IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 37, No. 4, August, 2023, pp. 955-962
Journal homepage: http://iieta.org/journals/ria

Smart Crowd Monitoring and Suspicious Behavior Detection Using Deep Learning

Chaya Jadhav, Rashmi Ramteke*, Rachna K. Somkunwar

Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune 411018, India

Corresponding Author Email: rashmiramteke1121@gmail.com

https://doi.org/10.18280/ria.370416

Received: 10 May 2023
Revised: 22 May 2023
Accepted: 31 May 2023
Available online: 31 August 2023

Keywords:
Long Short-Term Memory, Visual Geometry Group (VGG16), crowd monitoring, Fully Convolutional Networks, Internet of Things, deep neural networks, public safety, real time detection

ABSTRACT

In the face of burgeoning population growth, ensuring security during public events, familial gatherings, and in high-traffic areas has become increasingly challenging. The manual monitoring of these areas, though facilitated by closed-circuit television (CCTV) cameras, often proves laborious and error-prone, leading to potential oversight of suspicious activities within crowds. To ameliorate this issue, an intelligent system for crowd monitoring and suspicious activity detection has been developed, utilizing deep learning algorithms. Specifically, the combined use of Fully Convolutional Networks (FCN) and Long Short-Term Memory (LSTM) was employed in the analysis of crowd behavior. Although previous attempts have been made to address this issue, the accuracy of such systems has remained a concern, often marred by false alarms and overlooked incidents. However, the present system exhibits a marked reduction in both false positives and negatives, boasting an accuracy of 97.84%, a significant improvement over existing model. This research proposes an effective solution to the problem of manual crowd monitoring, offering enhanced security outcomes through intelligent, automated surveillance. The high accuracy achieved underlines the potential of deep learning techniques in revolutionizing the field of surveillance, with further implications for crowd management and public safety.

1. INTRODUCTION

Given the escalating global population, the complexity of human behavior, and the prevalence of densely populated environments, a prevailing sense of insecurity is often experienced. This unease has spurred a demand for vigilant security personnel, tasked with assuring the safety of the public. However, the continuous monitoring for suspicious activities, particularly within crowded areas, poses a formidable challenge even for the most adept security personnel. As a response to this exigency, closed-circuit television (CCTV) cameras have been developed and deployed. Nonetheless, the use of CCTV cameras presents its own set of challenges. Surveillance footage is typically stored in real-time within expansive databases, which are subsequently scanned when suspicious activities are suspected. Yet, the detection of anomalies within crowded environments remains an arduous task, even when using these extensive databases. These circumstances underscore the need for the development of a real-time system capable of detecting suspicious behavior or human anomalies. Such a system would facilitate continuous monitoring of densely populated areas, effectively manage crowds, and proactively identify anomalies [1]. This constitutes the primary focus of the current study, aimed at advancing the field of surveillance technology to effectively address the security challenges posed by our rapidly evolving society.

Researchers and practitioners from the fields of image processing, computer vision, machine learning, and deep learning collaborate on research initiatives to solve these disadvantages. The difficulty of spotting odd human behavior in public spaces must be thoroughly investigated by each person. Crowds are now a typical sight in public spaces such as highways, plazas, athletic arenas, bus terminals, train stations, and airport terminals. Again, public events like assemblies, concerts, sports, rallies, and marches may draw big crowds [2]. High levels of risk, insecurity, and management requirements are constant. However, it appears that there is poor administration and security of gatherings. We constantly search for anomalies in real-time, but our only resources are recorded movies [3, 4]. We check these stored database systems when anything unexpected occurs in public spaces to determine what transpired. How did it take place? and who should be accountable? However, the harm had already been done at this point. We constantly search for a crowd behavioral evaluation system for the automobile industry while taking into consideration all of these elements.

Monitoring a crowded area and location is a difficult task, which explains a wide variety of activities. A crowd's behaviour might be unpredictable for a time. Crowd management requires some time apart since the crowd's behavior cannot be restructured while it is engaging in the same act or for the same event [5]. Due to the fact that human behavior varies, even when it serves the same aim or does not acknowledge that the primary goal is risky [6]. Traffic congestion and human activity are risky. The viruses that cause corona, swine flu, avian flu, and other respiratory illnesses are also likely to spread because they are skin-to-skin transmissible [7]. Even to warn of riots and terrorist acts carried out in a public setting as a result of the crowd, it is crucial to comprehend their behavior. To safeguard the person, the surroundings, and the built environment, a number of strategies and procedures have been outlined. Tracking congestion behavior may be described by a number of

computer vision methods, although these algorithms occasionally fail for unknown reasons [8, 9].

