Deep Learning Models for Crime Intention Detection Using Object Detection

Abdirahman Osman Hashi[1*], Abdullahi Ahmed Abdirahman[2*], Mohamed Abdirahman Elmi[3*], Octavio Ernest Romo Rodriguez[4]
Faculty Member, Department of Computing, SIMAD University, Mogadishu Somalia[1, 2, 3]
Department of Computer Science-Faculty of Informatics, İstanbul Teknik Üniversitesi, İstanbul, Turkey[4]

Abstract—**The majority of visual based surveillance applications and security systems heavily rely on object detection, which serves as a critical module. In the context of crime scene analysis, images and videos play an essential role in capturing visual documentation of a particular scene. By detecting objects associated with a specific crime, police officers are able to reconstruct a scene for subsequent analysis. Nevertheless, the task of identifying objects of interest can be highly arduous for law enforcement agencies, mainly because of the massive amount of data that must be processed. Hence, the main objective of this paper is to propose a DL-based model for detecting tracked objects such as handheld firearms and informing the authority about the threat before the incident happens. We have applied VGG-19, ResNet, and GoogleNet as our deep learning models. The experiment result shows that ResNet50 has achieved the highest average accuracy of 0.92% compared to VGG19 and GoogleNet, which have achieved 0.91% and 0.89%, respectively. Also, YOLOv6 has achieved the highest MAP and inference speed compared to the faster R-CNN.**

Keywords—Object detection; deep learning; crime scenes; video surveillance; convolutional neural network; YOLOv6

I. Introduction

Due to the rising crime rate and offensive activities in recent times, it has become common to find CCTV cameras installed in public places such as shopping centers, avenues, banks, and so on. The purpose of these cameras is to enhance security and ensure the safety of the people in these areas. However, detecting weapons in surveillance videos still requires a lot of human intervention, which can result in errors [1]. Meanwhile, it is difficult for humans to constantly observe long videos or maintain multiple footage, and this can result in some rare crime scenes being missed out. As a result, there is a need to explore new technologies that can help improve the accuracy and efficiency of video surveillance in public places [2].

Recent surveillance cameras have the capability to record video only when motion is detected, unlike the older versions that record continuously irrespective of any activity. This feature enhances the efficiency of the system as it reduces processing time, search time, and the storage space required for the recorded videos [3,5]. Unfortunately, law enforcement agencies often follow a reactive approach, which results in delayed response times during crime incidents. In this approach, authorities rely on witness reports or CCTV footage to analyze the crime after it has occurred. This means when an incident takes place, investigators visit the site, manually retrieve the footage from the camera, and then try to locate the appropriate footage either by watching the entire video or using advanced algorithms to process it which takes a long time. For this case, to improve the efficiency of the security management system and minimize crime incidents and losses, an effective crime prediction analysis system is needed. Such a system would enable proactive crime prevention and ensure robust security management in public places such as banks, shopping malls, and avenues [3,6].

With the availability of large datasets, faster GPUs, enhanced machine-learning algorithms, and better computations, we can now efficiently prepare PCs and construct automated computer-based systems to differentiate and identify various things on a site with high accuracy. For instance, a remote embedded intelligent security monitoring system has been developed using computer vision modeling algorithms to proactively detect intruders. This system utilizes a camera to acquire background images, which are then modeled using the ViBe algorithm to perform object detection in the monitored area [7]. When a moving object, including humans, is identified, the system automatically triggers an alarm and sends a message or call to the user to take preventive measures. To get a better understanding of the situation and detect the intruder, users can log in to the server via a mobile application. The system was implemented on an ARM development board, which provides a platform for hardware and software development. This technology is useful in enhancing security and safety in public places such as shopping malls, airports, and banks, where quick detection of intruders can help prevent criminal activities [3].

In contemporary times, machine-learning and advanced image-processing algorithms have significantly contributed to the evolution of smart surveillance and security systems, as evidenced by recent developments [3,6,7,8]. In addition, the rise of smart devices and networked cameras has also boosted this field. However, detecting and tracking human objects or weapons still require cloud centers as real-time, online tracking is computationally expensive. Recent years have seen significant efforts in monitoring robot manipulators, which require high control performance in terms of reliability and speed [9,10].

