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

Advanced Deep Learning Model for Anomaly Detection Based Video Surveillance System

Jyoti Kukade[1], Prashant Panse[2]

Submitted: 24/09/2023 **Revised**: 26/10/2023 **Accepted**: 13/11/2023

Abstract: The information that is acquired from the beginning scenes of a movie is used to compile a dictionary of common occurrences. Although this method may perform well in terms of computing, it is not particularly effective at accurately identifying outliers in the data. Research is also being done to investigate the viability of using poorly supervised multi-instance learning (MIL) as a potential method for anomaly finding. The movies are segmented into easily digestible portions in order to make the training component of such procedures easier to carry out. It has been shown that regular CNNs are very effective at resolving problems of this nature. The new model, which requires a significant amount of computer resources as well as a significant amount of time for training, should be better suited for vocations that need video processing. The conditions under which a video abnormality manifests themselves are frequently illuminating. A store opening or a performance are both good examples of situations that can generate a crowd, yet it is unusual for people to keep their distance when there is the potential for the spread of an infectious disease. The great majority of algorithms designed to detect video anomalies are able to precisely localise the anomalies they find in both time and place.

Keywords: *Deep Learning Model, Anomaly Detection, Video Scrutiny System, CNN, EADN Network*

1. Introduction

Recently, security cameras have been installed in a variety of locations across the globe in an effort to increase public safety. It is reasonable to anticipate that this pattern will carry on. Because of the constraints of manual monitoring, law enforcement agencies are not particularly good at detecting or avoiding anomalous behaviour. This makes it difficult for them to do their jobs effectively. This is as a result of the constraints. With the use of a sophisticated computer vision system, abnormal behaviour can be identified and addressed. This system must be able to differentiate between normal and abnormal circumstances on its own, without requiring assistance from a human being. By utilising this kind of automated technology, the quantity of day-to-day human labour required to maintain manual observation can be cut down significantly. This makes it easier to keep an eye on everything that's going on.

An intelligent video anomaly detection system can recognise a broad variety of moving objects for use in video surveillance, or it can spot odd events such as fights, stampedes, accidents, or even homeless people with minimal training. This system can be used for either purpose. Additionally, it has the ability to identify entities whose behaviour is markedly distinct from the norm. In the context of video surveillance, for example, an intelligent video anomaly detection system might be able to recognise a large number of moving objects even without any prior instruction. This category of actions and things includes things like spotting specific events like fights, stampedes, automobile accidents, and vagrants.

The ability to recognise many moving objects with only less amount of training is a further useful feature for video surveillance systems. [1-2]

The increased use of surveillance cameras can be attributed to the growing need of maintaining city safety. These recordings are now readily available to a large section of the general audience. These recordings are utilised to monitor public activity and serve as a deterrent against undesirable occurrences such as accidents involving vehicles and illegal conduct. In the past, in order to detect irregularities on film, it required knowledgeable people to maintain a close check on it at all times. It is going to turn into a chore that is not only challenging but also time-consuming. Because an efficient detection strategy can result in significant cost savings, research into autonomous video anomaly detection is absolutely necessary from a pragmatic standpoint. Due to the fact that a dependable detection system has the potential to save considerable time and effort, this is the case. [3-6]

When attempting to define a video anomaly, it is sometimes helpful to consider the circumstances in which it was found. One good illustration of this would be the gathering of a large number of people in a single location, such as during a public event or in a crowded retail establishment. On the other hand, the unusual inclination of people to keep their distance when precautions are taken to limit the transmission of a virus is something to take note of. The majority of the algorithms that have been developed to

[1]*Student, Medi-Caps University, Indore, jyoti.kukade@medicaps.ac.in*
[2]*Professor, Medi-Caps University, Indore, prashant.panse@medicaps.ac.in*
* *Corresponding Author Email: jyoti.kukade@medicaps.ac.in*

detect outliers in video data are able to pinpoint both the temporal and geographical positions of these outliers. This capability can be found built into each one of these systems. To be more specific, while the method of detection is focused on identifying the video fragments that have abnormalities throughout all films, the localization method determines which frame is abnormal and explains which component of this bad frame is determined to be abnormal. This is done in contrast to the detection method, which is focused on identifying the video fragments that have abnormalities throughout all films. To restate, the approach identifies instances of abnormal video across a number of different films. Current research has the potential to overcome both of these challenges, and models that are built on deep learning have the capability of addressing both of these challenges in a complete manner. [7-9]