This work aims to develop a system that can classify normal and abnormal crowd behavior using a real-time video surveillance system to monitor highly populated metropolitan areas in order to cope with the challenges of these systems. This technology helps to prevent crime since the video that was obtained is powerful evidence against the offender and because crowd analysis will help security agencies stop crowd-related illegal behaviors like riots, etc. [10]. Since 1990, virtual environments have been a topic of discussion among computer vision researchers, however with certain restrictions due to a lack of data. The administration of intelligent settings has become a particularly difficult challenge for public security or safety due to the ongoing development of the population and digital information. As a result, computer vision challenges related to social and computational elements have been explored in relation to crowd monitoring and analysis using surveillance cameras [11]. Early on, computer vision algorithms provided great support for video surveillance systems; however, due to crowd density, there was an observable decline in the ability to record the system for identifying and following people from one group to another. There are three main procedures that are used while analyzing the behavior, estimating the density, and detecting crowd activities [12]: i) Pre-processing by segmentation; ii) Identifying an object's individual and group; iii) Recognizing an event or behavior. The old traditional inspection required a lot of time, therefore researchers applied artificial intelligence (AI) techniques to analyze and categorize the crowd scenario. AI makes it possible to track and estimate in real time the behavior of individuals present in congested areas quickly and accurately. Recently, governments have employed AI technology for a variety of purposes, including fraud detection, recruiting, advertising, and online dating monitoring. The deep learning (DP) model achieves extraordinarily high accuracy in a range of tasks, including detection, identification, and classification. It does this by using multi-layered artificial neural networks (ANNs). The bulk of researches have proven that DP types like CNN and recurrent neural network (RNN) are effective for classifying data. CNN and RNN are used to recognize digital images or movies and examine crowd behavior [13-18].

The proposed endeavor is driven by the COVID-19 pandemic, which sees crowds as its main enemy and social distance as a critical approach for slowing its expansion, and is based on the benefits of such deep learning systems. This work has developed a deep learning FCN + LSTM model to track social distance using digital photos and security footage. FCN LSTM is basically useful for prediction sequential problems like the one discussed in this research. FCN is useful and act as a feature extractor and FCN + LSTM helps for identifying the activities from the sequential image frames. The proposed technique consists of two steps: crowd analysis and suspicious crowd behavior identification using datasets with different densities. To verify the proposed FCN + LSTM model, assessment measures are used to the crowd sourcing data. The model provides better performance and in future it may led to better applications with the required and genuine changes in the architecture.

The rest of this article is divided into the following sections: Section 2 presents the pertinent work. In Section 3, FCN's and LSTM's outlining technique is explained. The suggested approach to crowd monitoring is described in Section 4. The suggested crowd monitoring strategy's findings are examined and discussed in Section 5. The research findings are concluded at Section 6.

2. EXISTING METHODS

Various machine learning and deep learning approaches have been used in numerous publications analyzing crowd behavior. The datasets, algorithms, and techniques employed by the authors, as well as the observed findings and future potential, are carried out in order to identify effective approaches for detecting anomalous crowd behavior in a variety of crowded scenarios. In this area, crowd behavior analysis research has been going on for a while. This section provides information on the numerous strategies employed as well as the system that is now in place.

A novel methodology for encoding video information was developed by Chong and Tay [19], using a deep learning technique and a large number of essential highlights that are inherently suggested from a lengthy video clip. A deep neural network was utilised to analyse video frames that had uncommon informational components that, when combined, constitute the video representation. The deep neural network was particularly built using a number of convolutional autoencoders. This representation is processed into a collection of convolutional short autoencoders in order to learn the regular temporal patterns. The neural network's data sources and yields are now completely independent of one another.

Albattah et al. [20] proposed an image classification, crowd control, and warning system for the Hajj. CNN, a deep learning method, is used to classify images. Recently, the scientific and industrial sectors have become interested in several applications of CNN for voice recognition and picture classification. The goal is to train the CNN model to classify crowds as densely packed, crowded, semi-busy, slightly congested, and normal using mapped picture data.

Convolutional neural networks (CNNs) are similar to basic neural networks (NNs) in that they are composed of neurons and receptive fields with trainable weights and biases. Each receptive field receives a group of inputs, performs a convolution, and then feeds the results into a nonlinearity function (such as ReLU or Sigmoid, for instance) [21]. The hidden layers gain rich properties that enhance the performance of the network as a whole (classifier and hidden layers), since CNN assumes that the input image is an RGB image. This structure offers benefits in terms of speed and accuracy because there are several things to be recognised in the crowd scene photographs. End-to-end networks are ones in which, upon receiving an input image, the network instantaneously produces the desired output.

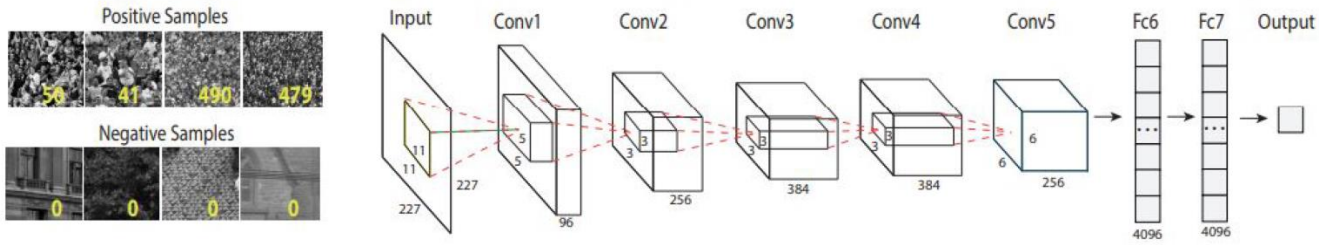

Figure 1. Convolutional Neural Network (CNN) architecture with positive and negative inputs [22]

Deep network pioneering work was put out in the study [22]. An end-to-end deep convolutional neural network (CNN) regression model was developed to count people in images of extraordinarily dense crowds. A dataset generated from Google and Flickr was annotated using a dotting technique. There is an average of 731 persons in each of the dataset's 51 photos. In this dataset, 95 counts are the lowest and 3714 counts are the greatest. On both positive and negative classes, the network was trained. The number of the items was identified on the positive photographs, whereas zero was labeled on the negative images. The network's structure consists of two fully connected layers and five convolutional layers. The network was trained on item classification using regression loss, as shown in Figure 1.

