International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

An AI Federated System for Anomalies Detection in Videos using Convolution Neural Network Mechanism

Sajeeda Shikalgar[1], Dr. Rakesh K. Yadav[2], Dr. Parikshit N. Mahalle[3]

Submitted: 26/10/2022 **Revised**: 14/12/2022 **Accepted:** 06/01/2023

***Abstract*—** Research on this topic has been going on for more than a decade at this point. It focuses on the detection of anomalies in video. Scholars' attention has been drawn to this topic in recent years due to the widespread applicability of its findings. As a direct consequence of this, a wide range of strategies have been utilised over the course of time. These methods range from matrix factorization to methods based on machine learning. Although there are already several studies being conducted in this subject, the purpose of this article is to provide an overview of the recent advancements in the field of detecting anomalies through the use of artificial intelligence and the internet of things (IIoT). Regarding the topic of detecting anomalies in video feeds of a single scene, this article provides a summary of previous research patterns. In this section, we discuss the various ways challenges might be formulated, databases that are open to the public, and criteria for evaluation. In this paper, we implement Convolution Neural Network for detection of anomalies in video. We provide the comparative analysis of these researches on the basis of its accuracy on standard dataset. Apart from that we implement the proposed system based on Convolution Neural Network for Video Anomaly Detection. We use the standard performance parameters like, precision, recall, F-Score and accuracy so as to evaluate the performance of our model. For better analysis we compared our model with state of art algorithms like, LR, NB and SVM.

Keywords— *Machine Learning, Deep Learning, Surveillance, Abnormal Event Detection, Video Anomaly Detection, Image Processing , Artificial Intelligence*

1. Introduction

In order to track human activities and deter violence from arising, video surveillance have indeed been widely present in different institutions. It comes as no surprise that every time anything comprise of various occurs, there must be someone behind monitoring the videos and signaling an alarm. With the advancement in technology, Artificial Intelligence and IIoT can be prove to be a game changer. However, these incidents do not occur too frequently and much of the time, nothing out of the ordinary can be observed by the individual watching these videos[1]. It is possible to classify these odd phenomena as deviations that can be described as patterns that do not correspond to what is considered natural. The role of detecting these nonconforming trends is called the identification of anomalies. The identification of video anomalies is the process of detecting anomalies in a video in space and/or time, where anomalies are actually events that are out of the norm. In other equivalent words, exceptions are often referred to as abnormalities, novelties and outliers. As a result, researchers have attempted to create a robust algorithm for the identification of irregularities that can simplify the tracking and detection process of suspicious activities in surveillance images.

Information derived from unsupervised airport luggage, to individuals falling down, to an individual loitering outside a hotel. We are following the description given in [2],

Definition 1 *Video abnormalities may be known to be the incidence of odd attributes of behavior or motion, or the occurrence of normal attributes of behavior or motion in uncommon locations or periods.*

This research focuses on the identification of single-scene video anomalies since it is the most frequent scenario for the detection of video anomalies in practical uses. A security camera observing a scene and an agent responsible for noticing any suspicious behaviour that happens are an encouraging example. This example illustrates the realistic value of designing real time video anomaly detection algorithms, and this is obviously a job that a computer can perform better because of the severe complexity for a human.

Take caution when capturing a long-term video stream (typically with little important happening). If a human is turned into a camera feed bank, the situation can better be interpreted as a series of problem detecting a single-scene video abnormality for two reasons: (1) localization-related anomalies have to be dealt with and (2) it is possible that no scenes of the directional microphones are reliable.

The majority of recent video anomaly identification research has failed to comprehend the significant differences between single-scene and multi-scene video classification algorithm. The use of destination anomalies in a single threat detection methodology but not in a multi-scene formulation is a significant distinction. The absence of a comprehension including its single-scene/multi-scene

[1]*Research Scholar, Shri Venkateshwara University, Amroha, Uttar Pradesh, India.*
[2]*Director Academics, Shri Venkateshwara University, Amroha, Uttar Pradesh, India.*
[3]*Professor and Head, Department of Artificial Intelligence and Data Science, Bansilal Ramnath Agarwal Charitable Trust's, Vishwakarma Institute of Information Technology, Pune, Maharashtra, India.*
sajeeda.shikalgar@mitwpu.edu.in[1], dir.academics@svu.edu.in[2], aalborg.pnm@gmail.com[3]

division line is largely due to the fact that most actual video detection datasets with the exception of the Streeter scene have comparatively low location-dependent anomalies, ensuring strategies that are not highly penalized that cannot handle location-dependent anomalies. A location-related anomaly is an entity or event that in some parts of a scene is unique, but not in others. Walking on the grass is an excellent example of this. There can be multiple points of uncertainty in a given scene.

Walking on the lawn and other areas where walking is prohibited is common and, as a result, uncommon. The only difference between these two businesses is their location. Normal footage is created for the development of a single type of normality from several different, unrelated scenes in multi-scene video image classification. The aim of this scenario is to determine the normalcy of a sequence of appearances and incidents that exist throughout all of the films. As the events in the multi-scene structure do not correlate, a model in which the action in some scenes is anomalous, but not in others, cannot be created. The detection of video anomalies in one scene is typical of location-based abnormalities (such as jaywalking, cycling on pedestrian sidewalks, driving in the complete opposite direction etc.) which include ordinary activities and objects in unusual locations. The quantity of scenes is also critical when working with several short films.

The usual training videos must be "consistent" in all scenes, in the sense that what is natural and what is anomalous must be the same. This does not extend to single-scene detection. This is due to the video from all the different scenes being combined to provide a common standard of normalcy. A car interfering to a home, for example, may be typical in one situation despite the presence of a boarding dock, while a truck confirming to a structure may be uncommon in another. There would not be any scenes like this in the future. As the single scene anomaly detection in video streams is practically more advantageous, we will focus our efforts in this research on this configuration of the issue.

