Al-Khwarizmi Engineering Journal

Al-Khwarizmi Engineering Journal
ISSN (printed): 1818 – 1171, ISSN (online): 2312 – 0789
Vol. 21, No. 3, September, (2025), pp. 55–65

Violence Prediction Estimation in Surveillance Cameras Using CNN with GAMMA Correction

Sara Ayad[1]*, and Yarub Alazzawi[2]

[1] *Iraqi Correctional Service, Ministry of Justice, Baghdad, Iraq*

[2] *Department of Mechatronics Engineering, Al-Khwarizmi College of Engineering, University of Baghdad, Baghdad, Iraq*

*Corresponding Author's Email: saraaagh785@gmail.com

(Received 24 June 2024; Revised 14 October 2024; Accepted 10 November 2024; Published 1 September 2025)

https://doi.org/10.22153/kej.2025.11.004

Abstract

In recent years, the deployment of surveillance cameras has significantly increased to enhance security in public and private spaces. Numerous businesses continue to employ individuals to monitor these cameras. However, unusual and suspicious activities in the video feeds are often overlooked due to the potential for human error. Consequently, manual monitoring of security cameras can be time-consuming and inefficient. This study investigates the application of deep learning techniques, particularly convolutional neural networks (CNNs) and support vector machines (SVMs), to predict violence in surveillance video streams. The proposed CNN model is optimised through the utilisation of gamma correction as a preprocessing step to extract essential spatial features from video frames, significantly enhancing the accuracy of violence detection. This study leverages the real-time capabilities of surveillance data by utilising the RLV dataset, which comprises a range of violent and non-violent scenarios. The CNN–SVM hybrid model developed in this study achieved an impressive 99% accuracy, outperforming traditional methods and demonstrating strong spatial feature extraction capabilities. Furthermore, this study addresses the challenges of real-time video surveillance by ensuring scalability and practical applicability, providing a robust solution for enhancing security measures in public and private spaces.

Keywords: *anomaly detection; surveillance cameras; Convolutional Neural Network (CNN); Deep Learning; Violence detection; Support Vector Machine.*

1. Introduction

The use of surveillance cameras to monitor public situations is becoming increasingly common due to the growing issues in public administration, security and safety. The task of monitoring security camera footage to detect unusual behaviour, extract patterns and promptly respond may appear simple for humans. However, a person finds it difficult to simultaneously observe multiple signals due to the inherent limitations of human capabilities [1]. The process is time-consuming and costly, necessitating the involvement of people and a workplace. Consequently, an automatic detection system is crucial. Anomalous event identification is an essential aspect of understanding human behavioural through surveillance cameras [2]. However, detecting such events in security footage presents several challenges: (1) A large database of anomalous events is difficult to obtain due to their rarity. The learning process may be adversely affected by the lack of data. (2) An 'anomaly' refers to anything that deviates from a pre-established pattern (or rule). We are unable to create a model specifically designed to handle uncommon events. (3) An activity may deem normal or abnormal depending on the situation. This notion suggests that a global anomalous event is likely to regularly occur in cases involving a gun club. Although this event is common in a shooting group, the practice of 'shooting' is often considered unusual. However, certain conduct may be considered an anomalous practice in a particular locale and context, referred to as a local abnormal event, even if it is not inherently anomalous. Varadarajan [3] characterised anomalous events as 'actions that are performed in an unusual location, at an unusual time'. Anomalous behaviour, a significant and recognised classification of learning methods, can be categorised into three approaches from a learning perspective: supervised, unsupervised and semi-supervised. Specifically, single-model learning focuses solely on either normal or abnormal events for training. Meanwhile, multi-model learning involves training of normal and abnormal events. The methods used in the single model learning process to distinguish abnormal events from normal ones include the following: learning the definition of normality [4-6], constructing a multidimensional model of abnormal events within the space of normal events [7-10] and learning rules that define abnormal events [11]. However, the multi-model learning strategy involves training each class either independently or in combination with other classes. This strategy is particularly effective when multiple groups of irregularities exist. In unsupervised learning, numerous clustering techniques assume that abnormal and normal events can be easily distinguished within the feature space [12-14]. Additionally, the semi-supervised method of detecting anomalous behaviour is not as accurate as unsupervised models and does not depend on pre-labelled data as the supervised approach does. However, anomalous behaviour detection is regarded as an unsupervised learning problem. The evolution of violence detection in video surveillance began with traditional methods that relied on handcrafted features and basic classifiers [15]. This approach was later enhanced by the integration of machine learning algorithms, which improved the classification of violent events by combining spatial and temporal features [16]. Finally, deep learning models, particularly convolutional neural networks (CNNs), have significantly improved the accuracy of violence detection in large-scale datasets and paved the way for highly sophisticated neural network architectures [17]. In this study, anomalous behaviour is considered a multiple-scene issue under a supervised learning scheme. The classification of numerous erratic behaviours in the real world as anomalies is based on the definition of an anomalous event. In this study, we focus on the dataset [18] that contains numerous anomalous, illegal and violent behaviours captured on video in public areas. These behaviours have a serious influence on individuals and the general public. Our proposed model utilised VGG16 as a CNN for feature extraction. Given the nature of the video dataset, we augment the model with an SVM, which can effectively handle this type of data. This methodology enhances the ability to detect irreversible damage whilst reducing the need for human labour and financial resources. Moreover, this strategy is of great importance to governments and the public because it can significantly reduce emergency response times. A summary of the primary contributions of the proposed method is provided below.

