Real-Time CCTV Video Analysis: Deep Learning for Weapon Detection

Subba Reddy Borra[1], B. Nikitha[2], B. Anjali[2], A. Sirichandana[2], D. Harshitha[2]

[1] Professor and Head, Department of Information Technology, Mallareddy Engineering College for Women, (UGC-Autonomous), Hyderabad, India, bvsr79@gmail.com.

[2] Student, Department of Information Technology, Mallareddy Engineering College for Women, (UGC-Autonomous), Hyderabad, India.

Abstract

CCTV cameras, you know, those surveillance cameras you often see in public places, stores, and important buildings, play a crucial role in keeping us safe and secure. They constantly record video footage to monitor what's going on around and ensure our safety. However, as safety concerns grow, it's becoming increasingly important to improve these CCTV systems, making them capable of detecting potential threats, like weapons, in real-time. The current way CCTV surveillance works is that human operators must manually watch all those live video feeds from multiple cameras, which is error-prone, and honestly, a person can only effectively keep an eye on so many cameras at once. With the huge amount of video data these cameras generate, it's practically impossible for human operators to watch everything continuously, which means security threats might get missed. The traditional approach is passive i.e., the security personnel or operators sit and watch the video feeds, hoping to spot anything suspicious, like someone carrying a weapon. But this approach has its limitations too. Humans can make mistakes, and they might not react quickly enough in a real-time situation. In addition, as the number of cameras increases, it becomes tough to scale this method, and the costs can go up significantly. So, to overcome these challenges and enhance public safety, a more advanced solution is required. Therefore, this project develops a real-time CCTV video analysis with deep learning for weapon detection. By using deep learning models, a sophisticated system can be built that quickly analyzes the video streams from CCTV cameras in real-time. This means it can detect weapons and potential threats as they happen. It's super-fast and accurate, so it reduces the chances of false alarms or missing something important. In addition, it's scalable, cost-effective and helps security agencies respond quickly to potential threats and keeps us all protected in a better and more efficient way.

1. Introduction

The crime rate across the globe has increased mainly because of the frequent use of handheld weapons during violent activity. For a country to progress, the law-and-order situation must be in control. Whether we want to attract investors for investment or to generate revenue with the tourism industry, all these needs is a peaceful and safe environment [1]. The crime ratio because of guns is very critical in numerous parts of the world. It includes mainly those countries in which it is legal to keep a firearm. The world is a global village now and what we speak or write has an impact on the people. Even if the news they heard is crafted having no truth but as it gets viral in a few hours because of the media and especially social media, the damage will be done [2]. People now have more depression and have less control over their anger and hate speeches can get those people to lose their minds. People can be brainwashed, and psychological studies show that if a person has a weapon in this situation, he may lose his senses and commit a violent activity [3]. High incidents were recorded in past few years with the use of harmful weapons in public areas. Starting with the past year's attacks on a couple of Mosques in New Zealand, on March 15, 2019 at 1:40 pm, the attacker attacks the

Christchurch AL-Noor Mosque during a Friday prayer killing almost 44 innocent and unarmed worshippers. On the same day just after 15 minutes at 1:55 PM, another attack happened killing seven more civilians [4]. Active shooter incidents had also occurred in USA and then in Europe. The most significant cases were those at Columbine High School (USA, 37 victims), Andreas Broeivik's assault on Uotya Island (Norway, 179 victims) or the Charlie Hebdo newspaper attack killing 23. According to stats provided by the UNODC, among 0.1 million people of a country, the crimes involving guns are very high i-e. 1.6 in Belgium, United States having 4.7 and Mexico with several 21.5 [5].

CCTV cameras play an important role to overcome this problem and are one of the most important requirements for the security aspect. [3]. CCTVs are installed in every public place today and are mainly used for providing safety, crime investigation, and other security measures for detection. CCTV footage is the most important evidence in courts. After a crime is committed, law enforcement agencies arrive at the scene and take the recording of footage with them [6]. If we look at the surveillance system of different countries around the world, UK has about 4.5 million cameras, which are used for surveillance. Sweden has about 50000 cameras installed around 2010. The government of Poland was able to reduce drug cases by 60% and street fights by 40% by installing just 450 cameras in the city of Poznan [7]. China has the world's biggest surveillance system and 170 million cameras around the nation, and these are expected to expand three times, through an additional 400 million to be connected by 2020. It took only seven minutes for Chinese officials to find and apprehend BBC reporter John Sudworth using their strong CCTV cameras network and facial recognition technology and put the criminal behind the bar [8]. In previous years, though having surveillance cameras installed, to use them for security purposes was not an easy and dependable method. A human has to be there all the time to monitor screens. CCTV operator has to monitor 20–25 screens for 10 hours. He has to look, observe, identify, and control the situation that can be harmful to the individuals and the property. As the number of screens increases, the concentration of the person decreases considerably to monitor each screen with time. It is impossible for the person monitoring the screens to keep the same level of attention all the time [9].

