Anomalous Detection System in Crowded Environment using Deep Learning

Dorcas Oladayo Esan
Department of Computer Systems Engineering
Tswhane University of Technology
Shoshanguve, South Africa
oladayojadesola10@gmail.com

Pius.A. Owolawi
Department of Computer Systems Engineering
Tshwane University of Technology
Shoshanguve, South Africa
OwolawiPA@tut.ac.za

Chuling Tu
Department of Computer Systems Engineering
Tshwane University of Technology
Shoshanguve, South Africa
duc@tut.ac.za

Abstract— **In recent years, surveillance systems have become very important due to security concerns. These systems are widely used in many applications such as airports, railway stations, shopping malls, crowded sports arenas, military etc., [1]. The wide deployment of surveillance systems has made the detection of anomalous behavioral patterns in video streams to become increasingly important. An anomalous event can be considered as a deviation from the regular scene; however, the distribution of normal and anomalous events is severely imbalanced, since the anomalous behavior events do not frequently occur, hence it is imperative to accurately detect anomalous behavioral pattern from a normal pattern in a surveillance system. This paper proposes a Convolutional Neural Network and Long Short-Term Memory (CNN-LSTM) technique. The CNN is used to extract the features from the image frames and the LSTM is used as a mechanism for remembrance to make quick and accurate detection. Experiments are done on the University of California San Diego dataset using the proposed anomalous behavioral pattern detection system. Compared with other existing methods, experimental analysis demonstrates that CNN-LSTM technique has high accuracy with better parameters tuning. Different analyses were conducted using the publicly available dataset repository that has been used by many researchers in the field of computer vision in the detection of anomalous behavior. The results obtained show that CNN-LSTM outperforms the others with overall F1-score of 0.94; AUC of 0.891 and accuracy of 89%. This result shows that the deployment of the proposed technique in a surveillance detection system can assist the security personnel to detect an anomalous behavioral pattern in a crowded environment.**

Keywords—Anomalous, Deep Learning, Surveillance, Convolutional Neural Network, Long Short-Term Memory Detection

I. INTRODUCTION

Anomaly detection refers to the techniques of finding specific patterns that do not conform to the normal patterns in the dataset. Anomaly detection has been one of the core research areas for a long time due to its ubiquitous nature[2]. A surveillance camera is one of the predominant security mechanism used by many public places such as airports, train stations, schools etc. [1], to protect lives and people's properties.

The conventional surveillance system records all activities that take place in the environment and are saved on the tape [3]. In the event of any crime being detected, the archived videotapes are retrieved, and the huge data are manually analyzed to detect the perpetrator [4]. This makes the current surveillance system limited in function since it lacks the intelligent capability to detect the probability of anomalous events before they occur. However, analyzing a huge dataset can be laborious and overwhelm the security experts who need the information to determine what is happening on the scene [5]. More specifically, they need to determine if the event is normal or anomalous so that quick action can be taken. Analyzing of behavioral patterns to differentiate between anomalous events from normal behavioral patterns in crowded environments are often done subjectively by security personnel to make a detection. However, in most cases this manual analyzing of voluminous data subjectively might not be accurate and thus makes effective and accurate detection of anomalous behavioral patterns to be challenging to the current surveillance systems [6].

With the fast-growing in crime rate, the deployments of the surveillance system have been ubiquitous and as such increases the market for surveillance cameras tremendously. According to the Information Handling Service (IHS) statistical estimation report on the amount invested on surveillance system worldwide from 2011 to 2016, the report revealed that of the total $8,5 billion was spent in 2010, approximately $11 billion was spent in 2011, in 2012 the total of $12.8 billion was invested on the surveillance system, $13.5 billion was spent in 2013, the total of $14 billion was invested worldwide on surveillance in 2014and approximately $15,5 billion was invested in 2016 [7] and with the continuous threats to the security in many organization, it is estimated that the world market for surveillance video cameras will increase significantly to $35 billion by 2021 [8]. These huge amounts invested in the surveillance system has had little or no effects on the accuracy of the current surveillance systems.

In fact, a lot of academic research work has been done on current surveillance security systems to determine anomalous behavioral patterns such as optical flow [8], spatiotemporal gradient [9], the social force model [10], chaotic invariant [11], and sparse representation [12] social force(SF)[9], a Mixture of Probabilistic Principal Component Analysis (MPPCA) [10], social force and a mixture of probability principal component analysis(SF-MPPCA)[10], Motion Deep Net(MDT)[11] have been studied and give satisfactory performance. These methods have contributed to the success of intelligent detection of anomalous behavior patterns through surveillance cameras. However, there is room for an improvement in terms of the degree of accuracy and model stability; hence, it is imperative to improve the existing state-of-the-art of surveillance systems in terms of detection accuracy when

978-1-7281-7624-6/20/$31.00 ©2020 IEEE
DOI 10.1109/CSCI51800.2020.00012

dealing with anomalous behavioral patterns in a crowded environment.