Another approach to detecting guns in surveillance films is to use pre-trained deep learning models. These models are intended to assist users in learning about algorithms or experimenting with current frameworks for better outcomes without explicit design. A deep learning neural network generally has five layers: input and output layers with Convolution, Max-Pooling, and Fully connected layers. Many individuals choose to employ pre-trained deep learning models due to constraints such as limited time, memory, and resources such as CPU and processors [6,35]. When opposed to machine learning, which involves explicit design, these pre-trained models produce better and more accurate outcomes. However, identifying firearms in surveillance films is difficult and subject to human mistake unless it is used a detection system. For instance, human guards may become fatigued or fall asleep when viewing huge volumes of recordings or maintaining several footages, resulting in missed opportunities to discover uncommon criminal intention scenarios that may be caught in many footages. To solve this issue, pre-trained deep learning models may be used to eliminate the need for human interaction while identifying possible threats in public venues [11,34].

For that reason, it is important to design an autonomous surveillance system capable of detecting firearms fast and accurately in order to prevent crime. Deep learning techniques play an essential role here. Hence, we are developing a system which takes advantage of pre-trained deep learning models that are VGG-19 and ResNet50 to detect the firearms with object detection. These models were chosen because recent object recognition models need a large number of parameters and a significant amount of time to train. VGG19 and ResNet50 can extract high-level feature maps from input, reducing its complexity. The Faster RCNN method is also used to build bounding boxes around objects in pictures. The objective is to use neural networks to identify anomalies such as weapons and firearms and determine whether or not the individual carrying them has a criminal intent.

The rest of the paper is organized as follows. The next section will provide some context for our issue and highlight pertinent related works. Section III describes our proposed technique. Section IV describes the experiments and the outcomes. Finally, Section V summarizes the work and offers some perspectives.

II. Related Work

The detection of metallic items that may represent a threat to public and homeland security is a major global concern [1]. Yet, screening for these devices might be difficult since it disrupts people's movements and creates a susceptible target for terrorist strikes [3,9]. In light of this, an automatic metallic item identification and categorization system is proposed. Two related areas are addressed in order to create and implement such a system; the development of a new metallic object detecting system and the establishment of a signal processing method to identify targeted signatures. The suggested approach is assessed by creating a database comprising of actual pistols and everyday items. Extraction of characteristics from four categories is used to examine system outcomes: time-frequency signal analysis, material composition, object form, and transient pulse response. Then, two categorization approaches are used to differentiate between hazardous and non-threatening things. The new system's feature combining and classification framework achieves a successful classification rate of more than 90% [8,12].

A. Handgun and Knife Detection in CCTV

Another method is to use (CCTV). The use of Closed-Circuit Television (CCTV) systems has become increasingly common in various settings, including offices, residential areas, and public spaces, and has been implemented in many countries. To enhance the effectiveness of these systems, image segmentation techniques are employed to track activities captured by CCTV cameras and apply machine learning algorithms [1,10]. Grega [13] for example, published an algorithm that recognizes knives and guns in CCTV images and warns the security guard or operator. The algorithm's specificity is 94.93% and sensitivity is 81.18% for knife detection, focused on reducing false alarms and delivering a real-time application. Furthermore, the specificity for the fire alarm system is 96.69% and the sensitivity is 35.98% for the various items in the movie. Author [14] developed a video classifier, also known as the Histogram of Directed Tracklets, that recognizes irregular circumstances in complicated sequences. In contrast to typical optical flow techniques that only assess edge characteristics from two consecutive frames, descriptors known as tracklets have been evolving across long-range motion projections. Spatiotemporal cuboid film sequences are statistically gathered on the tracklets that travel across them [15].

