Journal of Information Systems Engineering and Management
2025, 10(26s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

An Efficient Deep Learning Based Approaches for Crime Activities classification in Surveillance Videos

Raed Mageed[1], Hiyam Hatem[2]
[1,2]Affiliation Collage Of Computer Science And Information Technology, University of Sumer, Iraq

ARTICLEINFO

Received: 28 Dec 2024

Revised: 21 Feb 2025

Accepted: 01 Mar 2025

ABSTRACT

Video surveillance is mostly utilized in crowded areas to find and identify unusual activity in a complicated environment. Something out of the ordinary, or abnormal, is called an anomaly. Modeling and processing the results of the unusual situation might be a daunting and seemingly impossible task. Consequently, this study presents a deep-learning methodology for constructing a crime detection system. This advanced methodology encompasses many layers essential for feature extraction and classification, enabling the system to effectively and reliably identify criminal behaviors, the presented criminal activity classification models, where, two CNN methodologies employed (EfficientNet-B7 and ResNet50) trained and evaluated on the widely recognized UCF Crime datasets. The experiments uncover that the proposed system achieved outstanding results and outscored the other deep learning approaches with an accuracy of 99.48%, precision of 99.47%, sensitivity of 99.41%, and F1 score of 99.44% using the UCF crime dataset. Hence, this system is effectively capable of tracking criminals' trails and detecting crime events.

Keywords: Video surveillance, Deep learning, Crime Activity Detection, UCF-Crime dataset, Pre-trained CNN approach.

INTRODUCTION

Robbery, kidnapping, molestation, murder, vandalism, and violent assaults represent significant threats to public safety and societal order. There is an increase in the daily incidence of crimes. These actions encompass only those that are significant. Minor actions can lead to significant consequences over time; therefore, all activities should be carefully monitored. Governments globally are exploring innovative methods for crime identification and prevention [1]. Smart cities, encompassing educational institutions, retail centers, hospitals, and other facilities, are implementing surveillance cameras to prevent or identify theft. Effective management of the increasing volume of cameras and videos necessitates sufficient personnel resources. However, even with adequate staffing, video managers and filterers may experience mental health challenges as a result of heightened exposure to aggressive and criminal content [2].

Artificial intelligence-based computer vision methods for automated video crime detection are gaining increasing importance [3]. Deep learning (DL) algorithms, particularly convolutional neural networks (CNNs), has garnered significant interest in video crime detection, demonstrating promising outcomes [4]. CNNs are proficient in identifying intricate features within high-dimensional datasets, rendering them particularly suitable for tasks involving image and video classification and detection [5]. CNNs capable of processing complex criminal data and detecting anomalies. Crime activity detection systems can enhance flexibility, accuracy, and management of criminal activity progression through the effective application of CNN techniques [6].

Generally, crime activity detection can be categorised into three primary types: weakly supervised, supervised, and unsupervised. The deep learning method is trained in supervised detection of criminal activity using datasets that include labelled instances of both criminal and normal behaviors. This method learns to differentiate between criminal and normal instances based on the labelled data. In weakly supervised crime detection, a DL approach is applied to both labelled and unlabelled data concurrently. The DL method is trained on a dataset primarily consisting of labelled data (normal instances) and lacks explicit labels for crime instances. This approach learns normal patterns based on labelled data and identifies criminal patterns by analysing highly deviant instances from these patterns. The final category relies exclusively on unlabelled data to identify significantly deviant patterns without prior knowledge of what constitutes normal behaviour. The categories possess distinct advantages and disadvantages, with their selection influenced by factors such as data characteristics, availability of labelled data, and additional requirements pertinent to the crime detection task [7].

The subsequent sections of this paper are organised in the following manner. The subsequent section will present related works. Section 3 provides a detailed explanation of the crime-related dataset employed and the proposed system for detecting criminal activity. Section 4 presents a comprehensive overview of the experiments, evaluations, and discussions conducted. The final section highlights the key conclusions and potential future directions.

RELATED WORKS

Lorem ipsum Development of criminal detection technology is absolutely vital to prevent criminal activity. Deep learning has produced novel approaches and findings for crime detection. In this section we present some of the DL methods using for field of the research.
Khaire and Kumar [8] suggested Mobile-Net and Bidirectional Long Short-Term Memory, a pre-trained CNN-based criminal detection system for crucial situations like ATMs. We used raw RGB and depth pictures to extract rich visual characteristics and reduce training and classification computing complexity. This system excels because it trains solely on poorly labelled normal samples. The UCF crime2local dataset of six classes was used to train and test the proposed system. A subset of the UCF Crime dataset with 300 RGB video clips was used. The proposed approach has amazing 91.1% accuracy.