Few of the most recent datasets for anomaly detection like CUHK Avenue, Street Scene, UMN, Subway, and UCSD Ped1 & Ped2 are discussed in literature ([3][4][5][6][7]). A minor camera motion is currently present in the CUHK Avenue[4] and Street Scene[7] datasets. If the world position of each frame is maintained, a prototype for a particular scene will observe the camera movement, with each frame's overall overlap with the next ones (as might happen with a pan-tilt zoom surveillance camera). And these types of definition methods can sometimes be taken into account. There seem to be no standard datasets or structures available for this single-scene motion camera variant of the issue at the moment, so this is a promising area for future research.

A brief overview of common AI and IIoT based techniques for identifying single-scene video abnormalities is presented. Second, during in the training process, the features calculated from one or more recordings of a scene are used to create a model of acceptable behavior.

There was no effect on them by deviations. The detection method then creates new videos from the very same scene to evaluate the same features. The features are used in combination with the model to add anomaly results to each voxel in the input video. The results of the anomalies detected are then calculated to create binary spatial-temporal masks.

A. Other formulations of the problem

It's worth noting that many publications on video anomaly detection have provided different issue frameworks than the identify significant that is the focus of this survey. The above multi-view formulation ([8], [9], [10], [11]) has already been thoroughly addressed. Another potential approach for video anomaly detection is training-free video anomaly detection, which has been used in a number of publications ([12], [13]). The purpose of this formulation is either to identify changes in the test video or to recognize the most abnormal portions of the test video as indicators for anomalousness. There is no standard testing video provided in this formulation. The identification of discord in time series analysis is similar to the finding of the most unusual fragments of video testing. While these problem formulations are also useful, they are distinct from the classification of single-scene video irregularities in that they involve separate datasets and ground reality annotations.

The bulk of recent research relies on the RGB form of care when retrieving video features when it contributes to anomaly case detection. In this article, we use RGB and Flow machine learning techniques to extract video features and create a multi-model to overcome the issue of anomaly detection (ConvNets). The RGB stream looks for video frame anomalies, while the Flow stream looks for motion-related anomalies using dense optical flow. Furthermore, despite the fact that almost no extra calculation is added in inference, all of the frames in the video can be used in the proposed scheme, as opposed to[13]. The key explanation for this is that the overall amount of characteristics used in the anomaly model preparation is the same. One film is broken into multiple videos in total, with the video clip-level function integrating the attributes of all frames in a single clip.

The benefits of our two-stream anomaly event detection system are self-evident. In this article, instead of concentrating solely on RGB feature for MIL prototypes, we propose TAEDM, which can use both RGB and Flow modality data. In particular, the RGB modality information refers to the mathematical characteristics that underpin still images, such as colour, form, and the appearance of objects or individuals in the picture. As part of the event's mechanism, the information from the Flow system must be defined. As a result, TAEDM is capable of capturing complementary RGB got to witness from still frames as well as motion within images in a single comment. We have used two different and standard datasets for analysis of the implemented model: UCF-Crime[13] and ShanghaiTech[14], according to rigorous laboratory ablation studies. According to comprehensive laboratory ablation trials, our model's state-of-the-art success has been achieved.

The key contributions to this study are summarized as follows:

i) Detection and tracking in security videos is proposed using a novel two-stream-based anomalies approach. A comprehensive feature selection technique is also recommended for video-level functionality.

(ii) The proposed models are evaluated using the UCF-Crime[13] and Shanghai Tech[14] benchmark datasets, and the findings of both datasets indicate good performance relative to existing works.

B. Various Forms of Anomalies in the Video

It is our intention to present an exhaustive list about what we perceive as being the most prevalent video anomalies; nonetheless, a practical application may justifiably explain the statement of those other forms of anomalies in this regard.

i.Appearance-only Anomalies of appearance

These inconsistencies can be compared to the appearance of a suspected entity in a video. Examples are pedestrians on a footpath, or a wide vehicle on a no-entry path. Detecting these anomalies requires only a single video frame in a local region to be examined.

ii.Short-term Anomalies for Motion-Only

It is important to think of these irregularities in a scene as irregular object motion. Examples include a person working in a library or a car skidding sideways on the pavement. In general, these anomalies just need to be identified for a short amount of time to inspect the video in a local area. If they have this additional propriety, only appearance and abnormalities in the short run only may be pointed to as local abnormalities.

iii.Long-Term Abnormalities in Trajectory

It is possible to think of these variations in a scene as an irregular object trajectory. Examples involve people walking on a pavement in a zig-zag pattern, a vehicle flowing in and out of traffic, or loitering near international embassy buildings. The identification of trajectory deviations allows longer video sequences to be examined.

iv.Abnormalities in a Party

Community abnormalities can be thought of in a scene as an unexpected interaction of objects. A group of individuals walking in a line is an example (such as a marching band). The identification of group abnormalities includes the study of the interaction between two or more video regions.

v.Variations in the Time-of-Day

To any of the other types, this type of phenomenon is orthogonal. It's when they arise that what makes these events anomalous. While conceptually comparable to the location-dependent anomalies discussed previously, the "relevant contextual reference framework" for these anomalies is temporal rather than geographical, indicating that they are related to time rather than space. For example, folks who go to the movies in the early hours of the morning are a good example. When identifying these differences, it's usually only a matter of employing a different normalcy model during different points in time of day.

2. Related work

This segment would discuss the latest scientific findings in the identification of anomalies as well as the details concerning anomalies, ratings and two-way recognition.