B. Automatic Hanggun Detection using Machine Learning

Although there has been a recent advance in image-based machine learning, recognizing a knife-wielding assailant remains difficult. To address this issue, the authors [16] describe three approaches for automated threat detection utilizing various knife image datasets, with the goal of narrowing down plausible assault aims while decreasing false negatives and false positives. To begin, they employ a classification model based on Mobile Net in a sparse and pruned neural network that can notify an observer to the presence of a knife-wielding attacker with high accuracy (95%) and a low memory demand (2.2 MB). Second, they train a detection method (Mask RCNN) to segregate the hand from the knife in a single picture and give probable certainty to their relative placement, allowing for both bounding box localization and point threat inference. Finally, a Pose Net-based model assigns anatomical waypoints to narrow down threat features and decrease misconceptions of the attacker's objectives [4,17,18]. Furthermore, the authors identify and fix data gaps, such as the necessity to gather benign hands, which may impair the accuracy of the deployed knife threat detector. This study offers a thorough review of image-based warnings that may be used to prioritize and educate crime prevention strategies before any catastrophic results occur. Additional relevant study topics in this subject include, among others, Automated Handgun Detection Alarms in Videos Using Deep Learning, Automatic Visual Recognition of Armed Robbery, and Robust Item Detector Application for Visual Knife Detection [19,35].

Other authors proposed by analyzing the recorded films of the cameras. The author [8] describes a technology for automatically detecting handguns in surveillance films. By recognizing weapons in films, categorizing items as either a gun or not, and forecasting whether a crime has happened, the system hopes to control occurrences of crime. The system's performance is compared to a sliding window proposal technique, and it is discovered that FRCNN and RCNN-based models trained on a dataset perform well. The algorithm can reliably anticipate crime occurrences even in low-quality films, yielding good results [5,20,31].

The author [21] details a conventional approach for identifying the position of an armed robber. The method focuses on detecting individuals who are holding a knife in various positions relative to other people. To accomplish this, the system uses skeleton silhouette algorithms that segment the body into distinct parts and identify the position of a raised arm holding a knife at different angles. Through this process, the system is able to successfully detect the presence of a knife.

The author [22] describes a visual method for detecting automated weapons, specifically knives held in hands. This approach utilizes novel object detection algorithms to identify visual knives within a given video dataset. One of the primary challenges is detecting knife rotations at varying scales and positions, which can be difficult due to the multitude of possible orientations and positions of the knife in the dataset. To address this, the system is designed to detect all possible knife orientations and positions. Feature extraction is accomplished using foreground segmentation and FAST (Features from Accelerated Segment Test) for feature detection. Classification is then performed using MRA (Multi-Resolution Analysis).

C. Automatic Hanggun Detection using Machine Learning

Meanwhile, Convolutional Neural Networks (CNNs) have shown exceptional performance in image processing and object recognition during the last few years. CNNs are a sort of neural network that is specifically intended to recognize pictures [23,24]. These networks are made up of numerous layers, each having its own function. The first (input) layer receives an image as input. The next layer (the convolution layer) applies a collection of filters to the input picture, which are themselves tiny images. This layer takes characteristics from the input picture and extracts them. The following layer (the pooling layer) decreases the previous layer's output by pooling together all the pixels in a fixed-size square of the input picture. This layer reduces the number of parameters and makes the network more error-resistant [16,25]. CNN is trained on a large dataset of pictures containing the items to be recognized in order for the network to learn the characteristics that identify each object and correlate them with a given class. After trained, the network may be used to recognize the required items in new photos [26]. CNNs have demonstrated considerable potential for a variety of applications, including self-driving vehicles, face identification, crime detection, internet of things-based photovoltaics monitoring, and even COVID-19 detection, because to their capacity to identify a wide spectrum of objects and their error tolerance. The upcoming Fig. 1 elaborates a CNN model that has applied for a weapon detection obtained [9].