1.1 Deep Learning Based Techniques for Detecting Anomalies

The efficacy of deep learning approaches, including object classification, object recognition, activity recognition, and the detection of anomalies in video surveillance. Reconstruction error, future frame prediction, classifiers, and scoring are the four broad categories into which the many proposed solutions to this problem fall. These types of divisions had previously been described in the preceding paragraph. [10-12]

1.2 Complexities and Challenges in Anomaly Detection

Anomaly detection is unlike the vast majority of analytical and learning difficulties and undertakings in terms of problem complexity due to its unique qualities. The challenges and unanswered identification issues that come with complex anomalous data are compiled here. [13-15]

1.3 Unknown Ness

Anomalies sometimes involve several unknowns, such as data structures, distributions, and examples exhibiting unexpected abrupt behaviour. Terrorist attacks, fraudulent scams, and network hacks of unprecedented scale are only some examples of events that go unreported until after they have already taken place. [16-18]

1.4 Heterogeneous anomaly class categories

Since anomalies are not typical, it is possible for one category of anomalies to display characteristics that are radically different from those of another category of anomalies. For instance, in the context of video surveillance, the aberrant occurrences of robberies, road accidents, and burglaries appear very differently to the human eye.

1.5 Deep video anomaly detection, as detailed in Opportunities

We talk about deep video anomaly detection in the context of intelligent transportation and then point out some of the other places it could be useful, such as digital twins, smart homes, public health, and education. These are the topics we'll be discussing in depth.

1.6 Transportation Innovation

Manufacturing commodities, maintaining daily living, and growing the economy all require some form of transportation in order to function well. As a result of improvements in the transportation sector, the general population today has access to means of transportation that are not only quick but also pleasant, secure, and safe. On the other hand, the increasing number of automobiles and vehicles on the nation's roadways may have something to do with the rising human population. This is due to the fact that people demand more options for public transportation. Congestion and an increase in the number of accidents on the road could result from this. Because of this, what is now commonly referred to as the intelligent transportation system was eventually developed (ITS). The implementation of intelligent transportation systems (ITS) proven to be an effective remedy for the traffic issues that were brought about by the current pace of economic development. [19-20]

1.7 Online Learning

The traditional manner of offline teaching and learning has gradually migrated to online platforms as a result of advancements in information and communications technology over the course of the past decade. The COVID-19 outbreak pushed the process along at a faster pace. Because of the epidemic, traditional classroom learning is gradually becoming obsolete and being replaced by online education, which is rapidly becoming the major means of communicating new material. The demand for online tests has increased their frequency of administration. It is necessary to have the capacity to identify cheating and efficiently proctor online examinations in order to ensure that all test takers are treated in a fair manner. Regrettably, it is highly probable that the existing procedures for detecting cheating won't be as efficient as they formerly were in eradicating all cases of exam cheating. This is due to the fact that cheating is becoming increasingly sophisticated. [21-22]

1.8 Smart Home

Installing a home surveillance camera system is one of the most common things people do to make themselves feel safer in their own residence. Home automation systems that come equipped with built-in surveillance cameras as a standard feature are typically regarded as having a high level of security. Viewers can stay informed about current events by watching the video on their own computers or mobile devices, regardless of whether they are at home or on the move. Because it is ineffective to constantly gaze at the screen, it is essential for the system to be able to recognise unusual behaviours and send a warning signal as soon as a problem is identified. [23-25]

2. Objectives of the Study

- Researching Problem Complexities and Challenges in Anomaly Detection
- To research deep learning-based anomaly detection methods