Park et al. [9] introduced a deep learning-based surveillance video anomaly detection system that filters noisy predictions from poor supervision using an absorbing Markov Chain (MC). Graph convolutional networks were used to integrate this MC into deep learning network training. Segment-wise labels were improved using Gaussian mixture-based pseudo-labeling. This framework achieved 95.61% and 84.94% accuracy for Shanghai Tech and UCF Crime datasets.

Thakare et al. [10] developed a multi-stream deep learning surveillance video anomaly detection system. This approach targeted strong visual characteristics by integrating spatial and motion data. Multiple-instance learning-based advanced classifiers can manage feature variances. Using fuzzy aggregation, several feature stream scores were fused to improve segment recognition in long-duration movies. Eventually, only accident and fire activity were classified using a lightweight two-class classifier. For human and robbery classes, this classifier is unsuitable. With the UCF Crime dataset, our system achieved 84.48% classification accuracy in broad testing.

Qasim and Verdu [11] proposed a deep learning approach to classify normal or anomalous behaviors' in the Binary and UCF Crime datasets: violent and aberrant behaviors filmed by cameras in public settings. ResNet-18, ResNet-34, and ResNet-50 CNNs retrieve high-level spatial characteristics from video frames, whereas the simple recurrent unit (SRU) collects temporal features. ResNet-50 and SRU beat the other combined models on Binary and UCF-Crime datasets with 91.63% and 91.25% accuracy. Improving anomaly categorization, notably in the UCF Crime dataset, is necessary.

Gulati et al. [12] developed a criminal actions detection system using improved CNN with two convolutional layers, Batch normalization, max pooling, global average pooling, one dense layer, and sigmoid function for binary classification. This method employed a rolling averaging strategy to average the approach's predictions across "n" frames and use the highest probability forecast for these frames. The studies were done using Hockey Fight, Violent Flows, and DCSASS (a subset of the UCF crime dataset), and the findings were 97.94%, 95.75%, and 87.56% accurate.

Patwal et al. [13] attempted to identify damage, violence, burglary, assault, theft, explosion, and crowd abnormalities. This system used DenseNet-121, a powerful pre-trained CNN, to extract features and classify. It had 86.63% accuracy using the UCF Crime dataset. The classification accuracy of this system should be enhanced utilising different CNN methods.

Table 1: Methods used for criminal activates classification.

Ref.	Methodology	Key Features	Datasets	Achieved Accuracy
[8]	Mobile-Net and Bidirectional LSTM	Pre-trained CNN, raw RGB and depth images, reduced computational complexity	UCF Crime2local (6 classes, 300 RGB video clips)	91.1%
[9]	Deep Learning with Absorbing Markov Chain	Graph convolutional networks, Gaussian mixture-based pseudo-labeling	Shanghai Tech, UCF Crime	95.61% (Shanghai Tech), 84.94% (UCF Crime)
[10]	Multi-stream Deep Learning	Spatial and motion data integration, fuzzy aggregation	UCF Crime	84.48%
[11]	ResNet with SRU	ResNet-18, ResNet-34, ResNet-50 for spatial features, SRU for temporal features	Binary, UCF Crime	91.63% (Binary), 91.25% (UCF Crime)
[12]	Improved CNN	Two convolutional layers, Batch normalization, max pooling, global average pooling, dense layer, sigmoid function	Hockey Fight, Violent Flows, DCSASS (subset of UCF Crime)	97.94% (Hockey Fight), 95.75% (Violent Flows), 87.56% (DCSASS)
[13]	DenseNet-121	Pre-trained CNN for feature extraction and classification	UCF Crime	86.63%

PROPOSED DETECTION MODELS

DATASET DESCRIPTION

The UCF Crime dataset comprises a comprehensive compilation of 128 hours of video recordings obtained from actual security cameras [14]. The collection consists over 1,900 extensive and unedited recordings, illustrating the unrefined and authentic characteristics of security surveillance. The movies document 13 unique categories of aberrant behaviours that present considerable risks to public safety. The acts encompass abuse, arrest, arson, assault, road accidents, burglary, explosion, fighting, robbery, shooting, stealing, shoplifting, and vandalism.

PREPROCESSING STEPS

This step guarantees that the deep learning approaches become less sensitive to pixel intensity variations, contributing to enhanced convergence throughout training. We implement the subsequent preprocessing procedures to prepare the dataset for training the DL models, firstly the Surveillance videos were first transformed into consecutive frames with a frame rate of 30 frames per second.