In this regard, this study develops deep learning which integrates the CNN-LSTM model to improve surveillance system detection accuracy and develop a resilient model in the current surveillance systems. CNN is a machine learning model that extracts features from the image and learns from reconstructed image frames while LSTM utilizes the gating mechanism to selectively keep vital pieces of information for a long time, which are then used to find sequential patterns to remember during detection stage in image frames. The experiment carried out using publicly available datasets on the CNN-LSTM model achieved an accuracy of 88.5%, and when benchmarked with other existing state-of-the-art detection models.

The organization of this paper is as follows: Related works are suggested in Section II; Theoretical Technique in Section III, proposed anomalous detection system is explained in Section IV, experiments and result in Section V, and conclusion in Section VI.

II. RELATED WORK

The work done in[12] introduced abnormal event detection in video surveillance. The approach uses a modified version of pre-trained 3D residual convolutional network to extract Spatio-temporal features, and develop a robust classifier based on the selection of vectors of interest which can learn the normal behavioural model and detect potentially dangerous abnormal events. The result showed the effectiveness of using the proposed method for detecting abnormal events. It was able to detect abnormal events, prevent the marginalization of normal behaviour that rarely occurs during the training phase and adapt to the appearance of new normal events in the testing phase. However, the performance of the approach deteriorates significantly under the influence of environmental noise such as illumination, which results in false detection errors.

The contribution in reference [13] introduced abnormal behaviour detection in video surveillance. The approach used a mixture of dynamic textures to detect spatial and temporal anomalies. The experimental result shows that the proposed method using dynamic textures efficiently detects abnormal behaviours in crowded scenes. However, the quantitative evaluation of the approach to images affected by noise data was not presented.

The detection of anomalies in extremely crowded scenes was presented in [14]. The approach utilised a Spatio-temporal motion pattern to characterise the behavioural patterns in the scene. This approach successfully detects unusual motion patterns in complex scenes but fails to give an appropriate solution to the issue of false detection that is rampant in surveillance. According to research in [15], an approach for locating anomalies in a crowded scene for surveillance videos was introduced. The approach used path prediction using Support Vector Machine (SVM) to smooth and locate the trajectory in crowded scenes. The result obtained shows that the proposed approach has a better anomaly localisation performance compared to other state of art systems but fails in feature descriptors and the pursue of other trajectory detection techniques.

Abnormal activities detection in the surveillance video was introduced in [10]. The approach used a Space-time Markov Random Field (MRF) model to compute a maximum posterior estimate of the degree of normality at each local node. The result shows proposed model robustly detects abnormal activities both in a local and global sense and accurately localize the abnormal activities in a crowded video. However, the approach gives high false error result and this makes it not to be suitable for detection of anomalous in highly crowded environments.

However, despite various classic detection methods utilized by researchers in the literature to solve the challenging issues facing the detection of anomalies behaviour in the surveillance system, there are little works that have adequately addressed the issue of detection accuracy that arise due to high false detection error in the surveillance system. This study addressed the aforementioned issue by utilizing the characteristics features of CNN for obtaining features from the image and learn from reconstructed image features while the Long Short-Term Memory (LSTM) gating mechanism selectively keeps some vital features for a certain period, and this is used to find consecutive patterns to remember during the detection of the behavioural patterns in image frames. The next section discusses the theoretical background of CNN and LSTM.

III. THEORETICAL TECHNIQUES

This section gives a brief introduction to the theory and implementation of combined convolutional neural network and long short-term model, with emphasis on the detection of anomalies in noisy image frames.

A. CNN

Convolutional Neural Network (CNN) is an artificial neural network approach which is used for anomaly detection in time-series data. CNN's are mainly used in computer vision for tasks like object detection, classification and segmentation. CNN uses convolutional layers that are partially connected, reducing the number of parameters, enabling them to go deeper and train faster. In addition to the convolution layers, CNNs also use pooling layers for regularization to avoid over fitting.

The convolution layers have 256, 128 and 64 channels respectively, and the filter size of each layer is 3×3, a max-pooling layer with a filter size of 2×2 and a stride of 2 follows each convolution layer. The convolution layers are activated by a Rectified Linear unit (ReLu) activation function. Convolutional Neural Networks have the following layers: Convolutional, ReLU Layer, pooling and fully connected and this as shown in Fig.1.