1. We gather the essential data necessary to identify violence, including human bodies and their interactions, to minimise the complexity of the model's input. This approach facilitates the development of a real-time infrastructure for crime detection.
2. We uniquely combine features derived from human interactions with temporal changes in body postures to fully utilise both information sources.
3. We recommend the effective use of a support vector machine (SVM) in two-stream models designed for violent behaviour detection. This approach facilitates real-time operation whilst still maintaining accuracy.

Figure 1 presents a flow chart that illustrates the steps involved in implementing a video-based violence detection scheme. The first phase involves inputting a collection of movies that include violent and non-violent scenes. The second phase is key frame extraction, although not all methods utilise it. This process focuses on selecting frames that depict violent content, aiming to avoid the processing of large volumes of video, minimising the computational burden. In the third phase, the information is transformed to serve as an input for the violent behaviour detection algorithm; the type of input is based on the desired features.

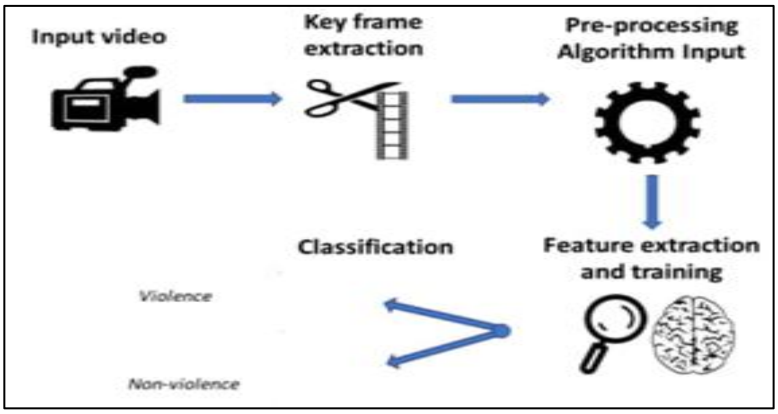


Fig. 1. Basic steps in the application of an algorithm for video violence detection

The remainder of this study discusses related projects that utilise different models and sub-models based on the overall concept of detecting anomalous behaviour in security camera footage. Section 2 presents an overview, followed by a detailed discussion of the proposed methodology in Section 3. Section 4 explores the findings from various experiments. Section 5 concludes the study with final remarks and future project directions.

2. Literature Review

Liu et al. [19] studied various experiments, among them one by reference [20], and it is an unsupervised deep representation approach for abnormal behaviour recognition in crowded environments. This scheme relies on hybrid deep learning models and two-channel framework to learn and generate anomaly scores. The validity of the approach has been proven experimentally with training data tests and with tests in different lighting conditions. Sreenu et al. [21] identified live public square CCTV stream videos as a useful big data to improve safety in crowded public areas. Our work is a comprehensive survey on a variety of deep learning based methods used for different crowd video analysis tasks ranging from features, action recognition, to violence detection. Farroq et al. [22] attempted to detect crowd divergence behavior for disaster prevention, such as stampedes. Their method presented a solution for crowd behaviour classification with a CNN on motion shaped images. The method performed better than a number of existing algorithms in terms of accuracy and employs a divergencelocalisation approach. The study also generates novel datasetsfor analyzing normal and abnormal crowd behavior in highly dense crowd environments.

in high-density settings. Behera et al. [23] proposed a deep learning method called PIDLNet, which combines 3D convolutional features with physics-based properties, to explain crowd dynamics. Varghese et al. [24] proposed an approach by integrating cognitive deep learning frameworks with a fuzzy computational model according to the sentiments, behavioural traits and visual examination. The methodology's objective lies in the estimation and anticipation of different types of collective actions from the crowd in surveillance devices to efficiently handle the crowd. B. Peixoto et al. [25] presented a new visual data-driven two-stage framework followed a data-driven approach for violent scene detection using two deep network model applied directly over raw image data. The two networks we adopt are a CNN–LSTM, which involves long short-term memory, and the 3D-based CNN (C3D). Instead of training the two networks separately, the authors combined C3D and CNN–LSTM, and trained the model for the violence detection task jointly. This method achieved 56% accuracy on the C3D and CNN–LSTM network where it was 63% for their proposed method in C3D and 61% for their proposed method in CNN–LSTM model. Soliman et al. [26] proposed a new model that takes a short clip as an RGB frame sequences and uses it end-toend with a deep neural network.