3. Research Method

What has been suggested (ADSV) and how its fundamental parts should be organised are explained. A graphical representation of proposed Anomaly Detection in Surveillance Video (ADSV) method. The first step in isolating individual keyframes from the rest of the shot is to formulate a plan for pinpointing the exact edges of each individual frame. Frames are sent to the model to guarantee it continues to receive spatial and temporal data from the intermediate layer. The goal is to keep the model's learning going as smoothly as possible. The next step involves analysing the spatial and temporal characteristics of each instance of anomalous behaviour across a set of photos by employing the LSTM cells and the Attention Network. We'll use the LSTM cells and Attention Network to do this. visual depictions of the peculiar actions observed. [26-28] The LWCNN was trained using a set of still images that were shown to it in the correct chronological order. Because of this, it was able to acquire the ability to recognise motion and action. Applying the recommended trained ADSV model to the video recordings will help find instances of potentially aberrant behaviour. Proposed corporate autonomous data network architecture is depicted graphically in Figure 1. Since anomaly detection and categorization is a more extensive task, it requires the successful completion of multiple simpler tasks before it can be accomplished. Segmenting images, extracting features, learning sequences, and classifying outliers are all examples of such tasks. In the initial step of the process, an algorithm that can identify the transitions between images is used to segment the crucial frames. In order to retrieve spatiotemporal information from the intermediate layer of the network, lightweight convolutional neural network (LWCNN) models use segmented frames as their input. Ultimately, this helps with image identification. Then, LSTM cells are used to extract typical spatiotemporal characteristics from a series of frames for each individual event. This phase follows on from the one that came before it. Overreactions have decreased dramatically as a result of this. To detect movement and action, the intended LWCN N used a sequence of frames in a time-based order. After the model has been trained from scratch, the category cross entropy loss function is used to fine-tune the model's settings. This step follows the completion of the model's initial training. [29] The loss function can be formally defined by doing the following operations:

$$Loss = -\sum_{i=1}^{N} y_i \cdot \log \hat{y}_i$$

The actual class label is represented by yi, and the estimated likelihood produced by the model for each outlier is also denoted using yi. The total number of outliers is given by N. After all of this has been set up, the model will compare the score to the key frames of the video that has been provided in order to derive the relevant anomaly score. The reality of the anomaly can be determined from the response given to this question. It is not possible to place enough emphasis on the fact that the key frames are a precise match for the film that was submitted. This is an undeniable fact that cannot be refuted. As a consequence of this, one can assume, with a high degree of confidence, that the out-of-the-ordinary incident that was spotted in the critical frames was also present in the original source material. [30-32]

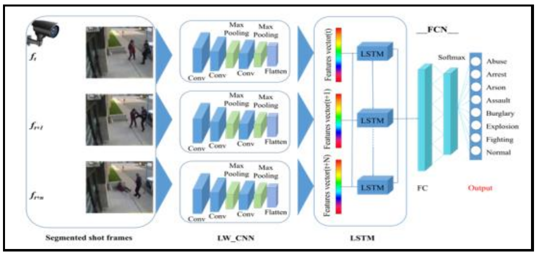

Fig 1. Anomaly identification in surveillance footage using the Proposed EADN framework.

3.1 Method proposed by LWCNN

CNNs are motivated by the anatomy of the visual cortex in animals, which serves as a source of inspiration for the principles upon which CNNs are founded. The synaptic connections between neurons in the convolution layers (CL) of a network's architecture are organised in a manner that is analogous to the visual pathways in the brains of animal species. The region of the input frame that is utilised by each cortical neuron in order to carry out the required processing is referred to as the "receiving field," and the term "receiving field" is used to characterise this region. This location does not come without its restrictions. During the training phase, CL applies filters with values that are learned, which enables it to learn feature representations and maintain spatial correlations between input frames when performing video analysis. While CL is learning feature representations, it is possible for the spatial relationships that exist between the input frames to be maintained. As a consequence of this, CL is able to maintain the spatial correlations that exist between the input frames during the entire process of conducting video analysis. The LWCNN architecture that has been presented has a total of five Time Distributed 2D CL layers in addition to two Time Distributed 2D max pooling layers. Table 1 has a complete listing of all of the following: [33]

Table 1: LWCNN approach descriptions that are applied with the suggested ADSV method.

LWCNN layer	**No. of channels**	**Kernel (h × w)**	**Padding**	**Stride (h × w)**	**Output**	**Parameters**
TimeDistributed-2D *CL*1 (ReLU)	64	3 × 3	Same	2 × 2	5,112 × 112 × 64	1792
TimeDistributed-2D *CL*2 (ReLU)	64	3 × 3	Same	2 × 2	5, 56 × 56 × 64	36928
TimeDistributed Max-Pooling-2D1	1	2 × 2	–	2 × 2	5, 28 × 28 × 64	0
TimeDistributed-2D *CL*3 (ReLU)	128	3 × 3	Same	2 × 2	5, 14 × 14 × 128	73856
TimeDistributed Max-Pooling-2D2	1	2 × 2	–	2 × 2	5, 7 × 7 × 128	0
TimeDistributed Flatten-1	–	–	–	–	5, 6272	0

3.2 Data Analysis

CUHKAvenue and UCF-Crime are two datasets that are available to the general public and are utilised in the process of evaluating the ADSV strategy that we have proposed. The event frames shown in Figure 2 come from a variety of datasets and have been annotated to indicate whether or not they are typical or unusual. The graphic contains some annotations for your convenience.

