Journal of Theoretical and Applied Information Technology
30th November 2022. Vol.100. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FUZZY AGGREGATION BASED DEEP LEARNING SYSTEM FOR ANOMALY IDENTIFICATION

[1]KALLEPALLI ROHIT KUMAR, [2]DR.NISARG GANDHEWAR

Department of Computer Science and Engineering
Dr A.P.J. Abdul kalam University, Indore-452010(INDIA)

ABSTRACT

Due to the rapid growth of urbanisation and the rise of industry, there is an increased demand for surveillance system that work in real time. Artificial intelligence-based anomaly identification system only addresses a portion of the difficulties, primarily ignoring the dynamic nature of aberrant or abnormal behaviour across time. Using a training dataset with established normality and known error values are two further drawbacks of anomaly identification systems. A new approach to identify anomaly and its localization of video stream in real time called the Step Incremental Learner (SIL), is presented in this study. As new anomalies and normalities emerge over time, the unsupervised deep learning technique known as SIL uses active learning with fuzzy aggregation to continually update and identify them. Three benchmark datasets are used to show and assess SIL in terms of accuracy, robustness, computing overhead, and contextual indicators. Experiments conducted by our team show that the proposed system is very effective for 24x7 surveillance of video systems.

Keywords: *Deep learning, Anomaly Identification, Localization, CNN, Fuzzy Aggregation, Active Learning*

1.INTRODUCTION:

All modern business and urban surroundings have video surveillance for its progress, operation and sustainability. Contribution of video surveillance lies in the direction of care, protection, security, organization, individuals, process and activities. Industrial domains are adapting towards new and complex machines, cyber world and energy proficient layouts. The surroundings are getting heavily populated, there is an increase in the number and usage of multi-storied constructions, increased pedestrian, crowd and motors.[1][2] The immense growth in modern surroundings both vertical and horizontal has resulted in an exponential positioning and arrangement of visual security systems like CCTVs. It is impractical for an observer to monitor and analyse the entire recording frame by frame. The machine learning methods for the study are grouped under object detection, activity detection and anomaly detection.

Anomaly detection is an integral part of video surveillance as it enhances the process of object and activity detection. The process of anomaly detection is very difficult as the anomalies are not known before, making it difficult for the human observer also. Any behaviour that deviates from expected or normal behaviour is termed as anomaly. The challenges of anomaly detection are computational complexity and the overheads of video processing in considerations with spatial and temporal dimensions. The margin of difference between normal and anomaly is more or less ill defined. The challenge that is very difficult to address is when the normal behaviour evolves with temporal and time [3].

The challenges of computational complexity and video processing overheads with spatial and temporal surroundings are addressed by most of the existing literature but the challenge where the normal behaviour evolves according to temporal surroundings and time is large unaddressed in any of the existing literature. In this study the authors address the problem of normal behaviour evolving with time.

