC and ROC. Additionally, we contrast our approach's outcomes with and without smoothness and sparsity requirements. The outcomes demonstrate that our strategy performs noticeably better than the current techniques. Specifically, when false positive rates are low, our strategy outperforms others in terms of true positive rates.

The findings of the binary classifier show that anomaly identification in actual surveillance footage is not possible using conventional action recognition techniques. This is due to the fact that our dataset consists of lengthy, uncut recordings, in which anomalies typically appear over brief periods of time. As a result, the characteristics taken from these training videos without trimming lack sufficient discrimination for the unusual occurrences.

For practically all test videos, the binary classifier in the studies yields extremely low anomaly scores. Dictionary learning is insufficiently reliable to distinguish between typical and unusual patterns. Apart from generating minimal reconstruction error for the average segment of the videos, it also generates minimal reconstruction error for the abnormal segment. Hasan et al. [14] pick up on typical patterns quite quickly.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134115

5.2. Analysis of the proposed method

Model Training. The fundamental premise of the suggested method is that the network can automatically learn to predict the location of the anomaly in the video if it is provided with a large number of both positive and negative films with video-level labels. To accomplish this, the network has to understand how to generate high results in training iterations for aberrant video portions. The network predicts high scores for both anomalous and typical video portions after 1,000 iterations. The network begins to generate low ratings for normal segments and retain high values for anomalous segments after 3,000 iterations. The network automatically learns to precisely localize anomaly as iterations grow and more videos are seen. It should be noted that the network can predict the temporal location of an anomaly in terms of anomaly scores even when we do not use any segment-level annotations.

Method	AUC
Binary Classifier	50.0
Hasan et al. [14]	50.6
Lu et al. [15]	65.51
Proposed w/o constraints	74.44
Proposed with constraints	75.41

Table 1. AUC comparison of various approaches on our dataset.

False Alarm Rate. A significant portion of a surveillance film is typical in the actual world. Low false alarm rates on typical videos should characterize a strong anomaly detection technique. As a result, we exclusively test our approach's and other methods' effectiveness on regular videos. The false alarm rates for several approaches at the 50% threshold are listed in Table 2. Compared to other approaches, ours has a substantially lower false alarm rate, suggesting a more reliable anomaly detection system in real-world scenarios. This demonstrates that our deep MIL ranking model learns more broad normal patterns when it is trained on both normal and abnormal films.

Method	[14]	[15]	Proposed
False Alarm Rate	27.2	3.1	1.9

Table 2. False alarm rate comparison on normal testing videos.

VI. CONCLUSIONS

We suggest using deep learning to find abnormalities in real-world situations found in surveillance footage. Anomaly detection may not be best served by relying solely on normal data because of the complexity of these real-world anomalies. We try to take advantage of regular and unusual security footage. Using a deep multiple instance ranking framework and weakly labeled data, we construct a general model of anomaly identification, avoiding the time-consuming task of temporally annotating anomalous portions in training films.

A new large-scale anomaly dataset with a range of real-world anomalies is introduced in order to validate the suggested method. The outcomes of our experiment on this dataset demonstrate that our suggested anomaly detection strategy outperforms baseline techniques by a significant margin. We also show how our dataset is helpful for the second task, which is the identification of abnormal activity.

REFERENCES

[1]. Zhang, Gaoyuan, Kai Chen, Congfang Ma, Sravan Kumar Reddy, Baofeng Ji, Yongen Li, Congzheng Han, Xiaohui Zhang, and Zhumu Fu. "Decision fusion for multi-route and multi-hop Wireless Sensor Networks over the Binary Symmetric Channel." Computer Communications 196 (2022): 167-183.

[2]. M. Sravan Kumar Reddy, and Dharmendra Singh Rajput. "Ternary-based feature level extraction for anomaly detection in semantic graphs: an optimal feature selection basis." Sādhanā 46 (2021): 1-16.

[3]. M. Sravan Kumar Reddy, and Dharmendra Singh Rajput. "Design and Development of Ternary-Based Anomaly Detection in Semantic Graphs Using Metaheuristic Algorithm." International Journal of Digital Crime and Forensics (IJDCF) 13, no. 5 (2021): 43-64.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134115

[4]. M. Sravan Kumar Reddy, and Dharmendra Singh Rajput. "Lion plus firefly algorithm for ternary-based anomaly detection in semantic graphs in smart cities." International Journal of Ad Hoc and Ubiquitous Computing 38, no. 1-3 (2021): 17-29.

[5]. M. Sravan Kumar Reddy and G. Padmaja on "Dynamic Rumor Influence Minimization with User Experience Model in Social Network" has been published in International Journal of Scientific Research and Review (IJSRR), ISSN NO: 2279-543X, Volume 8, Issue 3, March 2019, Page no. 583-590.

[6]. M Sravan Kumar Reddy and Dharmendra Singh Rajput. Improving ADA-boost as a Popular Ensemble in Classification Problems. International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-8 Issue-9S3, Jul, 2019. DOI: 10.35940/ijitee.I3043.0789S319

[7]. Dr. M. Sravan Kumar Reddy, B. Neelima Rani, "Food Category Prediction from Images Using CNN ", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.10, Issue 7, page no.g171-g175, July-2023, Available :http://www.jetir.org/papers/JETIR2307626.pdf

[8]. Dr. M. Sravan Kumar Reddy, L. Naga Swetha, " Deep Neural Network Based Paddy Crop Prediction Using Machine Learning", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.10, Issue 7, page no.g202-g206,July-2023,Available :http://www.jetir.org/papers/JETIR2307633.pdf

[9]. P. Kuppusamy, C.L. Hung, Enriching the multi-object detection using convolutional neural network in macro-image, in: 2021 International Conference on Computer Communication and Informatics (ICCCI), IEEE, Coimbatore, 2021, pp. 1–5, https:// doi.org/10.1109/ICCCI50826.2021.9402565.

[10]. j. Li, X. Jiang, T. Sun, K. Xu, Efficient Violence Detection Using 3d Convolutional Neural Networks, 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2019, pp. 1–8, https://doi.org/10.1109/ AVSS.2019.8909883.

[11]. K.E. Ko, K. Sim, Deep convolutional framework for abnormal behavior detection in a smart surveillance system, Eng. Appl. Artif. Intell. 67 (2018) 226–234, https:// doi.org/10.1016/j.engappai.2017.10.001.

[12]. W. Song, D. Zhang, X. Zhao, J. Yu, R. Zheng, A. Wang, A novel violent video detection scheme based on modified 3D convolutional neural networks, IEEE Access 7 (2019) 39172–39179.

[13]. C. Direkoglu, Abnormal crowd behavior detection using motion information images and convolutional neural networks, IEEE Access 8 (2020)80408–804

[14]. M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis. Learning temporal regularity in video sequences. In CVPR, June 2016.

[15]. C. Lu, J. Shi, and J. Jia. Abnormal event detection at 150 fps in matlab. In ICCV, 2013.