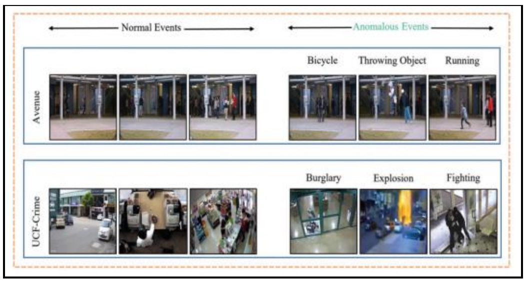

Fig 2: Examples from the dataset for anomaly detection

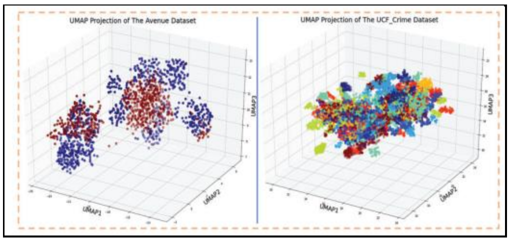

Fig 3: Avenue and UCF-Crime dataset's 3D-UMAP projection

Datasets CUHK-Avenue was captured with a 640x360 pixel resolution fixed video camera at Delhi University in New Delhi. The camera captured footage of the happenings on the street below. At Delhi University in New Delhi, you'll find a thoroughfare known as CUHK-Avenue. In total, there are 33 videos here; 21 of them show anomalous human behaviour and activities, while the other 16 show more typical behaviour. Examples of typical sidewalk behaviour include individuals taking long, leisurely strolls and community leaders forming walking clubs. [37-38] People that hang around suspiciously, get too close to the camera, roam aimlessly around the lawn, throw stuff away, waste time, or forget anything are participating in antisocial behaviour.

UCF-Crime: 13 unique real anomalous behaviours are represented in the 1900 long, unedited clips that make up the UCF-Crime collection. Abuse, assault, arson, burglary, explosion, brawl, traffic accident, robbery, theft, gunshot, vandalism, and traffic accidents are all examples of such actions. UCF-Crime is an acronym for the University of Central Florida's Center for Investigating and Combating Crime, the group responsible for creating the dataset. The remaining 150 typical movies and 140 outliers accounted for the remaining 270 in the testing set. The remaining 150 clips were the typical, everyday types. The table labelled "Table 2" below displays the outcomes of the analysis of the UCF-Crime dataset. [39]

Figure 3 shows real-life events from around the world, ranging from the ordinary to the extraordinary. Three-dimensional representations of the datasets are shown in Figure 3, and ROC and AUC metrics are used for quantitative evaluation. [40]

Table 2. The statistical information from the UCSD

Dataset	**No. of Videos**	**Training Set**	**Test Set**	**Average Frames**	**Data set Length**	**Example of anomalies**
UCSDPet 1	60	33	35	200	4 min	Walking across walkways,

						bikes, and tiny carts
UCSDPet 2	27	15	10	162	4 min	Bikers, tiny carts, and pedestrians crossing walkways
Avenue	36	15	20	833	4 min	Run, toss, and new thing

Table 3: The data from the UC-Crime dataset include statistical information.

Types of anomalies	**The number of videos**	**Training data**	***Testing data***
Arrest.	50	47.	*01*
Abuse.		42.	*03*
Arson		46	*02*
Assault		40	*08*
Fighting		26	*21*
Shooting		44	*4*
Vandalism		28	*20*
Explosion		44	*03*
Stealing		28	*20*
Shoplifting	100	94	*03*
Robbery		84	*12*
Burglary	150	144	*04*
Accident		125	*21*
Sum	*950*	*792*	*145*

Table 4: For the UCF-Crime Dataset, a frame-based (AUC) comparison of the proposed ADSV technique to existing methods

Method	**UCF-Crime Dataset**
Binary SVM classifier	50.0
LTR	50.6
Spatiotemporal	63.0
SCL	65.50
MIL-C3D with constraints	75.41
MIL-C3D without constraints	74.44
C3D	81.08
TSN-Optical flow	78.08
Proposed ADSV	97.0
TSN-RGB	82.12

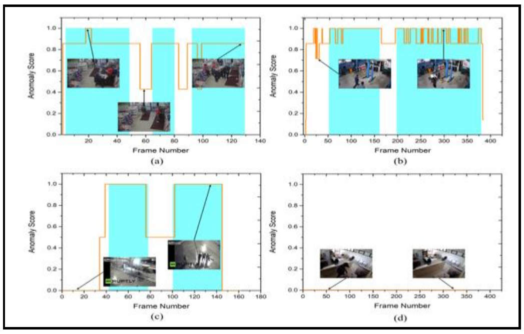

Fig 4. The strange score curves for each of the four test videos are represented here. The following videos are included in this package: The four subcategories are referred to as Normal-Videos-027-x264, Assault051-x264, Fighting047-x264, and Arrest048-x264 respectively.