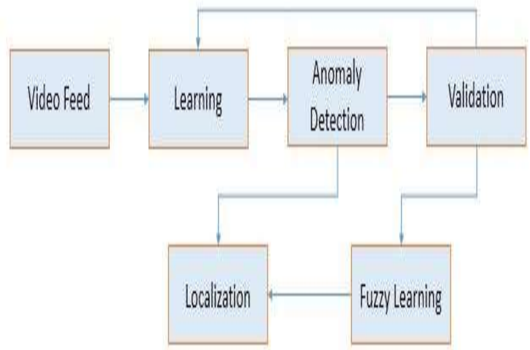

Figure 1: Overview of proposed SIL system

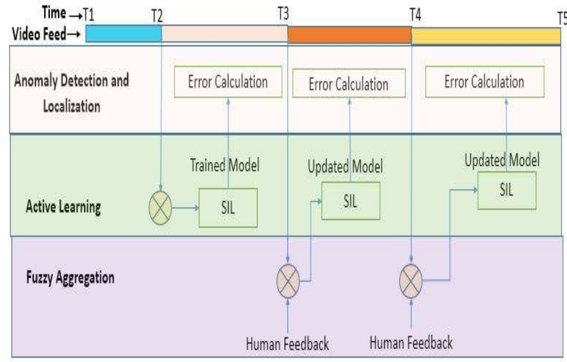

Figure 2: Functional system overview

The learning system is called Step Incremental Learner (SIL) that address the existing challenges mentioned before. SIL works in real time and detects anomalies based on the spatial and temporal values and patterns of normal. behaviour. SIL is motivated with the systematic learning process and active learning. SIL generates a fundamental knowhow from the available information to differentiate normal and anomalous behaviours and keeps upgrading the ongoing process with the change in surrounding.

Current deep learning algorithms for anomaly identification are limited by assessment based exclusively on errors in the reconstruction and need a training dataset with established normalcy. Briefly, the existing work focuses mostly on the complex computational challenges posed by the processing of extremely detailed information of data and recognising appropriate abnormalities from surveillance input. Since nature involving normal behaviour changes throughout time, a significant gap exists in our understanding of what constitutes normal behaviour.The proposel SIL method overcomes all the drawbacks of existing literatures.

Figure 1 illustrates an elevated overview of SIL. First, the real video surveillance stream is provided as input for the training of a spatiotemporal model of typical behaviour. Second, the learned model is used to detect and localise anomalies within the time span. The input and feedback from the human analyst is obtained to enhance the processing and detection rate of anomalies by SIL during fuzzy aggregation step. The feedback to the SIL system increases the learning rate and updates the normal behaviour.

Figure 2 shows an enhanced version of this overview as a functional perspective. The following is a computational model for anomaly identification in video surveillance. The training video stream (X_{train}) is made up of a series of videoframes that show typical behaviour in a particular camera view. In the actual world, R represents all of the video sequence of the camera view. During the testing step, a video feed (X_{test}) is used, with X_{test} R including both normal and abnormal video frames. The purpose is to use X_{train} to develop a representation of typical behaviour, which is then evaluated with X_{test} to identify abnormalities. Unlike earlier work, the SIL technique would (i)upgrade gained knowledge based on information from continually arriving streaming content, and (ii) The human observer gives feedback when an anomaly is detected and the SIL system learns from the feedback.

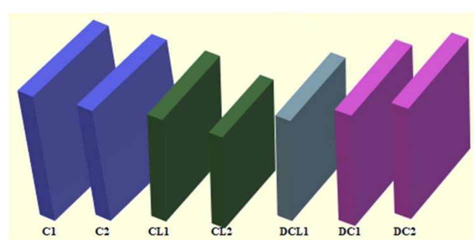

Figure 3: auto encoder architecture

2.LEARNING WITH AUTO ENCODERS

The video surveillance algorithm learns a spatial-temporal representation of typical behaviour from X_{train}as anticipated and approved behaviour. A spatiotemporal autoencoder is used in the SIL model to acquire the visual characterization of the input video stream. Autoencoders employ unsupervised backpropagation to establish target values equal to inputs while lowering reconstruction error. The spatiotemporal autoencoder in the proposed architecture is composed of a sequence of Convolution Neural Network CNN layers for learning the spatial and temporal representation is learned using sequence of ConvLSTM levels. The next sections detail the autoencoder's input and transformation layers.

Data Input Layer: This layer pre-processes the input. Pre-processing step helps increases the model's learning potential. Input frame is streamed and retrieved as successive frames. It is then converted to grayscale. The convertion minimizes the high magnitude of data to lower limits. It is changed to low pixel level of 224 x 224. Normalization is induced on this at a level between 0 to 1.[4][5][6] Input temporal cuboid is built by stacking consecutive frames of length T. The length T is the size of the sliding window used to control the temporal cuboid and it allows for longer-length motion to be included; but, as T increases, model convergence takes exponential time.

