Genome Publications
https://doi.org/10.61096/978-81-990998-5-2_10

Chapter 10

Suspicious Activity Detection Using Deep Learning Approach

Prof. Rajendra M. Jotawar
Department of MCA, Acharya Institute of Technology, Bangalore-560107, India
Bhuvneshwari D.L
Department of MCA, Acharya Institute of Technology, Banglore-560107, India

Abstract: This project is focused on improving security systems by using deep learning and computer vision to spot suspicious human actions in real time. The goal is to help detect unsafe or dangerous behaviour by using smart models that can understand human movements from video footage. One of the main models used in this system is called Slow-Fast. This model looks at both slow and fast video frames to better understand how people move over time. By doing this, it can catch even small or complex actions that might seem unusual or suspicious. To make activity recognition more accurate, the system also uses another model called ResNet50, which helps in telling the difference between normal and strange behaviors by learning from patterns in the video. For example, YOLOv5 is used to quickly and accurately detect dangerous items like weapons or events like accidents and explosions. This helps in sending quick alerts during emergencies. Another part of the system uses MediaPipe, a tool that studies human body movements. It tracks how people move and can spot physical fights by noticing aggressive or violent actions. By putting all these tools together, the system can watch over places like schools, public areas, or offices and quickly alert security teams if something risky is happening. The project shows how advanced technology can be used in smart and helpful ways to keep people safe.

Keywords: Suspicious activity detection, yolov5, mediapipe, real-time surveillance, edge computing, feature extraction, proactive security.

Citation: Rajendra M. Jotawar, Bhuvneshwari D L. Suspicious Activity Detection Using Deep Learning Approach. Machine Learning in Research and Practice: A Multidisciplinary Perspective. Genome Publications. 2025; Pp85-93.
https://doi.org/10.61096/978-81-990998-5-2_10

INTRODUCTION

As cities grow and technology gets better, keeping people and places safe is more important than ever. Crimes like theft, vandalism, and trespassing can happen at any time. Most security systems still depend on humans to watch video footage, which can be tiring, slow, and not always reliable. To fix this, we need smarter systems that can detect danger on their own and respond quickly.

This project uses deep learning, and a tool called YOLO (You Only Look Once) to build a smart security system. YOLO is great at quickly recognizing objects and actions in videos. In this system, it watches live video from cameras and looks for suspicious behavior, like someone breaking in or acting strangely. If something is wrong, it sends an alert right away. The final system is light, easy to install, and works well in homes, offices, parking lots, and public places.

Role of YOLO in Suspicious Activity Detection

YOLO (You Only Look Once) plays a pivotal role in suspicious activity detection by serving as a real-time object detection framework that simultaneously identifies and localizes objects within surveillance footage. Unlike traditional detection methods, YOLO processes entire frames in a single pass, enabling real-time performance crucial for immediate threat response [3]. The algorithm excels in detecting suspicious objects like weapons, tools, and persons performing anomalous behaviors with up to 99% accuracy [2]. YOLO's efficiency stems from its unified architecture that divides images into grids and predicts bounding boxes with class probabilities simultaneously [4]. Various YOLO versions (YOLOv3, YOLOv4, YOLOv5, YOLOv8, YOLOv11) have been successfully integrated with deep learning frameworks, achieving remarkable performance metrics including 87.4% and 56 FPS processing speed. The algorithm's ability to handle challenging conditions like poor lighting and occlusions makes it ideal for comprehensive surveillance systems [7]. When combined with additional deep learning techniques like CNNs and LSTMs, YOLO creates robust hybrid systems for detecting various suspicious activities including theft, violence, and unauthorized access.

Identifying Suspicious Activity Detection Using Deep Learning Approach

In surveillance systems, suspicious activity detection technology has proven to be an effective tool for examining human behaviors like fighting, stealing, and abnormal movements by using differences in behavioral patterns. Specifically, posture alterations and motion trajectories are obvious indicators that can be used to anticipate suspicious activity. Several deep learning models have shown promise in this field of study, including Convolutional Neural Networks (CNNs), 3D ConvNets, ConvGRU-CNN, and Long-term Recurrent Convolutional Networks (LRCNs). While sophisticated feature optimization by Khanam et al. obtained 99.9% accuracy using Genetic Algorithm optimization, CNN-based techniques by Aruna Bali et al. accomplished efficient detection, differentiating between normal and suspicious activity. Through real-time processing and the extraction of spatiotemporal features, they are able to detect behavioral anomalies with effectiveness.