In the figure 4, the UCF-Crime dataset is assessed with the assistance of these movies, which are included as part of the test set. Dots of cyan color indicate when a particular frame deviates from the standard. Because we wanted to make it easier to see the data, the video anomaly scores were normalized so that they fell inside the interval [0, 1]. This indicates that the severity of the abnormality is represented in a higher score as it gets more obvious. In other words, the score increases as the anomaly becomes more severe. The score increases in direct proportion to the degree of departure. If you want the finest experience possible, you should choose colour view, Studies Using Methods That Are Considered to Be State-Of-The-Art When compared to other EADN (Efficient Deep Learning model for Anomaly Detection) systems, The proposed model is evaluated using standard datasets.. The authors investigated a number of deep learning models, including ResNet-50 + multi-layer BD-LSTM, VGG-19 + multi-layer BD-LSTM, and InceptionV3 + multi-layer BD-LSTM. The ResNet-50 method combined with a bidirectional LSTM has the smallest model size among all the methods we've researched. When compared to other systems, the time and resources required by the EADN framework for anomaly detection are considerably less. This is true both in terms of the data storage and learning parameters. By comparing it to the most recent and cutting-edge approaches, we will be able to determine how effective the model is with regard to the size of the model, the complexity of the runtime, and the number of parameters. [41-43].

Table 5. Evaluation of EADN against state-of-the-art techniques in terms of parameters, model size, and time complexity.

Method	Constraint count in million	Model size	Latency per sequence
C3D	-	312	-
EADN(Proposed)	14.14	53.8	0.20
ResNet-50 + multi-layer BD-LSTM	24	142	0.20
Inception V3+ multi-layer BD-LSTM	22	148.4	0
VGG-19+multi-layer BD-LSTM	142	605.4	0.21

3.3 Discussion

In the past, low-level feature-based algorithms put a large amount of weight on the capability to recognise and eliminate outliers. This was due to the fact that there were no other algorithms that were suitable at the time. These processes can be broken down into three steps, which are as follows: Through the application of collaborative forces and optical flow, this model determined what constitutes normal and pathological behaviour. An attention-based LSTM was developed by Ullah et al. for the purpose of identifying activities in sports videos. Both methods can almost always be used interchangeably, depending on the circumstances. Utilizing a convolution block attention technique aided in enhancing the integrity of the spatial data. The improved feature maps can be given to a variety of sports by utilising a SoftMax, which is an algorithm that can be implemented in a neural network that is fully linked. The researchers Selicato et al. came to the conclusion that it would be more fruitful to search for anomalies in the genetic data rather than the visual data. [44-46] For instance, in order to distinguish normal and abnormal gene expression matrices, they used an ensemble-based method that involved principal component analysis (PCA) and hierarchical clustering. We were able to distinguish between them as unique individuals by using this strategy. Researchers made use of this method in order to classify genes into functional groups according to the amounts of expression they displayed. Riaz et al. suggested making use of a myriad of different detailed models in order to locate abnormalities in complex situations. Developing a model is the initial step that must be taken. The model is capable of predicting human posture by locating the bones and muscles in the appropriate places in the body. [47-49] The recently identified joints are then added to the features that are fed into a fully convolutional neural network (CNN) that has dense connections in order to identify outliers. One of the two ways that were described by Zhong et al. may be able to automatically discover anomalies without any assistance from a human being, but the other method, which requires human interaction to classify activities with noisy annotations, is unable to do so. In this essay, we will compare and contrast the two different approaches. As long as unexpected occurrences are unable to be forecast, the peculiar video annotations will continue to be puzzling. An activity classifier was used to classify the activities, and a graph convolution network was employed to reduce the noise caused by the annotations. In order to guarantee that the actions were categorised accurately, these two stages were carried out. The attention-based deep learning strategy that is discussed in this paper was proven to be the most effective method for detecting irregularities in surveillance footage after being compared to the several other methods that are currently being utilised for this purpose. It has been demonstrated that the attention-based deep learning strategy is superior to the vast majority of the other approaches that are now accessible. In this section, we will discuss the theoretical underpinnings of the ADSV approach that has been developed. [50-51]