Table I specifies the filters and kernel sizes for the SIL models. Layers of Convolutional LSTM (ConvLSTM) Recurrent neural network (RNN) processes input sequences using an internal memory to record dynamic temporal behaviour. In table 1 CV represents the convolution, CL of LSTM, DCV represents the process of deconvolution, T gives the temporal depth, F highlights the number of filters used, the kernel size is represented by K, the number of strides is given by S and the asterisk* highlights the process of multiplication. Long short-term memory (LSTM) units are an upgrade of the RNN's generic building pieces. The LSTM unit consists of a cell, an input gate, an output gate, and a forget gate. . The input gate is responsible for determining how far a value goes in the device. The forget gate decides how short the values from prior time stages continue in the system, while the gate that controls the output looks into how stretched the present input is used in the reckoning the start of unit is used to determine. There is no limit to how long the cell maintains its data.

TABLE 1
Spatiotemporal Autoencoder Architecture

ID	Input Tensor	Operations	Output Tensor
C1	T*224*224*1	CV-F:128;K:27*27;S:4	T*56*56*128
C2	T*56*56*128	CV;F:64;K:13*13;S:2	T*28*28*64
CL1	T*28*28*64	CL;F:64;K:3*3	T*28*28*64
CL2	T*28*28*64	CL;F:32;K:3*3	T*28*28*64
DCL1	T*28*28*32	CL;F:64;K:3*3	T*28*28*32
DC1	T*28*28*64	DCV;F:64;K:13*13;S:2	T*56*56*128
DC2	T*56*56*128	DCV;F:128;K:27*27;S:4	T*224*224*1

Because LSTM was designed and is used largely for modelling far-off sequential associations, it has a limitation in dealing with geographical data

3.ANOMALY IDENTIFICATION USING LOCALIZATION

The SIL model reconstructs pixel level normalcy of the input to the system. However, because such pictures were not given during the training phase, the trained autoencoder is unable to reliably reproduce anomalous or unseen sequences. This phenomenon is employed in order to analyse and find irregularities in the input footage. The reconstruction error (E) is

computed by squaring the values and taking their square The variables used for calculating the error are X represents the input, sliding window time is given by T, height and width are represented with h and w respectively. The temporal cube generated contains a value that identifies it abnormality. The value is known as reconstruction error.

The anomaly threshold is a criterion for determining whether or not a behaviour is anomalous.

When actual video surveillance applications are being used, a human observer may set rating depending on the level of sensitivity.

If the number was low, it would imply that there was a stronger reaction to the monitoring area, which would result in more warnings. More warnings lads to less sensitivity, which directly account for irregularities due to sensitivity in the respective domain being ignored.

In addition, we include a threshold based on temporal values which is identified according to the number of frames that directly can be used to identify an event is identified as an anomaly is more than. False-positive anomaly alerts are prevented by occlusion, motion blur and high-intensity illumination circumstances causing fast changes in the surveillance video stream. Figure 4 depicts an anomaly detection

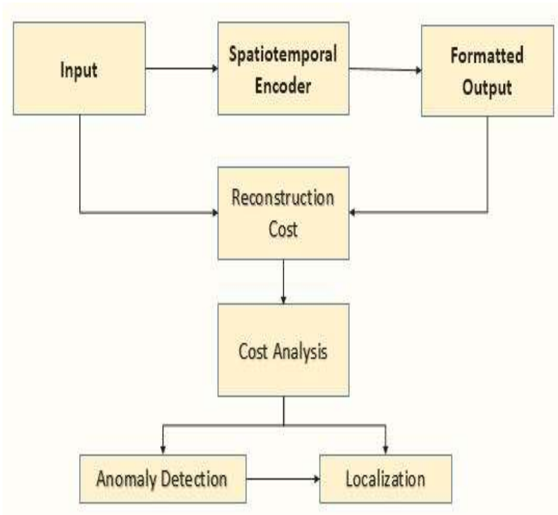

Figure 4: Anomaly Identification

3.1 Fuzzy Aggregation

Active learning in a real-world video surveillance environment is aimed at enabling the detection of abnormalities in rapidly changing conditions. We use the deep learning technique discussed previously to train the learning model to determine acceptable usual behaviour. The capacity of the detection system to adapt and recognise new instances in dynamic circumstances, such as those characterised by new normal behaviour that was not expected or by previously aberrant behaviour that was reformed too normal, is crucial.