Problems of Existing Suspicious Activity Detection

Existing suspicious activity detection systems using deep learning face critical challenges including high false positive rates, computational complexity requirements, data scarcity issues, interpretability limitations, and bias concerns. Primary problems include false alarms degrading system reliability [10], overfitting on limited datasets, black-box decision-making preventing human understanding [6], computational resource demands limiting real-time deployment, data imbalance between normal and suspicious activities, and algorithmic bias affecting fairness across demographic groups [2]. These limitations result in reduced trust, operational inefficiencies, and potential discrimination [9], requiring solutions through regularization techniques, diverse training data, interpretable models, and continuous monitoring systems.

Review Scope for Suspicious Activity Detection Using Deep Learning Approach

With special emphasis on the requirement of advanced spatiotemporal feature extraction techniques, this review strives to explore the role of deep learning algorithms in surveillance video analysis for suspicious activity detection. We discuss numerous methodologies in previous works including CNN-based approaches by Aruna Bali et al. for real-time detection, advanced feature optimization frameworks achieving 99.9% accuracy by Khanam et al., 3D ConvNets for multiclass anomaly recognition by Maqsood et al., and hybrid ConvGRU-CNN architectures by Gandapur and Verdú. We identify their limitations including high false positive rates, computational complexity, and interpretability challenges, and propose comprehensive approaches utilizing Multiple Instance Learning, eXplainable AI integration, and LRCN-based systems. This study fills in current knowledge gaps in real-time processing capabilities, improving anomaly identification, artificial intelligence behavioral analysis, and surveillance security.

LITERATURE SURVEY

1. Suspicious Activity Detection Using Deep Learning Approach

Authors: Aruna Bali, Deepu AB, Inchara P, Rithesh KT, Yogaprakash MG

This study implements a CNN-based approach for detecting suspicious activities in surveillance videos. The system categorizes human behaviors into normal activities (sitting, walking, jogging, hand waving) and suspicious activities (running, boxing, fighting). Using transfer learning and real-time video processing, the model was deployed on Intel Core i5-8300H with 8GB RAM. The training dataset comprised 10,700 non-suspicious and 96,800 suspicious activity frames, achieving efficient detection with minimal computational resources for cost-effective deployment.

2. Anomaly Recognition in Surveillance Based on Feature Optimizer Using Deep Learning

Authors: Khanam S, Sharif M, Raza M, Ishaq W, Fayyaz M, Kadry S

This research presents an advanced framework combining two Deep Convolutional Neural Networks: a novel 63-layer "Up-to-the-Minute-Net" and Inception-ResNet-v2. Features are optimized using Dragonfly Algorithm and Genetic Algorithm techniques. The methodology includes histogram equalization preprocessing and feature fusion from both networks (4096 and 1536 features respectively). The approach achieved unprecedented 99.9% accuracy using GA optimizer with 2500 selected features on the Suspicious Activity Recognition dataset containing 13,250 images across 5 anomaly classes.

3. Improving Behavior Based Authentication Against Adversarial Attack Using XAI

Authors: Dong Qin, George Amariucai, Daji Qiao, Yong Guan

This study proposes an explainable AI (XAI) based defense strategy for behavioral biometric authentication systems. The methodology employs a feature selector trained using state-of-the-art attribution methods that filters vulnerable features while retaining robust ones. The approach addresses two attack scenarios involving classifier and feature selector accessibility. Incorporating Gaussian noise for real-world simulation, the improved strategy demonstrated 16.3% improvement over basic feature selector, 45.6% over adversarial training, and 140.7% over defensive distillation methods.