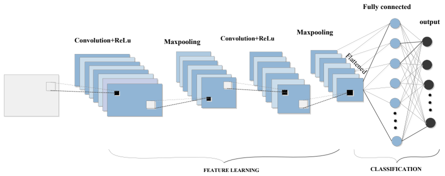


Figure 1. Convolution Neural Network (CNN) [16]

B. Convolution Layer

Convolutional Layer is the first layer to extract features (like edges, lines, and corners) from the input image by using the image-sharing weight and field accepting as shown in Fig.1 [16]. The convolutional layer conserves the relationship between pixels by learning image features using a small square of input data. The convoluted layer is used for learning filters that form feature maps, as in (1).

$$X_j^i = f(\sum_{i=1}^{l} x_l^{l-1} \times F_{i,j}^i + b_j^i), \;\; j \in [1, j] \qquad (1)$$

where x_j^i *is* jth is the layer for the channel, $l: F_{ij}^i$ is the i-th filter channel j in layer $l: b_j^i$ is filter bias for j of layer l; I *and* j are the amounts of channel layer $i-1$ and i, respectively. $f(.)$ represents Relu, known as the function for activation. The image output channel is based on the number of filters.

In image convolution, a filter is obtained from an image by multiplying the image pixel element using the input receptive field with the addition of each channel multiplication before moving to the appropriate stride. The output image feature maps are computed as in (2).

$$D_{out} = (D_{in} - C) / S_{conv} + 1 \qquad (2)$$

where D_{in} is the input layer side length, X^{1-1}:C is the filter side, D_{out} is the output layer side length, X^1, and S_{conv} is the height and width stride.

C. ReLu Layer

Rectified linear unit (ReLu) is one of the most notable non-saturated activation functions. This layer increases the nonlinear properties of the decision function and the overall network without affecting the receptive fields of the convolution layer. It effectively removes negative values from an activation map by setting them to zero. The ReLu activation function is defined as in (3).

$$\mathrm{Re}\,LU(x) = \max(0, x) \qquad (3)$$

$$\frac{d}{d_x} \mathrm{Re}\,LU(x) = \{1\, if\ x > 0; 0\, otherwise \qquad (4)$$

D. Pooling Layer

Pooling is an important concept of CNN. It lowers the size of the patch image features by down-sampling the pixels as shown in Fig.1. Also, it helps in obtaining deeper representations of consecutive layers as well as preventing over fitting and makes the features robust against noise and distortion [17]. This layer helps in Features are sub-sampled autonomously in each channel; the features are sub-sampled autonomously using selected value representatives from receptor fields as in (5).

$$X_j^i = X_i^{l-1 \times P_i^l}, \; i \in [1, l] \qquad (5)$$

where X_j^i is the $X_j^i\, j - th$ channel of layer l; P_i^l is the $i - th$ channel of the pooling operator in layer l; and I is the channel amount of both layer $l - 1$ and l. The in-plane size of the output feature maps is calculated as in (6).

$$D_{out} = (D_{in} - p_{pool}) / S_{conv} + 1 \qquad (6)$$

where P_{pool} is the side length of the pooling operator (assume the operator is a square f size $P_{pool} \times P_{pool}$) and S_{pool} is the stride (both in height and width).

E. Fully Connected (Dense) Layers

Fully connected (Dense) layers are used as the final layers of a CNN where the input from other layers are converted into the vector. Neurons in a fully connected layer have connections to all activations in the previous layer (these layers mathematically sum a weighting of the previous layer of features to determine a specific target output result). It transforms the output into any desired number of classes into the network.

One of the advantages of CNN is its ability to train autonomously from reconstructed image features and improve the training performance by weight sharing.

F. Long Short-Term Memory

Long Short-Term Memory (LSTM) is a unique class member of recurrent neural network (RNN) units with different gates connected to different neurons, it was designed to model temporal sequences and their long-range dependencies more accurately than conventional RNNs. The LSTM contains special units called memory blocks in the recurrent hidden layer. The memory blocks contain memory cells with self-connections which help in storing the temporal state of the network in addition to special multiplicative units called gates to control the flow of information. Each memory block in the original architecture contained an input gate and an output gate. The input gate controls the flow of input activations into the memory cell. The output gate controls the output flow of cell activations into the rest of the network. Later, the forget gate was added to the memory block [18].

Also, the modern LSTM architecture contains peephole connections from its internal cells to the gates in the same cell to learn the precise timing of the outputs [19]. LSTMs is a unique class member of RNN units with different gates connected to different neurons, as shown in Fig. 2.