A fuzzy aggregation technique is used to continually provide training to the sytem with all possible behaviours possible that is acceptable for the present monitoring situation. SIL approach is based on the cognitive ability to create a foundational knowledge base that can be continually expanded upon as new information comes to light. SIL is trained to identify abnormalities in a monitoring environment by observing the behaviour of people in the same situation. Because the training error of the input cuboid exceeds a threshold, an anomaly is detected when the video frame is recognised as such. A human observer then confirms the accuracy of the previously classified frames.
Proactively feeding the learning model with normality behaviour from human observers is a goal. The human observer can therefore classify an incorrectly identified video frame as 'normal,' which will be employed in the subsequent phase of continuous learning. The video frames identified as normal will be used to train the SIL model, continually improving its knowledge of the idea of normality after human observer feedback.

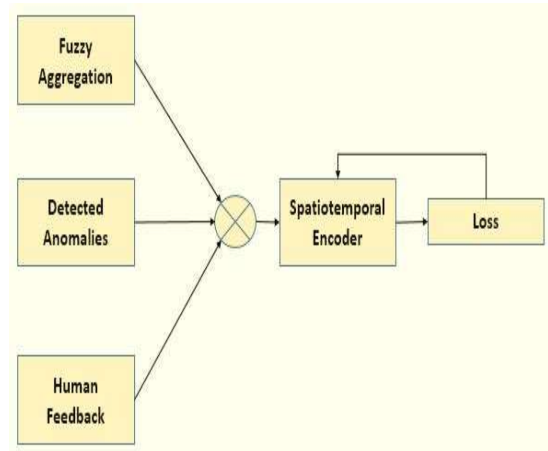

Figure 5: Active Learning using fuzzy aggregation

The CUHK consists of videos that are compiled using a stationary 640 360-pixel video camera that was capturing path movement at City University of Hong Kong. This compilation of videos contains 16 number of training videos of typical human behaviour and 21 videos of behaviour that are not considered normal. Common sidewalk behaviours include littering/throwing away items, coming in the direction of the recorder used as input device, strolling on the lawn and discarded items. The UCSD consiste of videos which are focused on two pedestrian paths, contains video of 238x158 pixels. This package comprises two datasets, ped 1 and ped 2, that represent a range of crowd circumstances, from sparse to dense. The train video examples only show people walking, whereas the test dataset contains samples that include walkers crossing the sidewalk or the grass and car activities. There are 34 training video samples and 36 test videos in Ped 1 dataset. The number of training videos in Ped 2 is 16 and the number of test videos is 12.

4.EXPERIMENT

1) First, the proposed spatiotemporal autoencoder anomaly detection efficiency and accuracy is tested by using three standard datasets.
2) Second, we assess the ISTL model's continuous learning potential for the UCSD Ped1 and Ped2 datasets, reclassifying a previously deemed anomalous event as normal.

3) Third, we do a runtime examination of our technique to demonstrate our algorithm's real-time processing capabilities.

Pre-processing of the input samples is done and the dimensionality of all the samples is set to 224x224 pixels. The pixels are normalized in a range from 0 to 1. On the basis of the frame rate of training data the T=8 is selected, which corresponds to a period of about less than a second by one third. Choice of temporal is based on maximising the speed to be caught between successive frames and reducing closure of the SIL caused by a high number of video frames..