A. The Identification of Irregularities

Anomaly activity recognition is one of the most difficult medical image analysis, and it has prompted a lot of study in recent decades[15-21], with the most widely used detection approaches falling into three categories.

The first set of techniques for identifying deviations is based on the premise that cases are uncommon and that unusual behavior is regarded as highly anomalous. Different mathematical models are used to encode standard patterns in these approaches, including Gaussian process models[22, 23], the interpersonal force model[24], Secret Markov-based models[25], the random field-based spatial-temporal Markov models[26, 27], the combination of dynamic models[21], and the treatment of deviations as outliers.

Sparse reconstruction [16, 28], which is used for habitual pattern learning, is the second category of approaches to anomaly detection. In particular, for normal actions, a dictionary is built using sparse representation, and exceptions are described as actions that have a high error rate. With the exciting breakthrough in deep learning, a number of researchers have recently developed deep neural networks for outlier detection, including video prediction learning[29] and abstraction attribute learning[30, 31].

The 3rd group is the hybrid approaches for modeling the normal behavior and anomalies[32, 33], using multi-instance learning in weak-checked movement models[33], for example Sultani et al. have created a MIL classifier[1] for the detection of abnormalities. In the meantime, a deep market head is employed to predict anomaly scores.

We reconstructed the model in this paper using a poor labelled supervised learning approach in order to take advantage of Sultani's work, which takes into account both regular and anomalous recordings.

B. Rating

Rank learning is a joint investigation that has been carried out with the task of machine learning and other research activities, including [34-38]. These approaches were designed to boost rather than individual ratings the relative ratings of the parts. With an intention to improvise the efficiency of recovery of the search engines, the author in [7] proposed a method called Rank-SVM. Whereas the author in [34] presented a detection mechanism so as to solve the problems related to multi-instance classification by incremental linear programming. The above mentioned methods are mainly used to resolve computational chemistry hydrogen abstraction problems.

Recently, researchers have suggested that computer-based vision-related applications would be successfully ranking in promising ways, for instance highlighting[35], individual identification[11], attribute training[36], graphics interchanging format (GIF) generation[37], face detection and testing[38], and metrical learning and image recovery[39]. Comprehensive annotations including both positive and negative specimens are needed for both of the above deep rating approaches. A regression issue in the method focusing on both normal and anomalous samples is unlikely to be identified as a correlation issue in the findings demonstrate the importance on both normal and anomalous samples by revisions in the topic of anomaly detection in this analysis. The proposed model employs MIL to train the anomaly model and classify abnormalities based on the effects of improperly controlled video segments. The suggested factor forces rankings include only two parts with either the highest positively and negatively anomaly scores, in contrast to the normal multi-instance learning (MIL).

C. Two-Stream Recognition of Behaviour

Recognition of video behaviour, which has recently received similar attention, has been carefully studied. In particular, it is superior to identify two stream-based steps[40-42]. In order to achieve the simultaneous extraction of the flow attribute and RGB using one class of action recognition techniques based on neuroscience, a two-stream neural network architecture has been proposed by the authors in study [40-42]. By integrating the results of these two ways, the final and optimal result of the event rating can be obtained. Wang et al. have developed a new Temporal Segment Network (TSN) to mimic the long-term visual structure in order to improve the performance of action detection even further. In addition, a large number of extension models[40] have been developed, with the goal of investigating convolutionary fusion[42] and residual connections[43, 44].

According to the model, there were still some links between the Flow streams and RGB[43]. By utilising dense networking interactions, the STDDCN[44] incorporates multi-scale information into the residual connections as well as a new knowledge exposing module. Different video tasks are often employed with two-stream approaches, including action recognition[40-43, 45, 46]. However, two-stream technology is seldom used to identify anomalies. We developed a modern paradigm for anomaly detection by two-stream events, based on the two-stream computational architectures, which use the complementary knowledge of the underlying RGB and Flow modalities. Anomaly detection shall determine the type of behavior (normal or abnormal) in respect of action recognition and determine the period of time for an exception. This makes it harder to discuss this subject than others.

3. Datasets used in the Literature

Two famous benchmark datasets widely used for anomaly detection tasks are UCF-Crime[1] and ShanghaiTech[9]. This paper shows how superior our built system can be when anomalous events are detected using both datasets. If the following steps are taken, the model presented in this paper can also be utilised to support other standard datasets: (1) The optical flow and RGB image are extracted from each available video in the dataset; (2) feature extraction is implemented to extract there features from each video; and (3) Flow streams and RGB are fed into their respective subnet branches in the training model proposed and predicted test results are obtained. A brief summary of the two benchmark problems will be given before the specifics are added.

A. UFC-Crime. [1] is a large-scale dataset comprising 13 phenomenon instances in total and 1900 recordings of real-word surveillance. In 950 of them, there are visible anomalies, while the rest are perceived as standard videos. The test collection has 1611 videos, (801 default videos, 809 anomalous videos) and 280 videos for the test set (150 normal, 150 anomalous videos).

Fig. 1. UCF-Crime Dataset

B. ShanghaiTech. [9] is a medium-scale dataset of 437 images, including a total of 130 abnormal incidents out of 13 scenes. The dataset will not be specifically used for anomaly case identification since there is no irregular video in the training collection. Zhong et al.[47] reconstructed the dataset to fix this issue by arbitrarily picking irregular test videos and placing them into the training data and vice versa. At the same moment, 13 scenes are used in both the preparation and evaluation datasets. The new dataset organization has made it relevant for the role of anomaly event detection. Therefore, before carrying out the tests, we conduct the same operation as in [48].