4. Conclusion

We offered tangible examples of the unique uses of deep video anomaly detection models in a variety of real-world scenarios and conducted a survey of the unresolved technical issues in this field. The process of putting the model into practise can be divided into the three steps that are outlined below. Image segmentation, feature extraction, and sequence anomaly learning and classification are the three processes that make up this process. The process of recovering and separating key shots from surveillance recordings by the application of an algorithm that recognises boundaries is referred to as shot segmentation. These are genuine film stills that were retrieved throughout the search. Then, proposed EADN infrastructure will be used to distribute a collection of frames for each primary shot sample that has been taken. After that, the frame sequence is sent into a lightweight convolution neural network, also known as a LWCNN, in order to have the spatiotemporal information from the intermediate layer retrieved. Then, the LSTM cells will learn the temporal and spatial properties of each extraordinary event by using a set of frames as input. This will allow the cells to differentiate between events. This indicates that the EADN may be utilised to identify and categorise irregularities that are present within the segmented clip of the video. Proposed model was evaluated in comparison to a wide variety of metrics, and the findings indicate that it is superior to recently published techniques in terms of performance. Experts in the field of computer vision are beginning to rely more and more on sophisticated monitoring technology. Consumer demand is a primary force behind this movement in the market. They assist the observer in responding swiftly and properly whenever something out of the norm occurs during the course of the observation. In order for these algorithms to function correctly and efficiently, they require not just a powerful data processing infrastructure but also a substantial volume of input data. Proposed model's performance for the detection of anomalies on a wide variety of datasets that are used to establish standards is comparable to that of state-of-the-art methods. In this paper, we describe the ADSV approach for finding anomalies that can be seen on video. The method of segmenting shots, extracting features, learning sequences, and classifying anomalies in surveillance footage essentially consists of three parts. These processes are: segmenting images, extracting features, and learning sequences.

5. Compliance with Ethical Standards

- On behalf of all authors, I, the corresponding author states that there is no conflict of interest.
- This article does not contain any studies with animals performed by any of the authors.
- This article does not contain any studies with human participants or animals performed by any of the authors.

Author contributions

Jyoti Kukade[1]: Conceptualization, Methodology, Software, Field study, Data curation, Writing-Original draft preparation, Software, Validation., Field study **Prashant Panse[2]:** Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Piza, E.L.; Welsh, B.C.; Farrington, D.P.; Thomas, A.L. CCTV surveillance for crime prevention: A 40-year systematic review with meta-analysis. *Criminol. Public Policy* **2019**, *18*, 135–159.

[2] Cheng, K.W.; Chen, Y.T.; Fang, W.H. An efficient subsequence search for video anomaly detection and localization. *Multimed. Tools Appl.* **2016**, *75*, 15101–15122.

[3] He, C.; Shao, J.; Sun, J. An anomaly-introduced learning method for abnormal event detection. *Multimed. Tools Appl.* **2018**,*77*, 29573–29588.

[4] Lu, C.; Shi, J.; Jia, J. Abnormal event detection at 150 fps in matlab. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, 1–8 December 2013; pp. 2720–2727.

[5] Sultani, W.; Chen, C.; Shah, M. Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6479–6488.

[6] Huo, J.; Gao, Y.; Yang, W.; Yin, H. Abnormal event detection via multi-instance dictionary learning. In *International Conference on Intelligent Data Engineering and Automated Learning*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 76–83.

[7] Zhang, D.; Gatica-Perez, D.; Bengio, S.; McCowan, I. Semi-supervised adapted hmms for unusual event detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 611–618.

[8] Mehran, R.; Oyama, A.; Shah, M. Abnormal crowd behavior detection using social force model. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 935–942.

[9] Nor, A.K.M.; Pedapati, S.R.; Muhammad, M.; Leiva, V. Abnormality detection and failure prediction using explainable Bayesian deep learning: Methodology and case study with industrial data. *Mathematics* **2022**, *10*, 554.

[10] Ullah, M.; Mudassar Yamin, M.; Mohammed, A.; Daud Khan, S.; Ullah, H.; Alaya Cheikh, F. Attention-based LSTM network for action recognition in sports. *Electron. Imaging* **2021**, *2021*, 302-1–302-6.