A sequence of fully connected layers was utilised for classification. Meanwhile, a pre-trained VGG-16 on ImageNet was used for the spatial feature extractor, whilst an LTSM was utilised for the temporal feature extractor. RealLife Violence Situations is a newly developed dataset consisting of 2000 videos, depicting either aggressive or peaceful scenarios. This dataset was used to improve the developed model, resulting in 88.2% overall accuracy. Abdali et al. [27] introduced a real-time violence detection that uses CNN to be the spatial feature extractor and LTSM to be the temporal feature extractor. The proposed model attained 98% success rate at 131 fps, meanwhile surpassing the state of the art methods. Individual verification data were constructed for the three datasets and perfect performance was achieved on the movie dataset, 96.33% on hockey battle dataset and 85.71% on the violent flow dataset. For training and fine-tuning the Inception-ResNetV2 architecture introduced by Jain et al. [28], they used a (CNN) architecture. The model was designed to learn patterns of motion in order to detect aggression. RGB video sequences were first converted to DIs by temporal averaging, which maps video streams into a single image by averaging all the background and static pixels over time. These DIs can enhance the motion-related characteristics efficiently while restraining discriminatory background information. The generated DI's were fed into the Inception-ResNetV2 model to acquire motion features and classification for aggressive or non-aggressive behavior. The below testing result is 86.78% with the violent real-life scenario dataset, 100.89% and 92.37% with the secondary testing dataset.

for the movie dataset and 93.33% for the hockey fight dataset. Sernani et al. [29] proposed three models for violent behaviour detection to evaluate the degree to which harmless gestures, such as claps, small hits or high fives, are considered violent (i.e. to reduce false positive values). These models were tested on the AIRTLab dataset, which was designed to assess the susceptibility of algorithms to false positives. When contrasted with 2D CNNs, Kang et al. [30] proposed a more advanced model that integrates LSTM and requires lower computational power than existing 3D-CNN-based methods. Specifically, the proposed models for MobileNetV3 and EfficientNet-B0 achieved tremendous success on six different violent datasets. Gadelkarim et al. [31] proposed two models for the recognition and categorisation of violence, which helps in preventing daily violence. The models were developed using the following datasets: XD-Violence, LAD2000 and UCF-Crime. Similar to the XD-Violence dataset, their research benefitted from transfer learning, which reduced the training period and improved accuracy. Islam et al. [32] proposed an effective two-stream deep learning architecture, with the first stream incorporating a pre-trained MobileNet and the second stream comprising a separate convolutional LSTM that processes frame content. They utilised three different fusion methods across three different datasets to combine the output maps from two streams. Their model demonstrated excellent performance using the RWF-2000 dataset, which is the most challenging among larger datasets, achieving over a 2% increase in accuracy compared with smaller datasets. Honarjoo et al. [33] proposed a convolutional-based approach with limited complexity and a series of frames with effective feature vectors. This approach successfully processed the violent waterfalls and hockey disputes. Guedes et al. [34] developed a method that utilised the DI approach to recognise violent activities that involve physical confrontations in video games. This method used handcrafted images processed through CNNs and an SVM, and it was successful when applied to the following datasets: 99.8% for movies, 97.5% for hockey games and 93.4% for crowd datasets. A summary of the papers is shown in Table 1.

Table 1,
Literature survey of all the referred papers

Year	Author	Point of Selecting the Research
2019	Liu et al. [19]	Explored hybrid deep learning methods for anomaly detection in crowded areas using unsupervised methods.
2020	Sreenu et al. [21]	Presented live CCTV streaming videos as a key big data source for security in crowded areas.
2020	Farroq et al. [22]	Proposed CNN-based approach to detect crowd divergence behavior and prevent stampedes, outperforming other techniques.
2020	Behera et al. [23]	Suggested PIDLNet, combining 3D convolutional features with physics-based properties for crowd action prediction
2020	Varghese et al. [24]	Developed a model combining deep learning and fuzzy logic for crowd behavior management
2021	Peixoto et al. [25]	Presented CNN-LSTM and C3D neural networks for violence detection, achieving 56% accuracy.

2021	Soliman et al. [26]	Used CNN-LSTM approach for violence detection, reaching 88.2% accuracy on the RealLife Violence Situations dataset.
2021	Abdali et al. [27]	Developed a real-time violence detector with 98% accuracy at 131 fps
2021	Jain et al. [28]	Utilized Inception ResNetV2 to transform RGB videos into dynamic images for motion recognition, achieving up to 100% accuracy.
2021	Sernani et al. [29]	Focused on reducing false positives in violent behavior detection models
2021	Kang et al. [30]	Proposed MobileNetV3 and EfficientNet-B0 architecture with less computational power than existing 3D-CNNs.
2021	Gadelkarim et al. [31]	Proposed models for violence detection benefiting from transfer learning, achieving higher accuracy on XD-Violence and UCF-Crime datasets.
2021	Islam et al. [32]	Proposed two-stream deep learning model with significant performance on RWF-2000 dataset.
2021	Honarjoo et al. [33]	Used convolutional-based approach to process violent datasets with feature vectors.
2021	Guedes et al. [34]	Developed method using dynamic images with CNN and SVM, achieving high accuracy in various datasets, including movies and crowd violence.

