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

DEEP LEARNIG MODEL IMPLEMENTING PIPELINED AUTO ENCODERS AND ONE CLASS LEARNING FOR ANOMALY IDENTIFICATION AND LOCALIZATION FROM SURVEILLANCE STREAM VIDEO

[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

The demand for greater security measures for monitoring and protecting operations has made video anomaly detection one of the most significant study areas in the field of computer vision. This is due to the fact that the detection of anomalies in video surveillance systems has increased, making it one of the leading focus areas in the current field of research. When looking for suspicious behaviours like aggression, robbery, and wrong U-turns, assigning real people to frequently analyse the surveillance footage is a process that is both time-consuming and prone to error. As a direct consequence of this, there is an urgent requirement for the development of advanced, fully automated video surveillance systems for use in public areas. The research uses pipelined deep encoders to find a solution to the problem of locating and identifying anomalies in surveillance videos. The authors especially combined the LSTM autoencoder with the convolutional autoencoder. This strategy helps gather spatial and temporal information from the input video stream. This allowed us to achieve optimal results. The authors used the one-class classification principle during the training phase of the model. The training is done on normal data, and when they were verifying or testing it, they used anomalous testing data. The analysis of the study is based on standard benchmark practises for error rate and the duration of time required for identifying anomalies in the sequence of the video stream. This study satisfies the requirements for operating in near real time.

Keywords: *Deep Learning, Auto Encoders, LSTM, Machine Learning*

1.INTRODUCTION

In recent years, surveillance cameras have grown in public spaces in order to develop security and safety measures and monitor current activities. Increased terrorist threats and other crimes have made intelligent video surveillance an essential component of security systems for identifying and mitigating potential dangers. As a result, experts predict the worldwide video surveillance market will be worth $106.98 billion by 2026. Many fields have seen success with the implementation of deep learning techniques.[1][2].

The security sectors that produce and sell products and services based on video surveillance have felt the effects of artificial intelligence, which has made ground breaking strides in the field of computer vision. Even though the application of artificial intelligence in the field of security is still in its infancy, artificial intelligence (AI)-based solutions are a huge boon to the security industry and the video surveillance industry. Here are a few reasons why AI-based solutions could make a big difference in the video surveillance industry.[3] In the past, humans had to watch security footage to make sure nothing suspicious was happening on the premises. On the other hand, it takes a lot of time and effort for a person to watch surveillance footage around the clock.

In addition, little effort was previously made to gain useful insights following a security incident. Due to this need, advanced systems for watching surveillance footage round-the-clock have become increasingly popular. However, AI-based solutions watch CCTV footage in real time and analyse live footage intelligently to spot unusual security breaches. [4][5] This makes it possible to keep an eye on the video feeds all the time, find out right away when something strange happens, and fix things right away.

Out of the ordinary data patterns are called anomalies. Anomaly detection has an extensive collection of submissions, depending on the type of data being processed. A very common usage of anomaly detection is in the detection of credit card fraud and the enforcement of two-step verification for email services. Its principal application in the networking world is to identify malicious intrusion attempts.

Items left behind, unlawful traffic violations, usage of parking in banned locations, fire occurrences, fatalities, burglaries, and acts of violence are some of the issues that need to be addressed. can all be detected with anomaly detection applied to surveillance videos documenting the events. Jaywalking, loitering, trespassing, crawling, and homicide are just some of the odd acts that can be uncovered with this technology. Developing a reliable algorithm that can locate the anomalies in a video stream in real time is a challenging task because of the framework-dependent and imprecise nature of anomalies. [6] [7] The difficulty of anomaly detection is compounded using low-resolution cameras, the variability of lighting conditions, and the presence of noise in the acquired footage. There must be a balance between precision and speed in surveillance programmes that operate in real time. Anomaly detection's use in the field of video surveillance has revealed that real-world occurrences are characterised by a wide variety of open-ended and intricate scenarios. [8] Due to the high cost of collecting labelled data for unusual events, the most common practise is to test the model on unconventional video content after training it on regular data. An aberrant pattern is defined as one that deviates significantly from the normative test data.[9].

