Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

Automated Unusual Event Detection

Mr. Praveen A Patil
Assistant Professor
Department of Electronics and Communications
ACS College of Engineering, Bangalore, India
praveenapatil143@gmail.com

Kusuma M N
UG Scholar
Department of Electronics and Communications
ACS College of Engineering, Bangalore, India
Kusumadeepu1006@gmail.com

Gouthami R
UG Scholar
Department of Electronics and Communications
ACS College of Engineering, Bangalore, India
gouthamigouthu0205@gmail.com

Fahad Sheriff
UG Scholar
Department of Electronics and Communications
ACS College of Engineering, Bangalore, India
fahadsheriff0923@gmail.com

Rahul S
UG Scholar
Department of Electronics and Communications
ACS College of Engineering, Bangalore, India
manojmanu282000@gmail.com

***Abstract*—The past few years brought big upgrades in monitoring tech thanks to progress in deep learning. Old-style camera setups mostly depend on people watching screens - a slow process prone to mistakes, especially where things get chaotic or packed. Because of these issues, this project introduces a smart setup powered by deep learning that spots odd actions live in footage. It's built to detect specific risky events like shoving crowds, large gatherings, fights, fires, crashes, thefts, and property damage. By using CNNs along with RNNs, it pulls details from both space and time across video frames. This mix - entered on an RNN structure - boosts how well it tells different behaviours apart. The proposed setup learns from lots of different examples, including everyday situations along with a fake crisis scenario make it strong and adaptable across different cases. To handle messy real-world settings, smarter data cleanup was used. Techniques combined with pre-trained models boost how fast threats get spotted - like in stores, roads, or packed areas. Results show it catches unusual actions very well, triggering quick alerts so police or rescue teams act faster when needed. A deep learning setup improves public safety, response times, plus supports city cameras working like a smart network.**

Keywords --Deep learning, abnormal activity detection, CNN, RNN, Surveillance, Fire detection, Accident detection, Shoplifting, Vandalism, Crowd analysis.

I. INTRODUCTION

As cities pack more folks into tight spaces, spotting odd actions fast becomes key to stopping trouble before it kicks off. With tech on the rise, how we watch for danger has shifted - so's how easily our freedom gets limited without us noticing. These setups catch unusual moves in camera feeds or signals, giving early heads-up when something sketchy might go down. Think of them like alarms built around weird patterns instead of sounds, helping flag crimes, crises, or risky moments real quick.

Regular security setups rely heavily on people watching more and more camera screens at once - this leads to tiredness and missed details. As HD cameras became common, the sheer volume of footage made it harder for humans to keep up effectively. That's where smart tech steps in, especially a type called deep learning. This approach is part of broader machine learning, using layered networks to spot patterns in massive amounts of info without constant guidance. Given sufficient labelled data, these systems are able to recognize minor irregularities or complicated scenarios more effectively than traditional methods.

Deep Learning has proven extremely effective for image and video analysis; therefore, it is most frequently selected when it comes to anomaly detection from surveillance footage. Instead of older techniques that rely on fixed rules or basic setups, these modern models pick up intricate details through layered learning steps. Because they figure out patterns on their own, there's no need to design features by hand - this gives detection tools more adaptability across different environments. Some odd actions catching most attention from public monitoring involve people shoving in groups, big gatherings piling up quickly, fights breaking out, sudden fires, crashes on roads or open areas, stealing small stuff from stores, also deliberate property damage. Each of these situations brings its own issues when trying to spot them or react properly. For instance, shoving and brawls often show rapid aggressive movement with unclear outlines making analysis tough; detecting flames or spotting wrecks might need combined sensor types plus speedy alerts to stop things getting worse. Likewise minor thefts like swiping items or trashing objects come with financial loss yet display subtle behaviour patterns - often missed by basic motion tracking systems.

II. LETURATURE REVIEW

The paper called "YOLOv5s-MEE" by Ping Yuan and others from 2024 is about finding behaviors in control rooms. They look at two things: making the pictures better and making the model work faster. First they use something called SRGAN to make the bad pictures from the CCTV cameras look better before they start training the model. Then they change the default YOLOv5 backbone to MnasNet to make the model

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

work better. They also add something called ECA-Net to help the model find the things in the pictures. The people who wrote the paper also change the default IoU loss function to EIoU. The YOLOv5s-MEE model can run fast at 75 FPS.Yolov5s-MEE needs good pictures to work well and it can only be used in control rooms. The YOLOv5s-MEE model is very good, at what it does. It has some limits.

