Effective Deep Learning Technique for Weapon Detection in CCTV Footage

1st Naresh Yeddula
Department of CSE(AI)
JNTUA College of Engineering
Ananthapuramu, India
nareshyeddula7@gmail.com

2nd B. Eswara Reddy
Department of CSE
JNTUA College of Engineering
Ananthapuramu, India
eswar.cse@jntua.ac.in

2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC) | 978-1-6654-9111-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICMNWC56175.2022.10031724

***Abstract*--The modern world is highly concerned with safety and security. The level of safety and security in a country's environment determines its capacity to draw international investment and tourism. Closed Circuit Television (CCTV) cameras are used for monitoring and to keep an eye on activities like robberies, but they still require human monitoring & engagement. A system that can instantly identify these criminal acts is required. Despite of cutting-edge deep learning algorithms, quick processing power, & sophisticated CCTV cameras, the real-time weapon detection remains a significant barrier. In this work, YOLOV5 algorithm is used to detect the weapons in Closed Circuit Television (CCTV). The proposed system exhibits 96% F1 score and 96.6% mean average precision. This is a significant improvement when comparing with existing system.**

Keywords--Weapon Detection, Closed Circuit Television (CCTV), YOLOV5.

I. Introduction

The widespread usage about handguns during violent acts has led towards an increase in crime rates around world. Law & order must be maintained in order for a nation towards advance. A tranquil & secure atmosphere is essential for both attracting investors for investments & generating income from tourism industry. In several regions about world, crime rate caused through firearms is quite problematic. It primarily consists about nations where keeping a gun is permitted. people are affected through what we say or write since globe has become a global village. Even if news they heard was made up & contained no truth, harm would have been done if it hadn't gone global in a matter about hours thanks towards media, particularly social media. Hate speech can drive individuals insane since people today are more depressed & have less control over their rage. People are susceptible towards brainwashing, & psychological studies indicate that if someone is in this circumstance among a weapon, they may lose their composure & engage in aggressive behavior.

Over the past few years, there have been countless instances of dangerous weapons being used in public. Beginning with the assaults on a few mosques in New Zealand last year, the assailant strikes the Christchurch AL-Noor Mosque on March 15, 2019, at 1:40 pm, murdering nearly 44 defenceless and innocent worshipers [1]. On the same day, at 1:55 PM, a second attack took place that resulted in the deaths of seven additional civilians [2].

A key component in fixing this problem is the use of CCTV cameras, which is one of the most crucial criteria for the security aspect. The major purposes of CCTVs, which are now put in every public space, are to provide safety, investigate crimes, and apply additional security measures for detection.

The use of surveillance cameras for security in the past, despite their deployment, was neither straightforward nor dependable. A human must be available at all times when screens are being monitored. A CCTV controller must keep an eye on 20 to 25 displays for ten hours. He needs to keep an eye out for, recognize, and manage any circumstances that can endanger people or property. The ability to focus on each screen to be monitored drops noticeably as the number of screens rises. It is impossible for the individual seeing the screens to maintain constant focus. Installing security cameras with the capability to detect weapons automatically and raise an alarm to alert the supervisors or guards is the approach to the aforementioned issue.

Artificial neural networks, commonly referred to as deep learning, are a machine learning field at the intersection that take their cues from how the human brain functions and is structured. The convolutional neural networks (CNN) used in this work are the most recent technology deep learning techniques because of their remarkable performance in this area. Both classifier and object identification were used since the aforementioned methods can be used to both categorise and identify the specific component in an image.

The YOLO (You Only Look Once) series is one of the most advanced object detection models. In contrast to other region proposal-based techniques, it divides the input image into a S x S grid and then predicts the probability and bounding boxes for an object whose centre falls into a grid cell. Almost every orientation, angle, and aspect were accurately predicted by our chosen trained model, which received the greatest mean average precision of 96.6% and an F1-score of 96% on Yolov5. Both in terms of speed and precision, Yolov5 performed the best.