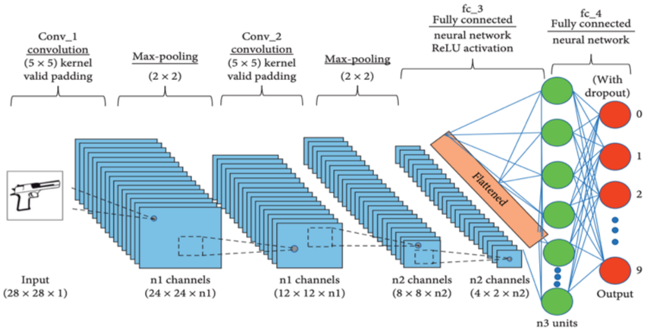


Fig. 1. Feedforward Convolutional Neural Network (CNN).

The cutting-edge YOLO V3 object identification model was applied and trained for obtained dataset for weapon detection in this work. Authors suggest a concept that gives a machine or robot a visionary sense to recognize dangerous weapons and can also inform a human administrator when a gun or a firearm is seen in the edge. +e experimental data demonstrate that the trained YOLO V3 model outperforms the YOLO V2 model and is less computationally costly. There is an immediate need to improve the present surveillance capabilities by providing better resources to enable monitoring the efficacy of human operators [27,32,33].

The phrase "backbones"[29] in the realm of object detection refers to the parameters or weights used to construct feature maps. These backbones are an essential component of the feature extractor since they generate the features that will be utilized for object detection [28]. There are several backbones that may be used for this purpose, each with its own set of benefits and downsides. The visual geometry group (VGG), (ResNet) residual neural network are among the most often used deep learning convolutional neural network (CNN) backbones in object detection techniques [30]. In terms of speed, efficiency, and accuracy, each of these designs offers various trade-offs. However, the proposed model will be applied VGG, ResNet and GoogleNet. The algorithms which we propose are able to detect the human operator when a gun is visible in the image.

III. Proposed Model

The aim of this study is to address the challenge of identifying indications of automated criminal intention and identifying hazardous circumstances using closed-circuit television (CCTV) systems. The primary objective of this research is to expedite the identification of weapons with improved precision and decreased false positives when compared to machine learning techniques, while also ensuring that convolutional neural networks (CNNs) maintain performance efficiency with fewer training samples. Pre-trained models such as GoogleNet and VGGNet-19 have been trained using millions of photographs, and possess the capability to recognize objects in new images with minimal errors. Owing to their superior training accuracy, we have opted to utilize the VGGNet19, GoogleNet, and ResNet50 models to effectively categorize and recognize objects.

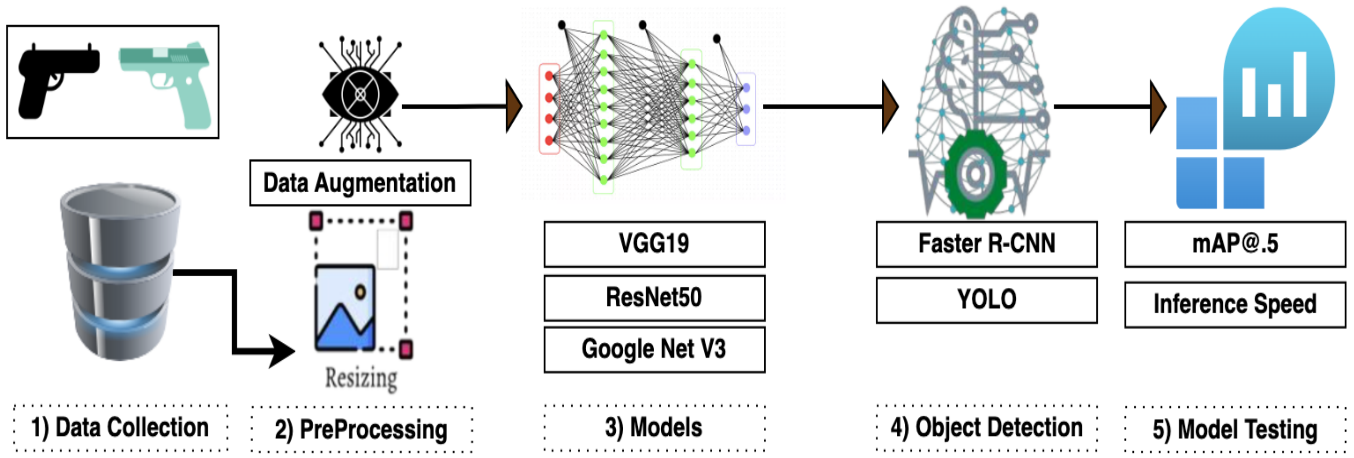


Fig. 2. Proposed model.