2. Literature Survey

The solution to problem is to install surveillance cameras with the ability to automatically detect weapons and raise alarm to alert the operators or security personals. However, there is not much work done on algorithms for weapon detection in surveillance cameras, and related studies are often considering concealed weapon detection (CWD), mostly using X-rays or millimeter waves images employing traditional machine learning techniques. In the past few years, deep learning in particular convolutional neural network (CNN) has given groundbreaking results in object categorizing and detection. It has achieved finest results thus far in classical problems of image processing such as grouping, detection and localization. Instead of selecting features manually, CNN automatically learns features from given data.

Bhatti, et al. [10] discussed a deep learning-based system for real-time weapon detection in CCTV videos. It likely presents a methodology and results related to weapon detection. However, this may not provide an extensive evaluation of the model's performance in various real-world scenarios. Qi, et al. [11] introduced a dataset and system for real-time gun detection in surveillance videos using deep learning. It focused on the dataset's characteristics and system performance. But the dataset used may be limited in diversity and generalization, impacting the system's real-world applicability. In addition, the results presented are very limited. In [12], authors discussed an algorithm for human pose estimation from CCTV images, likely focusing on the methodology used for posing estimation. However, the algorithm designed for specific types of CCTV images, limiting its generalizability. In addition, it has a lack of benchmarking. Arya described an automatic and accurate weapons detection model using an optimal neural network architecture [13]. But this work hasn't provided thorough validation on diverse real-world data, potentially limiting its practical utility. Moreover, the optimal architecture might require significant computational resources. Akhila, and Ahmed [14] focused on firearm detection using deep learning, likely presenting a method and its results. However, this work lacks the detailed information and the system's performance in real-world surveillance scenarios hasn't thoroughly discussed. In [15] authors discussed a computer vision-based framework for gun detection using the Harris interest point detector. But the use of the Harris interest point detector has a limitation that the model's performance compared to modern techniques. In addition, availability of suitable training data also limited.

3. Proposed System Model

The project aims to develop a system that can automatically and rapidly detect weapons in real-time surveillance video streams. This project leverages the YOLO (You Only Look Once) model, a popular deep learning framework for object detection, to achieve its objectives. Below is an overview of the key components and goals of this project as shown in Figure 1.