In this study, the learning rate of 0.04 was used with 1300 iterations for the SIL model. Utilizing the stochastic gradient descent approach, the spatiotemporal autoencoder model is optimised and the reconstruction loss is determined by using mean squared error. In order to prevent model overfitting, we used an early stopping regularisation strategy in which training is terminated when the loss stops improving. the training phase consisted of divideng the dataset into 60% for the first iteration, 20% for the second, and 20% for the third. During the active process of learning, the reconstruction error served as the fuzzy measure.

5.RESULTS - ANOMALY DETECTION

Three feature-based methods and four different deep learning methods were used for analysing anomaly detection. The first method is proposed by Brunetti et al. [19] for identifying the abnormal crowd behaviour identification implementing a social force model. The second method proposed by He et al. [12] uses the Markov field for anomaly detection. In the third method the assessment of anomaly detection is performed by combining the earlier two methods of using social force and Markov random field. In this approach the outliers identified are marked as anomalies. The deep learning models used for equating are listed here. First the feed forward model proposed by Tejedoret al. [16] It is an convolutional model with autoencoder that learns both properties and classifiers. As its foundation, the two-stream technique is utilised by the Gaussian Mixture Model – Deep Denoising Autoencoder. The latent space of each autoencoder is subjected to an application of a Gaussian Mixture Model in order to facilitate an improvement in the late fusing of both streams.[24]

Anomaly threshold and temporal threshold are involved in anomaly localization and identification stage. The optimal value of λ is identified for each of the three datasets in a range from 0 to 9. The value of λ was found to be different for all the three datasets. The difference is mainly due to the position of the camera in taking images for the three datasets. In some datasets the image is taken from close range and in some the position of the camera is far from the image. Table II gives the selection of anomaly threshold for the ped 1, ped2 and cuhk avenue datasets. The optimal accuracy of the SIL model is highlighted in Table II. The table gives the value of the model for the area under curve and equal error rate for the selected datasets. Table 3 gives the comparison for AUC and EER values for the three selected datasets for the study.

The efficiency of the proposed system is compared with some of the existing literature available online. The efficiency of the SIL model is very good in terms of Ped 1 and Ped 2 dataset. Only limited study have been performed with the cuhk avenue dataset. The performance of SIL is better when compared to the frest of the literature due to the use of deep learning and LSTM concepts. The SIL method stands out among all the existing literature when compared for all the 3 datasets.

Journal of Theoretical and Applied Information Technology
30th November 2022. Vol.100. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 6 Anomaly identification using Ped 1 dataset

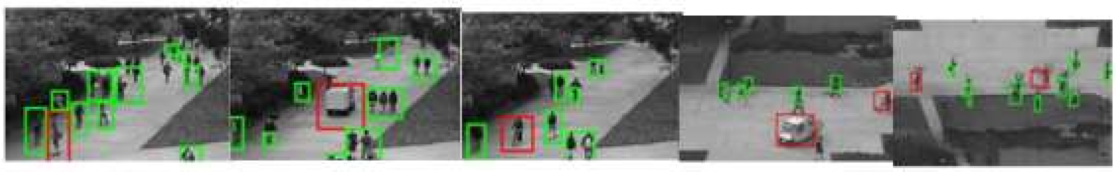

Figure 7: Anomaly identification using Ped 2 dataset

TABLE II

SELECTION OF ANOMALY THRESHOLD AND TEMPORAL THRESHOLD

Dataset	Optimal AUC/EER	Anomaly Threshold (μ)	Temporal Threshold (λ)
Ped 1	67.4/19.3	0.36	5
Ped 2	92.5/7.7	0.41	9
Avenue	79.8/19.2	0.31	2