3. Methodology

Section 3 presents an overview of the study's methodologies, including dataset acquisition, data pre-processing techniques (such as gamma correction) and the design of the CNN model for violence prediction. Moreover, this section explains the training, evaluation and performance metrics of these models, including the hybrid model (CNN–SVM) for feature extraction.

3.1 Datasets

The Real-Life Violence Situations dataset [35] consists of 2000 videos, which takes 1000 violence videos and 1000 non-violence videos crawled from YouTube. Each video ranges from 50-150 frames, and the violent and non-violent videos are clips of everyday life. Various images from the movies are shown in Fig. 2, showing images of violence with the head of the first row and non-violence of the watershed of the second row.

Fig. 2. RLVS dataset samples for (a) violence and (b) non-violence.

3.2 Proposed Approach

The block diagram provides a quick summary of all the procedures and stages involved in our methodology (Fig. 3). This diagram is divided into three phases: dataset preprocessing, model training and testing and validation. The initial stage, known as preprocessing, involves gathering raw data, cleaning it and adapting it for compatibility with a machine learning model. Thereafter, the models are trained for a specific task using either newly collected data or pre-trained modules. After the models have learned the fundamentals, they are tested across various applications to evaluate their performance in real-world scenarios. Each procedure is discussed in detail below.

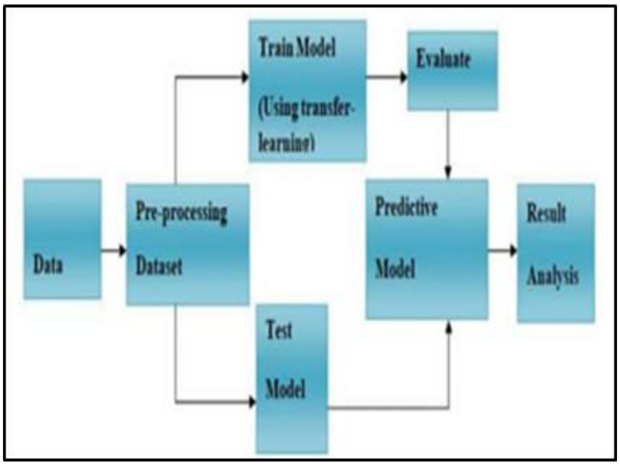


Fig. 3. Proposed model's block diagram

3.2.1 Data Pre-processing

Noise and missing values are unavoidable in the real world data, and ill-structured data will mislead machine learning models. Pre-processingThe pre-processing is intended to bring higher quality and accuracy to the dataset by refining the information in a descending order of priority and is empirical in nature to ensure the reliability in actual career The consistency of the data is assured. When introducing the proposed architecture, the shortcomings of skeletal detection may lead to the loss of information. Two pre-processing methods, histogram equalization and gamma correction, are applied to videos. Concretely, the contrast-limited adaptive histogram equalisation as implemented in Open CV (Open CV, 2022) is applied to compute histogram equalisation and γ = 1.5 for gamma correction. Performance is further enhanced by the addition of gamma correction. The results are shown in Fig. 4. The gamma correction of an image I containing pixel values in the range [0, 255] is given by the following formula:

$$I' = (I/255)^{Y} \times 255. \quad \ldots(1)$$

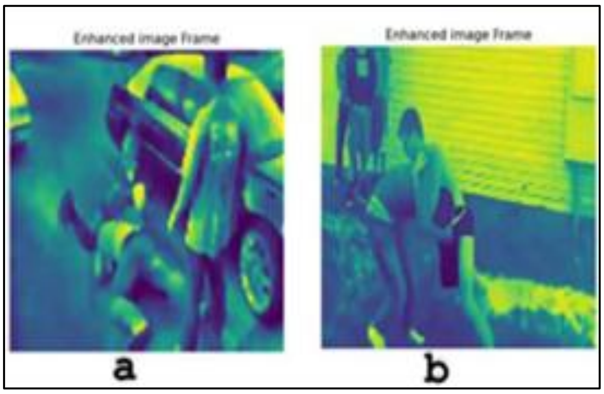


Fig. 4. a. Samples from the model with Gamma correction; b. Samples from the model without Gamma correction.

3.2.2 Data Partitioning

This dataset is split into testing set and training set (including validation set), which are 20% and 80% of the original photos respectively. The key difference between the two sets of images is that the training dataset has known ground truth labels for all the images, and the testing images are unmarked. This is a good opportunity to evaluate the model, test it with unseen data we have collected and execute experiments following our protocols.