The diagram presented in Fig. 2 illustrates the all-encompassing design of the system under consideration. At the initial stage, input frames are received through the input layer, which is also responsible for conducting pre-processing activities as data augmentation. After undergoing pre-processing, the images are transferred through various layers, such as Convolution, Max-pooling, and FC layers, which perform a range of operations including feature extraction, feature filtering, feature mapping, and classification. And, the object detection layer is responsible for detecting objects and classifying, and in the event of any detected criminal intentions, it utilizes a registered API to dispatch a security message after model testing. The explanation of each step will be explained in a detailed way.

The first step involves gathering a comprehensive collection of positive firearm images that depict guns of various sizes, angles, and colors. These images are then segregated into designated "weapons" folders. Additionally, negative images resembling guns are also compiled and stored in separate folders labeled "not weapons". The open pictures dataset V6, a widely utilized dataset containing over nine million photographs, is used to source the data. This dataset features several images of firearms and bladed weapons that can be used to train machine learning models to effectively identify criminal activity in images. For this purpose, three categories were extracted from the dataset, namely rifle (2072 images), handgun (607 photos), and shotgun (476 images), although some researchers have extracted six categories, we only extracted three categories. Subsequently, the three groups were forwarded to the pre-processing data level.

The second stage involved data pre-processing, which refers to the preparatory procedures that must be performed on images prior to analysis. This process may involve a range of activities, such as scaling or adjusting the image display. One of the critical tasks in data preparation is resizing the image to 256 x 256 pixels. This entails feeding input frames into the Input layer, which performs pre-processing operations on images of varying sizes (such as 256 x 256 width and height) and converts them to a standardized size of 224 x 224 x 3 (RGB values) by extracting RGB values from pixels. This guarantees that all data has a uniform size and can be easily compared. Furthermore, it simplifies working with images that are not big which can be advantageous for training computers to detect objects. Meanwhile, another crucial aspect of data preparation is data augmentation, which can be achieved through various methods such as flipping, rotating, or scaling the image. This technique aims to increase the quantity of data available for training and enhance the system's ability to recognize objects from various perspectives. Effective data preparation is a crucial stage in object detection, as it can determine the success or failure of the system. Therefore, by resizing data and augmenting it, we can increase the likelihood of achieving desirable outcomes.

In the third stage, it involved fitting three object detection models, namely VGG 19, ResNet, and GoogleNet. Each of these models has its own set of advantages and drawbacks, making it essential to conduct a comparative analysis to determine the most suitable model for our specific data. For our dataset, we employed transfer-learning to fine-tune the models. Ultimately, it should be noted that DarkNet is undoubtedly the foundational model for YOLO.

In the fourth step, the object detection model was trained using the most optimal backbone model on both the train and validation sets, comprising 70% and 20% respectively. This facilitated the development of a highly accurate and dependable model, capable of recognizing a diverse range of objects. The model provided a robust foundation on which to build the model, ensuring its precision and reliability. The R-CNN and YOLO v6 object detection models were utilized in this study. Faster-RCNN was preferred over R-CNN and Fast R-CNN due to its superior accuracy and efficiency in object detection, enabling it to process complex images and capture intricate details. The latest version of YOLO, YOLOv6, was found to be the most advanced and user-friendly option, offering higher speed and accuracy compared to previous versions. Labeled data, specifically bounding boxes, were employed to train the object detection algorithms. These bounding boxes were used to define the location of objects in the images, while the labels provided information about the type of objects.

The final stage involved model inference, wherein the object detection models were tested on a separate testing set comprising 10% of the data. This step was crucial in evaluating the accuracy and inference time of each model, as has been done in prior research. Testing the models on previously unseen data allowed us to determine their efficacy in practical scenarios.