The model that detects behaviour by Li Ying and others from 2024 is made for keeping an eye on safety in industries. This model looks for things like people smoking and not wearing the clothes for work. It uses something called YOLOv5. Makes it better by adding a thing called Triplet Attention module. This helps the model to understand what is happening in places. The model is better at finding problems in industrial areas.. It has a big limitation. The model can only find a few unsafe behaviours that it has been taught about. It might not be good at finding types of strange actions or working in different places. The unsafe behaviour detection model by Li Ying and others is still limited, in what it can do. The unsafe behaviour detection model has to be taught about each type of behaviour.

The driver abnormal expression detection method that Keming Yao and his team came up with in 2024 is used to keep an eye on how diverse doing like if they are tired or not paying attention by looking at their faces. They made their model better by using FasterNet-based modules and some other things like GSConv and VoV-GSCSP to help it work faster and smarter. They also got rid of some parts of the model that were not necessary which made it a lot smaller so it can work on devices that do not have a lot of power. This model is really small it is 4.6 MB and it still works pretty well it gets it right about 84.5 percent of the time.. The driver abnormal expression detection method only looks at faces and does not think about what the rest of the body is doing so it is not very good, at figuring out what people are doing in general. The driver abnormal expression detection method has some limits because it only looks at faces and does not look at the person.

The staff off-duty detection algorithm proposed by Wenrui Yan et al. (2024) seeks to determine whether there are any employees in their assigned posts, especially in control room settings. The framework optimizes the YOLOv5 framework by increasing feature extraction through attention techniques borrowed from Botnet and utilizing data enhancement approaches such as HSV changes and mosaic augmentation. These modifications result in slightly higher accuracy and F1-scores. Nonetheless, the technique only considers presence or absence detection and cannot evaluate complicated actions or consider temporal data like the duration of absence.

III. PROPOSED SYSTEM

The suggested system combines the progress in the field of deep learning, object detection, and action recognition to provide strong abnormal -activity detection in diverse settings. The fundamental component is the YOLO-based architecture (e.g., YOLOv5) which is trained on multimodal data containing annotated examples of such abnormal behaviours as crowd pushing, large crowd formation, fighting, fire, accidents, shoplifting, and vandalism. The following requirements are the prerequisites to the real-time deployment of the architecture

YOLOv5 is an advanced real-time object detection algorithm that detects and classifies multiple objects in one image or video frame using a single pass through a neural network. Due to its great efficiency and accuracy, it is more suitable for surveillance tasks, such as Unusual event detection.

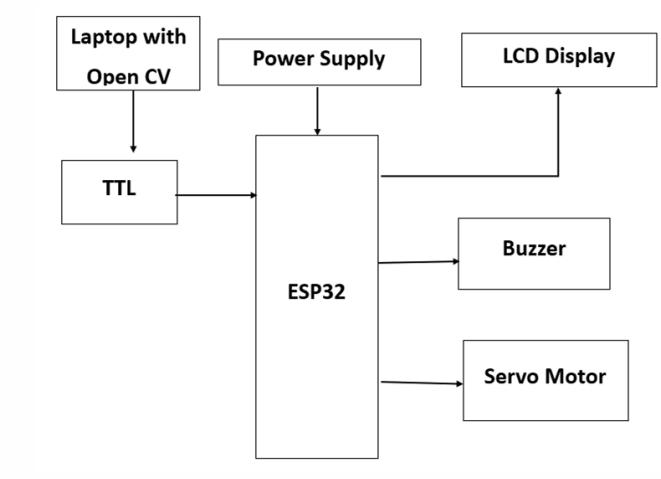

Fig.1 proposed system

It is important to note that the system we are looking at has parts that work together to detect and react to unusual events. A laptop is one of these parts. This laptop has OpenCV, which looks at video frames finds activity and starts the right actions. The ESP32 microcontroller is another part of this system. It gets signals from the laptop. Does something with them. Think of it as the controller of all the systems actions. It gets commands from the laptop. Carries them out. The good thing about the ESP32 microcontroller is that it has its power source so it can keep working without stopping.