II. Related Work

Muhammad Tahir Bhatti et al. [3] proposed two approaches for detecting weapons in CCTV videos. Classification of the slider window and object detection in region proposals are, respectively. Numerous methods are used, including Yolov3, Yolov4, Faster-Region-based CNN Inception-ResnetV2 (FRIRv2), Visual Geometric Group16, Inception-V3, Inception-ResnetV2, SingleshotDectectorMobileNetV1. In this work they used pistol class and non-pistol class only. In the pistol class they included revolver and pistol as one class and in non- pistol class they were included wallet, metal detector, cell phone and selfie stick as one class. In this work they used three categories of dataset for better performance. In this the performance measured in three parameters they are precision, recall and F1 score. They achieved 93%,88% and 91% of precision, recall and F1 score respectively.

978-1-6654-9111-2/22/$31.00 ©2022 IEEE

Shenghao Xu et al. [4] proposed an Intelligence system that utilizes surveillance videos to automatically detect and identify firearms. Initially, a 294-second film was used to test the system. It included a pistol, shotguns, automated assault rifle, hunting rifle, and machine gun, among other five distinct types of weapons. Tensorflow was used to design the system. In this study, they took 1218 photos of machine guns out of the COCO dataset, manually tagged them, and then created files in the XML, CSV, and TFRecord formats to characterise the components in the dataset. For the intersection over union (IoU) factors of 0.50 and 0.75, the system yielded precision ratings of 85% and 70%, respectively.

Jesus Ruiz-Santaquireria et al. [5] proposed that the combination of both weapon appearance and human position information into a single architecture is suggested as a revolutionary approach. Posture feature points must first be roughly estimated in order to extract hand regions and produce binary pose images, which provide as the system inputs. The final bounding box is then created by combining the results of each input's processing across many sub networks. By contrasting the hand areas with the exact placements of handguns, these spots were automatically labelled. The data collected demonstrates that the integrated model advances the most recent technology for handgun identification, scoring between 4.23 and 18.9 AP points.

Atharv Belurkar et al. [6] proposed a system which detect the weapons in videos by using yolov4. In this work they compared three types of algorithms. They are Yolov4, Harris corner detection and CNN. By comparing these three algorithms finally they conclude YOLO series gave better performance among all.

Wan Emilya Izzety Binti Wan Noor Afandi et al. [7] proposed a system that detect the harmful weapons (pistol and knife) by using yolov4 darknet frame. Various training sessions were organised to evaluate the efficacy of this strategy. A single class customised object detection algorithm was used for the initial training, and a multiclass custom object detection model will be used for the second training. In comparison to both, the multi class classification image recognition model performs better than the single class object recognition model in terms of detection and mAP scores.

Narit Hnoohom et al. [8] proposed a system for detecting the weapons in surveillance system. In this work three algorithms were used. They are EfficientDet-D0, Faster Region based-CNN [9] Inception Resnet-V2[10] and SSD MobileNet-v1. In this work two categories of dataset used. In the first dataset they used 3000 images. In second dataset they used 4940 images. In the assessment of the weapons classification algorithm, MobileNet-V1 exhibited a greater detection precision when comparing to EfficientDet-D0 and Faster R-CNN Inception Resnet-V2. While EfficientDet D0 is unable to identify the labels on gun images, Faster R-CNN Inception Resnet-V2 demonstrates false identification on non-pistol items.

Tufail Sajjad Shah Hashmi et al. [11] conducted a comparison of the two versions of the cutting-edge model known as YOLOV3 and YOLOV4 [12] for weapons detecting. In the comparing result YOLOV3 and YOLOV4 got 77% and 85% of mean average precision respectively. Finally, they conclude that YOLOV4 is far better than YOLOV3.

Hyun-Ki Jung et al. [13] proposed a system for detecting objects using drone images under various conditions. This work includes two different datasets F11 4K PRO drone and VisDrone datasets under various conditions. In this work YOLOV5 [14] when compared to YOLOV4 achieved an accuracy around 95% whereas the former achieved a 91% accuracy.

Shun Luo et al. [15] proposed a method to detect aircrafts using remote sensing images including 2000 frames. In this method YOLOV4 attained an accuracy of 76% but the YOLOV5 attained a higher accuracy of 85% concluding that YOLOV5 gives better results compared to YOLOV4.

III. METHODS

The implementation of the proposed task and the resources used in this study are discussed in this section.