IV. Results and Discussions

In this section, we will explain the dataset description and elaborate different categories that we extracted from the dataset. Moreover, we will discuss and compare the output result of the three models in term of the accuracy, recall and F1-score. Also, we will present the detection output and finally discuss the comparative analyse as a benchmark.

A. Dateset Description

Open Images dataset is a collection of nine million images that have been annotated with image-level names, object bounding boxes, object segmentation masks, visual connections, and localized narrative. It has the biggest existing dataset with object position annotations, with 16M bounding boxes for 600 item types on 1.9M photos. Professional annotators drew the boxes mostly by hand to guarantee accuracy and uniformity. The images are quite varied and frequently feature complicated scenarios with several items. For our models, we had extracted only three categories (as seen in Table I).

B. Identify Comparing VGG19, GoogleNet and ResNet50

As we mentioned in the methodology, we have applied three distinct algorithms and the upcoming tables will illustrate their performance. The applied algorithms were VGG19, ResNet50 and GoogleNet and their accuracy, loss, precision, and recall will be discussing in the upcoming tables. Notably, ResNet50 algorithm performed remarkably well in the classification task compared to the VGG19 and GoogleNet, achieving accuracy scores of 0.92%, 0.91% and 0.89%, respectively.

Although VGG 19 demonstrated marginally superior F1-score compared to GoogleNet, it was also associated with a lower support score. ResNet50, on the other hand, surpassed both VGG 19 and GoogleNet algorithms, yielding an accuracy score of 0.92%, along with higher F1-score (0.94%) and recall (0.91%) scores, and comparatively higher support (45%). GoogleNet exhibited the poorest performance amongst all three algorithms, with the lowest accuracy score of 0.89% and the weakest performance across all other metrics.

The upcoming Table II shows the training accuracy of the VGG19. It can be seen that the overall average of the accuracy is 0.91%, yet it has also scored a good performance in term of F1-score by achieving 0.93%.

The next table which is Table III shows the training accuracy of the ResNet50. It can be seen that the overall average of the accuracy is 0.92%, and it outperformed compared to other algorithms which we have also applied the same with this dataset. In term of F1-score, it has also outperformed other algorithms by achieving 0.94% in the average total.

The next table which is Table IV demonstrates the training accuracy of the GoogleNet. It can also see that the overall average of the accuracy is 0.89%, this makes the poorest performance that has been achieved compared to other algorithms. In terms of F1-score, it has achieved a good accuracy but yet it is still lower than VGG19 which also makes the lowest performance.

On the other hand, the upcoming Table V demonstrates the mAP and inference-speed results of two object identification methods, namely Faster CNN and YOLOv6, which are known for their fast-processing times. The mAP is a crucial metric for evaluating object detection models, representing the (AP) of each object class. The ratio of (TP) to the sum of TP and (FP) is computed to determine the AP (FP). The mAP is a measure of an object detector's ability to accurately identify and differentiate different types of objects in an image. The symbol @0.5 implies that an intersection over union (IoU) threshold of 0.5 was used. IoU is a measure of how well a predicted bounding box or mask aligns with the ground truth data. Inference speed, measured in frames per second (FPS), indicates how many frames the algorithm can process per second.

TABLE I. Dateset Description

	Extracted categories	
0	Handgun	607 Images
1	Rifle	2072 Images
2	Shotgun	467 Images

TABLE II. Traning Accuracy of VGG19

Accuracy 100%	***Training Accuracy of VGG19***			
	Accuracy	***Recall***	***F1-score***	***Support***
0	0.90	0.88	0.95	43
1	0.91	0.92	0.93	112
2	0.93	0.91	0.93	163
Avg/Total	0.91	0.90	0.93	106

TABLE III. Traning Accuracy of Resnet50

Accuracy 100%	***Training Accuracy of ResNet50***			
	Accuracy	***Recall***	***F1-score***	***Support***
0	0.92	0.91	0.96	45
1	0.91	0.90	0.94	108
2	0.94	0.93	0.92	170
Avg/Total	**0.92**	0.91	**0.94**	107