[11] Zhao, Y., Man, K.L., Smith, J., Siddique, K. and Guan, S.U., 2020. Improved twostream model for human action recognition. EURASIP Journal on Image and Video Processing, 2020(1), pp.1-9.

[12] Zhao, Y., Man, K.L., Smith, J. and Guan, S.U., 2021. A novel two-stream structure for video anomaly detection in smart city management. The Journal of Supercomputing, pp.1-15.

[13] Han, Z., Zhao, Y. and Man, K.L., 2019. Design and Implementation of the Enhanced Public Video Press for Unforeseeable Events Management. Journal of Industrial Information Technology and Application (JIITA), vol.3 no.4, pp.294-297.

[14] Zhao, Y., Gabriela, M. and Man, K.L., 2021. Video Anomaly Detection by the Combination of C3D and LSTM. in Proceedings of International Conference on Digital Contents: AICo (AI, IoT and Contents) Technology 2021.

[15] Zhao, Y., Zhang, J. and Man, K.L., 2020. LSTM-based Model for Unforeseeable Event Detection from Video

Data. in Proceedings of The International Conference on Recent Advancements in Computing in AI, IoT and Computer Engineering Technology(CICET) 2020. pp.41.

[16] Y. Zhu and S. J. A. P. A. Newsam, “Motion-aware feature for improved video anomaly detection,” ArXiv Preprint ArXiv:1907.10211, 2019.

[17] D. Zhang, D. Gatica-Perez, S. Bengio and I. McCowan, “Semi-supervised adapted HMMs for unusual event detection,” in *2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR’05)*, San Diego, CA, USA, pp. 611–618, 2005.

[18] R. Mehran, A. Oyama and M. Shah, “Abnormal crowd behavior detection using social force model,” in

[19] *2009 IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, FL, USA, pp. 935–942, 2009.

[20] K. M. Nor, S. R. Pedapati, M. Muhammad and V. J. M. Leiva, “Abnormality detection and failure prediction using explainable Bayesian deep learning: Methodology and case study with industrial data,” *Mathematics*, vol. 10, no. 4, pp. 554, 2022.

[21] P. Shukla, S. Nasrin, N. Darabi, W. Gomes, A. R. Trivedi *et al.*, “Compute-in-memory with monte-carlo dropouts for Bayesian edge intelligence,” ArXiv Preprint ArXiv:2111.07125, 2021.

[22] M. Ullah, M. Mudassar Yamin, A. Mohammed, S. Daud Khan, H. Ullah *et al.*, “Attention-based LSTM network for action recognition in sports,” *Electronic Imaging*, vol. 2021, no. 6, pp. 302–1, 2021.

[23] L. Selicato, F. Esposito, G. Gargano, M. C. Vegliante, G. Opinto *et al.*, “A new ensemble method for detecting anomalies in gene expression matrices,” *Mathematics*, vol. 9, no. 8, pp. 882, 2021.

[24] H. Riaz, M. Uzair, H. Ullah and M. Ullah, “Anomalous human action detection using a cascade of deep learning models,” in *2021 9th European Workshop on Visual Information Processing (EUVIP)*, Paris, France, pp. 1–5, 2021.

[25] B. Zhao, L. Fei-Fei and E. P. Xing, “Online detection of unusual events in videos via dynamic sparse coding,” in *CVPR 2011*, Colorado Springs, CO, USA, pp. 3313–3320, 2011.

[26] W. Luo, W. Liu and S. Gao, “Remembering history with convolutional lstm for anomaly detection,” in

[27] *2017 IEEE Int. Conf. on Multimedia and Expo (ICME)*, Hong Kong, China, pp. 439–444, 2017.

[28] W. Liu, W. Luo, D. Lian and S. Gao, “Future frame prediction for anomaly detection a new baseline,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 6536–6545, 2018.

[29] Y. Chang, Z. Tu, W. Xie and J. Yuan, “Clustering driven deep autoencoder for video anomaly detection,” in *European Conf. on Computer Vision*, Glasgow, United Kingdom, pp. 329–345, 2020.

[30] M. Sabokrou, M. Khalooei, M. Fathy and E. Adeli, “Adversarially learned one-class classifier for novelty detection,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 3379–3388, 2018.

[31] D. Tomar and S. Agarwal, “Multiple instance learning based on twin support vector machine,” *Advances in Computer and Computational Sciences*, vol. 1, no. 553, pp. 497–507, 2017.