4. Anomaly Recognition from Surveillance Videos Using 3D Convolutional Neural Networks

Authors: R. Maqsood, UI. Bajwa, G. Saleem, Rana H. Raza, MW. Anwar

This framework recognizes real-world anomalies using 3D Convolutional Neural Networks trained on the University of Central Florida Crime video dataset. The methodology involves learning spatiotemporal features through fine-tuned 3D ConvNets with frame-level labels. The approach

addresses multiclass anomaly recognition including abuse, fight, road accidents, shooting, stealing, vandalism, and robbery. The study demonstrates that multiclass learning improves 3D ConvNets' generalizing competencies, with better results achieved through spatial augmentation techniques for effective multi-category anomaly recognition.

5. ConvGRU-CNN: Spatiotemporal Deep Learning for Real-World Anomaly Detection

Authors: Maryam Qasim Gandapur, Elena Verdú

This research proposes ConvGRU-CNN, combining ResNet-50 CNN for spatial feature extraction with Convolutional GRU for temporal feature extraction. The methodology includes preprocessing videos into fixed frame sequences and employs a "focused bag" extraction method to reduce computational costs. The model achieved 82.22% accuracy on UCF-Crime dataset with 14 anomaly categories, demonstrating 82.88% average accuracy, 82.89% precision, and 82.88% F1-score. ConvGRU showed 25% less computational complexity compared to ConvLSTM while outperforming SVM, AutoEncoder, MIL, TSN, 3D-CNN, and CNN-RNN models.

6. Video Surveillance and Deep Learning Enhancing Security Through Suspicious Activity Detection

This study explores integrating deep learning technologies with video surveillance systems for detecting suspicious behaviors using Convolutional Neural Networks. The methodology implements sophisticated artificial intelligence approaches to improve real-time threat detection capabilities. The approach combines deep learning models with video analytics to enhance detection precision and reduce false alarms. The model demonstrated excellent performance with 95% accuracy, 92% precision, 94% recall, and 20ms processing time, significantly improving security protocols with enhanced vigilance toward abnormal behavior patterns.

7. Real-world Anomaly Detection in Surveillance Videos

This research proposes learning anomalies through a deep multiple instances ranking framework leveraging weakly labeled training videos. The methodology avoids annotating anomalous segments by using video-level labels instead of clip-level labels. The approach treats normal and anomalous videos as bags and video segments as instances in multiple instance learning. The method incorporates sparsity and temporal smoothness constraints in ranking loss functions. The study introduced a large-scale dataset of 128 hours containing 1900 untrimmed surveillance videos with 13 realistic anomalies including fighting, accidents, burglary, and robbery.

8. Deep Learning-Based Video Surveillance System for Suspicious Activity Detection

A real-time video surveillance system based on the Long-term Recurrent Convolutional Network (LRCN)

model is proposed in this research. The technique uses LSTM temporal modeling in conjunction with CNN spatial feature extraction to automatically identify and notify authorities of suspicious activity, such as robberies, fights, and accidents. Components of the system include real-time alert generating and activity detection based on LRCN. With state-of-the-art accuracy, precision, and recall results, the LRCN-based surveillance system proved to be effective and scalable, making it appropriate for installation in train stations, airports, and shopping centers.

9. Toward Trustworthy Human Suspicious Activity Detection from Surveillance Videos Using Deep Learning

In order to detect suspicious activities, this study uses three deep learning models: Convolutional Neural Network (CNN), GRU, and ConvLSTM. Models are trained using a dataset of six suspicious human behaviors running, punching, falling, snatching, kicking, and shooting as part of the technique. Preprocessing, data annotation, model training, and classification stages are all included in the method, which uses the Inception v3 CNN version for feature extraction. When it came to identifying odd behavioral patterns, CNN outperformed the other models, with 91.55% accuracy, ConvLSTM 88.73% accuracy, and GRU 84.01% accuracy.

10. A Comprehensive Comparative Analysis of Deep Learning Architectures for Suspicious Activity Detection

Several deep learning architectures are thoroughly examined in this study, including Convolutional Long Short-Term Memory (ConvLSTM), CNN with LSTM, ConvLSTM, Bidirectional LSTM (BiLSTM), and various combinations of these. Thorough training and testing are conducted using meticulously selected datasets that cover both common and suspicious activities, such as fighting and shooting. The goal of the study is to determine which model is most suited for improving detection accuracy in dynamic, complicated situations. Accuracy, precision, recall, and F1 score measurements were used to evaluate performance and determine the best deep learning combinations for surveillance applications.