Ashish Sharma et al. analyzed deep learning and machine learning monitoring methods in the study [23] and provided suggestions for a brand-new monitoring system. The main objective of deploying CCTV is to deter crime or property damage by recognizing suspicious or unusual behavior during the surveillance. An intelligent surveillance system that not only minimises the need for human monitoring but also promptly warns the appropriate authorities of impending issues is very necessary. Crime may frequently seem normal because individuals are virtually always aware that there are CCTV cameras around. But an excessive number of false warnings might potentially make people annoyed or lose trust in the system. A unique model with reduced training time, a smaller data set, high accuracy, and self-learning over time is therefore greatly needed. The approach needs to be clever, constantly monitoring the audience and alerting the authorities to any questionable behavior, if any.

A real-time crowd density estimation approach based on the multi-stage ConvNet was proposed [24] after the initial CNN-based methodology [25]. This strategy's basic premise is that certain CNN connections are unnecessary. As a consequence, related feature maps and their linkages from the second stage can be removed. Two multi-stage cascaded classifiers make up the network's design [26]. One convolutional layer and a sub sampling layer make up the first stage. The second stage utilizes the same architecture. The last layer classifies the crowd scenario as either very low, low, medium, high, or very high using a totally connected layer with five outputs. Since just 1/7 of the features came from the feature maps from the first stage, the authors optimised this step. The optimization process was based on comparing how similar the maps were. To reduce processing time, this map will be eliminated if the similarity is below a certain level.

Usman and Albesher [27] created an approach for aberrant crowd behaviour in the study by utilising motion awareness and heuristic search. The authors of this study offered a method for automatically spotting irregularities in crowd video sequences. The suggested methodology employs a gradient-based strategy with an activation function and a motion aware component to account for crowd dynamics. To generate the best classifier, a progressive upgrading approach based on GP-based training simulation is applied. The most effective mathematical expression is a general classifier that performs better when exploiting the decision space's hidden dependencies. The suggested method's effectiveness and superiority to the most recent approaches in terms of classification accuracy are supported by experimental findings. The authors have not made any comments about the system's timeliness or accuracy, both of which might be improved.

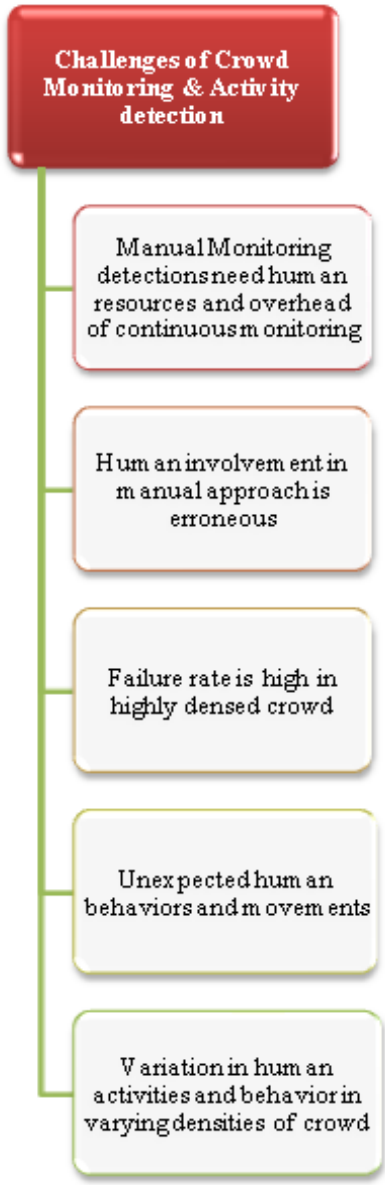

Figure 2. Challenges of existing crowd monitoring and activity detection systems

Sivalingan and Anandakrishnan [28] have focused their efforts on analysing suspicious behaviour to identify illicit behaviour. In order to measure suspicious pedestrians for in-flight (crime) detection, this study provided two different kinds of modules, each of which includes a speed calculation profile. The first module takes care of the validation task. This research tracks the speed of pedestrian gait and shows how it differs from the gait of the other n pedestrians using calculations that are presented. It might be difficult to research a topic like tracking pedestrian stride using video surveillance in real time. However, the proposed method would assess the pedestrian's gait speed using the walk ratio, AAC, gravity, as well as the dynamic, horizontal, and vertical components of the pedestrian, and would identify criminal activity based on the latter's suspicious activities. The effectiveness of the suggested research in comparison to existing methods for finding pedestrians, such RealBoost and DPM. In comparison to DPM and Real Boost approaches, the proposed work has a greater true positive rate and evaluates gait speed faster. The system's flaw is that it can only detect suspicious activity in still pictures; moving frames are not examined in real time. As a result, this system lacks the ability to accurately measure moving video frames. From the detailed literature survey, Figure 2 shows identified limitations of the existing crowd monitoring systems.