This article's introduction is followed by Related Work in Section 2. Section 3 highlights the methodology of the study. The experiment details are explained in Section 4. Section 5 shows the results of the study along with a comparison analysis of the existing literature.

2. RELATED WORK

Recently the gun shootings in campuses have made anomaly identification the top most priority for many academic institutions. In large organisations it is possible by 24x7 surveillance of their video. Due to the nature of the abnormalities and the necessity of monitoring the surveillance videos, this is the case. For video-based anomaly detection, prior studies have offered both a taxonomy of the criteria that must be examined and a hierarchy of the numerous approaches that can be used. Using the same taxonomy, we may classify anomaly detection techniques into two groups: those that rely on conventional techniques, and those that use deep learning.[10] Local and global feature modelling provide the basis of most established methods. Local feature modelling can make use of a wide variety of different methodologies, including Bayesian inference, dictionary learning methods, and pattern learning models, to name just a few of them.

Holistic techniques, on the other hand, focus most of their attention on studying density-based and trajectory-based methodologies for anomaly identification. Low-level features are typically gathered through traditional approaches, which rely on feature extraction procedures that have been hand-crafted based on a certain level of expertise. In a limited number of circumstances, these approaches will work. In contrast, deep learning-based techniques can be used in a broad variety of contexts.[11] Spatial-temporal learning models, generative models, temporal regularity models, and representation learning models are common examples of the types of deep learning-informed methods used for anomaly detection. This article is a component of a larger study on anomaly detection, and it discusses just a subset of the state-of-the-art techniques available at present.

The spatio-temporal characteristics of the video are faithfully represented by the content technique established by Cong et al. Using this strategy, data from a test sample is used to determine which training samples best match that data. Local spatio-temporal anomalies can now be detected inside a probabilistic framework developed by Aliev et al. (2019) [12]. In this setting, the connections between nearby nodes are used to produce anomaly score functions. In this way, the autoencoder can pick up on both the video's spatial and temporal details. Adding optical flow data to a temporal network helps it make more precise predictions. The anomaly value is determined by subtracting the predicted future frames from the actual ones. Errors are minimised when reconstructing training data using reconstruction-based methods, however the same techniques do not lead to many inconsistencies when applied to out-of-the-ordinary samples. Predictive methods, on the other hand, are premised on the idea that normal behaviour can be anticipated but deviant behaviour cannot. Predictive methods are useful for analysing surveillance footage, but they are susceptible to the background noise that is present in most of them.

Vosta and Yow (2021) [13] suggested a quick method for identification and localisation of anomalies in input video stream using of pipelined deep neural networks. This model employs a cascading network of 3D convolutional neural networks and 3D autoencoders. This work presents a method for detecting anomalies that combines both spatial and temporal information clubbed with one class learning a. Vasilopoulos et al. (22022) [4] provide evidence that autoencoders can be used to detect anomalies in movies and record the consistency found in extended videos. The re-building price tag is what makes this happen. In order to perform both learning and classification in a single framework, they proposed a fully convolutional end-to-end autoencoder.

Aboah et al. (2021) [15] learned activity patterns hierarchically inside the video. Using the collected activity patterns, the energy function for anomaly detection is computed. To learn visual and motion features, Esquivel and Zavaleta (2022)[16] recommended using a stacked denoising autoencoder. In addition, they provided a twofold fusion framework for the acquisition of data supplementary to behavioural and kinetic patterns. Roshtkhari et al. (2013) [17]have used probability theory for the early identification of anomalies. They are creating the codebook methodically, with a set of video phrases as their base. This approach involves incrementally modifying a likelihood density function in order to identify new baseline habits. Kim et al (2009) [18] used Markov model in random field as a means of detecting anomalies in video data. It deals with anomaly detection on two levels: locally and globally. Locally, it can distinguish out-of-the-ordinary behaviour inside a busy environment, while globally it can handle aberrant interactions between activities taking place in different parts of the scene. That is right; it's the one in charge of spotting irregularities. Mehran et al. (2009) [19] proposed the model of social forces. Assuming that the objects in the video are particles, this model calculates the interaction force between them by averaging the optical flow in the spatial and temporal dimensions. In order to identify the out-of-whack video frames, scientists used a technique dubbed "bag of words," which relies on the interaction forces between words to do so. Noe et al. (2012) suggested a method of anomaly identification using an MDT, or dynamic texture mixture. Time and space anomalies can be identified using this method since the method simultaneously depicts the scene's appearance and dynamics. While the discriminant saliency method is used to examine temporal anomalies, this one focuses on spatial ones on the assumption that they are rare occurrences.

