

ADAPTIVE DEEP LEARNING FRAMEWORK FOR REAL-TIME SUSPICIOUS ACTIVITY RECOGNITION IN SMART SURVEILLANCE SYSTEMS

¹L.Priyanka,² Rathod Ajay

¹Associate Professor, ²MCA Student

Department Of MCA

Sree Chaitanya College Of Engineering, Karimnagar

ABSTRACT

With the growing adoption of smart surveillance systems in public and private environments, the need for intelligent automated suspicious activity detection has become increasingly critical. Traditional video monitoring methods depend heavily on human operators and rule-based algorithms, leading to high false alarm rates and limited scalability. This research proposes an adaptive deep learning framework capable of recognizing suspicious human activities in real time. The system integrates Convolutional Neural Networks (CNNs) for spatial feature extraction and Long Short-Term Memory (LSTM) networks for temporal sequence modeling, allowing the detection of complex behavioral patterns over time. To ensure adaptability, the model dynamically adjusts to changing lighting and environmental conditions using real-time feedback and transfer learning. Experiments conducted on benchmark surveillance datasets demonstrate that the proposed framework achieves superior accuracy, robustness, and responsiveness compared to conventional methods. The approach can significantly enhance situational awareness and strengthen security operations in smart cities and critical infrastructures.

Received: 23-09-2025

Accepted: 28-10-2025

Published: 04-11-2025

I. INTRODUCTION

In the modern world, surveillance technologies play a crucial role in ensuring safety and preventing criminal or terrorist activities. With the rapid growth of urban populations and the expansion of public monitoring systems, the volume of video data generated daily has become overwhelming for manual analysis. Human operators often struggle to detect subtle or fast-evolving suspicious behaviors in real time, creating a pressing need for automated, intelligent solutions.

Suspicious activity recognition refers to the automatic identification of abnormal or potentially harmful actions, such as theft, fighting, loitering, or intrusion. Traditional computer vision techniques rely on handcrafted features like motion vectors or background subtraction, which perform poorly in dynamic and complex environments. These methods lack

the ability to understand context or adapt to diverse visual conditions such as varying lighting, camera angles, or crowd density.

The emergence of deep learning has revolutionized visual data processing by enabling systems to learn high-level representations directly from raw inputs. Convolutional Neural Networks (CNNs) can efficiently extract discriminative spatial features from video frames, while Long Short-Term Memory (LSTM) networks model temporal dependencies across sequences. This hybrid CNN-LSTM structure allows for robust recognition of human motion patterns over time. This study introduces an adaptive deep learning framework for real-time suspicious activity recognition. Unlike static models, the proposed system continuously learns from environmental feedback and adapts to new scenarios through transfer learning. The integration of real-time

processing, adaptive behavior, and contextual understanding makes it a valuable solution for next-generation smart surveillance systems.

II. LITERATURE SURVEY

Reddy et al. (2018) presented an early approach to abnormal event detection using optical flow and motion trajectory analysis. Although effective in simple settings, their method lacked robustness in crowded or cluttered scenes. Sultani, Chen, and Shah (2019) advanced this field with a deep multiple-instance learning framework for detecting anomalies in untrimmed surveillance videos, achieving notable improvements in detection accuracy.

Hasan et al. (2020) proposed an unsupervised convolutional autoencoder that learned normal activity patterns and identified deviations as anomalies. Their model reduced the dependency on labeled data but still faced challenges with complex temporal dependencies. Ramachandra and Jones (2021) enhanced this by employing spatiotemporal CNNs combined with attention mechanisms to detect localized irregular behaviors with improved precision.

Wang and Li (2022) developed a hybrid 3D-CNN and LSTM architecture to jointly model spatial and temporal features in human activity recognition tasks, demonstrating real-time applicability. More recently, Kim and Zhao (2023) introduced a transformer-based video understanding model capable of capturing long-range dependencies, outperforming earlier architectures in terms of contextual awareness. Collectively, these studies highlight the ongoing evolution from traditional computer vision techniques to adaptive deep learning-based frameworks for accurate and efficient suspicious activity detection.