1. Resizing: All frams (images) are adjusted to a uniform resolution of : (62×62) pixels in the UCF Crime dataset to maintain uniformity in input dimensions and minimise processing cost.

2. Normalisation: Pixel values of the video frames are normalized within [0, 1] using min-max normalization method to ensure that frames held the same scales. In the min-max normalization, every pixel intensity in the resized image (it's value ranging from 0 to 255) is processed individually, in which the normalized value can be attained via subtracting the pixel value from the minimum value in the non-criminal or criminal image divided by the subtraction of the maximum value from the minimum value, resulting to the scaled image.

DATASET SPLITTING

The UCF Crime datasets are divided into three subsets to support the training, validation, and testing of deep learning models.

1. Training Set (80%): This portion is employed for model training. It encompasses the bulk of the data and facilitates the model's comprehension of the fundamental patterns and characteristics essential for precise predictions.

2. Validation Set (10%): This subset is used to optimise the model's hyperparameters and mitigate overfitting. It offers an impartial assessment of the model's efficacy during the training phase.

3. Test Set (10%): This subset is utilised to assess the model's ultimate performance. It offers an impartial evaluation of the model's capacity to generalise to novel, unobserved data.

DATA AUGMENTATION

To improve the models' generalisation ability and mitigate overfitting, augmentation approaches substantially expand the effective size of the training dataset, allowing the models to acquire robust features that generalise effectively to novel data. We implement various data augmentation strategies on the training set:

1. Brightness Adjustment: The luminosity of frames is randomly modified within the interval [0.8, 1.2] to replicate diverse lighting situations.

2. Zooming: frames are enlarged by up to 10% to replicate varying proximities to the anomaly source.

THE PROPOSED DL MODEL

The proposed deep CNN approach encompasses two processes; extracting features and classification, Figure 1.

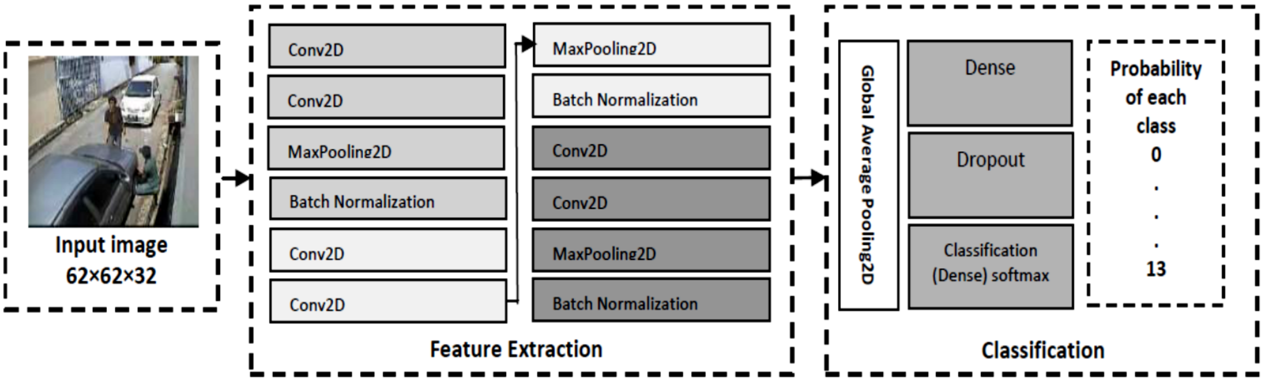


Figure 1: the proposed DL diagram

The description of the layers, their output shapes, and parameters for the proposed approach are depicted in Table 1.where , the Total parameters = 334,030 , Trainable parameters = 333,518 and Non-trainable parameters = 512 .

Table 2: Layers Details of the CNN Model.

Layer No.	Layer (Type)	Number of Filters	Output Shape	Number of Training Parameters
1	Conv2D	32	(None, 62, 62, 32)	896
2	Conv2D	64	(None, 60, 60, 64)	18496
3	MaxPooling2D	-	(None, 30, 30, 64)	0
4	Batch Normalization	-	(None, 30, 30, 64)	256
5	Conv2D	64	(None, 28, 28, 64)	36928
6	Conv2D	64	(None, 26, 26, 64)	36928
7	MaxPooling2D	-	(None, 13, 13, 64)	0
8	Batch Normalization	-	(None, 13, 13, 64)	256
9	Conv2D	128	(None, 11, 11, 128)	73856
10	Conv2D	128	(None, 9, 9, 128)	147584