Fig. 2. ShanghaiTech Campus

4. Implementation Details

System Architecture

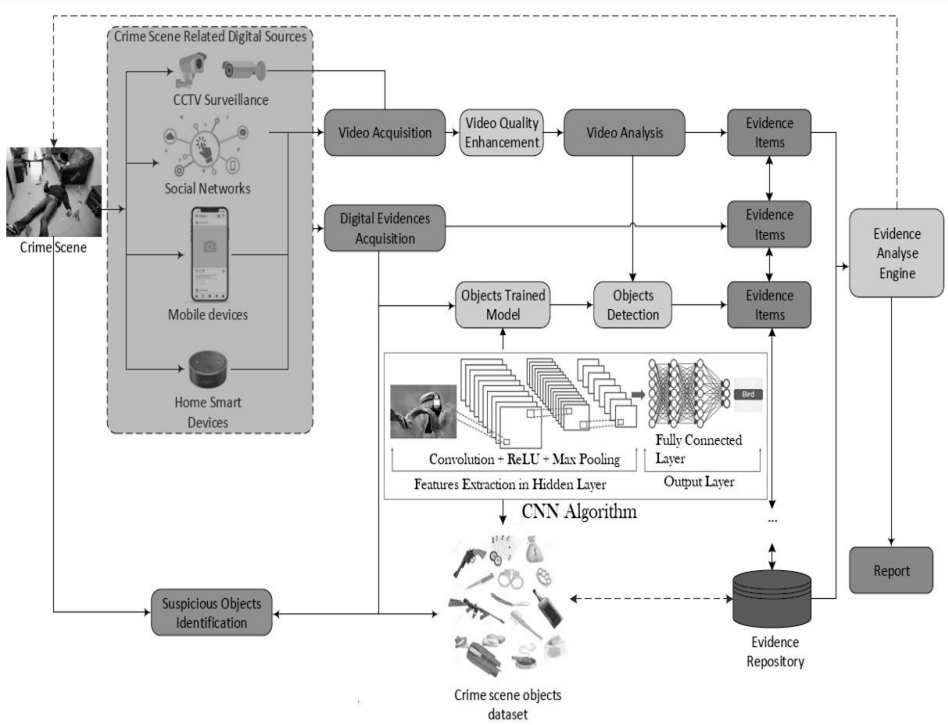


Fig.s 3. System Architecture

The proposed framework architecture is depicted in the following figure (Fig. 1). The first digital source connected to the crime scene is discovered, and information is taken from that source. These digital sources include mobile devices, smart devices, and surveillance cameras from closed-circuit televisions. The gathering of evidence from various sources then takes place through the use of either video or digital recordings. After that, an algorithm for enhancement is applied, which improves the overall video quality. Following that, evidence is gathered, and a CNN algorithm is utilised to identify the objects.

Convolutional Neural Networks

CNN is a kind of deep, forward-looking adaptive neural network used for accurate output in object detection algorithms such as the classification and identification of images [5]. CNNs are much richer in layers in contrast to the conventional neural network. It has non-linear activation weights, distortions and effects. Including height, width and depth, the neurons are volumetrically organised. Fig. 2 shows the architecture of CNN, it has a coalescent, uniform and fully linked layer. The collective layer and pooling layer are normally modified, and each filter's depth is increased from left to right and the performance size (height and width) is decreased. The fully-connected layer is the final step in the neural networking process.

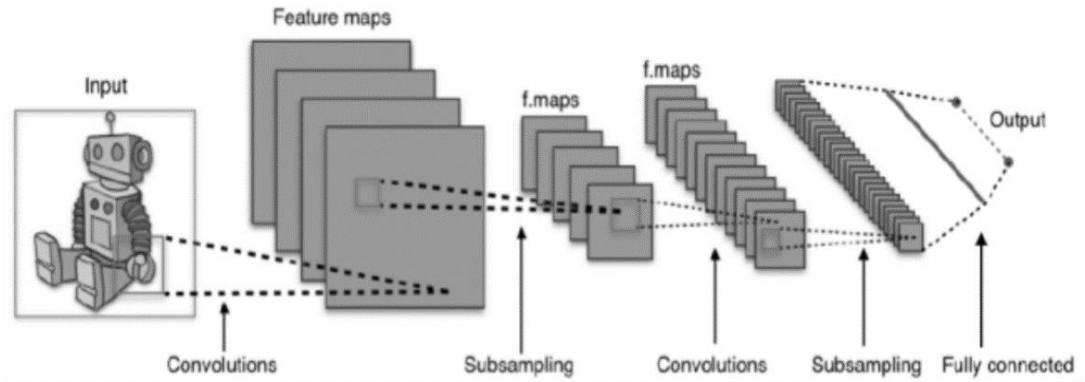


Figure 4. CNN Architecture

An image composed of individual pixels serves as the input. The three-dimensional model has the dimensions [50 x 50 x 3], which include the width, the height, and the depth (RGB channels). [13] The convolutional layer is responsible for measuring the output of the neurons that are connected to the inputs local regions. The layer parameters are made up of a sequence of learning filters, also known as kernels. These learning filters cover the width and height of the input volume, and they run vertically through the layer. This

creates a two-dimensional filter affine transformation to enable filters engagement when certain characteristics are recognized at a given point in the input. This map enables filter engagement because certain characteristics are discovered at a certain position. To activate the element in the most advantageous manner, the layer function known as Removable Linear Unit (ReLU) is utilised. ReLU is set in (1),

$$f(x)=max(0,x) \quad \text{.....(1)}$$

This equation does not apply to regression coefficients and increases for positive values linearly. The size of the volume is not changed. The layer of pooling leads to maximum activation of an area. This shows the dimensions of space such as size and shape. The output layer of the neural network is a fully connected layer. For the probability distribution function over the number of voltage groups in this layer, a softmax activation is used.