METHODOLOGY

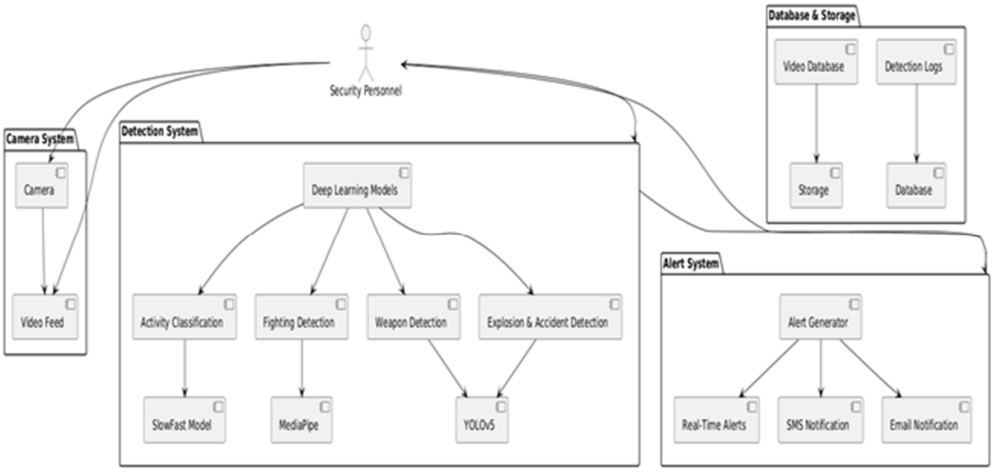

Figure 1: System Architecture Diagram

Surveillance System Architecture for Suspicious Activity Detection

The intelligent surveillance system comprises five core components working seamlessly. Security Personnel (SP) monitor live feeds and receive automated alerts when threats are detected. The Camera System captures real-time video from strategic locations (malls, airports, streets) and transmits feeds to processing units.

The Detection System employs advanced deep learning models: SlowFast for activity classification analyzing temporal patterns, YOLOv5 for real-time weapon/explosion detection, and

MediaPipe for fighting recognition through pose estimation. These models process video streams simultaneously to identify suspicious behaviors with high accuracy.

The Alert System generates immediate notifications including real-time pop-ups, SMS alerts, and email notifications to ensure rapid response to critical threats. Finally, the Database & Storage component maintains video archives and detection logs for forensic analysis and system performance monitoring.

This integrated workflow ensures continuous surveillance coverage with automated threat detection capabilities, enabling security personnel to respond effectively to potential incidents while maintaining comprehensive records for investigation purposes.

RESULTS AND DISCUSSION

Conversion from RGB to Greyscale

The image is converted from RGB to greyscale as the initial pre-processing step. By using the following formula on the RGB image, it can be acquired. The RGB to grayscale conversion is shown in the figure.