Literature	Method	Gap
Powar and Attar [22] (2021)	Pipelined deep encoders and one class learning	Hardware implementation not performed, Real time analysis needs to be performed
Esquivel et al.[17] (2022)	Distillation and joint temporal training on auto encoders.	Number of temporal features extracted for the study is less.
Bai et al. [25](2022)	Sensors and perception	Limited number of hardware resources explored
Paul et al.[24] (2022)	Five stage design pipeline	Temporal features not explored.

Vosta et al. [13](2022)	ResNet50 and Convolutional LSTM	Used only UCF dataset with anomalies
Aboah et al. [16](2021)	Decision tree and deep learning	Tested only single type of dataset
Ingle and kim[27](2022)	Resource constraint based on CNN	Hardware implementation not performed

The current study suggests pipeline-based auto encoders for the detection of anomalies in surveillance video streams. It identifies anomalies using probability statistics and the gaussian distribution. Further localization of the anomaly is also performed after its detection using deep learning. The use of LSTM and sliding window coupled with CNN is a novel method that has not been used in many studies. The study highlights the use of the un-supervisory method for the detection of anomalies.

3 | METHODOLOGY

An illustration of the suggested method with pipelined autoencoders for videos stream is shown in figure 2. The symbols used for the study are shown in Table 1. The anomaly identification process is accomplished with the usage of sequence-to-sequence LSTM autoencoders and convolutional auto encoders, and a single-frame convolution decoder for anomaly localisation. Figure 3 shows a convolutional autoencoder, as a first phase in the procedure, all video frames are scaled to uniform dimensions and normalised to lie inside the interval [0, 1]. The symbol x represents a single video frame. $(S_0$...n is shorthand for a video's sequence of n frames, which looks like this: $\{x_0; x_1;;x_n\}$

3.1 | Anomaly detection

The proposed system for identification of anomalies and localization is shown in figure 2. The system makes use autoencoders. Based on their field of research the authors have decided to use the sequence of convolutional autoencoder along with LSTM for identifying the anomalies existing in video stream. The localization process is performed with the use of single frame convolutional decoder. In the pre-processing stage the normalization of the video frames in a range of 0-1 is performed along with resizing of the dimensions of the frame to a standard level. X represents a single frame in the input video stream and the sequence of video frames is shown by $(S_{0,\ldots.n})$ for the sequence of input stream $(x_0,\ldots.\ x_n)$. The learning process of the single frame is performed by the convolutional encoder. It is a single frame convolutional encoder and it is used for spatial learning. The training phase is done by minimizing the distance between the input x and the output for the autoencoder. The distance is denoted by L_2. M(.) represents the autoencoder. Its function is to gather information about the temporal dependencies. This information helps to build expressive sequence modelling tasks.M(.) learns the encoded frames that are compressed and mapping in to a single vector. This is very helpful to the identification of anomalies existing in the video stream. The usage of K- sliding window is for selecting the sequence of compressed and encoded vectors. K in sliding window refers to the actual size of the sliding window.