The system detects something it can do several things. the ESP32 microcontroller can turn on the LCD display. The ESP32 microcontroller can activate a servo motor. This motor can help move a camera or open a gate. To sum up the laptop and the ESP32 microcontroller are crucial, to the system. The laptop detects activity and sends commands to the ESP32 microcontroller. The ESP32 microcontroller then carries out these commands. The ESP32 microcontroller and laptop work together to make the system react to events. The system uses the laptop and ESP32 microcontroller to detect and respond to abnormalities.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

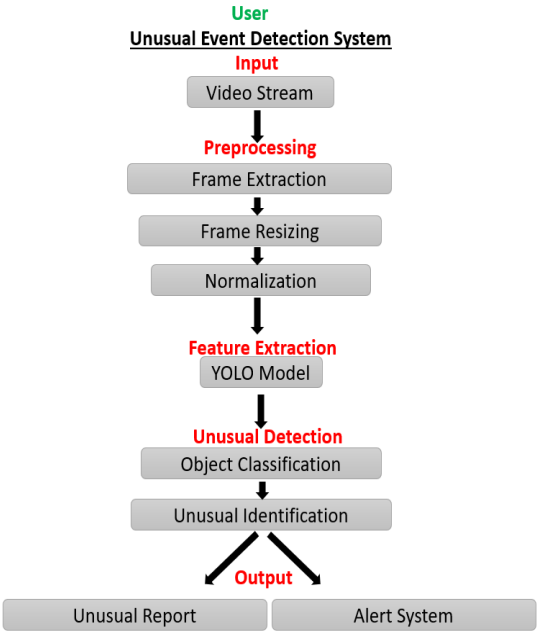

Fig.2 Flowchart of Event Detection

Anomaly detection workflow starts with dataset acquisition where video data can be obtained from public datasets such as UCF-Crime or Avenue Dataset or through custom CCTV. The dataset consists of videos containing both regular and abnormal activities like fighting, theft, fire, loitering, and normal walking. After collecting the videos, they have been pre-processed by converting videos to frames (i.e., 30 fps) and marking abnormal activities using Labelling and Roboflow. Also, images are resized into 640 x 640 and saved in YOLO format with classes and bounding boxes. In feature extraction, the YOLOv5 model applies convolutional neural networks that extract features from the videos including spatial and temporal features like human pose, motion, and object interaction. After extracting the above features, classification is done where activities are assigned classes such as Class 0: Normal, Class 1: Fighting, and Class 2: Theft. Also, YOLOv5 gives a confidence score to determine the correctness of detection. Results of the event detector are evaluated based on metrics such as precision, recall, F1 score, and mean Average Precision (mAP). this models yield approximately 90-95% accurate results. Finally, the model outputs a real-time bounding box with an activity label for detected actions. Whenever there are abnormalities in

IV. SYSTEM REQUIREMENTS

Hardware Requirements

Processor: Intel Core i5 or its higher or a similar AMD processor, to allow real time processing of video. Memory (RAM): 8GB RAM or more to run video analytics and YOLO model inference. Graphics Card: An NVIDIA GPU that supports CUDA (e.g. GTX 1050 Ti or later) is required to run accelerated inference of deep-learning and real-time detection. Storage: A 100 GB or lesser dataset, model, logs and video recording storage. Camera: IP camera with high resolution that can stream 720p or greater video feeds. Power Supply: It helps to keep the system running all the time.

Software Requirements

Operating System: we use Windows 10/11 and Ubuntu 18.04. It is recommended to use the deep learning approach in case Linux operating system is being used. Programming Language: we use Python. Deep Learning Tools: we use YOLO, PyTorch and TensorFlow. Computer Vision Tools OpenCV: We add Capture video data, process and visualize. YOLO Algorithm: we use YOLOv5.or a custom trained version. CUDA Toolkit and cuDNN: needed in the case of GPU acceleration (with Nvidia GPU).Other Libraries: NumPy, Matplotlib, pandas to work with data and visualise it. Database: A SQLite or MySQL to log events where required. IDE/Editor: Visual Code, PyCharm or Jupyter Notebook to develop.

Functional Requirements

Crowd Pushing Detection: This identifies and reports cases of crowd pushing in a crowd. Large Crowd Detection: Flag ration and detection of scenes with a crowd of a specific size. Fighting Detection: Fighting between people: Identify fighting and raise an alarm. Fire Detection: detect fire or smoke to provide early warning of fire. Detection of Accidents: on the road (people or vehicles): Recognition of accidents in cameras. Shoplifting Detection (S): Identifying suspicious behaviour that can be used to tell whether a theft has occurred in a retail setting. Vandalism Detection): Recognise property-destructive behaviours. Real-Time Alerts: Send notifications (audio/visual/email/message) on observed events. Video Feed Analysis: Process video feeds in groups of cameras. Event Logging: Record identified events and time stamping them to be later viewed.

Non-Functional Requirements.