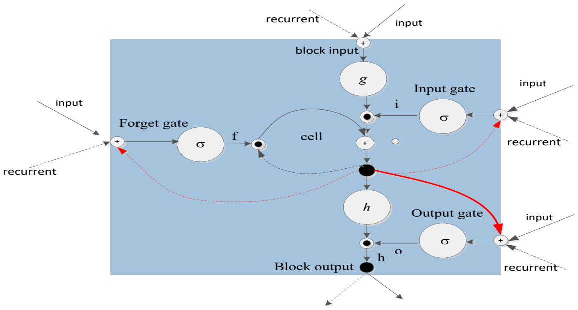


Figure 2. Long Short-Term Memory (LSTM) [20]

These four gates include the input state gate at time (X_t), forget (f_t), cell state gate (C_t), and hidden state gate at time (h_t). The four gates in the LSTM are represented as in (7) - (13).

$$X_t = \sigma_i (x_t W_{x_i} + h_{t-1} W_{h_i} + W_{c_i} \Theta + C_{t-1} + b_i) \qquad (7)$$

$$f_f = \sigma_f(x_t W_{xf} + h_{t-1} W_{h_i} + W_{cf} \Theta + C_{t-1} + b_f) \quad (8)$$

$$C_t = f_t \Theta C_{t-1} + i_t \Theta \ \sigma_c(x_t W_{x_c} + h_{t-1} W_{h_c} + b_c) \quad (9)$$

$$o_t = \sigma_0(x_t W_{x_0} + h_{t-1} W_{h_0} + W_{c_0} \Theta C_t + b_0) \quad (10)$$

$$h_t = (1 - U_t)\Theta h_{t-1} + U_t \Theta C_t \quad (11)$$

$$g = \tanh((x_t + s_{t-1}) W_g + b_g) \quad (12)$$

$$S_t = \tanh(C_t) \cdot o_t \quad (13)$$

where W is the weights matrices $(W_{x_i}, W_{h_i}, W_{c_i}, W_{c_f})$ is the matrix of weights from the input gate to the input), σ denotes tanh non-linearity, (X_t) is the input gate, (f_t) is the forget gate, (O_t) is the output gate, (C_t) represents the cell state, (h_t) is the hidden state.

From equation (7), the cell state adds the input of the current state with the previous hidden state to keep the current cell state informed. C_t is the memory cell that stores historical information. σ represents the sigmoid activation function. $W_{x_i}, W_{h_i}, W_{c_i}, W_{c_f}, W_{x_f}, W_{h_f}, W_{x_0}, W_{x_c}, W_{h_0}, W_{h_c}$ are Convolutional filter weights, b_i, b_f, b_0 and b_c represent the convolutional bias vector, i, f, o, c are respectively. The input gate, forget gate, output gate and cell activation vectors, g and h are the cell input and cell output activation functions, "σ", is the sigmoid activation function and the element-wise multiplication represents Θ.

One of the advantages of using LSTM for detection in image anomalies in its gating mechanism for keeping vital information in the image for a long time and using it to find sequential patterns [21].

IV. PROPOSED ANOMALOUS DETECTION SYSTEM

A. *CNN-LSTM Framework for Anomalous Detection*

This section focuses mainly on the development of the CNN-LSTM framework for anomalous detection. It uses the detected result for effective and efficient security planning. The components of this system model are divided into three phases: image acquisition stage, video image pre-processing stage and deep learning (CNN and LSTM) stage, as shown in Fig. 3.

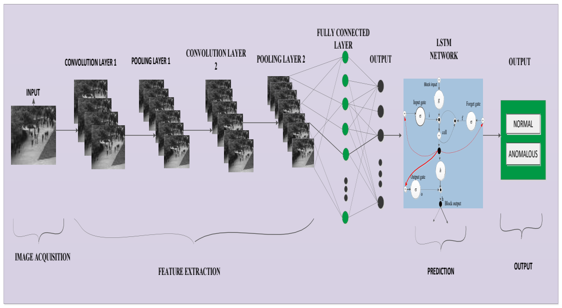


Figure 3. System Architecture of Anomalous Detection using CNN-LSTM

B. *Image Acquisition Stage*

The data used in this experiment was obtained from the publicly available dataset (UCSD) [22].

C. *Feature Extraction Stage*

The acquired image frames are passed into a convolutional layer and the pooling layer through the output of a max-pooling layer as in (14).

$$X_l^i = \tan\left(\sum_i x_j^{i-1} \times k_{ij}\right) + b_j^i \quad (14)$$

where X_j^i are feature maps generated by the convolutional layer 1, x_j^{i-1} are feature maps of the prior convolutional layer $l-1$, k_{i_j} are the trained convolution kernels and b_j^i is the additive bias. Finally, pooling max ⊙ is the max pooling operation and tanh ⊙ is the hyperbolic activation function.