TABLE IV. Traning Accuracy of Googlenet

Accuracy 100%	***Training Accuracy of GoogleNet***			
	Accuracy	***Recall***	***F1-score***	***Support***
0	0.88	0.90	0.92	26
1	0.90	0.89	0.91	94
2	0.89	0.91	0.90	160
Avg/Total	0.89	0.90	0.91	93

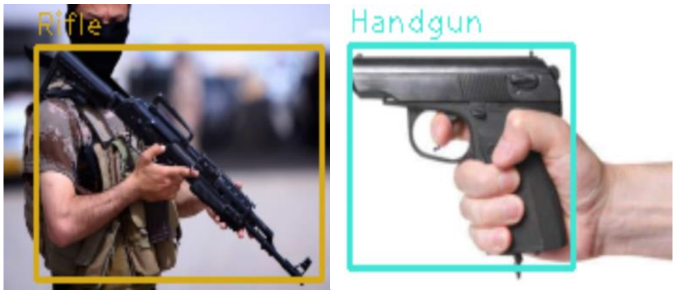
TABLE V. RESULT OF MODEL DETECTION

Algorithm	*MAP*	*Inference-speed*
Faster R-CNN	61.82	8
YOLOv6	**63.12**	68

The two algorithms exhibit different mAP scores and inference-speeds. (R-CNN), which is faster than the other method, achieved the highest score of 63.12%. However, it has the slowest inference speed of 8 frames per second (FPS). Conversely, YOLOv6 obtained the highest mAP@.5 score of 63.12%, but it has the slowest inference speed of 68 FPS.

C. Output of Gun Detection Result

The upcoming Fig. 3 demonstrates the output of the models for detecting the gun and classifying based on the object detected. Though we have not applied to send notification through API but it is important to make a separate comparing by which gun is detected.



(a) (b)

Fig. 3. (a) Rifle detection. (b) Handgun detection.

D. Comparative Analysis

Previous studies in this field have applied a range of models to detect firearms, as mentioned earlier. For instance, Author [1] developed a system utilizing VGG-19 to identify a weapon in a person's hand pointing at someone else, while Faster RCNN was used to draw bounding boxes around objects in images, resulting in an accuracy of over 80%. In comparison, our VGG19 model achieved an accuracy of 0.91%. It's worth mentioning that other researchers [4] used YOLOv5 and achieved a MAP of 56.92 and an inference speed of 61, but our model, utilizing YOLOv6, achieved a higher MAP of 63.12 and a faster inference speed of 68.

V. CONCLUSION

Deep learning object detection systems have the capability to offer significant benefits to officers and security professionals, as they can help in efficient way and not consuming much resources for forensic tasks. However, deploying artificial intelligence (AI) in this field raises concerns about possible misuse by law enforcement organizations, such as accusing innocent people or detecting bogus offenders. Our presented algorithms are capable of alerting human operators when a gun is detected in an image. In this study, we developed and assessed a system using a dataset of photos and videos obtained from the open images dataset V6, which contains over nine million images. Our results demonstrate that an early warning system in risky situations could lead to quicker response times, more efficient reaction times, and fewer potential casualties. Additionally, the proposed method outperforms other known approaches to crime detection. For further improvement, it would be novelty if it has incorporated other modalities such as audio or text data and this could improve the accuracy of the models.

REFERENCES

[1] Divya, S. M., Priya, G. S., Abitha, R., Sirisha, K., Manikanta, A., & Jayanth, K. AUTOMATED CRIME INTENTION DETECTION USING DEEP LEARNING.

[2] Sultana, T., & Wahid, K. A. (2019). IoT-guard: Event-driven fog-based video surveillance system for real-time security management. IEEE Access, 7, 134881-134894.

[3] Navalgund, U. V., & Priyadharshini, K. (2018, December). Crime intention detection system using deep learning. In 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET) (pp. 1-6). IEEE.