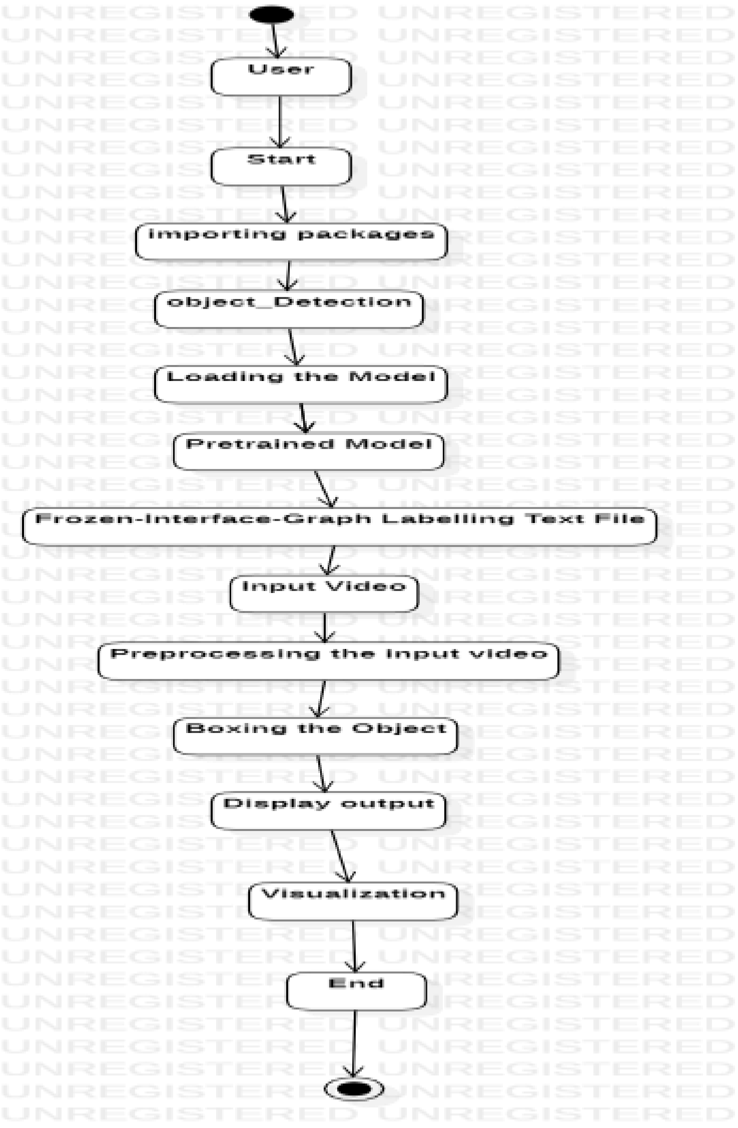


Figure 1: Overall design of proposed weapon detection model.

According to the facts, training and testing of proposed model involves in allowing every source image via a succession of convolution layers by a kernel or filter, rectified linear unit (ReLU), max pooling, fully connected layer and utilize SoftMax layer with classification layer to categorize the objects with probabilistic values ranging from [0,1]. Convolution layer as is the primary layer to extract the features from a source image and maintains the relationship between pixels by learning the features of image by employing tiny blocks of source data. It's a mathematical function which considers two inputs like source image $I(x, y, d)$ where x and y denotes the spatial coordinates i.e., number of rows and columns. d is denoted as dimension of an image (here $d = 3$, since the source image is RGB) and a filter or kernel with similar size of input image and can be denoted as $F(k_x, k_y, d)$.

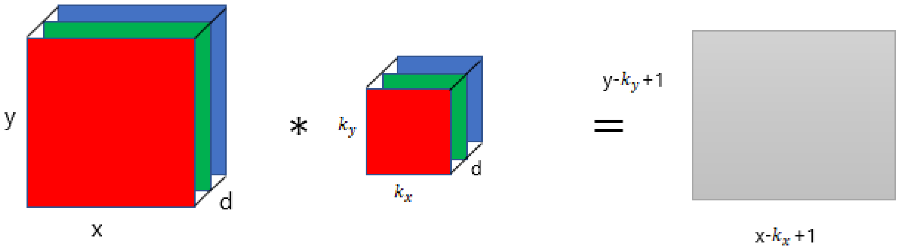
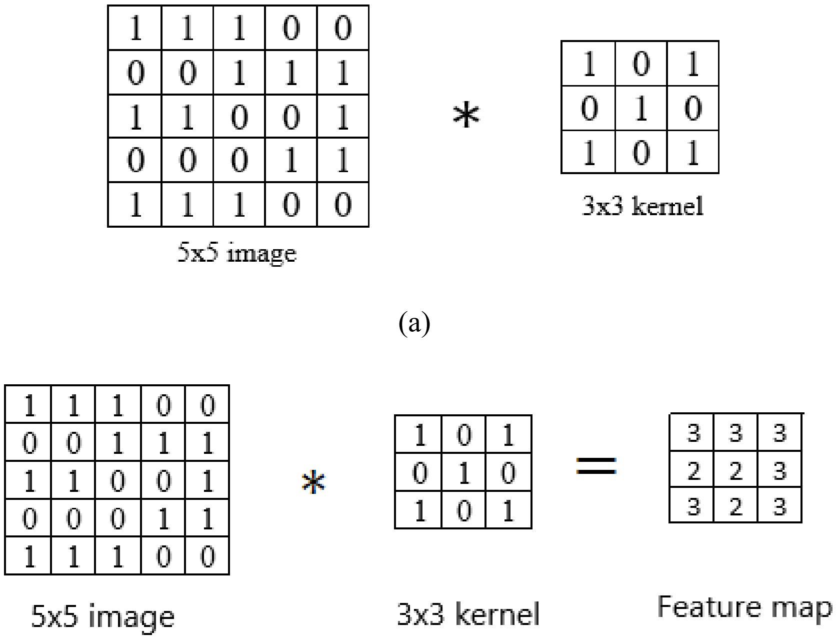


Fig. 2: Representation of convolution layer process.

The output obtained from convolution process of input image and filter has a size of $C\left((x - k_x + 1), (y - k_y + 1), 1\right)$, which is referred as feature map. Let us assume an input image with a size of 5×5 and the filter having the size of 3×3. The feature map of input image is obtained by multiplying the input image values with the filter values.