3.2.3 Training and Testing Models

Loss, cost function, and accuracy are computed at every epoch (passing all the training samples once) for training of the designed network layers at few epochs. The cost function called 'Sparse Categorical Cross entropy' is used in this situation to prevent overfitting. Training is carried out on VGG-16 with a batch size of 32 in order to create a head model for a fully connected classification layer using transfer learning methods. Classification report comprising the precision, F1 score and recall of the training models is created after evaluating them. Validation of these models is performed on real-world violent events dataset in order to evaluate their performance.

4. Analysis of Results

Once the model is created, and the desired outcome is achieved, whether the model is producing high-quality outcomes must be determined. A confusion matrix can be used to show the trained models' success rate to evaluate accuracy [36]. The variable value in Fig. 5 can be either positive or negative. The predicted data are shown in rows, and the actual data are presented in columns. True Positive: The actual data are positive and predicted as positive. Equation (2) can be used to determine the modelling accuracy. Equation (3) states that recall indicates the proportion of correctly predicted outcomes amongst all correctly predicted outcomes. The error rate (Equation (4)) indicates the percentage of the model's predictions that are erroneous. A low error rate indicates that the model effectively operates.

$$Accuracy{=}TP{+}TN/\ TP{+}TN\ TP{+}FP{+}TN{+}FN \quad \ldots(2)$$

$$Recall{=}TP/TP{+}FN \quad \ldots(3)$$

$$Errorrate{=}FP{+}FN/FP{+}FNTP{+}FP{+}TN{+}FN \quad \ldots(4)$$

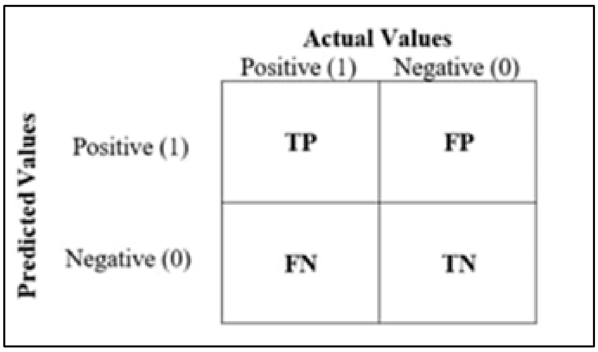


Fig. 5. Confusion matrix.

The method is evaluated by calculating the precision, recall and F1-score to illustrate how the proposed model worked on the dataset (Table 2).

Table 2,
Evaluation for classification on the RLV-Crime dataset.

Evaluation Metric	(Values%)
Precision	0.69
Recall	0.50
F1-score	0.58

Table 3 presents a comparison of the training and validation accuracy for both models. The CNN model with gamma correction achieves a training accuracy of 99.4% and a validation accuracy of 99.74%. Meanwhile, the model without gamma correction achieves a training accuracy of 97.5% and a validation accuracy of 97.13%. These results demonstrate that gamma correction enhances the model's overall performance, particularly in validation accuracy, which is critical for assessing how well the model generalises to new, unseen data. Furthermore, these results underscore the effectiveness of gamma correction in improving the robustness and reliability of the CNN model for violence detection.

Table 3,
The values of both train accuracy and validation accuracy for the RLV with CNN.

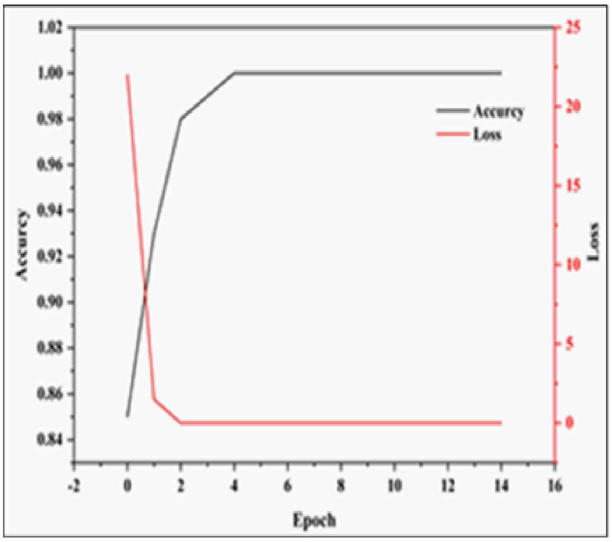
Model	Train accuracy	Validation accuracy
CNN with Gamma correction	99.4 %	99.74 %
CNN without Gamma correction	97.5 %	97.13 %

The proposed method is compared with a 3D convolutional network using accuracy (Acc) as the assessment parameter. Table 4 compares the Acc values for binary classification using our proposed method (CNN–SVM) with the different models for anomaly detection. We categorise all abnormal events as 'Anomaly' and non-anomalous data as 'Normal'. The test classifier displays the likelihood of a proper categorisation for aberrant events. Thus, our model outperformed the prior techniques (Table 4).