D. *Detection Stage Using CNN-LSTM*

The output of CNN is passed into the input of the LSTM units consist of memory blocks which are controlled by memory cell, input and output gate, and peephole connection as in equation (7) – (13) to detect the behavioral pattern from the image frames as either normal or anomalous.

E. *Performance Evaluation Mechanism*

This section presents the result of proposed CNN-LSTM methodologies as described in sections IV. The result of the proposed CNN-LSTM model is provided in detail. The performance of CNN-LSTM is evaluated using a confusion matrix as well as cross-validation, as explained in subsequent sections.

F. *Cross-validation Evaluation*

Cross-validation is employed in this case as a technique used to validate the performance of the CNN-LSTM model. It involves using k-folds of a cross-validation scheme; for instance, this experiment is set to use 90% for training and 10% for testing.

G. *Confusion Matrix*

A confusion matrix is a table that represents the number of behavioral patterns that are correctly classified or instances that are incorrectly classified by the model, as shown in Table 1.

TABLE I. CONFUSION MATRIX FOR THE DETECTION OF ANOMALIES IN A NOISY IMAGE

	Normal	Anomalous
Normal	*TP*	*FN*
Anomalous	*FP*	*TN*

In Table I, TP represents a true positive, TN is a true negative, FP is a false positive and FN is a false negative. The terminology used in Table I is further explained in equation (15) - (18).

[i] Precision: This is defined as the ratio of the correctly predicted positive observation (normal behavioral patterns) to the total number of positive observations correctly predicted [23]. This is computed in (15).

$$\Pr ecision = \frac{TP}{TP + FP} \quad (15)$$

[ii] Recall: This is also known as sensitivity and is defined as the ratio of correctly predicted positive observations to all observations of the actual class [23]. This is computed in (16).

$$\text{Re}call = \frac{TP}{TP + FP} \quad (16)$$

[iii] F1 Score: This is defined as the weighted average of precision and recall. This score takes both false positives and false negatives into consideration [23]. This is computed as in (17).

$$F_1 = 2 \times \frac{precision \times \text{Re}call}{\Pr ecision + \text{Re}call} \quad (17)$$

[iv] Accuracy: This is the ratio of behavioral patterns correctly detected by the model, reflected in (18).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (18)$$

The performance metric in equations (15) - (18) are used in section IV to evaluate the performance of the model to detect the anomaly detection system in a dynamic environment such as a university environment.

[v]Area under the Curve (AUC): this is calculated for the Sensitivity- Specificity curve and Decision Rate-False Alarm Rate was also used to compare the performance.

H. Dataset

The data set used is a publicly available UCSDped1. Pedestrian 1 dataset and only Pedestrian 1 is used in this work[22]. The data set consists of 200000 image frames with a resolution of158×238. The samples of frames with anomalous behavior are shown in Fig. 4.

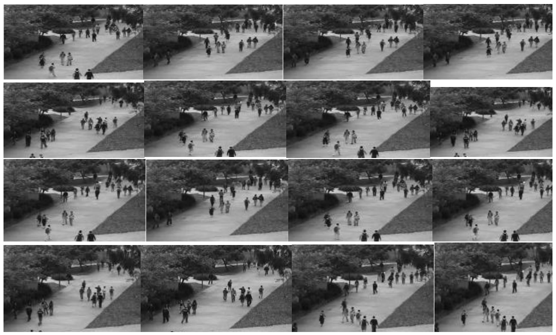


Figure 4. Samples of the anomalous dataset used [22]

V. Experiments and Results

This section introduces the experimental setup considered in this work, and the adopted integrated model using the CNN-LSTM algorithm. The experimental result of suspicious behavioral patterns is discussed. Also, the comparison results of the CNN-LSTM with other existing anomalous detection techniques is presented in a tabular format. The considered procedure and parameter selection used to optimize the experiment performance are explained in section C and D.