Mathematical Formulation

System S is represented as

S = {ID, P, F, T, CNN, M}

1. Input Dataset

ID = {i1, i2, i3…in}

Where ID is the input image dataset and i1, i2…in are the number of images.

2. Preprocessing

$\Phi = \{\Phi_1, \Phi_2, \Phi_3\}$

Φ is preprocessing and Φ_1, Φ_2 and Φ_3 are the steps to be carried out during preprocessing.

Φ_1 represents reading the input dataset

Φ_2 represents the improvement in the input image and

Φ_3 represents the cleaning of the image.

3. Feature Extraction

$F= \{f_1, f_2, f_3…f_n\}$

Here, F represents the set of extracted features from the input images. f_1, f_2, f_3… f_n represents the features such as color, thickness, border, etc., extracted from the input image.

4. Training and Testing file generation

$T = \{T_1, T_2\}$

Here T represents the set of Testing and Training file and T_1 is Training file and T_2 is Testing file both the files contains various extracted features values while training file contains class of each image as 0 or 1.

5. Convolutional Neural Network (CNN).

CNN = {C, RL, PO, FC, LS}

Where CNN is algorithm consisting of various stages as

C is convolutional operation

RL be the ReLU activation layer

PO be the Pooling layer

FC be the Full Connection layer and

LS be the Loss function.

6. Object Detection

O= {0, 1}

O is the set of Class having value 0 or 1

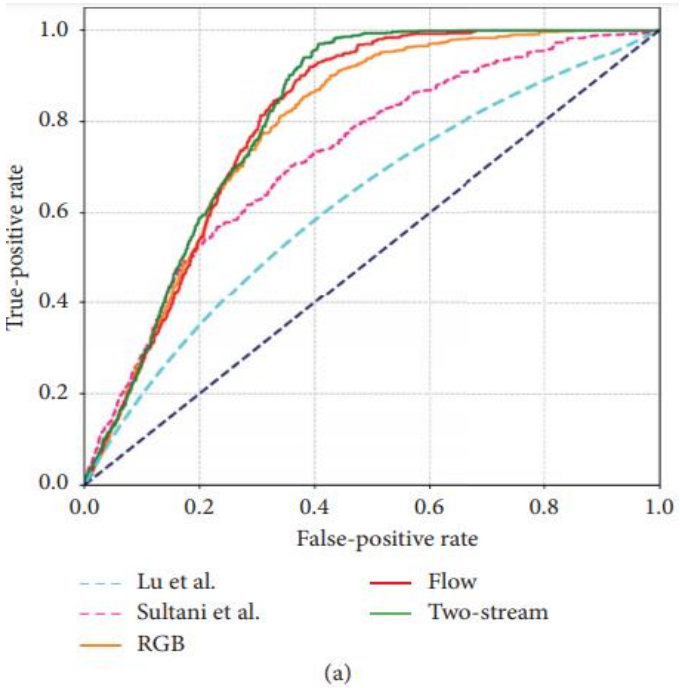
0 be the absent of object and 1 be the present of object

I. Comparitive Evaluation of the Various Model

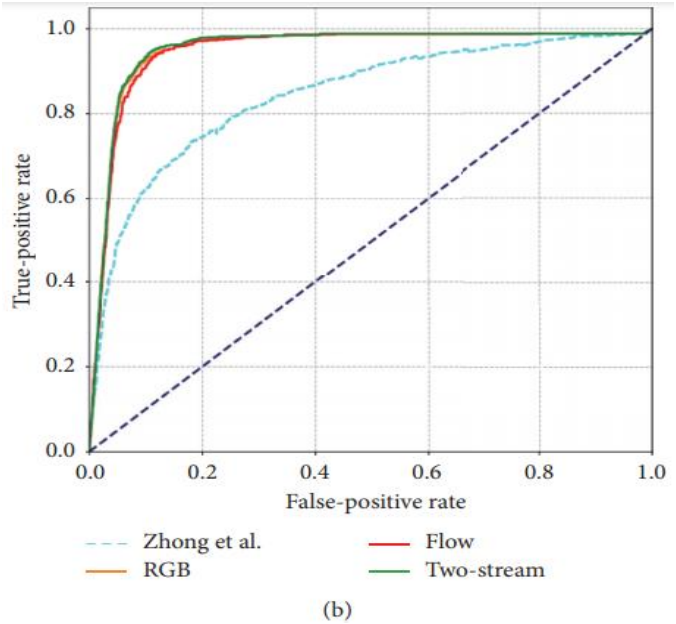
UCFCrime Dataset [1] helps to equate the current best practices with ShanghaiTech Dataset's [9] unique expertise to determine whether the suggested methodology is performing in line with recent and current state-of-the-the-art practices. You can see a comparison of the three ROC curves in Figure 3. In this figure 3, the events shown in red, blue, in yellow, and green are called a problem; they are a function of a separate RGB model being applied along with a Stream network, while yellow events are a combination of the two.

Figure 5 shows that RGB, Flow, and the Two models outperform the other models, indicating that the dense feature extraction is correct. Two gives two more positive outcomes than the RGB and Flow and positive cell numbers support the efficiency model's accuracy.

The findings are shown in Tables 1 and 2, which use the AUC models for UCF-Crime [1] and ShanghaiTech [9]. Figure 3 continues to show findings of various method are consistent with those of previous models, meaning the model is effective in generating results.



(a)



(b)

Fig. 5. ROC curves for various models implemented on UCF-Crime [1] and ShanghaiTech [9] datasets. (a) and (b) shows respective curves of ROC for UCF-Crime [1] and ShanghaiTech [9].