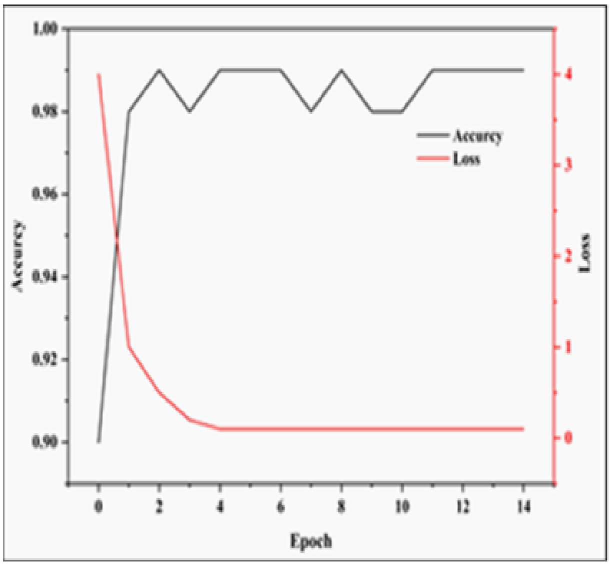
Table 4,
Acc for binary classification on the RLV-Crime dataset.

Model	Acc (%)
SVM Baseline	50.0
Zhong et al. (C3D) [37]	81.08
Sultani et al. (loss without constraints) [38]	74.4
our proposed model	99.7

Fig. 6(a) shows the training curve for binary classification in terms of accuracy and loss value using our suggested technique with gamma correction. Fig. 6(b) depicts the training curve for binary classification without gamma correction. The diagram illustrates the training curves of the CNN with an SVM classifier, comparing performance with and without gamma correction. In both cases, the model shows a rapid decrease in loss and a sharp increase in accuracy during the initial epochs, indicating effective learning. However, the gamma-corrected model (Fig. 6(a)) more quickly achieves lower loss and higher accuracy and stabilises at a slightly better performance level compared with the model without gamma correction (Fig. 6(b)). This notion suggests that gamma correction enhances the model's ability to learn and generalise from the training data, resulting in improved overall performance. The live detection system is applied, where a streaming video, either from the source camera or the web camera, will be converted into data frames and predicted (Fig. 7).



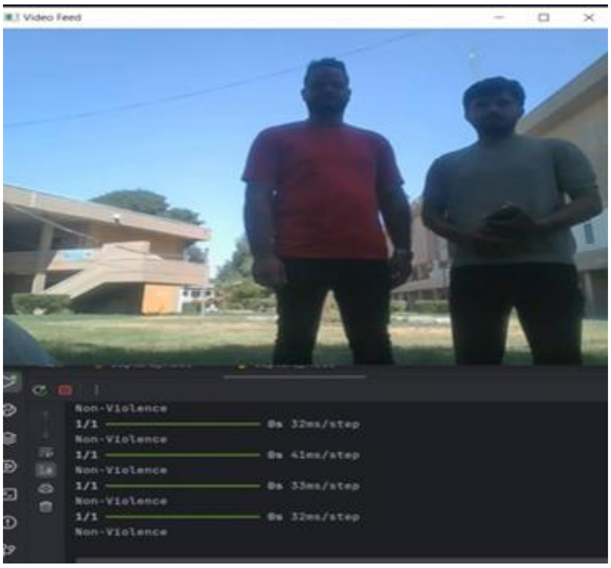
(a)



(b)

Fig. 6. Training curve for binary classification of RLV dataset with 2DCNN (a) with gamma correction, (b) without gamma correction.

(a)



(b)

Fig. 7. Output of the proposed method, for (a) violent, (b) nonviolence.

5. Conclusion

This study shows that deep learning methods, particularly CNN models combined with gamma correction, provide an effective method for identifying violent activities in surveillance footage. The model achieved a remarkable accuracy of 99% on the RLV dataset by leveraging the advantages of CNN's spatial feature extraction and integrating SVM for classification. The gamma correction step further enhances the quality of feature extraction, improving the model's reliability under varying lighting conditions. By contrast, traditional methods that rely on hand-crafted datasets and basic machine learning techniques exhibit low performance and limited generalisability. The model's ability to handle real-time data and its computational cost efficiency make it a promising solution for real-world applications. Future research may explore additional datasets from various video sources to further enhance the model's generalisability. Another potential direction involves integrating more advanced deep learning models, such as ResNet or LSTM, to improve the detection of spatial and temporal features.

References

[1] Geetha, R., Gunanandhini, S., Srikanth, G. U., & Sujatha, V. (2024). Human Stress Detection in and Through Sleep Patterns Using Machine Learning Algorithms. Journal of The Institution of Engineers (India): Series B, 1-23.