Performance: It will be Detection in real time or near real time. Accuracy: It will Accuracy and recall must be as high as possible to avoid both false negatives and positives. Scalability: It will Ability to handle more video streams and detections. Reliability: System must run without failure, and it must have strong error management. Security: The event data and video feeds should be stored and transmitted securely. Usability: Have a basic user interface in configuring detection thresholds and result viewing. Maintainability: Keep it modular and well documented to enable the future enhancements.

Requirement Specification

At least fifteen frames per second from each camera should be processed by the system. The YOLO detection algorithm needs to detect the following abnormal behaviors: crowd shoving, large crowds, fighting, fires, accidents, thefts, and vandalism. The alerts generated and logged include the behaviours type, time stamp, camera ID, and an optional

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

captured image of the occurrence. The application must allow the administrator to configure thresholds for what constitutes "large crowd" and detection sensitivity. All detections and related logs must be retrievable through an admin dashboard. System updates for example, model retraining and deployment should not cause more than 30 minutes of downtime.

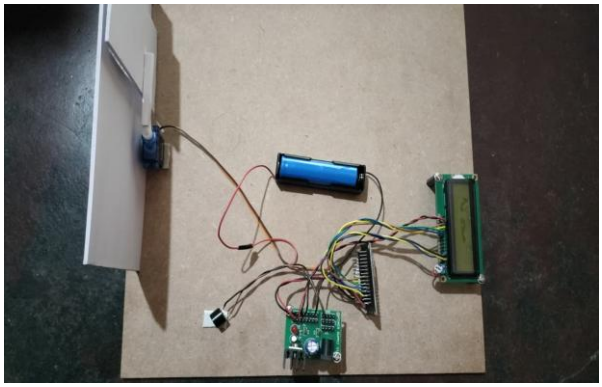

Fig.3 Hardware Components

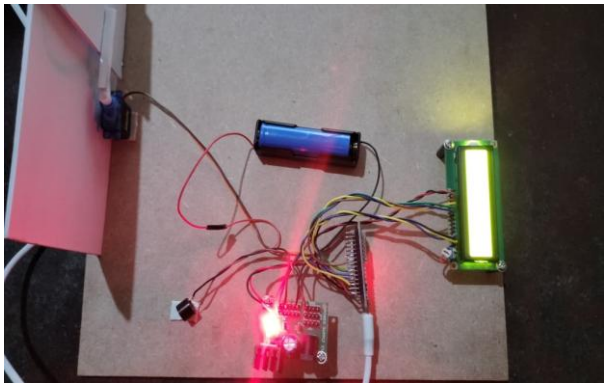

Fig.4 Hardware Connections

V. RESULT AND ANALYSIS

In this chapter, the empirical results of the suggested system on Automated Unusual Event Detection based on the YOLOv5 deep-learning framework are recorded. The effectiveness of the model is evaluated by a collection of metrics, which are precision, recall, F1 -score, accuracy, and mean Average Precision (mAP). The findings are evidence of the effectiveness of the YOLOv5 model in identifying abnormal behaviour of real-time surveillance images.

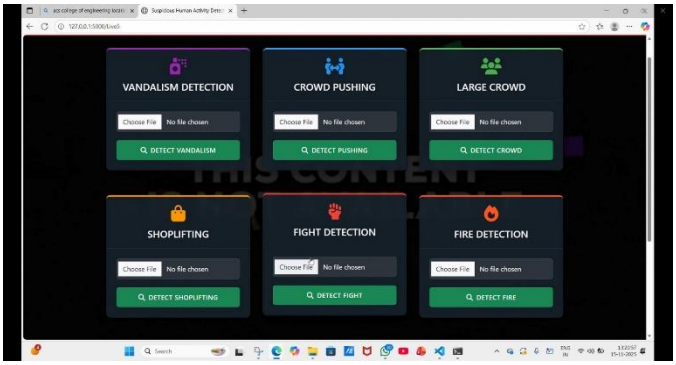

Fig.5 Working of Detection

This entire process you have implemented involves a multi-function AI surveillance system with real-time responses based on a combination of Computer Vision techniques. The system takes in the dataset from any of several datasets including UCF-Crime, Avenue or even your own CCTV footage which comprises both normal activities and some examples of abnormal activities such as fights, theft, fire, loitering, and just walking about. From these videos, frames are extracted, abnormal activities annotated, resized to 640*640 dimensions and saved in YOLO format. In YOLO v5, the features extraction process occurs through convolutional layers where spatial and temporal aspects such as body pose, body motion and object interaction are detected for activity classification into different classes like normal, fight and theft, all associated with confidence values. Model performance evaluation is done using parameters such as precision, recall, F1 score and mean average precision which have achieved up to 90-95 percent accuracy through fine tuning of the model. After all the processes, real time bounding boxes are produced with an activity label; and in the case of detection of abnormal activities, alerts will be produced. In connection with the hardware such as ESP32, LCDs, buzzer and servo motors.