3. OBJECTIVES

The limitations of the available crowd monitoring technology and the identified need for more research led to the development of the following objectives. These are the research objectives for developing a crowd surveillance system that can accurately detect suspicious behaviour before it begins and so protect society from undesirable criminal conduct.

1. To develop a system for crowd monitoring that uses deep learning methods to spot suspicious behaviour.
2. To develop a system for crowd surveillance that uses FCN and LSTM to precisely identify and monitor a person's positions in more crowded environments.
3. To develop an accuracy-based system with reduced false positives and false negatives to detect shady activities in crowded areas by properly analysing crowd behaviour using FCN and LSTM.

4. CONCEPTS AND DEFINITIONS

Here, in this research work, the two parts are important, in first part we talk about the preprocessing of the images ato extract the features and actual processing of the frames for detection of suspicious activities in the crowd.

4.1 Fully Convolutional network (FCN)

To our knowledge, the idea of extending a convnet to arbitrary-sized inputs first surfaced in Dabhi et al.'s [29] extension of the conventional LeNet [30] to recognise strings of digits. Because their net could only handle one-dimensional input strings, Matan et al. used Viterbi decoding to obtain their outputs. By extending convnet outputs [31] to 2-dimensional maps of detection scores for postal address block corners, Wolf and Platt. These two earlier works use detection using complete convolutional learning and inference. Rehman et al. [32] describe a convnet for coarse multiclass segmentation of C. elegans tissues using fully convolutional inference.

Fully convolutional computing has also been used in the present era of many-layered networks. The following methods achieve fully convolutional inference:

Picture restoration by Quadri and Katakdhond [33], semantic segmentation by Tripathi et al. [34], and sliding window detection by Kamthe and Patil [35] are other examples of related work. Hochreiter and Schmidhuber [36] apply completely convolutional training to effectively develop an end-to-end component detector and spatial model for posture assessment, despite the fact that they do not describe or examine this method.

Each data layer in a convolutional network has a size of h x w x d, where h and w are spatial dimensions and d is the feature or channel dimension. The first layer is the picture, which consists of color channels and pixels with dimensions of h x w. Higher layer positions correspond to their receptive fields, or the regions of the picture to which they have a path connection. Convnets are built on translation invariance. Convolution, pooling, and activation functions are their core components, which only need relative spatial coordinates and operate on local input regions. Using Eq. (1), these functions generate the outputs y_{ij}, writing x_{ij} for the data vector at position (i, j) in a particular layer and yij for the following layer.

$$y_{ij} = f_{ks} (\{x_{si}+\delta i, s_j+\delta j\} 0 \leq \delta i, \delta j \leq k) \quad (1)$$

where, f_{ks} specifies the type of layer: element-wise nonlinearity for an activation function, spatial maximum for max pooling, matrix multiplication for convolution or average pooling, and so on for various types of layers. The kernel size is referred to as k, while the stride or subsampling factor is referred to as s.

This functional form is retained during composition, and the transformation rule is followed for the kernel size and stride.

$$f_{ks} \circ g_{k's'} = (f \circ g)_{k'+(k-1)s', ss'}.$$

A net with only these types of layers computes a nonlinear filter, which we refer to as a deep filter or fully convolutional network, whereas a general deep net computes a generic nonlinear function. Any size input is automatically processed by an FCN to yield an output with corresponding (potentially resample) spatial dimensions.

A real-valued loss function created using an FCN defines a task. If the loss function is a sum across the spatial dimensions of that layer, the gradient of the final layer, l(x;)=ijl'(xij;), will equal the sum of the gradients of each of its spatial components. Stochastic gradient descent on ℓ computed on complete photos will be the same as stochastic gradient descent on ℓ' since all of the last layer receptive fields are taken into consideration as a minibatch.

When these receptive fields considerably overlap, feedforward computation and backpropagation are far more successful when computed layer-by-layer over the entire image as opposed to individually patch-by-patch.

The FCN is implemented to extract the features from the given data. It helps to work on the appropriate features during the system implementations. It is basically very helpful in the system implementation having sequential predictions.

4.2 Long short- term memory networks (LSTM)

One of the main contributions of the original long short-term memory (LSTM) model was the clever concept of introducing self-loops to make channels where the gradient may flow for extended periods of time [37]. Making the weight on this self-loop variable rather than fixed has been a substantial improvement [38]. The integration time scale may be dynamically changed by gating the weight of this self-loop, which is controlled by another hidden unit. We conclude that even for an LSTM with fixed parameters, the time scale of integration might fluctuate depending on the input sequence since in this situation, the time constants are generated by the model itself. The unconstrained handwriting recognition [39], speech recognition [40, 41], handwriting creation [40], machine translation [42], and picture captioning [43, 44] are only a few of the numerous areas where the LSTM has excelled. Figure 3 displays the block diagram of the LSTM. The relevant forward propagation equations for the design of deep recurrent networks are shown in the picture below.

Effective use of deeper structures has also been accomplished. "LSTM cells" in LSTM recurrent networks have an internal recurrence (a self-loop) in addition to the outer recurrence of the RNN. This is different from a unit that just applies an element wise nonlinearity to the affine transformation of inputs and recurrent units. The inputs and outputs of each cell are identical to those of a standard recurrent network, but they also contain extra parameters and a set of gating units to regulate the information flow.