A. System Setup and Packages

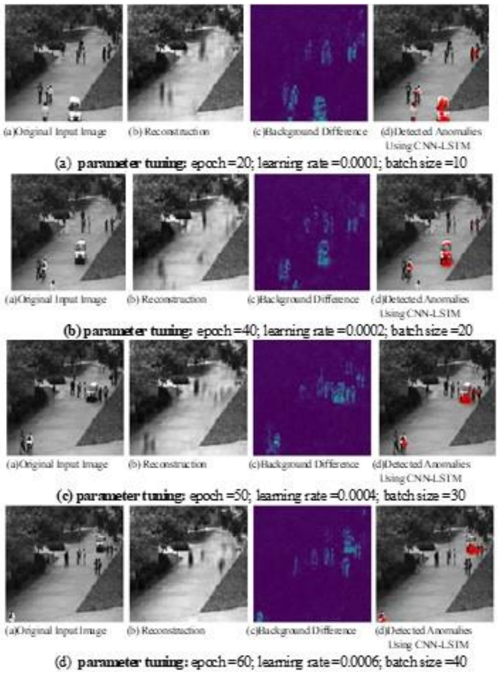
In this work, all programming was done using Python 3.6.9 (Anaconda 3) was used as a platform to implement the systems with necessary python libraries Spyder 3 IDE platform for running the python programs. For the implementation of the CNN-LSTM technique, the following libraries are used: Keras, Tensor flow library, Scikit-image, NumPy, SciPy Pillow/Pill, Matplotlib. Also, the hardware used for the CNN-LSTM model requires a high configuration of GPU for high performance to train maximum network size.

B. Parameter Selection for the Implementations

Parameter tuning becomes more and more expensive in terms of computing time with an increasing number of iterations. Hence, the objective of using the parameters tuning iterations to find a nearly optimal result for the CNN-LSTM model. The parameters that are embedded in this model consists of batch size and epoch which is used for training each image frame. Parameters tuning for batch size of 60 and epoch of 100 which is used for the training of the model. The details of the experiments are further explained in the next section.

C. Results and Analysis: Quantitative Observations of CNN-LSTM technique on anomalous behavioural patterns involving moving vehicle

Similar to experiment 1, this section evaluates the performance CNN-LSTM technique on the anomalous image frame as a moving vehicle. The quantitative analysis of the proposed technique on the normal behavioral pattern and anomalous behavioral pattern such as moving vehicle using varied parameters tuning (epoch, batch values and learning rate). The effect of the parameter tuning on the anomalous image frames is as shown in Fig. 5.



(a) parameter tuning: epoch=20; learning rate =0.0001; batch size =10

(b) parameter tuning: epoch=40; learning rate =0.0002; batch size =20

(c) parameter tuning: epoch =50; learning rate =0.0004; batch size =30

(d) parameter tuning: epoch=60; learning rate=0.0006; batch size =40

Figure 5. Quantitative observations of anomalous behavior involving a moving vehicle in a pedestrian walkway in the image frame

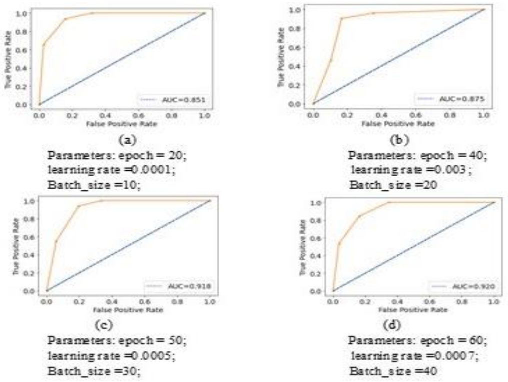
From Figs. 5(a)-(d) the effect of the parameter tuning on the anomalous image frames as a moving vehicle in a pedestrian walkway can be seen. From this Fig. 5, one can see that the anomalous detection from the image frames in Fig. 5(d) produces better visual detection accuracy compared to other parameter tuning result obtained in Fig 5(a)-(c).

To further show the performance accuracy of the CNN-LSTM on the behavioral patterns as a moving vehicle, a qualitative experiment was conducted using a cross-validation technique on the set of images obtained. During the cross-validation, 90% of the dataset was used for training while the remaining 10% was used for testing. The experiment is done four times using varied parameter tuning and the result of all the experiments conducted on pedestrian walkway dataset is shown in Table II.

TABLE II: RESULT OF AN EXPERIMENT CONDUCTED ON ANOMALOUS BEHAVIOR AS A MOVING VEHICLE WITH VARIED PARAMETERS

batch	epoch	TP	TN	FP	FN	F1	AUC	Acc
10	20	1650	92	8	250	0.92	0.851	87
20	40	1680	95	5	220	0.93	0.875	88
30	50	1700	97	3	200	0.94	0.918	89
40	60	1750	100	0	150	0.96	0.920	92
Total Accuracy						0.94	0.891	89

From this Table II, one can see that as the best combination of parameters (epoch = 60; batch number = 40; learning rate = 0.0006) produce FP of zero value with FN of 150 and the accuracy of 92%. To further interpret the result presented in Table II for the anomalous detection system results, the receiver operating characteristics (ROC) curve, which is the graph of the true positive rate (TPR) and false-positive rate (FPR) with varied values, is shown in Fig.6.