[32] F. Landi, C. Snoek and R. J. A. P. A. Cucchiara, “Anomaly locality in video surveillance,” ArXiv:1901.10364, vol. 1, 2019.

[33] S. Jayashri et al., “Video analytics on social distancing and detecting mask,” Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no. 9, pp. 2916–2921, 2021.

[34] Y. C. Hou, M. Z. Baharuddin, S. Yussof, and S. Dzulkifly, “Social distancing detection with deep learning model,” in 2020 8th International Conference on Information Technology and Multimedia (ICIMU). IEEE, 2020, pp. 334–338

[35] J. T. Zhou, J. Du, H. Zhu, X. Peng, Y. Liu, and R. S. M. Goh, “Anomalynet: An anomaly detection network for video surveillance,” IEEE Transactions on Information Forensics and Security, vol. 14, no. 10, pp. 2537–2550, 2019

[36] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. v. d. Hengel, “Memorizing normality to detect anomaly: Memoryaugmented deep autoencoder for unsupervised anomaly detection,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1705–1714

[37] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” Communications of the ACM, vol. 63, no. 11, pp. 139–144, 2020.

[38] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Muller, and M. Kloft, “Deep one-class classification,” ¨ in International conference on machine learning. PMLR, 2018, pp. 4393–4402.

[39] R. Chalapathy and S. Chawla, “Deep learning for anomaly detection: A survey,” arXiv preprint arXiv:1901.03407, 2019.

[40] S. W. Yahaya, A. Lotfi, and M. Mahmud, “Towards a data-driven adaptive anomaly detection system for human activity,” Pattern Recognition Letters, vol. 145, pp. 200–207, 2021.

[41] K. I. Withanage, I. Lee, R. Brinkworth, S. Mackintosh, and D. Thewlis, “Fall recovery subactivity recognition with RGB-D cameras,” IEEE transactions on industrial informatics, vol. 12, no. 6, pp. 2312–2320, 2016

[42] Nguyen Xuan Vinh, Jeffrey Chan, Simone Romano, James Bailey, Christopher Leckie, Kotagiri Ramamohanarao, and Jian Pei. 2016. Discovering outlying aspects in large datasets. Data Mining and Knowledge Discovery 30, 6 (2016), 1520–1555.

[43] Hu Wang, Guansong Pang, Chunhua Shen, and Congbo Ma. 2020. Unsupervised Representation Learning by Predicting Random Distances. In IJCAI.

[44] Hao Wang and Dit-Yan Yeung. 2016. Towards Bayesian deep learning: A framework and some existing methods. IEEE Transactions on Knowledge and Data Engineering 28, 12 (2016), 3395–3408.

[45] Siqi Wang, Yijie Zeng, Xinwang Liu, En Zhu, Jianping Yin, Chuanfu Xu, and Marius Kloft. 2019. Effective End-to-end Unsupervised Outlier Detection via Inlier Priority of Discriminative Network. In NeurIPS. 5960–5973.

[46] Xuanhui Wang, Nadav Golbandi, Michael Bendersky, Donald Metzler, and Marc Najork. 2018. Position bias estimation for unbiased learning to rank in personal search. In WSDM. 610–618.

[47] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. 2020. Generalizing from a few examples: A survey on few-shot learning. Comput. Surveys 53, 3 (2020), 1–34.

[48] Steve Webb, James Caverlee, and Calton Pu. 2006. Introducing the Webb Spam Corpus: Using Email Spam to Identify Web Spam Automatically.. In CEAS

[49] Peng Wu, Jing Liu, and Fang Shen. 2019. A Deep One-Class Neural Network for Anomalous Event Detection in Complex Scenes. IEEE Transactions on Neural Networks and Learning Systems (2019).

[50] Junyuan Xie, Ross Girshick, and Ali Farhadi. 2016. Unsupervised deep embedding for clustering analysis. In ICML. 478–487.

[51] Dan Xu, Elisa Ricci, Yan Yan, Jingkuan Song, and Nicu Sebe. 2015. Learning Deep Representations of Appearance and Motion for Anomalous Event Detection. In BMVC.

[52] Dhabliya, D., Sharma, R. Cloud computing based mobile devices for distributed computing (2019) International Journal of Control and Automation, 12 (6 Special Issue), pp. 1-4.

[53] Dhabliya, D., Parvez, A. Protocol and its benefits for secure shell (2019) International Journal of Control and Automation, 12 (6 Special Issue), pp. 19-23.