1	1	1	0	0
0	0	1	1	1
1	1	0	0	1
0	0	0	1	1
1	1	1	0	0

5x5 image

*

1	0	1
0	1	0
1	0	1

3x3 kernel

(a)

1	1	1	0	0
0	0	1	1	1
1	1	0	0	1
0	0	0	1	1
1	1	1	0	0

5x5 image

*

1	0	1
0	1	0
1	0	1

3x3 kernel

=

3	3	3
2	2	3
3	2	3

Feature map

(b)

Fig. 3: Example of convolution layer process (a) an image with size **5** × **5** is convolving with **3** × **3** kernel (b) Convolved feature map

ReLU layer: Networks those utilizes the rectifier operation for the hidden layers are cited as rectified linear unit (ReLU). This ReLU function $\mathcal{G}(\cdot)$ is a simple computation that returns the value given as input directly if the value of input is greater than zero else returns zero. This can be represented as mathematically using the function $max(\cdot)$ over the set of 0 and the input x as follows:

$$\mathcal{G}(x) = \max\{0, x\}$$

Max pooing layer: This layer mitigates the number of parameters when there are larger size images. This can be called as subsampling or down sampling that mitigates the dimensionality of every feature map by preserving the important information. Max pooling considers the maximum element form the rectified feature map.

Softmax classifier: Generally, as seen in the above picture softmax function is added at the end of the output since it is the place where the nodes are meet finally and thus, they can be classified. Here, X is the input of all the models and the layers between X and Y are the hidden layers and the data is passed from X to all the layers and Received by Y. Suppose, we have 10 classes, and we predict for which class the given input belongs to. So, for this what we do is allot each class with a particular predicted output. Which means that we have 10 outputs corresponding to 10 different class and predict the class by the highest probability it has.

4. Results description

This project uses the OpenCV (cv2) library and a pre-trained YOLOv3 model to perform object detection on either a video feed from a webcam or a video file. Specifically, it is designed to detect objects related to guns and knives, and it draws bounding boxes around those objects when detected. Figure 1 shows a still frame or a snapshot from a video captured by a closed-circuit television (CCTV) camera. The key characteristic of this figure is that it does not contain any visible weapons. The scene is free of objects like guns or knives. The purpose of this figure is to provide an example of a situation where no weapons are detected. Similar, to Figure 1, Figure 2 also shows sample snapshots from CCTV footage. This is another example of a scenario where the system does not detect any weapons.

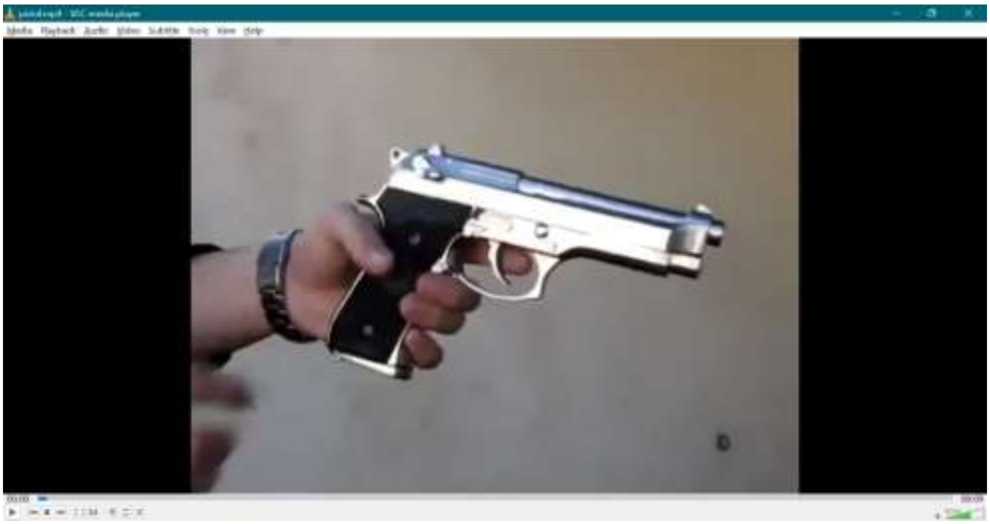


Figure 4: Sample video 3 collected from CCTV footage with a presence of weapon.

5. Conclusion

This project implemented a weapon detection system utilizing a pre-trained YOLO model. Its current functionality allows for the real-time detection of guns and knives in video streams, offering potential applications in security and safety monitoring. However, the system's accuracy hinges on the quality of the pre-trained model and the dataset used for training. To enhance detection accuracy and minimize false positives and negatives, further refinement through fine-tuning or dataset expansion is advisable. It's worth noting that, like any object detection system, it may not achieve perfect accuracy and could produce occasional inaccuracies. Here, this project also incorporates a user-friendly interface, permitting users to specify video sources, whether from a file or webcam. This simplicity ensures accessibility for a broad user base, including security personnel and administrators.