[2] Tian, B.; Morris, B.T.; Tang, M.; Liu, Y.; Yao, Y.; Gou, C.; Shen, D.; Tang, S. Hierarchical and networked vehicle surveillance in its:A survey. IEEE Trans. Intell. Transp. Syst. 2017, 18, 25–48. [CrossRef]

[3] Varadarajan, J.; Odobez, J.M. Topic models for scene analysis and abnormality detection. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, Kyoto, Japan, 27 September–4 October 2009;pp. 1338–1345

[4] Liu, M. C., Hsu, F. R., & Huang, C. H. (2024). Complex event recognition and anomaly detection with event behavior model. Pattern Analysis and Applications, 27(2), 51.

[5] Vosta, S., & Yow, K. C. (2022). A cnn-rnn combined structure for real-world violence detection in surveillance cameras. Applied Sciences, 12(3), 1021.

[6] Han, D., Wang, Z., Chen, W., Wang, K., Yu, R., Wang, S., ... & Yin, X. (2023).

[7] Anomaly Detection in the Open World: Normality Shift Detection, Explanation, and Adaptation. In NDSS.

[8] Park, H., Noh, J., & Ham, B. (2020). Learning memory-guided normality for anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 14372-14381).

[9] Dong, N.; Jia, Z.; Shao, J.; Xiong, Z.; Li, Z.; Liu, F.; Zhao, J.; Peng, P. Traffic abnormality detection through directional motion behavior map. In Proceedings of the 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance,Boston, MA, USA, 29 August–1 September 2010; pp. 80–84

[10] Loy, C.C.; Xiang, T.; Gong, S. Detecting and discriminating behavioural

[11] anomalies. Pattern Recognit. 2011, 44, 117–132. [CrossRef]

[12] Theves, S., Fernandez, G., & Doeller, C. F. (2019). The hippocampus encodes distances in multidimensional feature space. Current Biology, 29(7), 1226-1231.

[13] Cheng, L., Luo, S., Li, B., Liu, R., Zhang, Y., & Zhang, H. (2023). Multiple-instance learning for EEG based OSA event detection. Biomedical Signal Processing and Control, 80, 104358.

[14] Li, T., Wang, Z., Liu, S., & Lin, W. Y. (2021). Deep unsupervised anomaly detection. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 3636-3645).

[15] Nezamabadi, K., Sardaripour, N., Haghi, B., & Forouzanfar, M. (2022). Unsupervised ECG analysis: A review. IEEE Reviews in Biomedical Engineering, 16, 208-224.

[16] Usmani, U. A., Happonen, A., & Watada, J. (2022, July). A review of unsupervised machine learning frameworks for anomaly detection in industrial applications. In Science and Information Conference (pp. 158-189). Cham: Springer International Publishing.

[17] Fan, C., Liu, Y., Liu, X., Sun, Y., & Wang, J. (2021). A study on semi-supervised learning in enhancing performance of AHU unseen fault detection with limited labeled data. Sustainable Cities and Society, 70, 102874.

[18] Baradaran, M., & Bergevin, R. (2024). A critical study on the recent deep learning based semi-supervised video anomaly detection methods. Multimedia Tools and Applications, 83(9), 27761-27807.

[19] Ramírez-Sanz, J. M., Maestro-Prieto, J. A., Arnaiz-González, Á., & Bustillo, A. (2023). Semi-supervised learning for industrial fault detection and diagnosis: A systemic review. ISA transactions.

[20] Available online: https://visionlab.uncc.edu/download/summary/60-data/477-ucf-anomaly-detection-dataset (accessed on 12January 2018).

[21] Duong, H. T., Le, V. T., & Hoang, V. T. (2023). Deep learning-based anomaly detection in video surveillance: a survey. Sensors, 23(11), 5024.

[22] Liu, C. H., Chen, Z., & Zhan, Y. (2019). Energy-efficient distributed mobile crowd

sensing: A deep learning approach. IEEE Journal on Selected Areas in Communications, 37(6), 1262-1276.

[23] Sreenu, G., & Durai, S. (2019). Intelligent video surveillance: a review through deep learning techniques for crowd analysis. Journal of Big Data, 6(1), 1-27.

[24] Farooq, M. U., Saad, M. N. M., & Khan, S. D. (2022). Motion-shape-based deep learning approach for divergence behavior detection in high-density crowd. The Visual Computer, 1-25.

[25] Behera, S., Vijay, T. K., Kausik, H. M., & Dogra, D. P. (2021, November). PIDLNet: A physics-induced deep learning network for characterization of crowd videos. In 2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-8). IEEE.

[26] Varghese, E. B., Thampi, S. M., & Berretti, S. (2020). A psychologically inspired fuzzy cognitive deep learning framework to predict crowd behavior. IEEE Transactions on Affective Computing, 13(2), 1005 1022.

[27] Peixoto, B., Lavi, B., Martin, J. P. P., Avila, S., Dias, Z., & Rocha, A. (2019, May). Toward subjective violence detection in videos. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8276-8280). IEEE.

[28] Soliman, M. M., Kamal, M. H., Nashed, M. A. E. M., Mostafa, Y. M., Chawky, B. S., & Khattab, D. (2019, December). Violence recognition from videos using deep learning techniques. In 2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS) (pp. 80-85). IEEE.