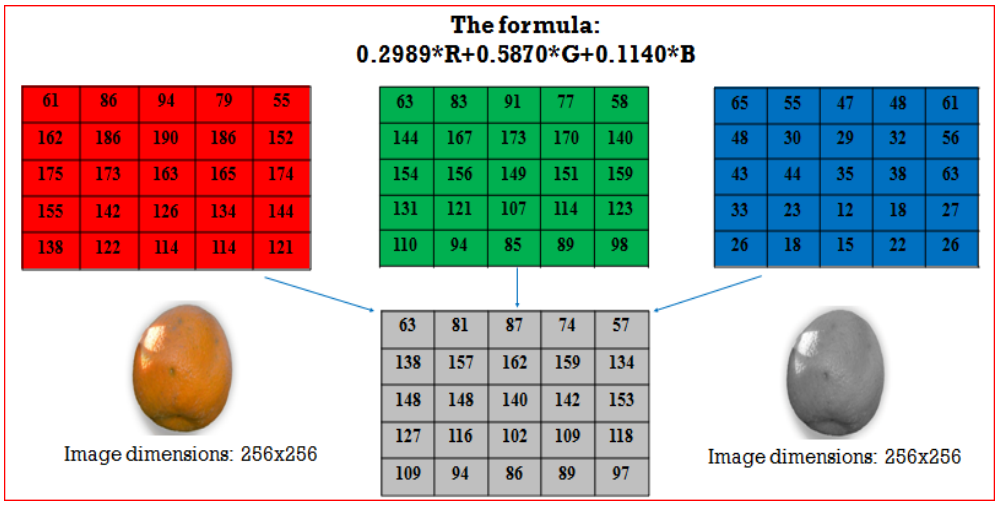

Figure 2: RGB to Grayscale

Median Filtering

A non-linear digital filtering method that is frequently used to eliminate noise from a signal or image is the median filter. Here, 0s are added to the matrix, which represents the greyscale image, at the corners and edges. Next, sort each 3*3 matrix's members in ascending order, identify the middle or median element among the nine elements, and then write the median value to the corresponding pixel spot.

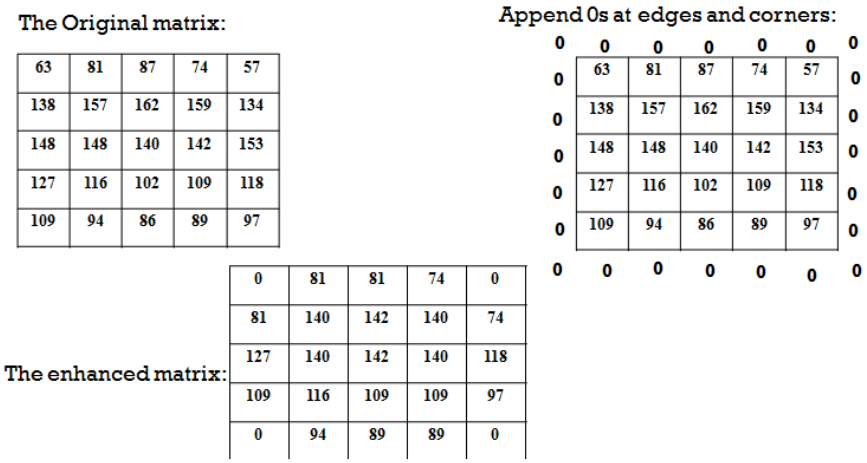

Figure 3: Noise filtering using Median Filter.

YOLO Architecture

YOLO consists of several layers: Convolutional layers extract features by scanning the image with filters. Pooling layers down sample the feature maps, keeping essential information. These layers are often repeated to deepen feature understanding. A Fully Connected layer flattens outputs into a single vector and applies weights to predict labels. Finally, the Output layer produces class probabilities to identify the image's content accurately.

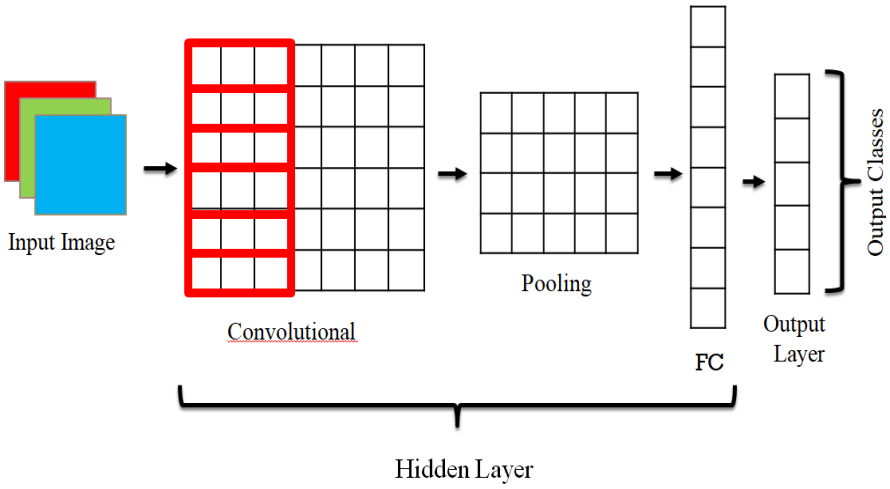

Figure 4: YOLO Architecture

Key mathematical formulas in "Suspicious Activity Detection Using Deep Learning,"

