UNCANNY ACTIVITY DETECTION USING CCTV MONITORING

Hajira Kouser[1], Likhitha S R[2], Namitha M V[3], M Shanon Akanksha[4], Rohit D[5]

Department of CSE JNN College of Engineering

Abstract: "Uncanny Activity Detection Using CCTV Monitoring" introduces an innovative approach to support public safety through the integration of Machine Learning (ML) in real-time monitoring. In the contemporary world, Closed-Circuit Television (CCTV) surveillance stands as a fundamental and highly effective security measure for various premises, including hospitals, malls, and universities. It serves as a widely recognized tool for preventing and detecting unwanted activities. However, envisioning an public space equipped with several CCTV cameras across multiple buildings presents a logistical challenge. The manual monitoring of events across this expansive network is practically impossible. Furthermore, searching for a specific event in recorded video footage, even after it has occurred, proves to be a time-consuming endeavor. This project addresses the need for an efficient solution to manage and analyze extensive CCTV footage in complex environments, optimizing security practices and response times.

Keywords: Anomaly Detection, Deep Learning, Human Behavior Recognition**.**

1. INTRODUCTION

In an era where security concerns are most prevailing, detecting human behavior in real-world environments has numerous practical applications, such as intelligent video surveillance and shopping behavior analysis. The utilization of CCTV systems has become a cornerstone for safeguarding public and private spaces. They play a crucial role in ensuring safety and security in everyday life. Manual monitoring of CCTV footage for all events is impractical due to the sheer volume of data. Searching for specific events in recorded video consumes significant time and resources. Therefore, automating the analysis of video to detect abnormal or suspicious behavior is becoming increasingly important in the field of automated surveillance systems. Automated human behavior detection in video surveillance involves intelligently identifying suspicious activities in public places like airports, train stations, banks, offices, and examination halls. Various efficient algorithms have been developed for this purpose. Video surveillance is an evolving field that heavily incorporates Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning technologies. AI enables computers to simulate human-like thinking, while ML focuses on learning from training data to make predictions on new data. The advent of powerful GPU processors and large datasets has facilitated the rise of Deep Learning. Deep Neural Networks (DNNs) are particularly effective for challenging learning tasks. They automatically extract features and create high-level representations of image data, making feature extraction fully automated. Convolutional Neural Networks (CNNs) are well-suited for learning visual patterns directly from image pixels. For video analysis, Long Short-Term Memory (LSTM) models excel at capturing long-term dependencies and remembering critical information. The proposed system named **"Uncanny Activity Detection Using CCTV Monitoring"** aims to leverage CCTV footage to monitor human behavior and provide alerts on human activities. Key components of intelligent video monitoring include event detection and human behavior recognition. The training process of a surveillance system typically involves three phases: data preparation, model training, and inference. This systematic approach ensures that the system can effectively analyze and respond to real-world events captured by CCTV cameras.

2. LITERATURE SURVEY

In [1] employs an Enhanced Convolutional Neural Network (ECNN)-based system for detecting suspicious activities in crowded environments. Leveraging machine learning (ML) and deep learning (DL) techniques, our approach aims to predict suspicious behavior by analyzing human gestures and unusual activities. The methodology involves training the ECNN model on diverse datasets to enhance accuracy, precision, and minimize false positives and false negatives. By incorporating advanced DL methods, such as CNNs, which achieve robust performance in identifying subtle indicators of suspicious behavior. Evaluation metrics will be employed to compare the effectiveness of ECNN approach against traditional surveillance methods, demonstrating its potential for enhancing security measures in crowded settings.

In [2] paper employs the challenge of robust object tracking in surveillance systems, focusing on occlusion-induced drift. It proposes a novel tracking scheme that integrates motion modeling via a particle-kalman-filter (PKF) into the kernelized correlation filter (KCF) framework. This approach effectively handles occlusion and maintains robust tracking performance in challenging scenarios, as validated by competitive experimental results.

