Weapon detection and crowd management using existing CCTV, AI and ML

Mohammed Tajuddin Azam[1], Sridhar Gummalla[2], Shaikh Shakeer Basha[3]
[1]*PG Scholar, Department of IT,*
[2,3]*Professor, Department of CSE,*
Shadan College of Engineering and Technology, Hyderabad, Telangana-500086

***Abstract*- In the world with the population the increase in crime rate has been evolved as well. To prevent crimes and for our safety it is important to have a device which help in recognizing the criminals. As urban security is constantly evolving the machine learning (ML) can be a game changing solution. Motivated by real-world challenges where conventional surveillance methods often fall short, our project aims to transform urban safety through innovative technological solutions. Our approach harness the potential of ML, involve YOLO which stands for "YOU ONLY LOOK ONCE" algorithm. With risks everywhere security is always an important aspect and due to rise in crime rate in a crowded events or suspicious area with less activity. With the growing demand for the protection of safety, security and personal property deploying video surveillance systems capable of recognizing and interpreting scenes and anomalous events plays a vital role in intelligent monitoring. This paper implement automatic gun or weapon detection using a faster region based convolution neural network algorithms. This paper works when in a crowd there are any weapon detected it can manage the crowd and locate any weapon if detected.**

I. INTRODUCTION

Owing to the increase of technology the urban security can be easily solved. Criminals usually chooses places which is highly crowded area or any area which has low activity and is lonely. These places are easy for the criminal to commit crime and escape easily. In a crowded place when crime takes place the panic occurs in the crowd and an uproar situation arises, the criminal take advantage of this situation and use it to escape and het vanished in the crowd. Whereas in the area which has less activity and is deserted area the criminal can commit a crime without getting captured by anyone and can easily leave the crime scene. When in a crowd the criminal can shoot the victim from a very close or may be a far range in situation like this investigating the criminal can be difficult as the criminal can get missed in crowd with naked eyes. Crime can be of many types robbery, murder, sexual assault, hit and run etc. in any of the case criminal do have a weapon. Especially if the criminal is associated with a gang or mafia keeping a weapon with them is a priority. The most used weapon by criminals is a gun. The gun can be of any type such as assault rifle (AR), pistol, shotgun, sniper, and sub machine gun. Using these weapon can give an advantage of range to the criminal.

Our project is a solution for the crime investigators to inspect the crime scene and get a good lead in the crime case. Our aim is to detect any malicious act captured by the CCTV in a crowded area. This project detect a weapon used by anyone in the crowd. To bring this work in action we have used Artificial Intelligence (AI) and Machine Learning (ML) algorithms.

To manage the crowd with existing CCTV we have used the YOLO algorithm. The YOLO algorithm have many versions here we are using the version 8 of the YOLO algorithm (YOLOv8).the YOLO algorithm is the real time weapon detection algorithm used in computer vision. Specialty of the YOLO algorithm is it is an amazing algorithm to detect weapon in a given image or in a given video. This algorithm finds and classifies in a single attempt through neural network. Weapon detection, localization and single pass are the key features of YOLO. Weapon detection in YOLO refers that yolo detect what the weapon is, such as (animal, weapon, human). Localization determines where the weapon is located and as the name of algorithm itself speak for its feature YOU ONLY LOOK ONCE which means the algorithm detects in single pass.

The second algorithm used here is the Faster RCNN. This is the faster approach for the CNN. Initially it was the CNN with the phenomenon of the neural network then an upgraded algorithm was RCNN and now the faster RCNN which is faster than the previous algorithms. The RCNN stands for "REGION-BASED CONVOLUTIONAL NEURAL NETWORK". The RCNN is an algorithm which also works for weapon detection but does have different working than the YOLO. The main difference between YOLO and RCNN is the RCNN focus on a particular weapon by giving a prior focus and detecting a particular area in the image. The RCNN select regions in the image and pass through neural network for a much detail detection. The faster RCNN adds a region proposal network to propose and entire pipeline is made end-to-end for faster output.

These are the major algorithms which work for the weapon detection and is used in our work widely. For better results in our work YOLO works with the crowd management and RCNN to detect a weapon.

II. BACKGROUND RESEARCH AND LITERATURE REVIEW

The CCTV was invented for the surveillance of the v-2 rocket tracking in the year 1942 by a German engineer known as Walter Bruch. Later on the CCTV was used by the American government surveillance of the troops and atomic experiments. The use of CCTV as a surveillance of public action was taken place in the year of 1960 where the CCTV Was used to prevent theft and robbery. Slowly every firm and organization started to use the CCTV for preventing the robbery. Gradually evolving for traffic management, home security and private security. Today the CCTV has evolved so much that every household do use a CCTV and in every organization installation of CCTV became a vital role.

In the 20th century the use of artificial intelligence was making a high demand previously the AI was used only for huge mathematical calculation. Many engineers gave their interest in AI and made AI a leading phenomenon in every aspect. AI is widely used in every field to make their work load easier and faster with the help of machine learning.