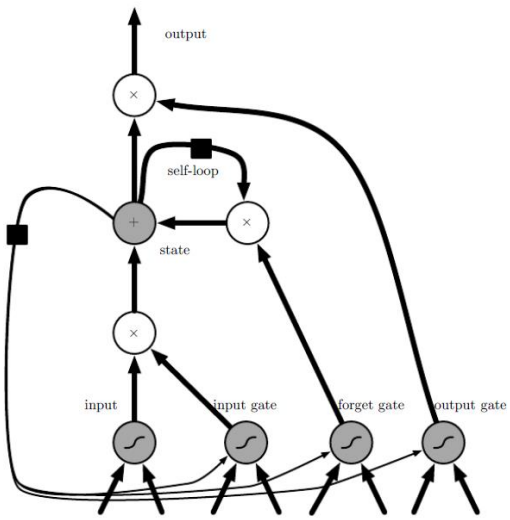

Figure 3. LSTM recurrent network block diagram of a "cell"

Recurrent connections between cells are employed instead of the typical hidden units present in conventional recurrent networks. A typical artificial neuron unit computes an input feature. If the sigmoidal input gate permits it, its value may be added to the state. The forget gate regulates the state unit's linear self-loop weight. Through the output gate, the cell's output can be disabled. Although the input unit may have any squashing nonlinearity, all gating units have a sigmoid nonlinearity. The gating units may additionally receive an additional input from the state unit. One time step of the delay is depicted by the black square.

The state unit $s_i^{(t)}$, which has a linear self-loop just like the leaky units described in the preceding section, is the most crucial element. However, in this case, the self-loop weight is controlled by a forget gate unit $f_i^{(t)}$ (for time step t and cell i, and is set by a sigmoid unit to a value between 0 and 1).

$$f_i^{(t)} = \sigma(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)})$$

where, b^f, U^f, and W^f are the corresponding biases, input weights, and recurrent weights for the forget gates, and $x^{(t)}$ is the current input vector and $h^{(t)}$ is the current hidden layer vector, which comprises the outputs of all LSTM cells. Consequently, the internal state of the LSTM cell is updated as follows, but with conditional self-loop weight fi(t):

$$s_i^{(t)} = (f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma(b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)})$$

where, b, U, and W, respectively, stand for the biases, input weights, and recurrent weights of the LSTM cell. The external input gate unit $g_i^{(t)}$ performs the same function with its own settings, however the forget gate employs a sigmoid unit to generate a gating value between 0 and 1:

$$g_i^{(t)} = \sigma(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)})$$

The output gate $q_i^{(t)}$, which likewise utilizes a sigmoid unit for gating, may also deactivate the output $h_i^{(t)}$ of the LSTM cell:

$$h_i^{(t)} = \tanh(s_i^{(t)}) q_i^{(t)}$$
$$q_i^{(t)} = \sigma(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)})$$

The parameters that make up its biases, input weights, and recurrent weights are b^o, U^o, and W^o. One of the variations, shown by the three gates of the i-th unit, may take the cell state $s_i^{(t)}$ and its weight as an additional input. Three more parameters must be included in order to do this. Both on challenging sequence processing tasks where state-of-the-art performance was attained [45] and on simulated data sets created for testing the ability to learn long-term dependencies, LSTM networks have been shown to learn long-term dependencies more quickly than simple recurrent architectures.

5. PROPOSED METHODOLOGY

The procedure is shown in Figure 4. The graphic shows how the input pictures are produced using video frames from an incoming stream. These photos are processed using the VGG16 model to extract the required image characteristics for crowd activity detection and person tracking. The preprocessed pictures that are above the LSTM recurrent layer provide input to the fully convolutional network. The FCN layer provides the input to the LSTM, which then generates the classification result to recognize crowd actions.

The input data is a video frame of arbitrary time frame which will be passed to the FCN+LSTM model. The VGG16 layers is applied for the useful feature extraction as a part of data preprocessing. This step is important to ensure the highly correlated features to be extracted from the frames of the video and then pass the input to the FCN layer which lies below the

VGG16 and above the LSTM layer. With extracted features, the FCN processes the input to its optimized accuracy extent and passes the processed input for classification to the LSTM. Finally, LSTM, a recurrent neural network component, aids in classifying incoming video frames as either regular video streaming or suspicious activity video. When a person or group of people shows conduct that is judged strange and unwanted in a crowd, the FCN+LSTM categorization recognizes suspicious behavior. Here, an intelligent monitoring system is created where the FCN+LSTM is continuously given incoming video streams of the crowd and, rather than manually visualizing the videos, the FCN+LSTM based model analyses them and only issues an alert if there is a chance of a crime or suspicious activity. As a result, manual surveillance is completely eliminated, and a sophisticated technology is in place to scan the crowd for signs of criminal activity.