References

[1] G. Arya. Automatic and Accurate Weapons Detection Model Using an Optimal Neural Network Architecture. SSRN 4172293 (2022). https://ssrn.com/abstract=4172293 (accessed 05 September 2022)

[2] K. Akhila, and K. R. Ahmed. Firearm Detection Using Deep Learning. Intelligent Systems and Applications: Proceedings of the 2022 Intelligent Systems Conference (IntelliSys) 544: 200-218 (2022).

[3] Ahmed, S.; Bhatti, M.T.; Khan, M.G.; Lövström, B.; Shahid, M. Development and Optimization of Deep Learning Models for Weapon Detection in Surveillance Videos. Appl. Sci. 2022, 12, 5772. https://doi.org/10.3390/app12125772

[4] Ahmed S, Bhatti MT, Khan MG, Lövström B, Shahid M. Development and Optimization of Deep Learning Models for Weapon Detection in Surveillance Videos. Applied Sciences. 2022; 12(12):5772. https://doi.org/10.3390/app12125772

[5] Ahmed, Soban, Muhammad Tahir Bhatti, Muhammad Gufran Khan, Benny Lövström, and Muhammad Shahid. 2022. "Development and Optimization of Deep Learning Models for Weapon Detection in Surveillance Videos" Applied Sciences 12, no. 12: 5772. https://doi.org/10.3390/app12125772

[6] Ratcliffe, J. Video Surveillance of Public Places. US Department of Justice, Office of Community Oriented Policing Services: Washington, DC, USA, 2006. Available online: https://www.ojp.gov/ncjrs/virtual-library/abstracts/video-surveillance-public-places (accessed on 05 September 2023).

[7] Cohen, N.; Gattuso, J.; MacLennan-Brown, K. CCTV Operational Requirements Manual 2009. Home Office Scientific Development Branch: St. Albans, UK, 2009. Available online: http://designforsecurity.org/downloads/CCTV_Requirements.pdf (accessed on 05 September 2023).

[8] Murthy, C.B.; Hashmi, M.F.; Bokde, N.D.; Geem, Z.W. Investigations of Object Detection in Images/Videos Using Various Deep Learning Techniques and Embedded Platforms—A Comprehensive Review. Appl. Sci. 2020, 10, 3280.

[9] Bhatti, M.T.; Khan, M.G.; Aslam, M.; Fiaz, M.J. Weapon Detection in Real-Time CCTV Videos Using Deep Learning. IEEE Access 2021, 9, 34366–34382.

[10] M.T. Bhatti, M.G. Khan, M. Aslam, and M.J. Fiaz. Weapon detection in real-time cctv videos using deep learning. IEEE Access 9: 34366-34382 (2021).

[11] D. Qi, W. Tan, Z. Liu, Q. Yao, and J. Liu. A Dataset and System for Real-Time Gun Detection in Surveillance Video Using Deep Learning. 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC) 667-672 (2021).

[12] S. U. Rahman, S. A. Khan, and F. Alam. An Algorithm for Human Pose Estimation from CCTV Images. International Journal of Education (IJE) 3: (2019).

[13] G. Arya. Automatic and Accurate Weapons Detection Model Using an Optimal Neural Network Architecture. SSRN 4172293 (2022). https://ssrn.com/abstract=4172293 (accessed 05 September 2023)

[14] K. Akhila, and K. R. Ahmed. Firearm Detection Using Deep Learning. Intelligent Systems and Applications: Proceedings of the 2022 Intelligent Systems Conference (IntelliSys) 544: 200-218 (2022).

[15] M.T. Bhatti, M.G. Khan, M. Aslam, and M.J. Fiaz. Weapon detection in real-time cctv videos using deep learning. IEEE Access 9: 34366-34382 (2021).

[16] D. Qi, W. Tan, Z. Liu, Q. Yao, and J. Liu. A Dataset and System for Real-Time Gun Detection in Surveillance Video Using Deep Learning. 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC) 667-672 (2021).

[17] S. U. Rahman, S. A. Khan, and F. Alam. An Algorithm for Human Pose Estimation from CCTV Images. International Journal of Education (IJE) 3: (2019).

[18] T. R. Kumar, and G. K. Verma. A computer visionbased framework for visual gun detection using harris interest point detector. Procedia Computer Science 54: 703-712 (2015).