(a)
Parameters: epoch = 20;
learning rate =0.0001;
Batch_size =10;

(b)
Parameters: epoch = 40;
learning rate =0.003 ;
Batch_size =20

(c)
Parameters: epoch = 50;
learning rate =0.0005;
Batch_size =30;

(d)
Parameters: epoch = 60;
learning rate =0.0007;
Batch_size =40

Figure 6. ROC curve with varied parameters for anomalous behavior as a moving vehicle

Fig.6(a) shows the corresponding ROC curve for anomalous behavior as a moving vehicle using the CNN-LSTM model at Epoch 20. Fig.6 (b) shows the corresponding ROC curve for performance accuracy of CNN-LSTM technique on anomalous behavior as moving vehicle with parameter tuning of an epoch of 40; learning rate of 0.0002 and batch size of 20. Fig.6(c) shows the corresponding ROC curve for the performance accuracy of CNN-LSTM technique on anomalous behavior moving vehicle with parameter tuning of an epoch of 50; learning rate of 0.0004 and batch size of 30. Finally, Fig.6(d) shows the corresponding ROC curve for the performance of CNN-LSTM on anomalous behavior as moving with parameter tuning of an epoch of 60; learning rate of 0.0006 and batch size of 40. One can see that at Fig. 6(d) gives optimal F1-score of 0.96, the accuracy of 92% and Area under Curve (AUC) of 0.920.

The accuracy obtained from Fig. 6 confirms that CNN-LSTM approach can easily be used to augment the work of security personnel by assisting them to raise quick awareness in a situation where there is anomalous behavior in a university environment to prevent hazardous events. The proposed CNN-LSTM approach gives an overall average of F1-value of 0.94, AUC of 0.891 and accuracy of 89%. This performance verifies that CNN-LSTM can effectively be used for anomalous detection in a crowded environment such as university.

D. Benchmarking the Proposed CNN-LSTM with other Existing Detection Techniques

This section compares the accuracy of the proposed technique with other methods to detect behavioral anomalies on UCSDped1 and University of Minnesota (UMN) benchmark datasets. The methods include social force(SF)[9], a Mixture of Probabilistic Principal Component Analysis (MPPCA) [10], social force and a Mixture of Probability Principal Component Analysis(SF-MPPCA)[10], Motion Deep Net(MDT)[11]. The comparison was also done in terms of the method used, the dataset used, F1-score, AUC and the accuracy drawn from each method. A summary of existing suspicious detection techniques and resulting characteristics are depicted in Table III.

Table III: PERFORMANCE COMPARISON OF THE COOPERATIVE MEDIAN FILTERING AND KNN TECHNIQUE WITH OTHER EXISTING TECHNIQUES

Method Used	Dataset Used	F1-score	AUC	Accuracy (%)
SF[9]	UMN	-	0.73	79%
MPPCA [10]	UCSD (ped1)	-		82%
MDT [11]	UCSD (ped1)	-		55%
SF-MPPCA [10]	UCSD (ped1)	-		65.51%
Proposed CNN-LSTM	UCSD (ped1)	0.94	0.891	89%

The literature cited in Table III presents several methods used for the detection of anomalous behavioural patterns compared with the proposed method, and this contains the sparse combination learning framework (SCLF) in [36], the mixture of probabilistic principal component analysis (MPPCA) approach in [10], the social force model (SF) in[9], and their extension (SF+MPPCA) in [10], a mixture of dynamic texture (MDT) in[11]. The proposed model integrates the CNN on LSTM and uses CNN to extract features while the LSTM is used for remembrance and detection, making it more suitable for the problem addressed in this research. Also, the technique yields better performance when used for detection in crowded and big data samples. This result is evidence that the CNN-LSTM approach can be a useful application that can assist security officers in the detection of anomalous behaviour in a university environment before it leads to crime.

VI. Conclusion

The integration of CNN-LSTM for the detection of anomalies in a university environment was presented. CNN is used to learn image features and LSTM for remembering the vital features for future use. The CNN-LSTM approach is evaluated on the publicly available UCSDped1 dataset consisting of 20000 image frames. It provides a detailed comparison with some other existing techniques for detecting anomalous behavior. From the result obtained, the CNN-LSTM gives overall accuracy of 88.5%, F1-score of 0.935 and AUC of 0.887 which outperforms the others existing state of the art of anomalous detection techniques as shown in Table II. With these results one can see that CNN-LSTM technique can be used to augment the work of security personnel in the detection of anomalous behavior in a university environment and hotspot regions. Furthermore, the qualitative result reveals the anomalous and normal region within the frame. Further work could be done to investigate the time it takes to apply the proposed approach in the detection of anomalous behavior.