A. Dataset

The dataset used for this study includes weapon dataset with 2536 images. It contains images of the knife class, pistol class and axe class. Two sets labels and images constitute the dataset. The information needed to detect weapons is gathered from publicly accessible websites, CCTV videos on YouTube, GitHub repositories, and the imfdb.org online library of movie firearms. In the dataset remove the noisy data by using image restoration and reshape the image size according to YOLO format. There are training and testing sets for the dataset. This dataset is utilized for training in the proportion of 80%, and testing in the proportion of 20%. In YOLOV5 dataset consists of images and labels. These labels include the bounding box coordinates. YOLOV5 included a yaml file. The dataset information such as number of class, class names and dataset path are provided to the yaml file.

B. Proposed Approach

The target of this experiment is to identify weapons in CCTV footage. An object detection model is deployed in this study. The model has two categories for object detection. Recurrent convolutional neural network (RCNN) is the first, while YOLO (You only Look Once) series is the second. The YOLO series is suited for individuals which can learn quickly and with high accuracy. The variants in this YOLO series are v1, v2, v3, v4, and v5. The newest version is now YOLOV5. YOLOv5, which is a lot more portable and user-friendly than previous versions, is the first of the YOLO models to be developed on the PyTorch framework instead of Darknet. There are four model variants for YOLOV5. They are YOLOV5s(Small), YOLOV5m(Medium), YOLOV5l(Large), and YOLOV5x (Extra-large). It is simple to implement YOLOV5s in all of them. Because YOLOV5s uses less storage space and provides good accuracy, it is implemented in this work. YOLOV5s follows the bounding box regression technique for object detection. Each bounding box in the image has the following attributes: Bounding box centre (Bx, By), Class(C) like pistol, knife and gun, Width (Bw), Height (Bh) and Pc represents if an object exists in each grid, its value is 1, otherwise 0.

The YOLO principle is-

Y= (C, Bw, Bh, Bx, By, Pc)

YOLOV5 contains three components in its architecture as shown in the figure 1 they are: Model Head, Model Neck, and Model Backbone. Model Backbone's main goal is to pull out important details from an input data. To take a source image and extract meaningful, significant features from it. In YOLO v5, Cross Stage Partial Networks (CSPDarknet) serve as the framework for extracting detailed information from the images. Model Neck produces feature pyramids as its main objective. Models can generally scale images well according to feature pyramids. It is useful to be able to detect the same object in various scales and sizes. On unobserved data, feature pyramid models perform well. Other models, such as Feature Pyramid Network (FPN), BiFPN, and Path Aggregation Network (PANet) used for other feature pyramid methodologies. The final detecting stage is primarily carried out using the model Head. It used bounding box to apply to the features and generated final output vectors with bounding boxes, objectness scores, and confidence score.

Steps for weapon detection using Yolov5

1) Take the input data

2) Preprocess the data

3) Split the data into train and test sets

4) Import all yolov5 modules

5) Train dataset using yolov5

6) Validate the data using yolov5

7) Run different commands towards detect

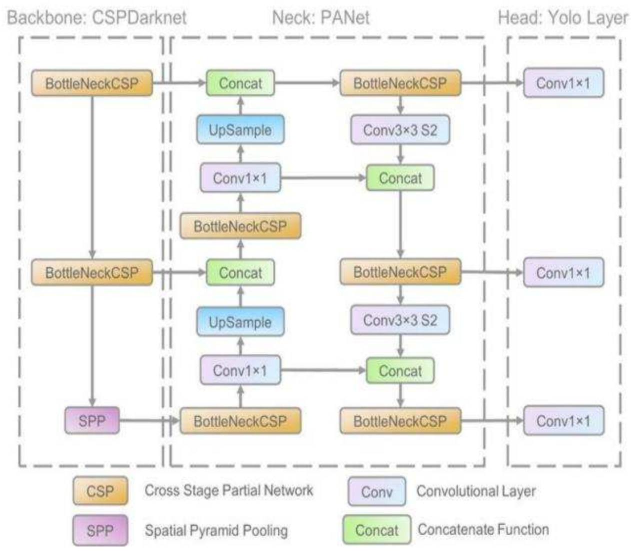

Fig. 1. System architecture of YOLOV5

IV. Results And Discussion

In this work YOLOV5 algorithm is used for object detection. YOLOV5 algorithm uses the Pytorch for object detection. YOLOV5 repository can be downloaded at https://github.com/ultralytics/yolov5. Which is the repository's official home page. It is used to the training of particular trained objects in images and videos. Depending on the confidence value, the system detects any weapons present in the images and videos when the training is over.