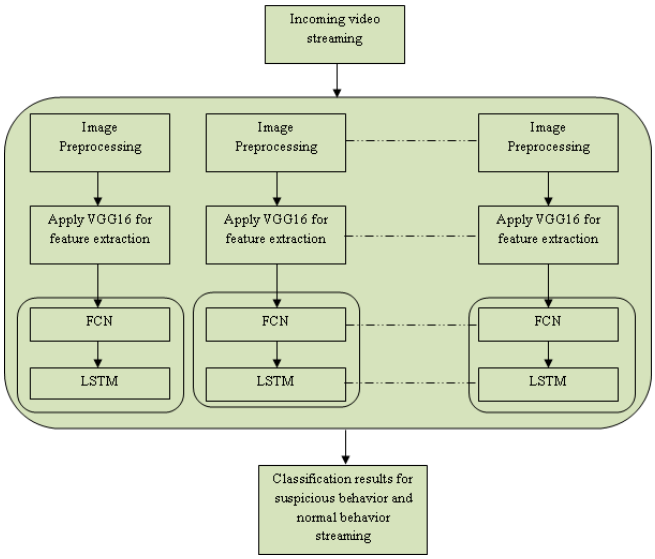

Figure 4. Proposed architecture

The proposed model is trained and tested for validation of the accuracy of the model using the crowd monitoring and crowd counting datasets. The FCN and LSTM techniques will be implemented for sequential predictions, where FCN will be first to act as a feature extractor and predictor for sequential predictions and then LSTM will be applied for further classifications.

6. RESULTS AND DISCUSSION

The model's input data for training and testing is picture data with photographs of occasions, crowds, or crowd zones. The following chart displays the graph of the total loss vs the total validation loss for the LSTM model. The graph demonstrates how the overall loss amount visibly drops until it reaches 10. The value of the FCN + LSTM model's overall loss is then steadily increased and calculated.

The FCN + LSTM model type spans a wide range of time and space. In order to produce natural language strings and to encode deep spatial elements, this uses an LSTM (decoder) and a convnet (encoder). LSTM allows for the modeling of sequential data with various periods. However, there is a constant rise and fall in the value of the overall validation loss. Additionally, the value increases in comparison to the other points as the tidal validation graph reaches 30. The overall loss in this case is less than the overall validation loss.

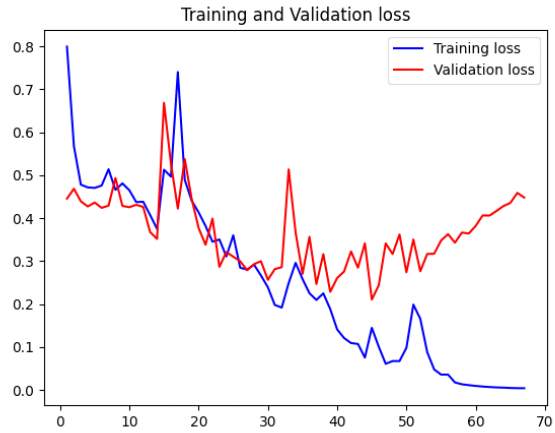

Figure 5. Performance of LSTM

The graph shown in Figure 5 is of total accuracy versus total validation accuracy for the FCN + LSTM model is shown in the Figure 6. In this graph, it can be seen that there is a small fluctuation in the overall accuracy for the FCN + LSTM model, and that this fluctuation is significantly greater before the value reaches 30. On the other hand, it can be observed that there is a large decline at first and subsequently an increase to a certain point in the graph of the overall validation accuracy for the lrcn model. The overall validation accuracy graph varies continually after crossing the value of 10. The graph also shows that validation accuracy is lower than overall accuracy. The overall accuracy of the FCN + LSTM model is found to be 97.84%. This number beats the existing models and also it has fully automation support.

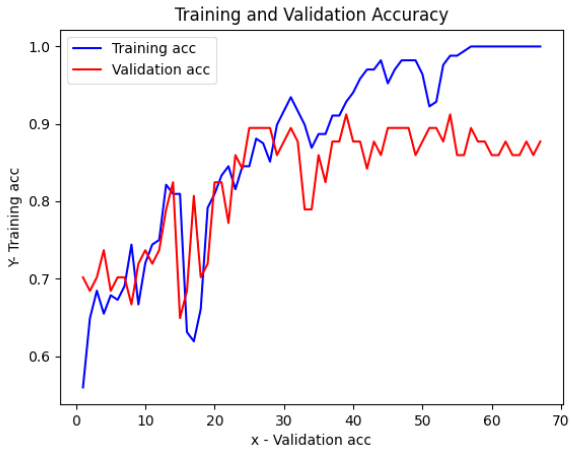

Figure 6. Performance of FCN + LSTM

7. CONCLUSION

The goal of this study is to develop a completely automated system for observing crowds and identifying suspicious behavior. It employs LSTM, FCN, and deep learning to find suspicious crowd behavior. It presents a technique that accurately identifies suspicious behavior without the involvement of a human by utilizing the VGG16 and FCN +

LSTM. By eliminating the need for human video traffic analysis from the system's CCTV footages, this technology lessens the workload of the government forces and security agencies. After the model is instantiated in the actual environment, the automation in the model automatically performs the alerts in case of questionable activities in the crowd situations. The data on accuracy has shown that the model's accuracy, which is 97.84%, is better than that of the present systems and that the quantity of false alarms has been greatly reduced. The system will eventually employ real-time input footages and build effective deep learning methods to increase accuracy while reducing development durations.