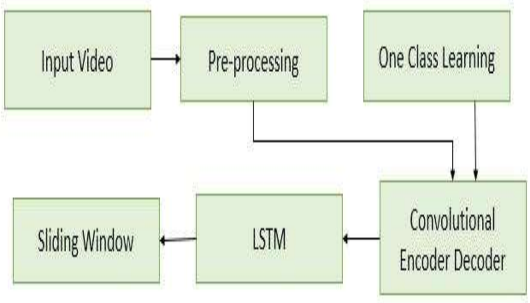

Figure 1: Architecture Model Of Proposed System

The frames labelled as $S_{0\ldots.n}$ are part of an input video stream denoted as $(X_0,x_1\ldots., x_n)$. This input frames are given to an encoder depicted as F_{enc}(.). The first output of the encoder is the sequence {Fenc(S0….n)}. The temporal information is learned by the seq2seq LSTM autoencoder. The sequence LSTM learns is from the sequence The sequence of representation is now {Fenc(S0….n)} and not as the original input signal F_{enc}(.). The training objective is to minimize the L2 distance between sequence (G_{enc}(Fenc(S0….n))) and (Fenc(S0….n)). The primary principle of using LSTM is that it is trained on supervisory data that is labelled with huge sample size and test on data without any labels or with samples having less records and are labelled.

When dealing with anomaly identification in surveillance footage, unusual occurrences typically last only a fraction of a second compared to typical ones. Furthermore, the ambiguity anomalies, and the nature of anomalies and large-scale anomaly patterns, to obtain labelled data in this setting anomalies. As a result, the majority of methods use a monolithic approach to learning for discriminating between typical and out-of-the-ordinary instances. We used multivariate Gaussian distribution to aid in LSTM and obtain the decision boundary for fitting normal data. To do this, an RBF layer is placed in front of the LSTM encoder's bottleneck.

The goal is to model all M_{enc}(Fenc(S0….n) training samples as Gaussian distributions, making it easy to spot outliers.

In order to effectively identify the outlier and label it as anomaly, the input video stream is converted to a form of Gaussian distribution. Training samples of the input video stream of the form of G_{enc}(Fenc(S0….n) are traines as gaussian distribution. The seq2seq encoded input video stream is tailored for gaussian distribution. The fitting process of the LSTM is performed with the help of RBF kernel. The logic of the entire process is to construct a hyperplane to receive the data that don't belong to the Gaussian distribution

Eqn 1 gives the definition of the kernel described in expressions of feature samples in mathematical form

$$\mu[y, y'] = e^{\left[-\frac{\|y-y^1\|}{2\sigma^2}\right]}$$

1

A dimensional vector for fitting the Gaussian distribution in the hyperplane is given by eqn. 2.

$$k[y, y^1] = e^{[-y\|y-y^1\|]}$$

2

The list $S^1_{0\ldots n}$ is the list of all sequences available at the end of training. The final list has a dimension of [N,n] if the sequence consists of N training input streams to the autoencoder. The elementwise μ and σ standard deviation are computed as

$$\sigma = \sqrt{\frac{\sum_{j-1}^{d}(S0\cdot.n^1)\mu^2}{N}}$$

3

After the completion of the training phase, the test of the video stream is performed. For testing the stream is given to the proposed algorithm. The algorithm determines whether the test data contains anomalies or not. The regularity metrics consists of two hyperparameters namely α and β .α is the limit of variation that the anomalies can contain. The majority of a population with a Gaussian distribution has its centre somewhere close to the distribution's mean.

The standard deviation is multiplied by the size of the test sample to determine how far off the mean it is. In the event that this difference is larger than, we label the sample as abnormal. The value of indicates the fraction of the latent space that contains outliers.

If the computed value of β * n has a greater number of anomalous values, in that case the entire test sample is taken as anomalous

```
Algorithm Anomaly Detection
Input (Video Frame, α, β)
Start
Compute α for video frame S
Compute β for video frame S
If α > sample mean && α > β
Check
{ if α >set limit
Print anomalous
}
Else
Print Not anomalous
End
```

4. EXPERIMENT

The proposed system was developed on a Intel Core i10- windows 11 operating system 16 GB RAM, Python, and TensorFlow installed workstation.

4.1. Dataset

The CUHK Avenue dataset was 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 camera, strolling on the grass and discarded items. The UCSD pedestrian Dataset, which 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.

The experimental setup is explained as follows

1) First, the proposed spatiotemporal autoencoder anomaly detection efficiency and accuracy is tested by using three standard datasets.