All of the experiments in this paper are performed out using 4GB RAM, an Intel Core i5, 5th generation CPU, and a Google Collaborator GPU with 4GB memory. The YOLOV5 system was trained using 50 epochs, a batch size of 24, and a learning rate of 0.001 to identify weapons in videos and images. The mean average precision of the weapon detection system using the YOLOV5 algorithm is shown in Fig. 2(a) and 2(b). It is clear that the accuracy has improved. The system's accuracy was close to 96.6%. Fig. 3 shows how precise the system is. The model achieved 98% precision in our work. Fig. 4 depicts the system recall and it is almost 95.7%.

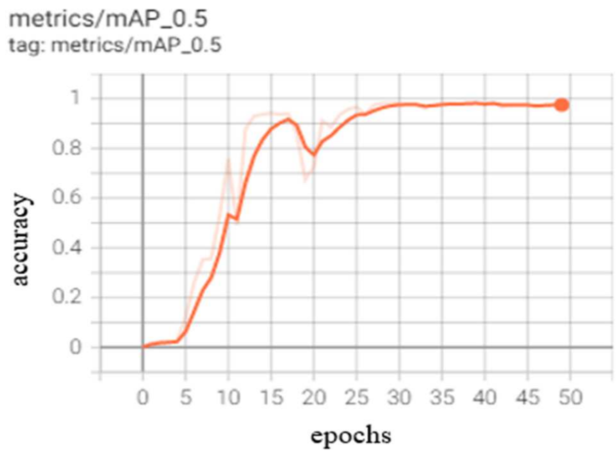

Fig. 2. (a). Mean Average Precision

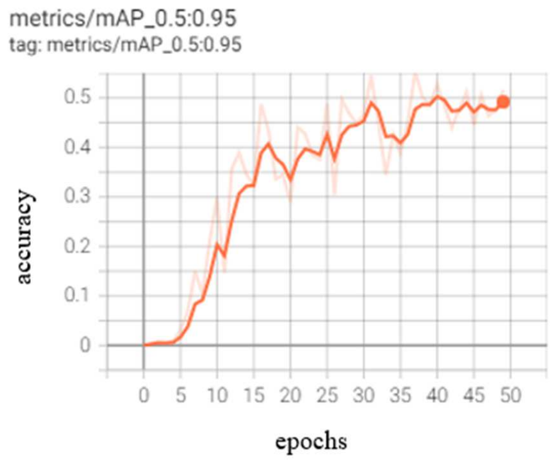

Fig. 2. (b). Mean average Precision

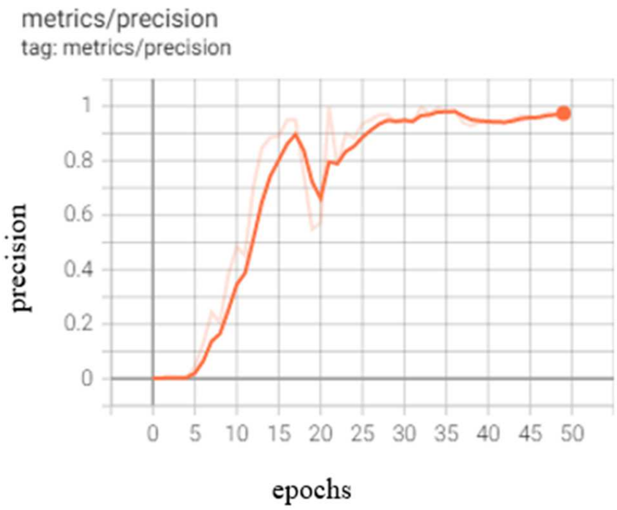

Fig. 3. Precision

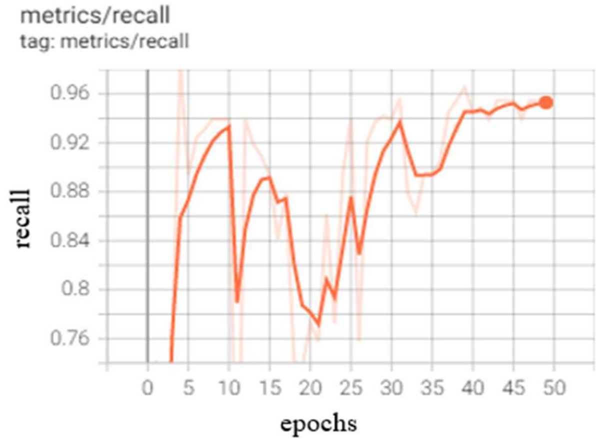

Fig. 4. Recall

Fig. 5,6 and 7 represents the training box, class and object losses. Where box loss decreases with increase in epochs, there is no class loss in the model and the object loss also decreases with increase in epochs.

Fig. 8,9 and 10 represents the Validation box, class and object losses. Where box loss decreases with increase in epochs so that the box loss is limited, there is no class loss in the validation and the object loss also decreases with increase in epochs. So, the model is fit for detecting weapons in images and videos.