11	MaxPooling2D	-	(None, 4, 4, 128)	0
12	Batch Normalization	-	(None, 4, 4, 128)	512
13	Global Average Pooling2D	-	(None, 128)	0
14	Dense	-	(None, 128)	16512
15	Dropout	-	(None, 128)	0
16	Classification (Dense)	-	(None, 14)	1806

The procedure of feature extraction identifies the most pertinent image features to generate a sequence of feature vectors. This process is achieved through three fundamental blocks. Each block consists of four layers.

a) Two layers of convolution are employed to execute the primary operations within the proposed CNN, facilitating the extraction of essential features by processing a predetermined number of images to generate output 2D feature maps. Each layer is subjected to the Rectified Linear Unit (ReLU) activation function to enhance the speed of the training process and facilitate faster convergence.

b) One MaxPooling: This layer is designed to down-sample the resulting feature maps, effectively reducing the feature maps by specific regions within the image. Max pooling mitigates overfitting by abstracting the visual representation of the image.

c) One Batch normalisation significantly improved the efficiency of CNN training. Batch normalisation mitigates the issues of fading and exploding gradients by stabilising activations throughout training, thereby reducing the dependency on careful initialisation. Furthermore, it functions as a regularizer, thereby eliminating the necessity for employing dropouts. Dropout serves as an effective method for addressing overfitting in convolutional neural networks (CNNs), enabling better generalisation to unseen data and enhancing overall performance.

In the classification process, numerical image features are categorised into 14 classes for crime activity utilising the UCF Crime dataset.

The classification process for the UCF crime dataset is executed using the following steps:

a) One Global Average Pooling: This pooling method offers several advantages that enhance the performance of the approach. The method selects the most significant elements from the input to capture and learn the key features within the image, thereby improving the accuracy of the approach.

b) A dense layer: comprising fully connected neurones transforms the feature maps obtained from Global Average Pooling into a one-dimensional vector, thereby influencing the entirety of the feature information.

c) Dropout: involves the random selection of neurones that are excluded from the training process.

d) The Softmax layer facilitates classification by ensuring that the sum of probabilities across all classes equals one, thereby providing a normalised output suitable for image classification tasks. The normalisation attribute may enhance classification reliability and stability, particularly in the context of imbalanced datasets.

PRE-TRAINED CNN APPROACH IMPLEMENTATION

The proposed system implemented two of the most efficient pre-trained CNN models EfficientNetB7 and Resnet50. EfficientNet-B7 [15] is accurate and computationally efficient. In this approach, a novel compound scaling method was introduced in which the count of layers (depth), channels (width), and input image size (resolution) were uniformly scaled using a collection of scaling coefficients, scaling the whole network dimension instead of one at a time for better performance with fewer resources. Depthwise separable convolutions reduce computing costs and improve efficiency without losing accuracy in EfficientNet-B7. EfficientNet-B7 has convolution (64-filter, 2-stride, and 3x3 kernel size), batch normalization, and Swish activation layers. Finally, classification uses global average pooling and fully linked layers. Widely acknowledged for its efficacy in image classification applications is the 50-layer deep convolutional neural network ResNet50 model. This work investigates the use of ResNet50 for crime categorization utilizing a large-scale dataset intended for criminal activity in the frames extracted from the dataset video [16]. This shows how well ResNet50 distinguishes crime-related activity in surveillance recordings. The accuracy achieved by EfficientNet-B7 approach was 97%, precision was 97%, sensitivity was 96%, and F1 score was 96%. While the accuracy achieved by the Resnet50 approach was 98%, precision was 99%, sensitivity was 98%, and F1 score was 98%.

EXPERIMENTAL RESULTS

Amet To investigate the system's performance various experiments have been conducted on the datasets. The UCF Crime dataset [14] involves 1900 surveillance videos associated with crime scenes (anomalies) in public locations. UCF Crime dataset contains 14 activities classes. In every full-length video, each tenth frame is extracted and merged for each video in that class. The count of extracted PNG images (frames) is 1,377,653 of size 64×64. Balancing the dataset required so (900,000) images are removed from the category of normal videos. The resulted balanced UCF Crime dataset is demonstrated in Figure 2.