REFERENCES

[1] Veerapathiran, S., Ramachandran, S. (2022). A multi task allocation based time optimization framework using social networks in mobile crowd sensing. Instrumentation Mesure Métrologie, 21(6): 237-241. https://doi.org/10.18280/i2m.210605

[2] Albayrak, A. (2022). Artificial intelligence based social distance monitoring in public areas. Traitement du Signal, 39(3): 961-967. https://doi.org/10.18280/ts.390323

[3] Gupta, A., Satpute, V.R., Kulat, K.D., Bokde, N. (2016). Real-time abandoned object detection using video surveillance. In: Afzalpulkar, N., Srivastava, V., Singh, G., Bhatnagar, D. (eds) Proceedings of the International Conference on Recent Cognizance in Wireless Communication & Image Processing. Springer, New Delhi. https://doi.org/10.1007/978-81-322-2638-3_94

[4] Pawade, A., Anjaria, R., Satpute, V.R. (2021). Suspicious activity detection for security cameras. In: Kumar, R., Dohare, R.K., Dubey, H., Singh, V.P. (eds) Applications of Advanced Computing in Systems. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-33-4862-2_22

[5] Sindagi, V.A., Patel, V.M. (2017). A survey of recent advances in CNN-based single image crowd counting and density estimation. Pattern Recognition Letters, 107: 3-16. https://doi.org/10.1016/j.patrec.2017.07.007

[6] Mohammed, A., Shaery, A., Khozium, M.O. (2019). Crowd management challenges: Tackling Approach for realtime crowd monitoring. International Journal of Scientific Engineering and Research, 7(1): 84-88.

[7] Yamin, M., Ades, Y. (2009). Crowd management with RFID and wireless technologies. 2009 First International Conference on Networks & Communications, pp. 439-442. https://doi.org/10.1109/NetCoM.2009.14

[8] Hu, W.M., Tan, T.N., Wang, L., Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34(3): 334-352. https://doi.org/10.1109/TSMCC.2004.829274

[9] Tapas, B., Nain, N., Ahmed, M., Sharma, V. (2015). An adaptive codebook model for change detection with dynamic background. 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Bangkok, Thailand, pp. 110-116. https://doi.org/10.1109/SITIS.2015.89

[10] Medel, J.R., Savakis, A. (2016). Anomaly detection in video using predictive convolutional long short-term memory networks. https://doi.org/10.48550/arXiv.1612.00390

[11] Zhou, B.L., Wang, X.G., Tang, X.O. (2012). Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. IEEE Conf. on Computer Vision and Pattern Recognition, Providence, RI, USA, pp. 2871-2878. https://doi.org/10.1109/CVPR.2012.6248013

[12] Liu, J., Gao, C.Q., Meng, D.Y., Hauptmann, A. (2018). Decidenet: Counting varying density crowds through attention guided detection and density estimation. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 18-23, https://doi.org/10.1109/CVPR.2018.00545

[13] Ito, R., Tsukada, M., Kondo, M., Matsutani, H. (2019). An adaptive abnormal behavior detection using online sequential learning. 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), New York, NY, USA, pp. 436-440. https://doi.org/10.1109/CSE/EUC.2019.00087

[14] Khan, A., Shah, J., Kadir, K., Albattah, W., Khan, F. (2020). Crowd monitoring and localization using deep convolutional neural network: A review. Application of Science, 10(14): 4781. https://doi.org/10.3390/app10144781

[15] Saqib, M., Khan, S., Sharma, N., Blumenstein, M. (2018). Person head detection in multiple scales using deep convolutional neural networks. 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, pp. 1-7. https://doi.org/10.1109/IJCNN.2018.8489367

[16] Zhang, Y.M., Zhou, C.L., Chang, F.L., Kot, A. (2019). Multi-resolution attention convolutional neural network for crowd counting. Neurocomputing, 329: 144-152. https://doi.org/10.1016/j.neucom.2018.10.058

[17] Zhu, L.P., Li, C.Y., Yang, Z.G., Yuan, K., Wang, S. (2020). Crowd density estimation based on classification activation map and patch density level. Neural Computing and Applications, 32: 5105-5116. https://doi.org/10.1007/s00521-018-3954-7

[18] Basalamah, S., Khan, S., Ullah, H. (2019). Scale driven convolutional neural network model for people counting and localization in crowd scenes. IEEE Access, 7: 71576-71584. https://doi.org/10.1109/ACCESS.2019.2918650

[19] Chong, Y.S., Tay, Y.H. (2017). Abnormal event detection in videos using spatiotemporal auto encoder. https://doi.org/10.48550/arXiv.1701.01546

[20] Albattah, W., Khel, M.H.K., Habib, S., Islam, M., Khan, S., Kadir, K.A. (2020). Hajj crowd management using CNN-based approach. Computers, Materials & Continua, 66(2): 2183-2197. https://doi.org/10.32604/cmc.2020.014227

[21] Wu, B., Nevatia, R. (2005). Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors. International Journal of Computer Vision, 75: 247-266. https://doi.org/10.1007/s11263-006-0027-7

[22] Karlik, B., Olgac, A.V. (2011). Performance analysis of various activation functions in generalized mlp architectures of neural networks. International Journal of Artificial Intelligence and Expert Systems (IJAE), 1(4): 111-122.