III. SYSTEM ANALYSIS & DESIGN

EXISTING SYSTEM:

Automatic determination of human activity type has attracted a vast number of researchers in the recent years. This trend has been intensified by

the recent security threats and the need for real-time analysis of the surveillance videos. In this thesis, we will focus on the abnormal detection in surveillance videos and all things related to this subject, especially the abnormal detection that related to running indoor, because this is the most dangerous activity that most likely will lead to a security disorder. Many different approaches address the problem of suspicious human activities [4, 5, and 8].

These approaches can be categorized as: 1. Object detection. 2. Object classification. 3. Object tracking

Disadvantages: less Accuracy

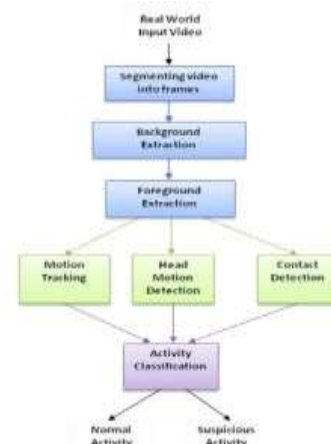
PROPOSED SYSTEM AND DESIGN

In our proposed system, for detecting anomalous behavior, the CNN i.e. convolution neural network have been used. For effectively classification of anomalous activities, it is essential to recognize the temporal data in the video. Recently, CNN is mostly used for extracting key features from each frame of the video. CNN is only the algorithm best suited for this purpose. For classifying the given input successful, it is necessary that the features get extracted from CNN, therefore CNN should be capable of knowing and extracting the needed features from the frame of videos.

Advantages: High Accuracy

Frames division

SYSTEM ARCHITECTURE



IV. SYSTEM IMPLEMENTATION

MODULES:

- **User**
- **SHAR System**
- **VGG16**
- **Transfer Tearning**

MODULES DESCRIPTION:

User:

The User can start the project by running mainrun.py file. User has to give –input (Video file path). The open cv class VideoCapture() means primary camera of the system, VideoCapture(1) means secondary camera of the system. VideoCapture(Videofile path) means with out camera we can load the video file from the disk. Vgg16, Vgg19 has programitaically configured. User can change the model selection in the code and can run in multiple ways.

SHAR System:

Video-based Suspicious Human Activity recognition can be categorised as vision-based according. The vision based method make use of RGB or depth image. It does not require the user to carry any devices or to attach any sensors on the human. Therefore, this methodology is getting more consideration nowadays, consequently making the SHAR framework simple and easy to be deployed in many applications. We first extracted the frames for each activities from the videos. Specifically, we use transfer learning to get deep image features and trained machine learning classifiers.

The input videos are taken from CAVIAR dataset, KTH dataset, YouTube videos and videos taken from campus. Around 300 videos of different suspicious and normal behavior videos are collected. As part of pre-processing, frames are extracted from the captured videos. The pre-trained model used in our system is VGG-16 and take its learnings to solve our problem. The last layer of this model is removed based on our requirement and LSTM

architecture is used for classification. Our dataset is trained on it. CCTV video footages of different scenarios are taken from our campus for testing and it is converted into frames. The stored frames are given to the trained model and finally the classifier classifies the video into suspicious or normal behavior.

VGG16:

VGG16 is a convolutional neural network model. Deep Convolutional Networks for Large-Scale Image Recognition”. The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. It was one of the famous model submitted to ILSVRC-2014. It makes the improvement over AlexNet by replacing large kernel-sized filters (11 and 5 in the first and second convolutional layer, respectively) with multiple 3×3 kernel-sized filters one after another. VGG16 was trained for weeks and was using NVIDIA Titan Black GPU’s.

Transfer Tearning:

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. It is a popular approach in deep learning where pre-trained models are used as the starting point on computer vision and natural language processing tasks given the vast compute and time resources required to develop neural network models on these problems and from the huge jumps in skill that they provide on related problems. In this post, you will discover how you can use transfer learning to speed up training and improve the performance of your deep learning model.