Table 1: Analytical comparison of various techniques on UCF-Crime Dataset

Method	Accuracy (%)
Sultani et al. [1] with w/o constraints	74.57
Sultani et al. [1] with w constraints	76.71
Hasan et al. [10]	51.6
Lu et al. [29]	66.51
RGB [48]	77.51
Flow [48]	79.39
Two [48]	80.62

Table 2: Analytical comparison of various techniques on ShanghaiTech [9] Dataset

Method	Accuracy (%)
Zhong et al. [47]	86.78
RGB [48]	92.90
Flow [48]	93.89
Two [48]	95.43

The proposed model of a convolutional neural network was applied to the UCF-crime Dataset that we developed. On the basis of the following performance parameters, we conduct an analysis to determine how well the suggested model performs:

1. Precision - The number of correctly predicted positive cases as a percentage of the total number of true positive that were correctly predicted makes up the denominator.

$$Precision = \frac{[TP]}{[TP + FP]}$$

2. Recall / Sensitivity - The proportion of all positive occurrences that are counted toward the overall amount of positive instances that can be considered positive. This amount serves as the denominator for the overall count of positive occurrences.

$$Recall/Sensitivity = \frac{[TP]}{[TP + FN]}$$

3. $F1$ Score - The sum total of all of the component sums is equal to the harmonic mean of recall and precision. The final F1 score is determined by adding up all of the individual factors. In most cases, a model will have a high F1 score if it accurately predicts that the results will be positive.

$$F1\ Score = \frac{2 * [Recall * Precision]}{[Recall + Precision]}$$

4. Accuracy - Classification, the term "classification accuracy" refers to the degree of accuracy that is typically taken into consideration. To determine it, divide the total number of accurate predictions by the total number of samples fed into the model.

$$Accuracy = \frac{[TP + TN]}{[TP + TN + FP + FN]}$$

We also built some other traditional machine learning methods such as LR, NB, and SVM so that we could conduct a comparative analysis. In addition, we came up with the identical settings for the

algorithms that were mentioned. The conclusion that can be drawn from all of the data is presented in Table 3.

Table 3. Performance Parameters for Proposed CNN Model and Comparison with LR, NB and SVM

Algorithms	Precision	Recall	F-Measure	Accuracy
LR	80.96	85.96	83.38	83.32
NB	83.01	82.6	84.81	83.6
SVM	81.94	88.92	85.28	85.06
CNN	90.34	89.22	89.78	90.13

Figure 6. Shows the comparative analysis of the above mentioned algorithm with proposed CNN model based on Performance Parameters.

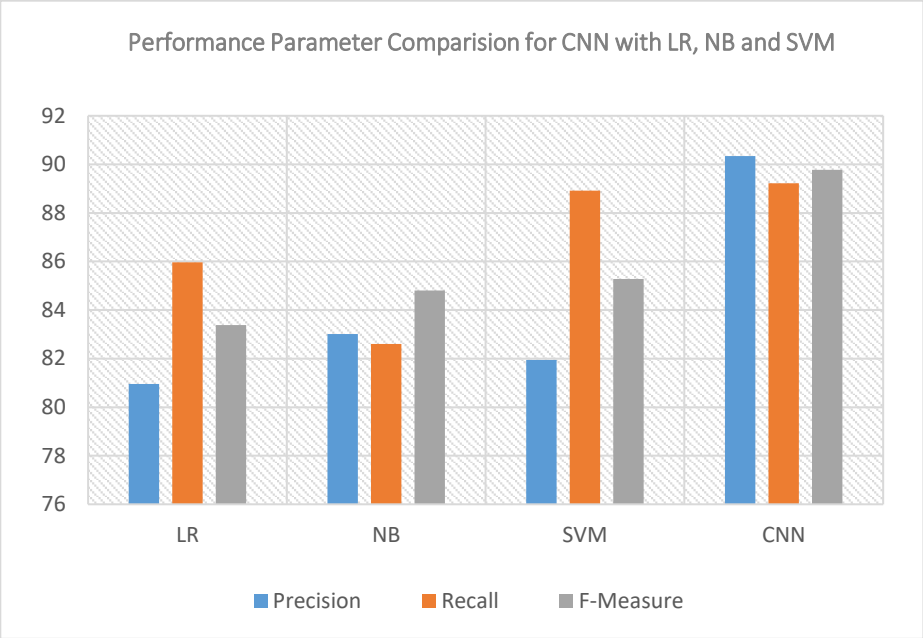


Fig. 6. Comparison of Performance Parameters for CNN, LR, NB and SVM

The comparison of the accuracy of all the algorithms is shown in figure 7.