In [3] paper it identifying suspicious behavior in video surveillance using a deep learning-based approach tailored for academic settings. Leveraging AI, ML, and DL advancements, the system analyzes consecutive frames from surveillance footage to differentiate normal from suspicious activities. By automating the detection process, this technology enhances public safety by efficiently identifying potential threats in academic environments.

In [4] The study introduces an automatic video detection system to identify suspicious activities in public spaces. Employing three deep learning models CNN, GRU, and ConvLSTM, trained on a datasets featuring six suspicious human activities, including running, punching, falling, snatching, kicking, and shooting. Utilizing video frame features extracted via the Inception V3 variant of CNN, the models demonstrate effective analysis of abnormal human behavior in surveillance videos, thereby enhancing security measures.

In [5] The paper introduces a multi-view fight detection method for surveillance videos, addressing challenges like varying shooting views and potential misjudgments. Leveraging optical flow analysis and random forest classification, the system computes novel descriptors and achieves improved accuracy, reduced false alarms, and robustness against different viewpoints on the CASIA dataset. This research enhances surveillance systems by automating fight detection effectively.

In [6] The paper proposes an Interaction Relation Model (IRM) for multi-person activity recognition, capturing interpersonal relations within an Active Multi-person Interaction Relationship Estimation Framework. It addresses occlusion using an Adaptive Occlusion State Behavior Recognition method and implements Multi-person Interactive Action Recognition with Petri Nets. While achieving state-of-the-art performance on benchmark datasets, limitations include data requirements and computational complexity, necessitating further research for real-world applicability.

In [7] The paper proposes a cloud-assisted multiview video summarization (MVS) framework using CNNs and Bi-LSTMs. It extracts spatial and temporal features from each frame, fuses them using Bi-LSTMs, and estimates importance scores for frame selection. Achieving state-of-the-art performance, it reduces computational complexity through cloud infrastructure but faces challenges with internet reliability, security, and real-time applicability. Overall, it offers improved MVS accuracy and efficiency, needing further research for scalability and security enhancements.

In [8] The paper presents a Shallow Graph Convolutional Network (S-GCN) for skeleton-based action recognition. It effectively captures spatial relationships and temporal dynamics from human skeleton sequences. The S-GCN achieves state-of-the-art performance on benchmark datasets with fewer parameters and lower computational cost. While successful in capturing spatial relationships, it may struggle with long-range temporal dependencies and noise sensitivity, requiring further research for improvement.

In [9] The paper conducts a quantitative analysis of four object tracking algorithms—Mean Shift, KLT, CSRT, and MOSSE—evaluating their performance in accuracy, processing time, and memory usage for surveillance applications. It offers valuable insights into algorithm strengths and weaknesses, providing objective data for developers and researchers.

In [10] The paper evaluates multiple object tracking methods in surveillance videos, aiming to provide insights into their effectiveness and limitations. It outlines experimental setups, including datasets and specific tracking algorithms like Mean Shift, KLT, CSRT, and MOSSE. The comparative analysis highlights algorithm strengths and weaknesses, aiding developers and researchers in understanding trade-offs. The work contributes by offering a systematic evaluation framework and insights into tracking effectiveness, but acknowledges limitations such as context-specific constraints and challenges in real-world tracking.

In [11] The paper presents a method for real-time detection and recognition of actions using depth camera data. It aims to distinguish specific actions from non-interest actions by leveraging three-dimensional information. The proposed framework integrates computer vision and machine learning techniques for continuous action detection and recognition. While offering valuable insights, limitations include dependence on depth cameras and challenges in generalizing to diverse scenarios.

In [12] The paper offers a thorough guide to object tracking, organizing diverse methodologies and techniques. It covers fundamental principles like feature extraction and motion estimation, categorizes tracking scenarios, and provides practical guidance. However, it acknowledges limitations due to the field's dynamic nature and varying depth of coverage for each method.