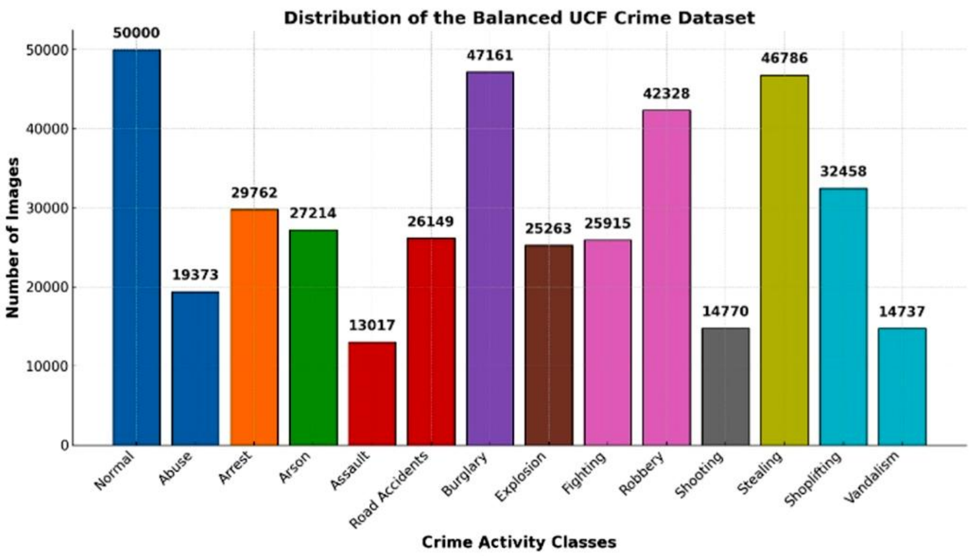


Figure 2. The balanced UCF Crime Dataset Distribution.

Figure 3: samples of the UCF-Crime dataset.

Generally, video crime detection systems are assessed by calculating the accuracy of crime activity classification for individual surveillance videos. Nevertheless, this measure is not capable of completely explaining the systems' performance. Other measures are also implemented that are more concise and can supply more precise information concerning the assessment of the classifier's performance.

The performance of the proposed DL classification models is by checking its confusion matrix. This matrix depicts the variance between the ground truth of the utilized dataset and the approach's predictions. Figure 4 illustrate the confusion matrices for the proposed CNN and the pre-trained models using the balanced UCF Crime.