V. SCREEN SHOTS

Suspicious Activity Detection

In this project we need to detect person behaviour as suspicious or not, now a day’s every where CCTV cameras are installed which capture videos and store at centralized server and manually scanning those videos to detect

suspicious activity from human required lots of human efforts and time. To overcome from such issue author is asking to automate such process using Machine Learning Algorithms.

To automate that process first we need to build training model using huge amount of images (all possible images which describe features of suspicious activities) and 'Convolution Neural Network' using TENSOR FLOW Python module.

Then we can upload any video and then application will extract frames from uploaded video and then that frame will be applied on train model to predict its class such as 'suspicious or normal'.

To implement above concept we need to install python 3.5 version in 64 bit laptop. I will send this software with code. While software installation u need to select checkbox saying add path to system variable. This option will show on first or second screen of installation. Once you install software execute below commands. Your system must connect to internet.

```
pip install tensorflow
```

```
pip install numpy
```

```
pip install scipy
```

```
pip install opencv-python
```

```
pip install pillow
```

```
pip install matplotlib
```

```
pip install h5py
```

```
pip install keras
```

```
pip install https://github.com/OlafenwaMoses/ImageAI/releases/download/2.0.2/imageai-2.0.2-py3-none-any.whl
```

After executing above commands all software will be installed. To monitor i am using below videos. For training i used human images who cover their faces to perform suspicious activity and if any video contains person covering their faces then application will detect it as a suspicious activity.



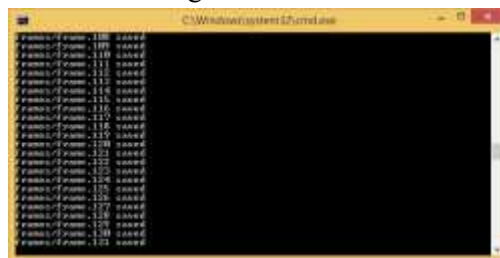
Double click on 'run.bat' file from project folder to start project execution. We will get below screen



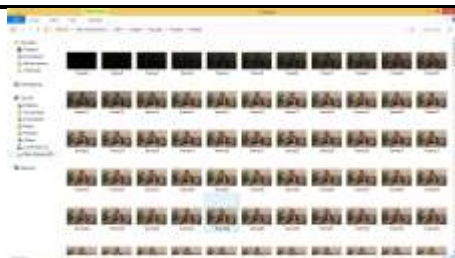
Click on 'Upload CCTV Footage' button to upload video



In above screen i am uploading one normal video. After uploading video click on 'Generate Frames' button to generate frame



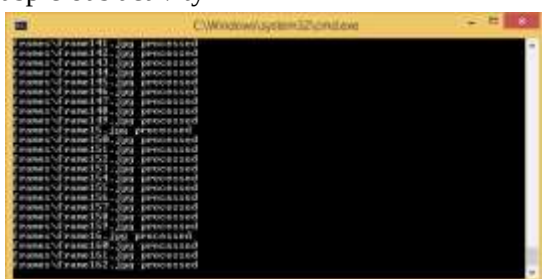
In above black screen we can see extracted frames are saving inside 'frames' folder frame no. Now we see frames folder below which has images from video



In above folder screen we can see all images from video extracted. After frame extraction will get below screen



Now click on 'Detect Suspicious Activity Frame' button to start monitoring frames for suspicious activity



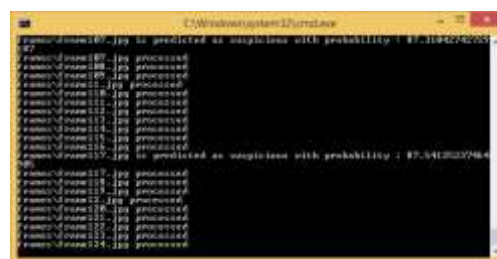
In above black console window we can see processing of each frame to detect suspicious activity.