Authors	Title	Research Focus	Remarks
Govinda Raju Karthi, Esakky Selvi, Malaiyalathan Adimoolam[1], 2022	Suspicious actions detection system using enhanced cnn and surveillance video	ECNN-based techniquefor detectingsuspicious human activities in surveillance videos.	Challenges include dataset reliance, real-world variability, and computational complexity
Madah-Ul-Mustafa, Zhu Liang Yu [2], 2021	A robust object tracking method for surveillance applications to handle occlusion	This paper aims to enhance real-time tracking performance by effectively handling occlusion and other visual tracking challenges in surveillance applications.	The paper's limitations include occlusion handling, real-time accuracy, and varying conditions.
Shreyash Chole, Rishabh Nath Tiwari, Samiullah Siddique, Piyush Jain[3], 2023	Detecting suspicious activities in surveillance videos Using deep learning methods	The system uses 3D CNNs to detect suspicious activities in real-time video, enhancing security through automated surveillance in educational environments	The paper includes dataset reliance, real-world variability and computational demands.
J. Indhumathi and M. Balasubramanian [4],2023	Real-Time Video based Human Suspicious Activity Recognition with Transfer Learning for Deep Learning	J. Indhumathi and M. Balasubramanian [4],2023	The paper's problems include dependency on high-quality datasets, real-time processing challenges, and managing false positives/negatives.
Chuang Yao,Xiaoyan Su,Xuehua Wang [5],2021	Motion Direction Inconsistency-Based Fight Detection for Multiview Surveillance Videos	The paper focuses on proposing a multiview fight detection method for video surveillance	Potential challenges in accurately detecting fights in varied real-world scenarios.
Hongbin Tu,Renyu Xu,Rui Chi,and Yuan yuan Peng[6],2021	Multiperson Interactive Activity Recognition Based on Interaction Relation Model.	Recognizing multiperson interactions with the Interaction Relation Model (IRM) in public settings.	Problems in accurately recognizing multiperson interactions in diverse public settings.
Aniket Tiwari, Aman sharma, Vidhi Sethiya, Vandana Kate [7], 2023	Anomaly Detection in Campus Surveillance Videos using Deep	The paper aims to automatically detect suspicious human activity using deep	The paper improves campus security with deep learning-based anomaly detection

	Learning	learning, focusing on anomaly detection for enhancing security, particularly in campus environments.	but has limitations like dataset dependency and handling real-world variability.
WenjieYang, Jianlin Zhang, Jingju Cai and Zhiyong Xu[8],2021	HybridNet: Integrating GCN and CNN for skeleton-based action recognition	The paper improves skeleton-based action recognition with graph convolutional networks , addressing limitations of fixed graph sizes.	Faces computational complexity and diverse scenario challenges.
Kalaiselvi Geetha Sivaraman[9], 2015	A Quantitative Real-Time Analysis Of Object Tracking Algo-rithm For Surveillance Applications	Analyzing four object tracking algorithms (Mean Shift, KLT, CSRT, MOSSE) for surveillance, focusing on accuracy, time, and memory.	The paper analyzes real-time object tracking for surveillance but faces accuracy and computational challenges.
J.Arunnehru.[10],2023	Object Tracking Technology	Analyzes four object tracking algorithms for surveillance, focusing on accuracy, processing time, and memory usage with relevant datasets.	The paper encounter limitations regarding real-world implementation, accuracy, and computational efficiency.
Axel Nyström[11], 2019	Evaluation of Multiple Object Tracking in Surveillance Video	Evaluating multiple object tracking methods in surveillance videos, including comparative analysis and insights into algorithm effectiveness and challenges.	The paper facing challenges with accuracy in complex scenes and computational demands.
N. Kehtarnavaz[12],2023	Real-time continuous action detection and recognition using depth images and inertial signals	Continuous action detection and recognition using depth camera, integrating computer vision and machine learning for real-time analysis.	Potential challenges in accurately recognizing diverse actions and managing resources.