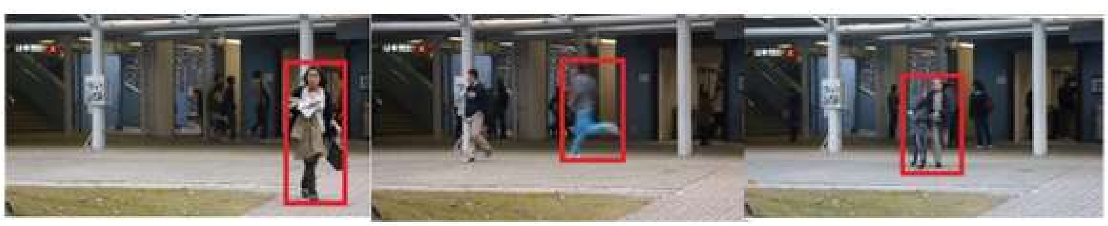

Figure8: Anomaly identification using CUHKAvenue dataset

TABLE III

COMPARISON OF AUC AND EER

Model	Ped 1 AUC/EER	Ped 2 AUC/EER	Avenue AUC/EER
SF (2009)	57.5/31.0	45.6/39.0	NA
MPCCA (2009)	63.8/38.0	59.3/28.0	NA
MPCCA + SF (2010)	72.2/33.0	63.3/39.0	NA
Conv-AE (2016)	78.0/25.9	88.0/22.5	69.2/22.1
S-RBM (2017)	73.3/38.4	78.4/19.5	75.8/29.2
ConvLSTM-AE (2017)	73.5/NA	82.1/NA	72.0/NA
Unmasking (2017)	66.4/NA	84.2/NA	79.3/NA
GMM-DAE (2020)	NA/NA	96.3/NA	89.3/NA
GS (2022)	71.1/31.2	87.3/14.1	83.3/21.8
Proposed (SIL)	79.2/18.8	94.1/7.7	72.8/21.2

Table 4 gives the anomaly identification ratio. Using the Ped1 dataset, before the process of active learning the anomaly detected were 11 and on collection of feedback and further training the number of cases of anomaly got reduced to 3. There was a significant increase in the ratio for anomaly detection using the fuzzy process. In the second sample using the ped2 dataset 5 anomaly were detected out of 7 but with the use of fuzzy aggregation learning the anomaly detection in real time got reduced to just 2 and table 5 gives the processing details.

TABLE IV

ANOMALY DETECTION FOR CYCLING SCENARIO

Dataset	Prior to Active Learning	After Active Learning
UCSD Ped 1	11 / 14	3 / 14
UCSD Ped 2	5 / 7	2 / 7

Values presented as detected anomalies / total test sample

The accuracy of the deep learning models are tested based on Equal Error Rate (EER) and computation of area of ROC. Overall, our method beats all handmade approaches, and we get comparable

performance to deep learning representation-based methods on the Ped 1 and Avenue datasets. Our suggested SIL technique outperforms all comparable models, including the benchmark GMM-DAE (2020) approach, on the Ped 2 and Ped 3 dataset and the current GS(2022) on the Ped1, Ped 2 and Avenue datasets.

Figure 6, figure 7 and figure 8 show the qualitative study of the localised anomaly patches for the three-dataset selected for the study. In the UCSD ped 1 dataset, SIL locates abnormalities such as bikes and automobiles on the paths, pedestrians strolling across the pathways, crowd lingering, and pedestrians dragging trolleys. We chose riding on pedestrian walkway scenarios from the UCSD Ped 1 and Ped 2 datasets to show SIL's active learning capacity. We considered riding on pedestrian routes as usual, thus we categorised all bicycle anomaly detections as normal.

CONCLUSION

This article is a study based of deep and active learning; we have developed a novel method for video surveillance systems identifying anomaly identification in real time. Anomaly identification in real time scenario has three major issues that our proposed SIL technique overcomes. First it grips with real-time processing of massive amounts of multidimensional surveillance data. Second it structures the process of anomaly identification using the concept of normality and third it uses the concept of fuzzy aggregation to learn from human feedback thus implementing active learning concept. The SIL technique that has been presented uses a spatiotemporal autoencoder model as its foundation. This model identifies spatial regularities and is comprised of convolution layers.