Fig.6 Telegram Notification

In essence, what we have here is a perfect example of an Automated Unusual Event Detection in real time that combines computer vision and communication. The process starts with the input of video feed and the subsequent analysis of that feed via a pre-trained YOLOv5 detector. Suspicious events detected in the video feed include fire, smoke, vandalism, shoplifting, or a fight. When an anomaly is detected, the system provides a confidence score in addition to outlining the anomaly with the help of bounding boxes in the video feed. Next comes the transmission of information to the alerting component, a Telegram bot, for instance, that will promptly notify the user of the presence of the suspicious event along with its type, confidence score, and time of detection. As an additional option, users receive the Google Maps address of the site with the anomaly such as ACS College of Engineering. In certain situations, the user also gets the annotated image that helps identify the anomaly. Hence, the process becomes fully

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

automated and consists of four stages – detection → classification → alerting → response, which makes the system very practical and effective for use.

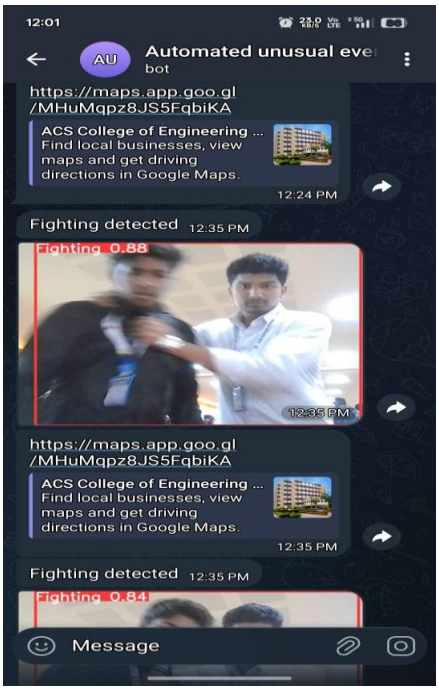

Fig.7 Telegram Notification

VII. CONCLUSION

The proposed project is titled Automated Unusual Event Detection Using YOLOv5 and aims to advance the issue of ensuring the safety and security of people by utilizing the deep learning algorithms to provide the real-time detection of abnormal activities in the video surveillance. With the inclusion of the YOLOv5 object detector, the system can identify and classify abnormal behaviours, such as fighting, loitering, theft and other suspicious behaviours with high accuracy and minimal latency. YOLOv5 has significant strengths including high processing speed, light architecture and better performance in comparison to predecessors and other detection models. A carefully selected list of normal and abnormal human behaviours was used to train the model, which is reliable and robust in different real-world situations. In general, the obtained system shows promising outcomes with high precision, recall and F1-score. Its relevance to the context of smart surveillance e.g. retail settings, pedestrian streets, educational and transport interchange hubs gives it a powerful means of automated surveillance and early warning of threats.

REFERENCES

[1] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, You Only Look Once: Unified, Real-Time Object Detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[2] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, YOLOv4: Optimal Speed and Accuracy of Object Detection, arXiv preprint arXiv:2004.10934, 2020. [Paper Link]

[3] Glenn Jocher et al., YOLOv5 by Ultralytics, GitHub Repository, 2020.

[4] Sultani, Waqas, Chen, and Mubarak Shah, Real-world Anomaly Detection in Surveillance Videos, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[5] Hasan, Mahmudul, et al., Learning Temporal Regularity in Video Sequences, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[6] Cheng-Chun Lee, Yi-Hsuan Tsai, Wei-Chen Chiu, Yu-Chiang Frank Wang, Abnormal Event Detection via Recurrent Neural Networks, IEEE International Conference on Image Processing (ICIP), 2018.

[7] UCSD Anomaly Detection Dataset, University of California, San Diego (UCSD).

[8] Avenue Dataset, City University of Hong Kong.

[9] ShanghaiTech Campus Dataset for Abnormal Event Detection, ShanghaiTech University.

[10] Ultralytics YOLOv5 Documentation, Official Docs – Object Detection and Training Guide.