Acknowledgment

The authors acknowledge the contributions and financial support of the Department of Computer Systems Engineering, Tshwane University of Technology.

Reference

[1] V. A.Kotkar and V.Sucharita, "a comparative analysis of machine learning based anomaly detection techniques in video surveillance " *journal of engineering and applied sciences,* pp. 9376-9381, 2017.

[2] C. C. Aggarwal, "An introduction to outlier analysis, in Outlier Analysis.Cham," pp. 1-40, 2016.

[3] S. V. Rajenderan and K. F. Thang, "Real-Time Detection of Suspicious Human Movement " *international Conference on Electrical Electronics Computer Engineering and their Applications,* 2014.

[4] K. C. Baumgartner, S. Ferrari, and C. G. Salfati, "Bayesian Network Modelling of Offender Behaviour for Criminal," *Master of Science, Department of Mechanical Engineering and Material science, Duke University,* 2005.

[5] J. Lepon and R. RPopkin, *A study of CCTV at Harvard*, 2007.

[6] T. Zhang, Y. Y. Tang, Z. Shang, and X. Liu, "Face Recognition Under Varying Illumination using Gradientfaces," *IEEE Transactions* vol. vol.18,no.11, pp. 2599 – 2606, 2009.

[7] J. Cropley, "video surveillance growth set to exceed 10 percent in 2015," ed: IHS Markit, 2015.

[8] J. Cropley, "Global Professional Video Surveillance Equipment Market Set for Third Year of Near Double-Digit Growth in 2019," in *IHS Market Video Surveillance Intelligence Service*, ed, 2019.

[9] R. Mehran, A. Oyama, and M. Shah., "Abnormal crowd behaviour dtection using social force model. In Computer Vision and Pattern Recognition (CVPR)," 2009.

[10] J. Kim and K. Grauman, "Observe locally, infer globally:I. A space-time MRF for detecting abnormal activities with incremental updates. ," *IEEE Conference on Computer Vision and Pattern Recognition,,* pp. 2921–2928., 2009.

[11] V. Mahadevan, V. B. W. Li, and N. Vasconcelos, "Anomaly detection in crowded scenes. In Computer Vision and Pattern Recognition (CVPR)," 2010.

[12] S. Bouindour, H. Snoussi, M. M. Hittawe, N. Tazi, and T. Wang, "An On-Line and Adaptive Method for Detecting Abnormal Events in Videos Using Spatio-Temporal ConvNet," *Applied Sciences,* 2019.

[13] N.Divya, S. A. Begum, and Dr.A.Askarunisa, "Unsupervised Video Anomaly Detection," *Karpagam Journal of Engineering Research,* vol. Volume No.: II, , 2015.

[14] L. Kratz and K. Nishino, "Anomaly Detection in Extremely Crowded Scenes Using Spatio-Temporal Motion Pattern Models," *IEEE,* 2009.

[15] T. Zhang, A. Wiliem, and B. C. Lovell, "Region-Based Anomaly Localisation in Crowded Scenes via Trajectory Analysis and Path Prediction," *IEEE,* 2013.

[16] S. Albelwi and A. Mahmood, "A Framework for Designing the Architectures of Deep Convolutional Neural Networks," 2017.

[17] Z. Tang, Z. Chen, Y. Bao, and H. Li, "Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring," *WILEY,* vol. 1, pp. 1-12, 2018.

[18] F. A. Gers, J. Schmidhuber, and F. Cummins, " "Learning to forget: Continual prediction with LSTM," Neural Computation," vol. 12, pp. 2451-2471, 2000.

[19] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," *Journal of Machine Learning Research,* vol. 3, pp. 115-143, 2003.

[20] T. Fischer, C. Krauss, and E. J. Oper, " Deep learning with long short-term memory networks for financial market predictions," pp. 270, 654–669, 2018.

[21] C. Guggilla, T. Miller, and I. Gurevych, "CNN and LSTM based Claim Classification in Online User Comments," in *International Conference on Computational Linguistics*, 2016, pp. 2740-2751.

[22] "http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm," ed.

[23] P. Chouhan and V. Richhariya, "Anomaly Detection in Network using Genetic Algorithm and Support Vector Machine " *international Journal Computer science and Information technologies* vol. 6, pp. 4429-4433, 2015.