[29] Abdali, A. M. R., & Al-Tuma, R. F. (2019, March). Robust real-time violence detection in video using cnn and lstm. In 2019 2nd Scientific Conference of Computer Sciences (SCCS) (pp. 104-108). IEEE.

[30] Jain, A., & Vishwakarma, D. K. (2020). Deep NeuralNet For Violence Detection Using Motion Features From Dynamic Images. In 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT). https://doi.org/10.1109/icssit48917.2020.9214153

[31] Sernani, P., Falcionelli, N., Tomassini, S., Contardo, P., & Dragoni, A. F. (2021). Deep Learning for Automatic Violence Detection: Tests on the AIRTLab Dataset. IEEE Access, 9, 160580–160595. https://doi.org/10.1109/access.2021.3131315

[32] Kang, M., Park, R., & Park, H. (2021). Efficient Spatio-Temporal Modeling Methods for Real-Time Violence Recognition. IEEE Access, 9, 76270–76285. https://doi.org/10.1109/access.2021.3083273

[33] Gadelkarim, M., Khodier, M., & Gomaa, W. (2022). Violence Detection and Recognition from Diverse Video Sources. In 2022 International Joint Conference on Neural Networks (IJCNN). https://doi.org/10.1109/ijcnn55064.2022.9892660

[34] Honarjoo, N., Abdari, A., & Mansouri, A. (2021). Violence Detection Using OneDimensional Convolutional Networks. https://doi.org/10.1109/ikt54664.2021.9685835

[35] Guedes, A. R. M., & Cámara-Chávez, G. (2020). Real-Time Violence Detection in Videos Using Dynamic Images. https://doi.org/10.1109/clei52000.2020.00065.

[36] Real life Violence Situation dataset details,[Availablonline]:https://paperswithcode.com/dataset/real-life-violence-situations-dataset.

[37] Kim J H, Song J H and Lim D H 2020. CT Image Denoising Using Inception Model. Journal of the Korean Data And Information Science Society, 31 (3), pp. 487–501.

[38] Ryoo, M. S., & Aggarwal, J. K. (2009, September). Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities. In 2009 IEEE 12th international conference on computer vision (pp. 1593-1600). IEEE.

[39] Zhong, J. X., Li, N., Kong, W., Liu, S., Li, T. H., & Li, G. (2019). Graph convolutional label noise cleaner: Train a plug-and-play action classifier for anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1237-1246).

[40] Sultani, W., Chen, C., & Shah, M. (2018). Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6479-6488).

تقدير التنبؤ بالعنف في كاميرات المراقبة باستخدام CNN مع تصحيح جاما

سارة اياد ربيع[1]*، يعرب عمر ناجي[2]

[1] *دائرة الإصلاح العراقية، وزارة العدل، بغداد، العراق*

[2] *قسم هندسة الميكاترونيكس، كلية الهندسة الخوارزمي، جامعة بغداد، بغداد، العراق*

*البريد الالكتروني: saraaagh785@gmail.com

المستخلص

شهدت السنوات الأخيرة ارتفاعًا كبيراًفي نشر كاميرات المراقبة لتعزيز الأمن في كلّ من الأماكن العامة والخاصة. لا تزال الكثير من الشركات تستأجر شخصًا لمراقبة الكاميرات، ولكن نظرًا لحدوث خطأ بشري في بعض الأحيان، فمن الراجح أن يتغاضى الفرد المستأجر عن بعض الأحداث الغريبة في موجزات الفيديو. وبالتالي قد يكون من غير المجدي قضاء الوقت والجهد في تتبع كاميرات المراقبة. في هذه الدراسة أبحث في تطبيق تقنيات التعلم العميق، وخاصة الشبكات العصبية التلافيفية (CNNs) وآلات الدعم المتجهة(SVMs) ، للتنبؤ بالعنف في تدفقات فيديو المراقبة. من خلال استخدام تصحيح جاما كخطوة للمعالجة المسبقة، تم تحسين نموذج CNN المقترح لاستخراج السمات المكانية الأساسية من إطارات الفيديو، مما يعزز بشكل كبير دقة اكتشاف العنف. تستفيد الدراسة من قدرات الوقت الفعلي لبيانات المراقبة، مع التركيز على مجموعة بياناتRLV ، والتي تشمل مجموعة من السيناريوهات العنيفة وغير العنيفة. وحقق نموذج CNN-SVM الهجين الذي تم تطويره في هذا البحث دقة مذهلة بنسبة ٩٩٪، متفوقًا على الطرق التقليدية ومظهرًا قدرات استخراج السمات المكانية القوية. فضلاً عن ذلك، يتناول هذا البحث تحديات المراقبة بالفيديو في الوقت الفعلي من خلال ضمان قابلية التوسع والتطبيق العملي، مما يوفر حلاً قويًا لتعزيز تدابير الأمن في الأماكن العامة والخاصة.