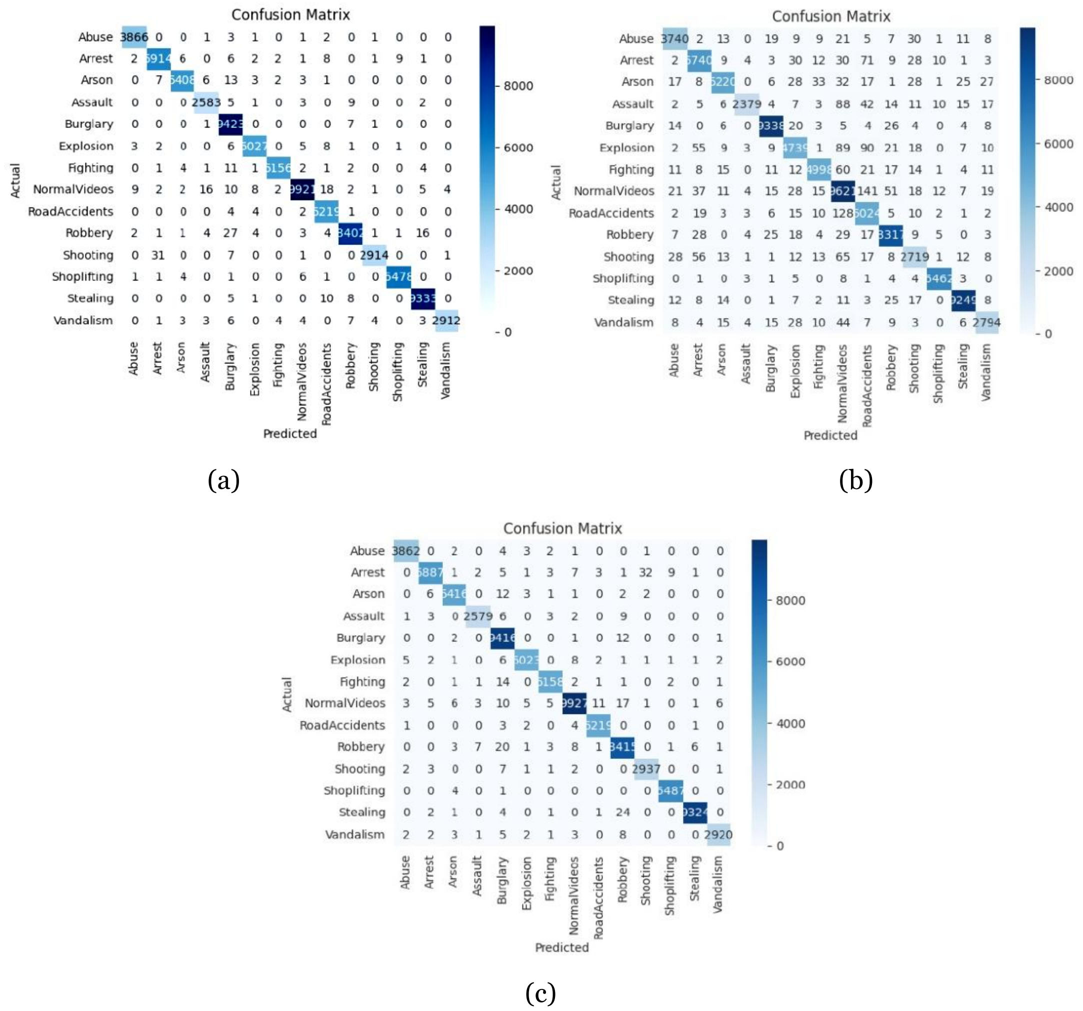


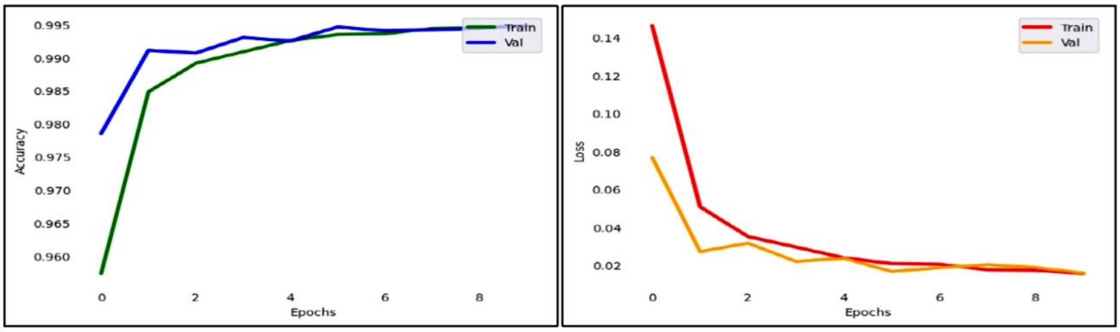
Figure 4: the Confusion matrix of UCF Crime dataset for (a) The proposed classification model, (b) EfficientNet-B7, (c) Resnet50.

Table 3 provides a comprehensive overview of the performance of the proposed system on the UCF Crime dataset, showcasing high precision, sensitivity, and F1 scores across various classes.

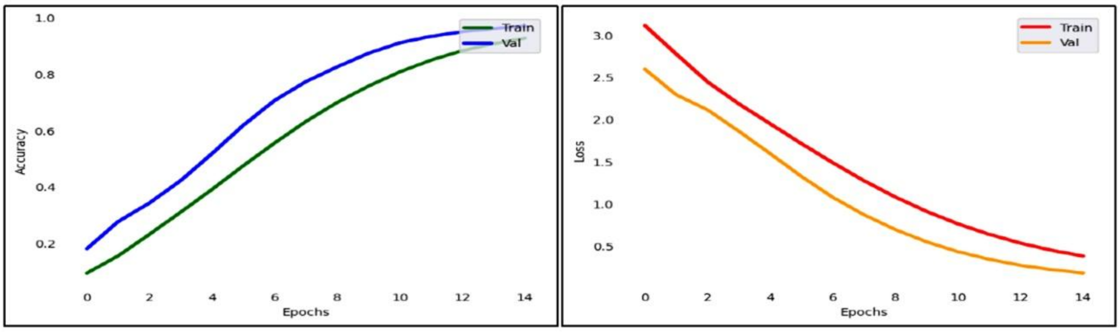
Table 3: the proposed models results.

Classes	**Precision**	**Sensitivity**	**F1 Score**
Abuse	0.995622	0.997677	0.996649
Arrest	0.992282	0.993616	0.992948
Arson	0.996315	0.993570	0.994941
Assault	0.987763	0.992317	0.990034
Burglary	0.989085	0.999152	0.994093
Explosion	0.995051	0.994855	0.994953
Fighting	0.998064	0.994791	0.996425
Normal Videos	0.996885	0.992000	0.994436
Road Accidents	0.989947	0.997897	0.993906
Robbery	0.995616	0.992440	0.994025
Shooting	0.996921	0.986459	0.991662
Shoplifting	0.998305	0.997843	0.998074
Stealing	0.996689	0.997435	0.997062
Vandalism	0.998286	0.988124	0.993179
Accuracy		0.994806	