2) Second, we assess the OCC model's calculation of anomalies using the Gaussian distribution on UCSD Ped1 and Ped2 datasets, reclassifying a previously deemed anomalous event as normal.

3) Third, we implement k sliding window for finding the anomaly localization

4.2. Evaluation methodology

As the performance parameter, we employed the equal error rate (EER), to compare the suggested method with the state-of-the-art methodologies.

$$\text{EER}=\frac{\text{FP+FN}}{\text{TC}}$$

where False Positive (FP) indicates that the actual frame is classified correct but the result is a false positive, projected as normal, False Negative (FN) shows that the actual frame is abnormal even though it was forecasted as normal, and TC reflects the predicted amount of time that an event would take place. Count of all of the frames contained in the dataset being tested. Bring down the EER. value, the model is the superior choice.

Frame Level: - Predicting whether a given frame contains an abnormality is the focus of frame-level evaluation. Each frame's prediction is derived by calculating its anomaly score, and then, depending on the value of the defined threshold parameter the classification as anomaly or not anomaly id performed. It is often used as a yardstick by which anomaly detection models can be evaluated. After the model classifies a frame as normal or abnormal, it is compared to a ground-truth label to determine if the model was accurate.

Pixel Level: - The model can foretell which image pixels represent out-of-the-ordinary occurrences. To count the amount of genuine and erroneous positive frames, it is equated to ground-truth anomaly mark. The concept of assessing a model's ability to localise an abnormal event down to the pixel level has gained widespread acceptance among researchers.The calculation of precision, recall and FScore is as follows [29][30]

Literature	Frame level EER	Pixel level EER
Aboah et al. (2021)	29.7	33
Aliev et al. (2019)	43	18.3
Esquivel and Zavaleta (2022)	28	29
Proposed	25	27
Vasilopoulos et al. (2022)	33	43
Vosta and Yow(2021)	41	36

$$\text{Precision}=\frac{\text{TP}}{\text{(TP+FP)}}$$

$$\text{Recall}=\frac{\text{TP}}{\text{(TP+FN)}}$$

$$F\ Score = 2x\frac{Precision * Recall}{Precision + Recall}$$

5. RESULTS DISCUSSION AND COMPARATIVE ANALYSIS

Here, we present the model's performance on three different benchmark datasets and draw comparisons between them. The most problematic aspect of all three datasets is that only video streams marked as normal without any anomalies are freely available. The proposed system has to be built on this dataset, which consists of only normal records. As a result, the suggested strategy requires the detection and localization model to only use normal data for training. In all of the experiments performed, the k value of the sliding window was set to 5 to get uniform results across all three datasets. The value deviation more than the set value is used to find out the anomalies. The gaussian mean is calculated using eqn 3 and ifthe variance of α is greater than the gaussian mean, the object is marked as an anomaly.

5.1. UCSD Peds1 Data Set

For the UCSD Peds1 data set, we used parameters α and β. The value is set to 3 and β is set to 0.6 for the study. Frame-level and pixel-level ROC curves were obtained for the Peds1 dataset, with AUCs of 25% and 27%, respectively. Although the Peds1 dataset's elevated camera angle makes it challenging to distinguish normal from anomalous entities, we were still able to achieve an AUC of 27 when classifying pedestrians and skateboarders based on their appearance at the pixel level.