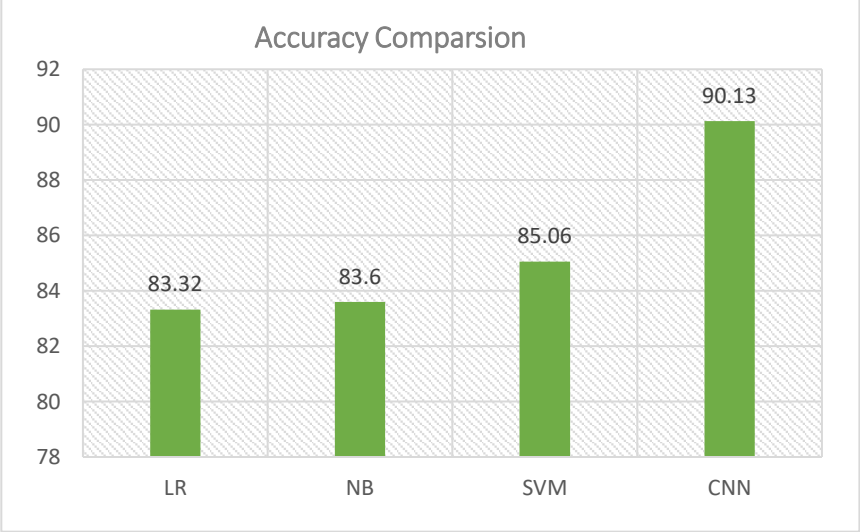


Fig. 7. Comparison of Performance Parameters for CNN, LR, NB and SVM

From the above resulted represented in figure 4 and figure 5, we can observe that CNN performs better in terms of precision, recall and F-Score as compared to the other standard algorithms. We can also observe that the accuracy of CNN comes to be 90.13% which is considerably better than other algorithms like SVM with accuracy of 85.06%, NB with 83.6% and LR with accuracy of 83.32%. Thus we can say that our proposed CNN model outperforms other state of art algorithms in terms of accuracy for video anomalys detection.

5. Conclusion

This study is based on current discoveries in video anomaly detection, and its goal is to present an overview of the fundamentals as well as the most recent advancements in this field. These new methods help in anticipating frames, as well as in assigning frame

sort, and a number of additional features, such as image reconstruction or classification, have been added to the present four phases of detecting anomalies. These classifications are reflective of a variety of techniques, and they also encourage employees and researchers to think creatively and beyond the box. This section of the paper included a variety of different datasets, including the display resolution and video/example information to illustrate how these commonly-used datasets differ. Additionally, it reported a comprehensive exploration of two parallel study datasets from UCF and ShanghaiTech, which showed that the utilization of the proposed method has a substantial evidence of ablation. When given sufficient time, the datasets have a tendency to grow to life-like levels of complexity. This is especially true as they continue to collect more information regarding more practical matters about everyday life. There is always going to be some sort of manual annotation work for all of these films that needs to be done by the uploader. In addition to motion, the study of unsupervised and poorly supervised methods ought to be given priority in the realm of new research domains, on account of the fact that earlier studies had produced positive results and big SC datasets as a result of their application. Experimentation with earlier published datasets, researching the large-scale, experimental construction of deep learning, and the significance of a single video frame of film should all be included in the research that is conducted on a smaller scale. Investigation, advancement of single video construction strategies, design rules, research results, and findings, and popularity prediction are all aided by experiments, which are preferable to the study of recently conducted and smaller-scale datasets. Experiments also contribute to the investigation process. Instead than relying on previous research and studies, scale up based on the results of experiments.

References

[1] Sultani W, Chen C, Shah M (2018) Real-world anomaly detection in surveillance videos. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 6479–6488

[2] V. Saligrama, J. Konrad, and P.-m. Jodoin, “Video Anomaly Identification,” IEEE Signal Processing Magazine, vol. 27, no. 5, pp. 18–33, Sep. 2010.

[3] Weixin Li, V. Mahadevan, and N. Vasconcelos, “Anomaly Detection and Localization in Crowded Scenes,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 1, pp. 18–32, Jan. 2014.

[4] C. Lu, J. Shi, and J. Jia, “Abnormal Event Detection at 150 FPS in MATLAB,” in IEEE International Conference on Computer Vision (ICCV). Sydney, Australia: IEEE, Dec. 2013, pp. 2720–2727.

[5] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, “Robust RealTime Unusual Event Detection using Multiple Fixed-Location Monitors,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 3, pp. 555–560, Mar. 2008.

[6] Unusual crowd activity dataset of University of Minnesota, available from http://mha.cs.umn.edu/proj events.shtml.

[7] B. Ramachandra and M. Jones, “Street Scene: A new dataset and evaluation protocol for video anomaly detection,” in Winter Conference on Applications of Computer Vision (WACV), 2020.

[8] R. Morais, V. Le, T. Tran, B. Saha, M. Mansour, and S. Venkatesh, “Learning regularity in skeleton trajectories for anomaly detection in videos,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 11 996–12 004.

[9] W. Liu, W. Luo, D. Lian, and S. Gao, “Future frame prediction for anomaly detection–a new baseline,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 6536– 6545.

[10] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, “Learning Temporal Regularity in Video Sequences,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 733–742.

[11] R. T. Ionescu, F. S. Khan, M.-I. Georgescu, and L. Shao, “Objectcentric autoencoders and dummy anomalies for abnormal event detection in video,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 7842–7851.

[12] A. Del Giorno, J. A. Bagnell, and M. Hebert, “A Discriminative Framework for Anomaly Detection in Large Videos,” in European Conference on Computer Vision (ECCV). Springer International Publishing, 2016, vol. 9909, pp. 334–349.

[13] B. Zhao, L. Fei-Fei, and E. P. Xing, “Online detection of unusual events in videos via dynamic sparse coding,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, CO, USA: IEEE, Jun. 2011, pp. 3313–3320.

[14] W. Luo, W. Liu, and S. Gao, “A revisit of sparse coding based anomaly detection in stacked RNN framework,” in Proceedings of the International Conference on Computer Vision, Venice, Italy, October 2017.

[15] L. Kratz and K. Nishino, “Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models,” in Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, June 2009.

[16] B. Zhao, L. Feifei, and E. P. Xing, “Online detection of unusual events in videos via dynamic sparse coding,” in Proceedings of the Computer Vision and Pattern Recognition, Providence, RI, USA, June 2011.

[17] S. Wu, B. E. Moore, and M. Shah, “Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes,” in Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, June 2010.