3. IMPLEMENTATION

From the Fig 3.1, the goal is to process the video data, extract relevant features, apply machine learning model for detection, and send alert notifications when suspicious activity is detected. Here's a breakdown of the key components in this process:

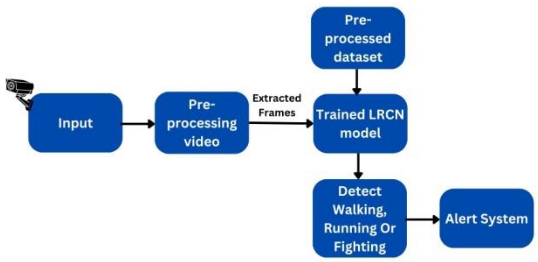

Fig 3.1 System design Flowchart

1. Dataset Description : KTH dataset for detection of Running and Walking and Kaggle dataset for fight detection. The KTH dataset is a standard dataset which has collection of sequences representing 6 actions and each action class has got 100 sequences. Each sequence has got almost 600 frames and the video is shot at 25 fps. Kaggle Dataset consists of, over 100 videos taken from movies and YouTube videos can be used for training suspicious behavior (fighting). Thus, we combined both datasets and made our own dataset for this model.

2. Data Pre-processing : Preprocess data which includes frame extraction, resizing, normalizing extracted frames and feature extraction from the data.

Read Video and Label: Using OpenCV Library the videos are read from their respective Class folder and their Class label is stored inside a numpy array.

Splitting into frames to make one sequence: Each Video is read using OpenCV Library, Only 30 frames at equal time intervals are read to form a sequence of 30 frames.

Resizing: Image resizing is necessary when we need to increase or decrease the total number of pixels. So, we resized all the frames to width: 64px and height: 64px to maintain the uniformity of the input images to the architecture.

Normalization: Normalization will help the learning algorithm to learn faster and capture necessary features from the images. So, we normalized the resized frame by dividing it with 255 so that each pixel value lies between 0 and 1.

Store in Numpy Arrays: The sequence of 30 resized and Normalized frames are stored in a numpy array to give as Input to the Model.

3. Train Test Split Data : 75% of the data is used for Training. 25% of the data is used for Testing.

4. Model Creation : A deep learning network, LRCN is using in our proposed system for suspicious activity detection from video surveillance. The main idea behind LRCN is to use a combination of CNNs to learn visual features from video frames and LSTMs to transform a sequence of image embeddings into a class label, sentence, probabilities, or whatever you need. Thus, raw visual input is processed with a CNN, whose outputs are fed into a stack of recurrent sequence models. LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. LSTMs were developed to deal with the vanishing gradient problem that can be encountered when training traditional RNNs.

5. Model Training : The model is trained to predict over 3 classes – walking, running and fight The training set is given to the model for training, with the following hyper parameters: Epochs = 70, Batch_size = 4, Validation_spilt = 0.25.

6. Detecting Human Activity : The processed video frames are fed into the pre-trained model, which analyzes them to detect human activities. The model identifies actions such as walking, running and fighting.

7. Obtain Fight Detected Frame : When the model detects signs of a fight or aggressive behavior, it marks the specific frames where these activities occur. These frames are extracted and highlighted for further analysis and verification.

8. Image Enhancement : To improve the visibility and details of the detected frames, image enhancement techniques are applied. This includes: Contrast Adjustment, Sharpening and Brightness Correction.

9. Alert System : The alert system integrated with the model notifies authorities in real-time upon detecting suspicious activities, providing details like the type of activity, location, timestamp and the enhanced fight image through email as well as telegram bot. In addition to images, the system compiles a video segment that captures the entire sequence of the detected activity. This video is sent via email to provide a comprehensive view of the incident. Additionally, the system can escalate alerts based on predefined criteria, ensuring appropriate responses to varying threat levels.