Table 1 Analysis for Ped 1 dataset

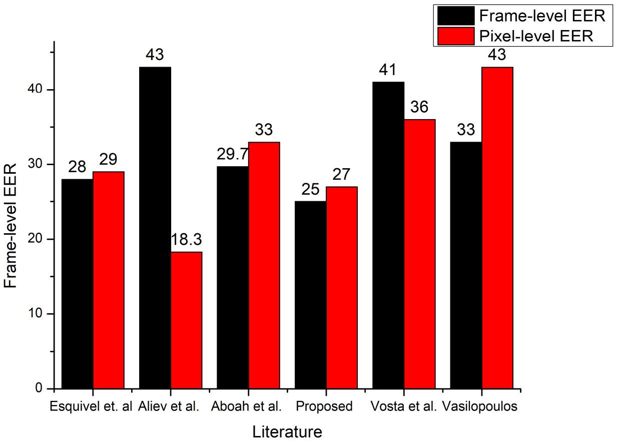

Figure 2: Result Analysis For UCSD Peds1 Dataset

5.2. CUHK Avenue Dataset

The scenarios used in the instructional films are realistic. During the testing process, both typical and unusual occurrences are filmed and analysed. The authors include both timebased and pixellevel annotations. The CUHK Avenue dataset is more difficult than the Peds1 dataset. The following are some of the difficulties encountered when working with this dataset. An average camera can detect a handshake every 50 frames. The training data contains a few extreme examples. In the context of training data, certain typical patterns arise seldom. we were able to acquire AUC values of 72% and 23%.

Table 2: Analysis for CUHK dataset

Literature	Frame level EER	Pixel level EER
Aboah et al. (2021)	27.2	79
Aliev et al. (2019)	24.1	77
Esquivel and Zavala (2022)	25	65
Proposed	23	72
Vasilopoulos et al. (2022)	24.6	82
Vosta and Yow(2021)	23.2	69

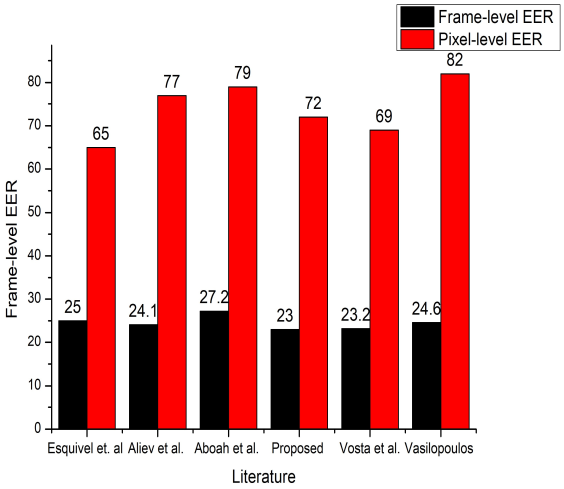

Figure 3 : Result Analysis For CUHK Dataset

As with other forms of supervised learning, the discriminative learning approach suggested in is one that starts with a training set. To differentiate between two successive video sequences, Ionescu used anrevealing strategy by training a classifier that is binary with the most discriminating features removed at each stage.

Table 3: Analysis For UMN Ped 2 Dataset

Literature	Frame level EER	Pixel level EER
Aboah et al. (2021)	88	52
Aliev et al. (2019)	97	47

Esquivel and Zavaleta (2022)	98.3	68
Proposed	89	87
Vasilopoulos et al. (2022)	75	42
Vosta and Yow(2021)	91	38

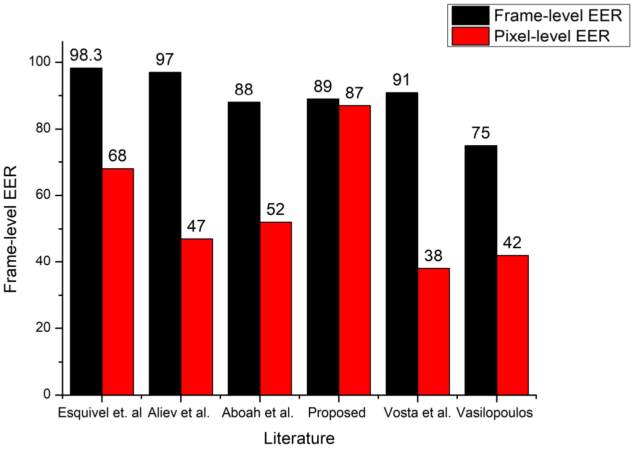

Figure 4: Result Analysis For UMN Ped 2 Dataset

The dataset from UMN contains three different settings where unusual occurrences took place. All of the scenes include pedestrians who suddenly break into a run. Two of the three scenarios take place in outdoor settings, while the third takes place in an indoor setting. Every one of the three sequences begins with a regular behaviour (walking) and ends with an abnormal behaviour (quick running). frames in the UMN dataset have a resolution of 320x240. Only temporal annotations, such as whether or not a frame has an anomaly, are present in the dataset. We trained the model exclusively on typical data and then put it to the test on outlier data. As shown in the UMN dataset, there are three distinct types of group anomalies. Sensitivity analysis on the UMN dataset yielded the same results as those on the previously stated datasets: $\alpha = 3$, $\beta = 0.7$.