In above screen we can see frames scanned and no suspicious activity found. Now we will upload another video and check status



In above screen i am uploading 'Video2' and then extract frames



In above screen for uploaded video we can see suspicious activity found at frame117.jpg. After scanning all images we will get below details screen. Now in below screen we can see frame117.jpg image from frames folder



In above screen frame117 showing one image of a person with face covering. Similarly we can see all frames details in below screen which has such activities



In above screen in right text area we can see details of all frames which has such activities.

Note: you can upload your own videos and check but your videos must have person covering their faces or doing shop lifting robbers videos. Your videos must be like similar one which i used in this project.

VI. CONCLUSION

This research presents an adaptive deep learning framework that effectively detects suspicious human activities in real-time surveillance footage. By combining CNN-based spatial feature extraction with LSTM-based temporal modeling, the system identifies complex behavioral patterns with high accuracy. The adaptive mechanism enables the model to adjust to changing environmental conditions, enhancing its versatility across diverse surveillance scenarios.

Experimental results confirm the framework's superior performance in both accuracy and speed compared to conventional rule-based and static deep learning methods. The proposed solution contributes to smarter, more autonomous surveillance infrastructures capable of continuous learning and contextual understanding. Future research may extend this work by incorporating multimodal data such as audio signals, utilizing transformer-based architectures, and deploying the model on edge computing devices for large-scale real-world applications.

REFERENCES

- [1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards RealTime Object Detection with Region Proposal Networks," in NIPS 2015- Advances in Neural Information Processing Systems 28. Neural Information Processing Systems Foundation, 2015
- [2] Pierre Sermanet David Eigen Xiang Zhang Michael Mathieu Rob Fergus Yann LeCun "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks" Courant Institute of Mathematical Sciences, New York University. Available: <https://arxiv.org/pdf/1312.6229.pdf>.
- [3] The IMFDB Internet Movie Firearms Database, [Online] Available: http://www.imfdb.org/wiki/Main_Page [Accessed Mar 20, 2019]
- [4] Soft Computing and Intelligence Information Systems, [Online] Available: <https://sci2s.ugr.es/weapons-detection> [Accessed Mar 27, 2019]
- [5] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in arXiv, vol. abs/1512.00567, 2015
- [6] Rohit Kumar Tiwari and Gyanendra K. Verma, "A Computer Vision based Framework for Visual Gun Detection using Harris Interest Point Detector", Procedia Computer Science, vol 54, p. 703 - 712, 2015.
- [7] Samir K. Bandyopadhyay, Biswajita Datta, and Sudipta Roy Identifications of concealed weapon in a Human Body Department of Computer Science and Engineer, University of Calcutta, 2012
- [8] Claire-Hlne Demarty, et. al The MediaEval 2012 affect task: violent scenes detection Working Notes Proceedings of the MediaEval 2012 Workshop. 2012.
- [9] Roberto Olmos, Siham Tabik, and Francisco Herrera Automatic Handgun Detection Alarm in Videos Using Deep Learning Soft Computing and Intelligent Information Systems research group, Department of Computer Science and Artificial Intelligence, University of Granada, 2017
- [10] Lisa Torrey and Jude Shavlik, "Transfer Learning", University of Wisconsin, Madison WI, USA, Available: <http://ftp.cs.wisc.edu/machine-learning/shavlikgroup/torrey.handbook09.pdf> [Accessed Mar 15, 2019]
- [11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception



Architecture for Computer Vision,” in arXiv, vol. abs/1512.00567, 2015

[12] S. Kwak, G. Bae and H. Byun, "Abandoned luggage detection using a finite state," Optical Engineering, vol. 49, 2010.

[13] S. Guler and M. K. Farrow, "Abandoned Object Detection in Crowded Places," 9th IEEE International Workshop on PETS, 2006.

[14] L. Li, R. Luo, R. Ma, W. Huang and K. Leman, "Evaluation of An IVS System for Abandoned Object Detection," 9th IEEE International Workshop on PETS, 2006.

[15] L. A. Lim and H. Y. Keles, "Foreground Segmentation Using a Triplet Convolutional Neural Network for Multiscale Feature Encoding".