The high need of artificial intelligence in the market encouraged us to use artificial intelligence (AI) and machine learning (ML). During a survey we got to know the importance of CCTV in the urban security. In the case study of previous criminal records the investigators got substantive evidence regarding the case. And in some cases the CCTV do had crucial evidence but the evidence was not easily found due to crowd. The criminals where right in front of the eyes but it took repeated attempts of watching the footage and took a much time to get to the evidence.

A case from the Spain was listed in which the victim was shot dead in a carnival which was crowded by people. The moment fire tool place a chaotic moment appeared and crowd started to run along. The officer investigated the CCTV footage and missed the evidence, 2 days later the officer reviewed the footage with focus and found the evidence which was right in front but due to chaotic situation the evidence was hard to be found. In such cases a system which can detect the weapon and manage the crowd can be useful to easily find the evidence through the CCTV footage.

III. SYSTEMATIC ASSESSMENT OF DETECTION METHODS

YOLO and RCNN are the most used and trusted algorithms for the weapon detection from the any image. These algorithms are widely used for the detection of the weapon. CCTV is used to record the 24x7 footage of a particular area which can be used to detect the weapon passing through algorithms such as YOLO and RCNN. Both the algorithms have their own working techniques and both algorithms are reliable in detection of weapon.

i. YOLO (YOU ONLY LOOK ONCE)

YOLO is a reliable, efficient and suitable for real-time weapon detection. The YOLO have different variants such as YOLO (v1, v2, v3, NAS…v11). The YOLO recently got a new version on February 20th 2025. YOLO treats weapon detection as a single regression problem. In the process the image is divided into grids and these grids are directly predicted using bounding boxes. Unified model structure of YOLO allow it to process images extremely quickly, this phenomenon of the YOLO make it suitable for real-time applications such as autonomous driving and video surveillance. YOLO have been upgrading its versions with speed, accuracy and detecting smaller weapons. The YOLO became a popular framework for its balance of accuracy and speed. We have used the YOLOv8 version due to its easier interface and a clean

python interface. The yolo works on python version 3.8 and extending. The YOLO is installed using pip with the command "pip install ultralytics" and later the YOLO can be imported from ultralytics.

ii. FASTER RCNN

The Faster RCNN stands for "Faster Region Based Convolutional Neural Network" which was initiated in the year 2015 by Shaoqing ren. Shaoqing builds it upon earlier models like CNN, RCNN, and FAST RCNN. There are two main stages from which the images passes for weapon detection. The very first stage is RPN (Region Proposal Network) which scans the image using a convolutional backbone such as ResNet, which generate region proposals bounding boxes where weapon might be located. The second stage is where the proposal are fed into Fast RCNN that classifies the weapon and refine the bounding box coordinates. Faster RCNN set a benchmark in medical imaging and surveillance

Summary for YOLO and RCNN

Model	Faster RCNN	YOLO
Detection style	Two-stage (proposal and classification)	Single stage
Speed	Moderate (5-10 FPS)	Very fast (up to 100+ FPS)
Accuracy	High for small and dense weapon	Good to excellent in modern version
complexity	More complex training	Easier to deploy and train
Best use case	High accuracy offline application	real time, embedded and resource constrained

IV. CHALLENGES AND LIMITATION

As said the term "every power do have its limitation" we can say that a CCTV is a powerful medium to get evidence from a crime scene, with its power there are few challenges and limitation. CCTV footages are used for collecting evidence and during this process there are certain problems faced by the investigators. The initial limitation is the quality of picture in the footage. Due to the average quality of the CCTV footage it is difficult to detect any weapon in the captured picture or a video footage. Larger weapon are easier to detect the weapon, when the weapons are smaller it is hard to detect the weapon. To search a weapon which is smaller the image captured by the CCTV needs to be zoomed in to search the weapon in the image or video. Due to the average quality of pictures when the image is zoomed in the image gets blurry and the weapon in the image will not be as clear as expected.

To think about challenge faced during weapon detection is when the area which is under surveillance is crowded with a huge amount of peoples. Manually detecting any weapon in such crowd is not an easy job, investigators must go through every minute pixel of an image to search for an weapon. The time taken to search for an weapon in the image is higher due to a huge crowd. Time taken in detecting an weapon in the image will be more than expected. In some cases where investigator miss the weapon during the investigation had to give multiple attempts which will double the time consumed in search of a weapon. Due to the limitation of quality in the image taken by the CCTV some weapons do not have clear appearance which leads to invisibility of a weapon during the weapon detection.

Another challenge and limitation is the angle of the CCTV, previously the CCTV were meant to surveillance only a particular angle. In such limitation of CCTV the challenge faced is to detect a weapon when it is place at the edge or border line of the CCTV angle. When a weapon is in the corner there are high chances that half of the weapon is cropped and is out of the frame, owing to this manually detecting a weapon can be challenging. The weapon can be confused with other weapon because of the similarity between two weapons.

V. PROPOSED ENHANCEMENT

To work around the challenges and limitation of weapon detection using CCTV we have proposed this work which can make weapon detection easier.