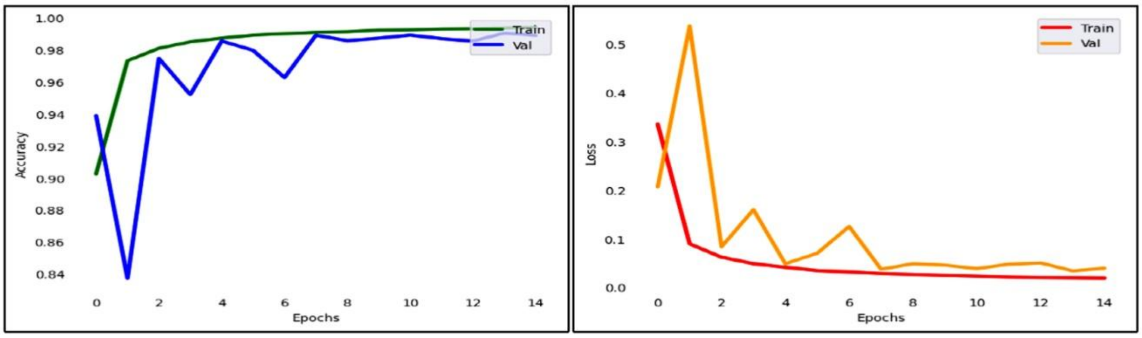
The proposed crime activity detection system was run at 10-15 epochs, 32-batch size, and 0.00003 learning rate. The final validation/training accuracies alteration using the proposed CNN approach were observed in epoch number 9, and their values were 0.9948 and 0.9938. The approach is over-fitted when the validation loss is decreased and increased once again, and it is under-fitted when the validation loss becomes extremely high. While the ideally lined up curve depicts the CNN approach performance is ideal. The final validation/training losses alteration was observed in epoch number 9, and their values were 0.0159 and 0.0156, Figure 5 reveal these modification until the best epochs were reached. The training approach performed better than the validation approach. In the validation approach, the accuracy increased and loss decreased since it referred to how the actual results were turned from the predicted results.



(a)



(b)



(c)

Figure 5: Validation/training accuracy & loss using (a) The proposed model, (b) EfficientNet-B7, (c) Resnet50.

Evidently for the pre-trained models, that the Resnet50 achieved higher accuracy than EfficientNet-B7, mainly for three reasons, the UCF datasets contain simpler patterns which efficiently learned by a smaller model, the datasets used do not contain enough variance, thus EfficientNet-B7 could not generalize well enough and Resnet50 had better pre-trained weights on the datasets used. The DL approach compared with the pre-trained models using the same count and size of frames, the experiments were accomplished using the same (conditions, epoch's count and batch size). Furthermore, a comparison with the other pertinent works presented in Table 4.

Table 4: Comparison with other Crime Classification Systems using UCF Crime dataset.

Ref. no.	Method implemented	Precision	Sensitivity	F1 Score	Accuracy
[17]	MobileNet-V2, Residual Attention-based LSTM	87%	78%	81%	78.43%
[10]	Multi-stream deep learning with fuzzy aggregation	89%	86%	81%	84.48%
[11]	ResNet-50, SRU models	91.54%	-	91.94%	91.25%
[9]	Deep neural network with absorbing Markov Chains	-	-	-	84.94%
[13]	DenseNet-121	-	-	-	86.63%
Proposed Approaches	Resnet50	99%	98%	98%	98%
	EfficientNet-B7	97%	96%	96%	97%
	Proposed DL model	99.47%	99.41%	99.44%	99.48%

We should refer to few facts: that Resnet50 approach provided higher performance than EfficientNet-B7. EfficientNet-B7 requires high computation which means slower inference speed making it not recommended for real-time criminal activity classification. The Resnet50 provided an intermediate solution for crime detection system based on some factors such as: computational efficiency, quick image processing, and limited storage capacity. Its main weaknesses come from the need for deeper feature extraction.

The accuracy achieved by EfficientNet-B7 approach was 97%, precision was 97%, sensitivity was 96%, and F1 score was 96%. While the accuracy achieved by the Resnet50 approach was 98%, precision was 99%, sensitivity was 98%, and F1 score was 98%. The proposed models achieved the highest results and outscored the other implemented DL approaches. The accuracy achieved by the proposed CNN approach was 99.48%, precision was 99.47%, sensitivity was 99.41%, and F1 score was 99.44%. These values of the measurements indicate the presented models are qualified for classify the crime activities and may offer a quit promising service for future implementation.