SIL addresses the tightly coupled reliance on a predefined normality training dataset by including a fuzzy aggregate of operator input to enhance the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

process of learning from mistakes or from anomalies that the system was not able to identify. It human operator further strengthens the training and learning process of the proposed system. For the purpose of overcoming sparse assessment, which is primarily dependent on reconstruction error, proposed SIL system makes use of two thresholds, namely the anomaly threshold and the temporal threshold, which are both determined by the context of the video surveillance stream. The efficiency of the system in terms of accuracy, robustness, and low computational overhead of the proposed approach, in addition to the contextual indicators of the proposed approach, have been demonstrated by the results of experiments conducted on three datasets namely ped 1 ped 2 and chukka avenue, confirming its wide applicability in surveillance systems. SIL guarantees that a human observer is not necessary round the clock to continually watch surveillance footage in order to identify unusual behaviour. The proposed system requires the help of human support only for the verification of reported abnormalities in real circumstances and the further development of the learning model need human intervention. As our future work we would like to eliminate the human support feedback and make it an autonomous end to end system.

REFERENCES

[1]B. Son, Y. Her, and J.-G. Kim, "A Design and Implementation of Forest-Fires Surveillance System based on Wireless Sensor Networks for South Korea Mountains," p. 8, 2006.

[2]M. Carvajal-Yepes et al., "A global surveillance system for crop diseases," Science, vol. 364, no. 6447, pp. 1237–1239, Jun. 2019, doi: 10.1126/science.aaw1572.

[3]K. K. Verma, B. M. Singh, and A. Dixit, "A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system," Int. j. inf. tecnol., vol. 14, no. 1, pp. 397–410, Feb. 2022, doi: 10.1007/s41870-019-00364-0.

[4]F. T. Al-Dhief et al., "A Survey of Voice Pathology Surveillance Systems Based on Internet of Things and Machine Learning Algorithms," IEEE Access, vol. 8, pp. 64514–64533, 2020, doi: 10.1109/ACCESS.2020.2984925.

[5]K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, "A Survey on Machine Learning Techniques for Cyber Security in the Last Decade," IEEE Access, vol. 8, pp. 222310–222354, 2020, doi: 10.1109/ACCESS.2020.3041951.

[6]K. A. Joshi and D. G. Thakore, "A Survey on Moving Object Detection and Tracking in Video Surveillance System," vol. 2, no. 3, p. 5, 2012.

[7]H. V. Padullaparti, P. Chirapongsananurak, M. E. Hernandez, and S. Santoso, "Analytical Approach to Estimate Feeder Accommodation Limits Based on Protection Criteria," IEEE Access, vol. 4, pp. 4066–4081, 2016, doi: 10.1109/ACCESS.2016.2589545.

[8]R. Sagar, R. Jhaveri, and C. Borrego, "Applications in Security and Evasions in Machine Learning: A Survey," Electronics, vol. 9, no. 1, p. 97, Jan. 2020, doi: 10.3390/electronics9010097.

[9]A. Brunetti, D. Buongiorno, G. F. Trotta, and V. Bevilacqua, "Computer vision and deep learning techniques for pedestrian detection and tracking: A survey," Neurocomputing, vol. 300, pp. 17–33, Jul. 2018, doi: 10.1016/j.neucom.2018.01.092.

[10]W.-T. Chen, P.-Y. Chen, W.-S. Lee, and C.-F. Huang, "Design and Implementation of a Real Time Video Surveillance System with Wireless Sensor Networks," in VTC Spring 2008 - IEEE Vehicular Technology Conference, Marina Bay, Singapore, May 2008, pp. 218–222. doi: 10.1109/VETECS.2008.57.

[11]S. N. Jyothi and K. V. Vardhan, "Design and implementation of real time security surveillance system using IoT," in 2016 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, Oct. 2016, pp. 1–5. doi: 10.1109/CESYS.2016.7890003.