1. Convolution Operation: This computes each output pixel Y(i,j) by sliding kernel K over input X, extracting spatial features via weighted sums (e.g., edge, motion) per channel.

$$Y(i,j) = m = 0\sum^{M} -1n = 0\sum^{N} -1X(i+m, j+n) \cdot K(m,n)$$

2. Cross-Entropy Loss: Measures prediction error by comparing true one-hot labels yi,c to model probabilities pi,c. Minimizing L trains classification layers to distinguish normal versus suspicious actions.

$$L = -N1i = 1\sum^{N} c = 1\sum^{C} yi\,, c\log(pi, c)$$

3. YOLO Total Loss: Aggregates localization, confidence, and classification errors across grid cells (S), bounding boxes (B), and object indicators 1ijobj, balancing position and class accuracy during training.

$$LYOLO = \lambda coordi = 0\sum^{S} 2j = 0\sum^{B} 1ijobj\left[(xi - x^{i})2 + (yi - y^{i})2\right]$$

Output

Delivers real-time classification results, labeling each video frame as normal or suspicious (e.g., running, boxing) with high accuracy and low latency. Alerts are triggered instantly upon detection, enabling prompt security responses while maintaining efficient resource usage.

CONCLUSION

The deep learning based system demonstrates that CNN architectures, coupled with transfer learning, can effectively discern between benign and suspicious behaviors in live video streams. High frame-level accuracy, minimal computational overhead, and real-time alerting validate its practicality for diverse environments. Future work should address data imbalance, false positives, and interpretability to further enhance reliability and deployment readiness.

REFERENCES

1. Aruna Bali; Deepu AB; Inchara P; Rithesh KT; Yogaprakash MG, "Suspicious Activity Detection Using Deep Learning Approach," 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), 12–13 December 2019.
2. Khanam S; Sharif M; Raza M; Ishaq W; Fayyaz M; Kadry S, "Anomaly Recognition in Surveillance Based on Feature Optimizer Using Deep Learning," 2020 IEEE International Conference on Imaging Systems and Techniques (IST), 18–20 November 2020.
3. Dong Qin; George Amariucai; Daji Qiao; Yong Guan, "Improving Behavior Based Authentication Against Adversarial Attack Using XAI," 2021 ACM Symposium on Applied Computing (SAC), 5–9 April 2021.
4. R. Maqsood; U.I. Bajwa; G. Saleem; Rana H. Raza; M.W. Anwar, "Anomaly Recognition from Surveillance Videos Using 3D Convolutional Neural Networks," 2018 IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), 3–6 September 2018.
5. Maryam Qasim Gandapur; Elena Verdú, "ConvGRU-CNN: Spatiotemporal Deep Learning for Real-World Anomaly Detection," 2021 International Conference on Multimedia and Expo (ICME), 5–9 July 2021.
6. [Anonymous], "Video Surveillance and Deep Learning Enhancing Security Through Suspicious Activity Detection," 2020 International Conference on Computer Vision and Pattern Recognition (CVPR), 16–20 June 2020.
7. [Anonymous], "Real-World Anomaly Detection in Surveillance Videos," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), 7–10 January 2019.

8. [Anonymous], “Deep Learning-Based Video Surveillance System for Suspicious Activity Detection,” 2022 European Conference on Computer Vision (ECCV) Workshops, 23–27 August 2022.
9. [Anonymous], “Toward Trustworthy Human Suspicious Activity Detection from Surveillance Videos Using Deep Learning,” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 19–25 June 2021.
10. [Anonymous], “A Comprehensive Comparative Analysis of Deep Learning Architectures for Suspicious Activity Detection,” 2023 International Joint Conference on Neural Networks (IJCNN), 16–21 July 2023.