6. CONCLUSION

The evaluation of the reference datasets has been demonstrated in this work. In order to demonstrate the use of real-world data and the execution of anomaly detection in an indoor setting. In this paper, for the purpose of anomaly identification within the framework of deep learning, we provide a pipelined autoencoder as well as an intelligent surveillance model. The system takes as input a video stream and performs analysis based on auto encoders. The combination of pipelined autoencoders with LSTM enhances the results at both bit and frame levels on the three datasets the experiment was carried upon.The frame-level and pixel-level EER for anomaly recognition on standardised datasets was obtained with good significance. The challenges included the absence of normal data alone, varying environmental factors, and diverse camera angles.When compared to alternative methods, the suggested model's low per-frame processing time makes it suitable for use in real-time video processing for anomaly identification.

6.1. Future Work

The authors, after having implemented the pipelined auto encoders with LSTM for detecting anomalies, would like to go further and add more features to the one class approach and implement it on a hardware device.

REFERENCE

[1] Saligrama, V., Chen, Z.: Video anomaly detection based on local statistical aggregates. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2112–2119. (2012)

[2] Sabokrou, M., et al.: Deep-cascade: cascading 3D deep neural networks for fast anomaly detection and localization in crowded scenes. IEEE Trans. Image Process. 26(4), 1992–2004 (2017)

[3] Hasan, M., et al.: Learning temporal regularity in video sequences. In:2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 733–742. (2016)

[4] Xu, D., et al.: Video anomaly detection based on a hierarchical activity discovery within spatio-temporal contexts. Neurocomputing. 143, 144–152 (2014)

[5] Xu, D., et al.: Detecting anomalous events in videos by learning deep representations of appearance and motion. Comput. Vis. Image Underst. 156, 117–127 (2017)

[6] Roshtkhari, M.J., Levine, M.D.: An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions. Comput. Vis. Image Underst. 117(10), 1436–1452 (2013)

[7] Kim, J., Grauman, K.: Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2921–2928. (2009)

[8] Xiao, T., Zhang, C., Zha, H.: Learning to detect anomalies in surveillance video. IEEE Signal Process. Lett. 22(9), 1477–1481 (2015)

[9] Marsden, M., et al.: Holistic features for real-time crowd behaviour anomaly detection. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 918–922. (2016)

[10] Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 935–942. (2009)

[11] Mahadevan, V., et al.: Anomaly detection in crowded scenes. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1975–1981. (2010)

[12]R. A. Aliev, J. Kacprzyk, W. Pedrycz, M. Jamshidi, M. B. Babanli, and F. M. Sadikoglu, Eds., 10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions - ICSCCW-2019, vol. 1095. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-35249-3.

[13]S. Vosta and K.-C. Yow, “A CNN-RNN Combined Structure for Real-World Violence Detection in Surveillance Cameras,” Applied Sciences, vol. 12, no. 3, p. 1021, Jan. 2022, doi: 10.3390/app12031021.

[14] E. Vasilopoulos, G. Vosinakis, M. Krommyda, L. Karagiannidis, E. Ouzounoglou, and A. Amditis, “A Comparative Study of Autonomous Object Detection Algorithms in the Maritime Environment Using a UAV Platform,” Computation, vol. 10, no. 3, p. 42, Mar. 2022, doi: 10.3390/computation10030042.

[15]M. Al-Sa’d, S. Kiranyaz, I. Ahmad, C. Sundell, M. Vakkuri, and M. Gabbouj, “A Social Distance Estimation and Crowd Monitoring System for Surveillance Cameras,” Sensors, vol. 22, no. 2, p. 418, Jan. 2022, doi: 10.3390/s22020418.