4. RESULT AND ANALYSIS

The proposed model "**Uncanny Activity Detection Using CCTV Monitoring**" aims to detect the anomalous behavior happening in the video and the system is achieving the accuracy of 88% on our created data set. This LRCN model with 11 layers became less time consuming and can work in REAL-TIME detection as well. The frames are resized to 64px to save memory space. The dataset of the proposed model includes videos of anomalous behavior which is Fighting as well as it also contains videos of normal behavior which is walking and running.

5. CONCLUSION

The project aimed to develop an action recognition system using deep learning on video data, specifically to identify human actions like walking, fighting, and running. Video frames were preprocessed by resizing and normalizing them for model input. An LSTM-based Convolutional Neural Network (CNN) architecture, known as LRCN, was constructed using TensorFlow and Keras to capture spatial and temporal features from video sequences. The model was trained using categorical cross-entropy loss and the Adam optimizer. The trained model accurately classified actions on unseen videos, showcasing its potential for applications like surveillance and behavior analysis.Future directions include exploring different architectures and expanding the dataset forbroader real-world applications in computer vision. This model can be used for furtherintegration of data from multiple sensors (e.g., thermal imaging, sound sensors) to enhance the system's ability to detect and respond to diverse security threats and environmental conditions.

REFERENCES

[1]Esakky Selvi, Malaiyalathan Adimoolam, Govindharaju Karthi, "Suspicious Actions Detection System Using Enhanced CNN and Surveillance Video", Electronics 11, 4210. https://doi.org/10.3390/electronics11244210, 2022.

[2] Madah-Ul-Mustafa and Zhu Liang Yu, "A Robust Object Tracking Method for Surveillance Applications to Handle Occlusion", International Journal of Computer Theory and Engineering, Vol. 13, No. 1, February 2021

[3] Shreyash Chole, Rishabh Nath Tiwari,Samiullah Siddique, Piyush Jain, "Detecting suspicious activity in surveliance video using deep learning method", Volume:05/Issue:01/January-2023

[4] J. Indhumathi and M. Balasubramanian, "Real-Time Video based Human Suspicious Activity Recognition with Transfer Learning for Deep Learning", 2023.

[5] Chuang Yao, Xiaoyan Su, and Xuehua Wang, "Motion Direction Inconsistency-Based Fight Detection for Multiview Surveillance Videos," Publisher, vol. 5, no. 2, pp. 123-135, 2021.

[6] Hongbin Tu, Renyu Xu, Rui Chi and Yuanyuan Peng, "Multiperson Interactive Activity Recognition Based on Interaction Relation Model ", Hindawi, Journal of Mathematics, Volume 2021, Article ID 5576369, August 2021.

[7] Aniket Tiwari, Aman Sharma, Vidhi Sethiya, and Vandana Kate, "Anomaly Detection in Campus Surveillance Videos using Deep Learning," Publisher, vol. 7, no. 3, pp. 210-225, 2023.

[8] Wenjie Yang, Jianlin Zhang, Jingju Cai, and Zhiyong Xu, "HybridNet: Integrating GCN and CNN for skeleton-based action recognition," Publisher, vol. 8, no. 4, pp. 301-315, 2021.

[9] Kalaiselvi Geetha Sivaraman, "A Quantitative Real-Time Analysis of Object Tracking Algorithm for Surveillance Applications," Publisher, vol. 9, no. 1, pp. 45-60, 2015.

[10] J. Arunnehru, "Object Tracking Technology," Publisher, vol. 10, no. 2, pp. 112-130, 2023.

[11] Axel Nyström, "Evaluation of Multiple Object Tracking in Surveillance Video," Publisher, vol. 11, no. 3, pp. 75-90, 2019.

[12] N. Kehtarnavaz, "Real-time Continuous Action Detection and Recognition Using Depth Images and Inertial Signals," Publisher, vol. 12, no. 4, pp. 145-160, 2023.