The work we proposed here is to make weapon detection and crowd management faster and accurate using artificial intelligence (AI) and machine learning (ML). Using the models like YOLO and faster RCNN

we can come to a solution which can resolve the issue in weapon detection. Our work can make the weapon detection faster than manually detecting the weapon and need no repeated attempts.

As listed above the initial challenge is to find small weapons in the CCTV images due to its bad quality when zoomed in. The Faster RCNN will help us in detecting these small weapons in CCTV images. The faster RCNN passes the image through RPN which detect smaller images with more effective and accuracy rate. The model is trained in such a way that it detect any weapon and a bounding box marks the weapon and mark a word around the weapon. The marked bounding box show us where a weapon is placed.

In the second challenge listed above, we know that the weapon detection is difficult in the crowded place and CCTV fails to show smaller weapon in the crowded place to counter this challenge we are using the YOLO model which will help us in managing the crowd. The crowd management help us to bound people with a bounding box and assign them with numbers so that we can detect people and manage the crowd. In a crowd when we are detecting smaller weapon through CCTV images we have to initially manage a crowd to make our weapon detection easier. When the crowd is managed by numbering and bounding we can use faster RCNN to detect weapon.

For the third challenge we can use the faster RCNN and YOLO together we can pass previous frames of surveillance captured by CCTV in these models and any of those frames. As a result the frames can be matched to detect the final weapon

VI. CONCLUSION AND FUTURE WORK

By this research work we conclude that the CCTV surveillance can be an important medium to detect weapon. CCTV installation can help in urban security. The footage captured need artificial intelligence and machine learning to make urban security and further investigation easier. The time take by the manual detection of weapon can be much longer than done through artificial intelligence and machine learning. Smaller weapon in the crowd can barely be seen which can lead in missing the critical evidence. CCTV plays a major role in surveillance and using YOLO and faster RCNN smaller weapons in a heavy crowded area can be detected easily.

Future work can be done with upgraded versions of YOLO and the faster RCNN. Adding advance feature such as real time detection advancement of weapon through the CCTV. When the YOLO and faster RCNN models are advanced and directly connected to CCTV instant weapon detection can be possible. The instant detection of weapon through CCTV will help in taking precautions. In capturing footages memory allocation can be higher, to manage the storage future work can be done by adding YOLO and RCNN directly to CCTV. This future work will help in reducing time taken for collecting multiple frames of CCTV footage and passing each frames to the YOLO and faster RCNN process.

This work is beneficial for the smart homes by IOT. The door bell security camera can also adopt this work and make the security system smarter and increase the security for any individual. Instant detection of weapon through CCTV can help people to get alert and take necessary precautions. Smarter CCTV combined with YOLO and faster RCNN and further versions of these models can reduce the crime rate and help people to prevent the upcoming crime. With the increase in technology the future plan for this work will be to directly combine any sort of surveillance medium with AI and ML models like YOLO and faster RCNN to detect any weapon at the very moment and take security measures before the crime take place.

REFERENCES

[1]. Wei Liu et al., “SSD: Single Shot MultiBox Detector”, European Conference on Computer Vision, Volume 169, pp 20-31 Sep. 2017.

[2]. D. Erhan et al., “Scalable Weapon Detection Using Deep Neural Networks,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[3]. Ruben J Franklin et.al., “Anomaly Detection in Videos for Video Surveillance Applications Using Neural Networks,” International Conference on Inventive Systems and Control, 2020.

[4]. H R Rohit et.al., “A Review of Artificial Intelligence Methods for Data Science and Data Analytics: Applications and Research Challenges,”2018 2nd International Conference

on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), 2018.

[5]. Abhiraj Biswas et. al., "Classification of Weapons in Video Records using Neural Network Framework," International conference on Smart Systems and Inventive Technology, 2018.

[6]. Pallavi Raj et. al.," Simulation and Performance Analysis of Feature Extraction and Matching Algorithms for Image Processing Applications" IEEE International Conference on Intelligent Sustainable Systems, 2019.

[7]. Mohana et.al., "Simulation of Weapon Detection Algorithms for Video Surveillance Applications", International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), 2018.

[8]. W. Halboob, H. Altaheri, A. Derhab and J. Almuhtadi, "Crowd Management Intelligence Framework: Umrah Use Case," In IEEE Access, vol. 12, pp.6752-6767, 2024.

[9]. S. Kilic and M. Tuceryan, "Crime Detection from Pre-crime Video Analysis with Augmented Pose Information," 2023 IEEE International Conference on Electro Information Technology (eIT), Romeoville, IL, USA, 2023, pp. 1-6.

[10]. A. Goenka and K. Sitara, "Weapon Detection from Surveillance Images using Deep Learning." 2022 3 rd International Conference for Emerging Technology (INCET), Belgaum, India, 2022, pp. 1-6.

[11]. A. Kumari Sirivarshitha, K. Sravani, K.S. Priya and V. Bhavani, "An approach for Face Detection and Face Recognition Libraries in Python," 2023 International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 1274-1278

[12]. A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi, and A. Pentland, "Once upon a crime: towards crime prediction from demographics and mobile data," Proc. of the 16th Intl. Conf. on Multimodal Interaction, pp. 427-434, 2014.