[16]A. Aboah, M. Shoman, V. Mandal, S. Davami, Y. Adu-Gyamfi, and A. Sharma, “A Vision-based System for Traffic Anomaly Detection using Deep Learning and Decision Trees,” in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, Jun. 2021, pp. 4202–4207. doi: 10.1109/CVPRW53098.2021.00475.

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

[18]S. S. Sarikan and A. M. Ozbayoglu, “Anomaly Detection in Vehicle Traffic with Image Processing and Machine Learning,” Procedia Computer Science, vol. 140, pp. 64–69, 2018, doi: 10.1016/j.procs.2018.10.293.

[19]D. Kim and T.-Y. Heo, “Anomaly Detection with Feature Extraction Based on Machine Learning Using Hydraulic System IoT Sensor Data,” Sensors, vol. 22, no. 7, p. 2479, Mar. 2022, doi: 10.3390/s22072479.

[20]S. M. Noe, T. T. Zin, P. Tin, and I. Kobayashi, “AUTOMATIC DETECTION AND TRACKING OF MOUNTING BEHAVIOR IN CATTLE USING A DEEP LEARNING-BASED INSTANCE SEGMENTATION MODEL,” p. 10, 2022.

[21]A. Osipov et al., “Deep Learning Method for Recognition and Classification of Images from Video Recorders in Difficult Weather Conditions,” Sustainability, vol. 14, no. 4, p. 2420, Feb. 2022, doi: 10.3390/su14042420.

[22]K. Pawar and V. Attar, “Deep learning model based on cascaded autoencoders and one-class learning for detection and localization of anomalies from surveillance videos,” IET Biometrics, vol. 11, no. 4, pp.

289–303, Jul. 2022, doi: 10.1049/bme2.12064.

[23]T. Sharma, B. Debaque, N. Duclos, A. Chehri, B. Kinder, and P. Fortier, "Deep Learning-Based Object Detection and Scene Perception under Bad Weather Conditions," Electronics, vol. 11, no. 4, p. 563, Feb. 2022, doi: 10.3390/electronics11040563.

[24]A. Paul, Md. A. S. Tajin, A. Das, W. M. Mongan, and K. R. Dandekar, "Energy-Efficient Respiratory Anomaly Detection in Premature Newborn Infants," Electronics, vol. 11, no. 5, p. 682, Feb. 2022, doi: 10.3390/electronics11050682.

[25]Z. Bai, G. Wu, X. Qi, Y. Liu, K. Oguchi, and M. J. Barth, "Infrastructure-Based Object Detection and Tracking for Cooperative Driving Automation: A Survey." arXiv, Mar. 19, 2022. Accessed: Sep. 26, 2022. [Online]. Available: http://arxiv.org/abs/2201.11871

[26]T. Ahmed, B. Oreshkin, and M. Coates, "Machine Learning Approaches to Network Anomaly Detection," p. 6.

[27]P. Y. Ingle and Y.-G. Kim, "Real-Time Abnormal Object Detection for Video Surveillance in Smart Cities," Sensors, vol. 22, no. 10, p. 3862, May 2022, doi: 10.3390/s22103862.

[28]A. A. Alsulami, Q. Abu Al-Haija, A. Alqahtani, and R. Alsini, "Symmetrical Simulation Scheme for Anomaly Detection in Autonomous Vehicles Based on LSTM Model," Symmetry, vol. 14, no. 7, p. 1450, Jul. 2022, doi: 10.3390/sym14071450.

[29] Nair, V., Kosal Ram, P. G., &Sundararaman, S. (2019). Shadow detection and removal from images using machine learning and morphological operations. *The Journal of Engineering*, *2019*(1), 11–18. https://doi.org/10.1049/joe.2018.5241

[30] Rajesh Banala, D.Upender,: "Remote Home Security System Based on Wireless Sensor Network Using NS2", International Journal of Computer Science and Electronics Engineering, India, Vol. 2 Issue 2 (2012).