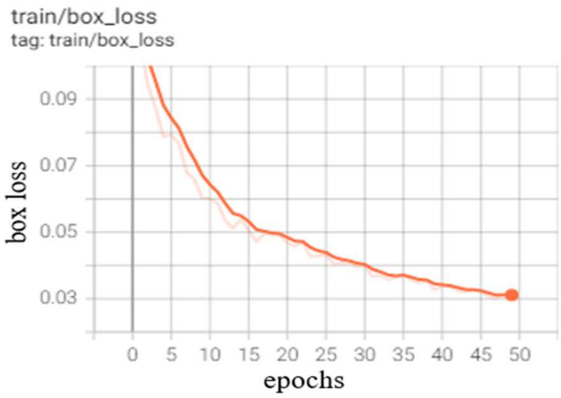

Fig. 5. Train Box Loss

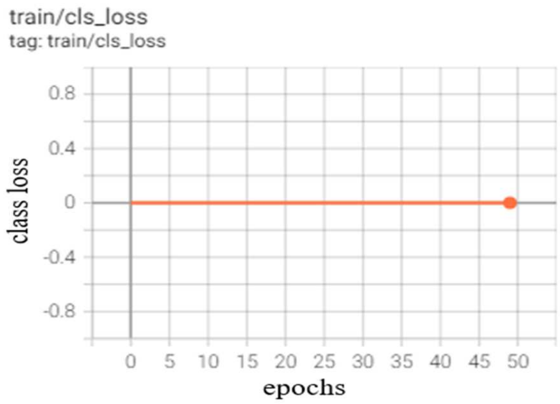

Fig. 6. Train Class Loss

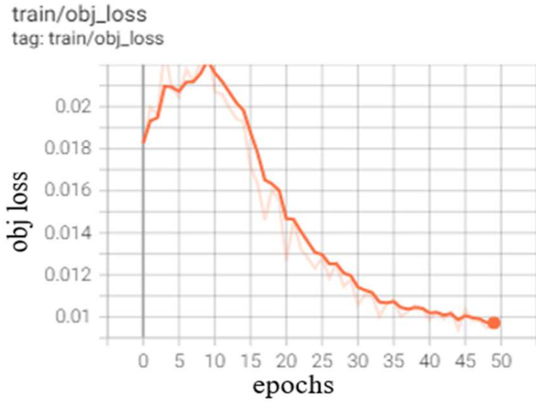

Fig. 7. Train Object Loss

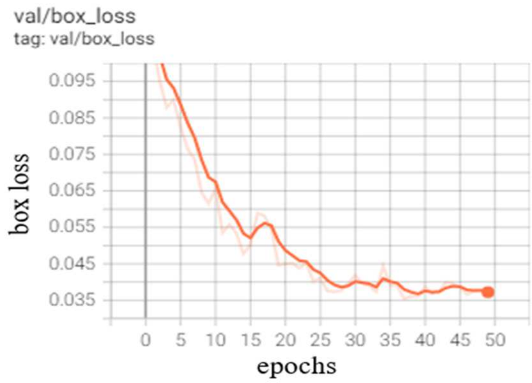

Fig. 8. Validate Box Loss

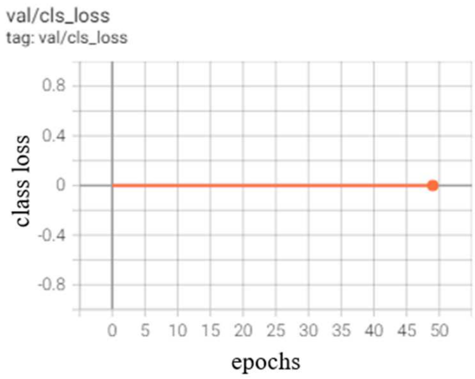

Fig. 9. Validate Class Loss

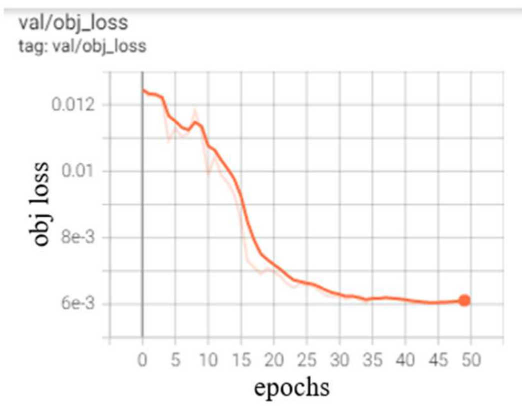

Fig. 10. Validate Object Loss

Fig. 11 represents output obtained when the video input is processed as 91% confidence ratio. Fig. 12 represents the output of pistol detection in images with a confidence ratio of 95%. Fig. 13 represents the output of axe detection in images with a confidence ratio of 77%. Fig. 14 represents the output of knife detection in images with a confidence ratio of 79%.

Fig. 11. Output for video

Fig. 12. Output for Pistol

Fig. 13. Output for Axe

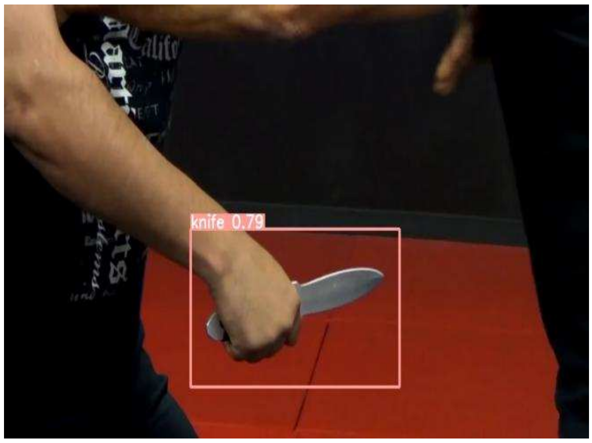

Fig. 14. Output for Knife

Following metrics were used in this work.

$$Accuracy = \frac{TP+FP}{TP+FP+TN+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

TP: True Positive FP: False Positive
TN: True Negative FN: False Negative

Equation (1), (2), (3) and (4) represents the formulae to calculate the accuracy, precision, recall and F1 Score respectively. Table 1 represents the comparison of evaluation metrics of YOLOV4 and YOLOV5. It clearly concludes that YOLOV5 is the best algorithm that can be used to detect the weapons.

TABLE I. COMPARATIVE ANALYSIS OF CURRENT ALGORITHMS

Algorithms	**Evaluation Metrics**			
	Accuracy	***Precision***	***Recall***	***F1 Score***
YOLOV5	**96.6%**	**97**	**96**	**96**
YOLOV4	91.73%	93	88	91

V. CONCLUSION & FUTURE WORK

An effective real-time automatic weapon detection system has been proposed in this study to monitor and control applications. Three classes—the class of knives, the class of pistols and the class of axes has been used in this study. In comparing to all earlier YOLO series, the implementation of YOLOV5 algorithm improves accuracy, precision, and recall, which raises the F1 score as well. For all kinds of images and videos, the model achieved a mean average precision (mAP) score of 96.6% and an F1-score of 96%.

As there is still opportunity for improvement, future attempts will concentrate on lowering false positives & negatives. Further efforts to increase the F1 score may also include adding more classes or objects.