[18] N. Li, H. Guo, D. Xu, and X. Wu, “Multi-scale analysis of contextual information within spatio-temporal video volumes for anomaly detection,” in Proceedings of the International Conference on Image Processing, Paris, France, October 2014.

[19] B. Antic and B. Ommer, “Video parsing for abnormality detection,” in Proceedings of the International Conference on Computer Vision, Barcelona, Spain, November 2011.

[20] H. Mobahi, R. Collobert, and J. Weston, “Deep learning from temporal coherence in video,” in Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada, 2009.

[21] W. Li, V. Mahadevan, V. NJIToPA, and M. Intelligence, “Anomaly detection and localization in crowded scenes,”

IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, pp. 18–32, 2014.

[22] N. Li, X. Wu, H. Guo et al., "Anomaly detection in video surveillance via Gaussian process," International Journal of Pattern Recognition and Artificial Intelligence, vol. 29, no. 6, Article ID 1555011, 2015.

[23] K. Cheng, Y. Chen, and W. Fang, "Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression," in Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, June 2015.

[24] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, June 2009.

[25] T. M. Hospedales, S. Gong, and T. Xiang, "A Markov clustering topic model for mining behaviour in video," in Proceedings of the International Conference on Computer Vision, Kyoto, Japan, October 2009.

[26] C. Wang, Z. Chen, K. Shang, and H. Wu, "Label-removed generative adversarial networks incorporating with KMeans," Neurocomputing, vol. 361, pp. 126–136, 2019.

[27] T. Meng, K. Wolter, H. Wu, Q. Wang, and M. Computing, "A secure and cost-efficient offloading policy for Mobile Cloud Computing against timing attacks," Pervasive and Mobile Computing, vol. 45, pp. 4–18, 2018.

[28] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB," in Proceedings of the International Conference on Computer Vision, Sydney, Australia, December 2013.

[29] W. Liu, W. Luo, D. Lian, and S. Gao, "Future frame prediction for anomaly detection—a new baseline," 2018, https://arxiv.org/abs/1712.09867.

[30] Y. S. Chong and Y. H. Tay, "Abnormal event detection in videos using spatiotemporal autoencoder," 2017, https://arxiv.org/abs/1701.01546.

[31] W. Luo, W. Liu, and S. Gao, "Remembering history with convolutional LSTM for anomaly detection," in Proceedings of the International Conference on Multimedia and Expo, Hong Kong, China, July 2017.

[32] K. P. Adhiya, S. R. Kolhe, and S. S. Patil, "Tracking and identification of suspicious and abnormal behaviors using supervised machine learning technique," in Proceedings of the International Conference on Advances in Computing, Communication and Control, Mumbai India, January 2009.

[33] C. He, J. Shao, and J. Sun, "An anomaly-introduced learning method for abnormal event detection," Multimedia Tools and Applications, vol. 77, no. 22, pp. 29573–29588, 2018.

[34] C. Bergeron, J. Zaretzki, C. M. Breneman, and K. Bennett, "Multiple instance ranking," in Proceedings of the 25 th International Conference on Machine Learning, Helsinki, Finland, 2008.

[35] T. Yao, T. Mei, and Y. Rui, "Highlight detection with pairwise deep ranking for first-person video summarization," in Proceedings of the 016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 2016.

[36] J. Wang, Y. Song, T. Leung et al., "Learning fine-grained image similarity with deep ranking," 2014, https://arxiv.org/abs/1404.4661.

[37] M. Gygli, Y. Song, and L. Cao, "Video2GIF: automatic generation of animated GIFs from video," 2016, https://arxiv.org/abs/1605.04850.

[38] S. Sankaranarayanan, A. Alavi, and R Chellappa, Triplet Similarity Embedding for Face Verification, https://arxiv.org/abs/1602.03418, 2016.

[39] A. Gordo, J. Almazan, J. Revaud, and D. Larlus, "Deep image retrieval: learning global representations for image search," 2016, https://arxiv.org/abs/1604.01325.

[40] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," 2014, https://arxiv.org/abs/1406.2199.

[41] L. Wang, Y. Xiong, Z. Wang et al., "Temporal segment networks: towards good practices for deep action recognition," 2016, https://arxiv.org/abs/1608.00859.14 Security and Communication Networks

[42] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," 2016, https://arxiv.org/abs/1604.06573.

[43] C. Feichtenhofer, A. Pinz, and R. P. Wildes, "Spatiotemporal residual networks for video action recognition," 2016, https://arxiv.org/abs/1611.02155.

[44] H. Kwon, Y. Kim, J. S. Lee, and M. Cho, "First person action recognition via two-stream ConvNet with long-term fusion pooling," Pattern Recognition Letters, vol. 112, pp. 161–167, 2018.

[45] L. Sevillalara, Y. Liao, F. Guney, V. Jampani, A. Geiger, and M. J. Black, "On the integration of optical flow and action recognition," 2018, https://arxiv.org/abs/1712.08416.

[46] Z. Qiu, T. Yao, and T. Mei, "Learning spatio-temporal representation with pseudo-3D residual networks," 2017, https://arxiv.org/abs/1711.10305.

[47] J. Zhong, N. Li, W. Kong, S. Liu, T. H. Li, and G. Li, "Graph convolutional label noise cleaner: train a plug-and-play action classifier for anomaly detection," 2019, https://arxiv.org/abs/1903.07256.

[48] Wangli Hao, Ruixian Zhang, Shancang Li, Junyu Li, Fuzhong Li, Shanshan Zhao, Wuping Zhang, "Anomaly Event Detection in Security Surveillance Using Two-Stream Based Model", Security and Communication Networks, vol. 2020, Article ID 8876056, 15 pages, 2020. https://doi.org/10.1155/2020/8876056