CONCLUSIONS AND FUTURE WORKS

Crime detection systems in surveillance films employing deep learning methodologies provide a compelling research domain aimed at reducing criminal activities and fostering a peaceful community. Numerous solutions have previously developed to mitigate crime rates; yet, significant enhancements in their efficacy remain possible. This study presents a modified deep learning model for the categorisation of anomalies and criminal behaviour, utilising the UCF Crime dataset for experimental purposes. The suggested system effectively attained the lowest loss and greatest accuracy on both the training and validation datasets. The suggested CNN approach successfully identifies and distinguishes criminal behaviours. This indicates that the structural modelling parameters significantly enhanced the classification of different criminal categories. The experimental findings indicate that the suggested method surpassed the performance of previously implemented techniques (EfficientNet-B7 and ResNet50) as well as other pertinent studies in identifying illegal behaviours. Future endeavours focus on gathering additional annotated datasets that accurately represent diverse criminal actions and creating methodologies that can process and capture temporal information from video feeds in real time without sacrificing precision.

REFRENCES

[1] K. Hu, L. Li, X. Tao, J. D. Velásquez, P. Delaney, "Information fusion in crime event analysis: A decade survey on data, features and models", Information Fusion, vol. 100, 2023.

[2] Hatem, H., Beiji, Z., & Majeed, R. (2015). Human facial features detection and tracking in images and video. Journal of Computational and Theoretical Nanoscience, 12(11), 4242-4249.

[3] K. B. Sahay, B. Balachander, B. Jagadeesh, G. A. Kumar, R. Kumar, L. R. Parvathy, "A real time crime scene intelligent video surveillance systems in violence detection framework using deep learning techniques", Computers and Electrical Engineering, vol. 103, 2022.

[4] F. L. Sánchez, I. Hupont, S. Tabik, F. Herrera, "Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects", Information Fusion, vol. 64, pp. 318-335, 2020.

[5] A. J. Yousif and M. H. Al-Jammas, “Real-time Arabic Video Captioning Using CNN and Transformer Networks Based on Parallel Implementation ”, Diyala Journal of Engineering Sciences, vol. 17, no. 1, pp. 84–93, Mar. 2024.

[6] S. Vosta, K.-C. Yow, "A CNN-RNN Combined Structure for Real-World Violence Detection in Surveillance Cameras", Applied Sciences, vol.12, no. 3, 2022.

[7] B. Asal, A. B. Can, "Ensemble-Based Knowledge Distillation for Video Anomaly Detection", Applied Sciences, vol. 14, no. 3, 2024.

[8] P. Khaire, P. Kumar, "A semi-supervised deep learning based video anomaly detection framework using RGB-D for surveillance of real-world critical environments", Forensic Science International: Digital Investigation, vol. 40, 2022.

[9] J. Park, J. Kim, B. Han, "End-to-end learning for weakly supervised video anomaly detection using Absorbing Markov Chain", Computer Vision and Image Understanding, vol. 236, 2023.

[10] K. V. Thakare, N. Sharma, D. P. Dogra, H. Choi, I.-J. Kim, "A multi-stream deep neural network with late fuzzy fusion for real-world anomaly detection", Expert Systems with Applications, vol. 201, 2022.

[11] M. Qasim, E. Verdu, "Video anomaly detection system using deep convolutional and recurrent models", Results in Engineering, vol. 18, 2023.

[12] A. Gulati, S. Pawar, P. Sheoran, C. Sali, "SUSPICIOUS BEHAVIOUR DETECTION USING CNN AND ROLLING AVERAGE", International Research Journal of Modernization in Engineering Technology and Science, vol.3, no.5, 2021.

[13] A. Patwal, M. Diwakar, V. Tripathi, P. Singh, "An investigation of videos for abnormal behavior detection", Procedia Computer Science, vol. 218, pp. 2264-2272, 2023.

[14] 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). IEEE. https://doi.org/10.1109/CVPR.2018.00680

[15] M. Tan, Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", International Conference on Machine Learning, 2019.

[16] Joshi, M., & Chaudhari, J. (2022). Anomaly Detection in Video Surveillance using SlowFast Resnet-50. International Journal of Advanced Computer Science and Applications, 13(10). http://dx.doi.org/10.14569/IJACSA.2022.01310112

[17] W. Ullah, A. Ullah, T. Hussain, Z. A. Khan, S. W. Baik, "An Efficient Anomaly Recognition Framework Using an Attention Residual LSTM in Surveillance Videos", Sensors, vol. 21, no. 8, 2021.