[23] Sharma, A., Varshney, N. (2020). Identification and detection of abnormal human activities using deep learning techniques. European Journal of Molecular &

Clinical Medicine, 7(4): 408-417.
[24] Wang, C., Zhang, H., Yang, L., Liu, S., Cao, X.C. (2015). Deep people counting in extremely dense crowds. In Proceedings of the 23rd ACM International Conference on Multimedia, ACM: New York, NY, USA, pp. 1299-1302. https://doi.org/10.1145/2733373.2806337
[25] Fu, M., Xu, P., Li, X.D., Liu, Q.H., Ye, M., Zhu, C. (2015). Fast crowd density estimation with convolutional neural networks. Engineering Applications of Artificial Intelligence, 43: 81-88. https://doi.org/10.1016/j.engappai.2015.04.006
[26] Sermanet, P., Kavukcuoglu, K., Chintala, S., LeCun, Y. (2013). Pedestrian detection with unsupervised multi-stage feature learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, pp. 3626-3633. https://doi.org/10.1109/CVPR.2013.465
[27] Usman, I., Albesher, A.A. (2021). Abnormal crowd behavior detection using heuristic search and motion awareness. International Journal of Computer Science and Network Security, 21(4): 131-139. https://doi.org/10.22937/IJCSNS.2021.21.4.18
[28] Sivalingan, H., Anandakrishnan, N. (2021). Analysing the suspicious behaviour in video surveillance for crime detection using gait speed monitoring. ICTACT Journal on Image & Video Processing, 12(1): 2502-2507. https://doi.org/10.21917/ijivp.2021.0355
[29] Dabhi, M., Shah, M., Bharti, P., Puvar, P., Prajapati, B. (2020). Unusual activity detection in crowd using deep learning. International Journal of Emerging Technologies and Innovative Research, 7(6): 470-476.
[30] Gawande, U., Hajari, K., Golhar, Y. (2023). Real-time deep learning approach for pedestrian detection and suspicious activity recognition. Procedia Computer Science, 218: 2438-2447. https://doi.org/10.1016/j.procs.2023.01.219
[31] Buttar, A.M., Bano, M., Akbar, M.A., Alabrah, A., Gumaei, A.H. (2023). Toward trustworthy human suspicious activity detection from surveillance videos using deep learning. Soft Computing. https://doi.org/10.1007/s00500-023-07971-x
[32] Rehman, A., Saba, T., Khan, M.Z., Damaševičius, R., Bahaj, S.A. (2022). Internet-of-things-based suspicious activity recognition using multimodalities of computer vision for smart city security. Security and Communication Networks, 2022: 8383461. https://doi.org/10.1155/2022/8383461
[33] Quadri, S.A., Katakdhond, K.S. (2022). Suspicious activity detection using convolution neural network. Journal of Pharmaceutical Negative Results, 13(1): 1235-1245. https://doi.org/10.47750/pnr.2022.13.S01.151
[34] Tripathi, R.K., Jalal, A.S., Agrawal, S.C. (2018). Suspicious human activity recognition: A review. Artificial Intelligence Review, 50: 283-339. https://doi.org/10.1007/s10462-017-9545-7
[35] Kamthe, U.M., Patil, C.G. (2018). Suspicious activity recognition in video surveillance system. 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, pp. 1-6. https://doi.org/10.1109/ICCUBEA.2018.8697408
[36] Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8): 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
[37] Gers, F.A., Schmidhuber, J., Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. Neural Computation, 12(10): 2451-2471. https://doi.org/10.1162/089976600300015015
[38] Graves, P.M., Richards, F.O., Ngondi, J., Emerson, P.M., Shargie, E.B., Endeshaw, T., Ceccato, P., Ejigsemahu, Y., Mosher, A.W., Hailemariam, A., Zerihun, M., Teferi, T., Ayele, B., Mesele, A., Yohannes, G., Tilahun, A., Gebre, T. (2009). Individual, household and environmental risk factors for malaria infection in Amhara, Oromia and SNNP regions of Ethiopia. Transactions of the Royal Society of Tropical Medicine and Hygiene, 103(12): 1211-1220. https://doi.org/10.1016/j.trstmh.2008.11.016
[39] Graves, A. (2013). Generating sequences with recurrent neural networks. https://doi.org/10.48550/arXiv.1308.0850
[40] Graves, A., Jaitly, N. (2014). Towards end-to-end speech recognition with recurrent neural networks. 31st International Conference on Machine Learning, ICML 32(2): 1764-1772.
[41] Idrees, H., Saleemi, I., Seibert, C., Shah, M. (2013). Multi-source multi-scale counting in extremely dense crowd images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, pp. 2547-2554. https://doi.org/10.1109/CVPR.2013.329
[42] Chen, K., Loy, C.C., Gong, S., Xiang, T. (2012). Feature mining for localised crowd counting. In Proceedings of the British Machine Vision Conference, Surrey, UK, p. 3. https://personal.ie.cuhk.edu.hk/~ccloy/project_feat_mine_count/index.html#:~:text=A%20multi-output%20regression%20framework%20for%20localised%20crowd%20counting,edges%20and%20texture%20features%2C%20from%20each%20cell%20region.
[43] Kiros, R., Salakhutdinov, R., Zemel, R.S. (2014). Unifying visual-semantic embeddings with multimodal neural language models. https://doi.org/10.48550/arXiv.1411.2539
[44] Sutskever, I., Vinyals, O., Le, Q.V. (2014). Sequence to sequence learning with neural networks. https://doi.org/10.48550/arXiv.1409.3215
[45] Graves, A. (2012). Supervised sequence labelling with recurrent neural networks. Studies in Computational Intelligence. https://www.cs.toronto.edu/~graves/preprint.pdf.