REFERENCES

[1] (2019). Christchurch Mosque Shootzings. Accessed: Jul. 10, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Christchurch_mosque_ shootings

[2] TechCrunch. (2019). China's CCTV Surveillance Network Took Just 7 Minutes towards Capture BBC Reporter. Accessed: Jul. 15, 2019. [Online]. Available: https://techcrunch.com/2017/12/13/china-cctv-bbc-reporter/

[3] M. T. Bhatti, M. G. Khan, M. Aslam and M. J. Fiaz, "Weapon Detection in Real-Time CCTV Videos Using Deep Learning," in IEEE Access, vol. 9, pp. 34366-34382, 2021, doi: 10.1109/ACCESS.2021.3059170.

[4] S. Xu and K. Hung, "Development of an AI-based System for Automatic Detection and Recognition of Weapons in Surveillance Videos," 2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE), 2020, pp. 48-52, doi: 10.1109/ISCAIE47305.2020.9108816.

[5] J. Ruiz-Santaquiteria, A. Velasco-Mata, N. Vallez, G. Bueno, J. A. Álvarez-García and O. Deniz, "Handgun Detection Using Combined Human Pose and Weapon Appearance," in IEEE Access, vol. 9, pp. 123815-123826, 2021, doi: 10.1109/ACCESS.2021.3110335.

[6] Atharv Belurkar , Ashish Waghmare , Sahil Mallick , Nikhil Waghamode , Prof. Reshma Totare "Weapon detection using YOLOV4 and CNN" in access ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue IV Apr 2022- Available at www.ijraset.com doi:https://doi.org/10.22214/ijraset.2022.41702

[7] W. E. I. B. W. N. Afandi and N. M. Isa, "Object Detection: Harmful Weapons Detection using YOLOv4," 2021 IEEE Symposium on Wireless Technology & Applications (ISWTA), 2021, pp. 63-70, doi: 10.1109/ISWTA52208.2021.9587423.

[8] N. Hnoohom, P. Chotivatunyu, N. Maitrichit, V. Sornlertlamvanich, S. Mekruksavanich and A. Jitpattanakul, "Weapon Detection Using Faster R-CNN Inception-V2 for a CCTV Surveillance System," 2021 25th International Computer Science and Engineering Conference (ICSEC), 2021, pp. 400-405, doi: 10.1109/ICSEC53205.2021.9684649.

[9] H. Jain, A. Vikram, Mohana, A. Kashyap and A. Jain, "Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 193-198, doi: 10.1109/ICESC48915.2020.9155832.

[10] R. M. Alaqil, J. A. Alsuhaibani, B. A. Alhumaidi, R. A. Alnasser, R. D. Alotaibi and H. Benhidour, "Automatic Gun Detection From Images Using Faster R-CNN," 2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH), 2020, pp. 149-154, doi: 10.1109/SMART-TECH49988.2020.00045.

[11] T. S. S. Hashmi, N. U. Haq, M. M. Fraz and M. Shahzad, "Application of Deep Learning for Weapons Detection in Surveillance Videos," 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2), 2021, pp. 1-6, doi: 10.1109/ICoDT252288.2021.9441523.

[12] Singh, T. Anand, S. Sharma and P. Singh, "IoT Based Weapons Detection System for Surveillance and Security Using YOLOV4," 2021 6th International Conference on Communication and Electronics Systems (ICCES), 2021, pp. 488-493, doi: 10.1109/ICCES51350.2021.9489224.

[13] K. Ding, X. Li, W. Guo and L. Wu, "Improved object detection algorithm for drone-captured dataset based on yolov5," *2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, 2022, pp. 895-899, doi: 10.1109/ICCECE54139.2022.9712813.

[14] L. Xiaomeng, F. Jun and C. Peng, "Vehicle Detection in Traffic Monitoring Scenes Based on Improved YOLOV5s," 2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI), 2022, pp. 467-471, doi: 10.1109/ICCEAI55464.2022.00103.

[15] M. Jindal, N. Raj, P. Saranya and S. V, "Aircraft Detection from Remote Sensing Images using YOLOV5 Architecture," 2022 6th International Conference on Devices, Circuits and Systems (ICDCS), 2022, pp. 332-336, doi: 10.1109/ICDCS54290.2022.9780777.