[12]T. He et al., "Energy-efficient surveillance system using wireless sensor networks," in Proceedings of the 2nd international conference on Mobile systems, applications, and services -MobiSYS '04, Boston, MA, USA, 2004, p. 270. doi: 10.1145/990064.990096.

[13]C. Clavel, T. Ehrette, and G. Richard, "Events Detection for an Audio-Based Surveillance System," in 2005 IEEE International

Conference on Multimedia and Expo, Amsterdam, The Netherlands, 2005, pp. 1306–1309. doi: 10.1109/ICME.2005.1521669.

[14]G.Sreenu and M. A. Saleem Durai, “Intelligent video surveillance: a review through deep learning techniques for crowd analysis,” J Big Data, vol. 6, no. 1, p. 48, Dec. 2019, doi: 10.1186/s40537-019-0212-5.

[15]S. Tiwari, H. Ramampiaro, and H. Langseth, “Machine Learning in Financial Market Surveillance: A Survey,” IEEE Access, vol. 9, pp. 159734–159754, 2021, doi: 10.1109/ACCESS.2021.3130843.

[16]J.Tejedor, J. Macias-Guarasa, H. Martins, J. Pastor-Graells, P. Corredera, and S. Martin-Lopez, “Machine Learning Methods for Pipeline Surveillance Systems Based on Distributed Acoustic Sensing: A Review,” Applied Sciences, vol. 7, no. 8, p. 841, Aug. 2017, doi: 10.3390/app7080841.

[17]J. Hou, C. Wu, Z. Yuan, J. Tan, Q. Wang, and Y. Zhou, “Research of Intelligent Home Security Surveillance System Based on ZigBee,” in 2008 International Symposium on Intelligent Information Technology Application Workshops, Shanghai, China, Dec. 2008, pp. 554–557. doi: 10.1109/IITA.Workshops.2008.223.

[18]Dr. E. Baraneetharan, “Role of Machine Learning Algorithms Intrusion Detection in WSNs: A Survey,” JITDW, vol. 02, no. 03, pp. 161–173, Aug. 2020, doi: 10.36548/jitdw.2020.3.004.

[19]B.Lavi, M. F. Serj, and I. Ullah, “Survey on Deep Learning Techniques for Person Re-Identification Task.” arXiv, Jul. 19, 2018. Accessed: Jul. 01, 2022. [Online]. Available: http://arxiv.org/abs/1807.05284

[20]C. N. T. Siebel and S. J. Maybank, “The ADVISOR Visual Surveillance System,” p. 9.

[21]T. Zhang, A. Chowdhery, P. (Victor) Bahl, K. Jamieson, and S. Banerjee, “The Design and Implementation of a Wireless Video Surveillance System,” in Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, Paris France, Sep. 2015, pp. 426–438. doi: 10.1145/2789168.2790123.

[22]N. H. Tandel, H. B. Prajapati, and V. K. Dabhi, “Voice Recognition and Voice Comparison using Machine Learning Techniques: A Survey,” in 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, Mar. 2020, pp. 459–465. doi: 10.1109/ICACCS48705.2020.9074184.

[23]Z. Guan, L. Bian, T. Shang, and J. Liu, “When Machine Learning meets Security Issues: A survey,” in 2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR), Shenyang, Aug. 2018, pp. 158–165. doi: 10.1109/IISR.2018.8535799.

[24] Y. Ouyang and V. Sanchez, ``Video anomaly detection by estimating likelihood of representations," 2020, *arXiv:2012.01468*.

[25] E. Cruz-Esquivel and Z. J. Guzman-Zavaleta, "An Examination on Autoencoder Designs for Anomaly Detection in Video Surveillance," in *IEEE Access*, vol. 10, pp. 6208-6217, 2022, doi: 10.1109/ACCESS.2022.3142247