[4] Boukabous, M., & Azizi, M. (2023). Image and video-based crime prediction using object detection and deep learning. Bulletin of Electrical Engineering and Informatics, 12(3), 1630-1638.

[5] Kaushik, H., Kumar, T., & Bhalla, K. (2022). iSecureHome: A deep fusion framework for surveillance of smart homes using real-time emotion recognition. Applied Soft Computing, 122, 108788.

[6] Mathur, R., Chintala, T., & Rajeswari, D. (2022, January). Detecting criminal activities and promoting safety using deep learning. In 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (pp. 1-8). IEEE.

[7] Ahmed, S., Bhatti, M. T., Khan, M. G., Lövström, B., & Shahid, M. (2022). Development and optimization of deep learning models for weapon detection in surveillance videos. Applied Sciences, 12(12), 5772.

[8] Venkatesh, S. V., Anand, A. P., Gokul Sahar, S., Ramakrishnan, A., & Vijayaraghavan, V. (2020). Real-time Surveillance based Crime Detection for Edge Devices. In VISIGRAPP (4: VISAPP) (pp. 801-809).

[9] Narejo, S., Pandey, B., Esenarro Vargas, D., Rodriguez, C., & Anjum, M. R. (2021). Weapon detection using YOLO V3 for smart surveillance system. Mathematical Problems in Engineering, 2021, 1-9.

[10] Qin, Z., Liu, H., Song, B., Alazab, M., & Kumar, P. M. (2021). Detecting and preventing criminal activities in shopping malls using massive video surveillance based on deep learning models. Annals of Operations Research, 1-18.

[11] Shah, N., Bhagat, N., & Shah, M. (2021). Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. Visual Computing for Industry, Biomedicine, and Art, 4, 1-14.

[12] Kaya, V., Tuncer, S., & Baran, A. (2021). Detection and classification of different weapon types using deep learning. Applied Sciences, 11(16), 7535.

[13] Arunnehru, J. (2021). Deep learning-based real-world object detection and improved anomaly detection for surveillance videos. Materials Today: Proceedings.

[14] Grega, M., Łach, S., & Sieradzki, R. (2013, February). Automated recognition of firearms in surveillance video. In 2013 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA) (pp. 45-50). IEEE.

[15] Sung, C. S., & Park, J. Y. (2021). Design of an intelligent video surveillance system for crime prevention: applying deep learning technology. Multimedia Tools and Applications, 1-13.

[16] Mehta, P., Kumar, A., & Bhattacharjee, S. (2020, July). Fire and gun violence based anomaly detection system using deep neural networks. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 199-204). IEEE.

[17] Hussain, S. A., & Al Balushi, A. S. A. (2020). A real time face emotion classification and recognition using deep learning model. In Journal of physics: Conference series (Vol. 1432, No. 1, p. 012087). IOP Publishing.

[18] Amrutha, C. V., Jyotsna, C., & Amudha, J. (2020, March). Deep learning approach for suspicious activity detection from surveillance video. In 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (pp. 335-339). IEEE.

[19] Kaliappan, J., Shreyansh, J., & Singamsetti, M. S. (2019, March). Surveillance Camera using Face Recognition for automatic Attendance feeder and Energy conservation in classroom. In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN) (pp. 1-5). IEEE.

[20] Shirsat, S., Naik, A., Tamse, D., Yadav, J., Shetgaonkar, P., & Aswale, S. (2019, March). Proposed system for criminal detection and recognition on CCTV data using cloud and machine learning. In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN) (pp. 1-6). IEEE.

[21] M. Grega, A. Matiolanski, P. Guzik, and M. Leszczuk, “Au- ´ tomated detection of firearms and knives in a CCTV image,” Sensors, vol. 16, no. 1, p. 47, 2016.

[22] Verma, G. K., & Dhillon, A. (2017, November). A handheld gun detection using faster r-cnn deep learning. In Proceedings of the 7th international conference on computer and communication technology (pp. 84-88).

[23] Nagayama, I., Miyahara, A., & Shimabukuro, K. (2019). A study on intelligent security camera system based on sequential motion recognition by using deep learning. Electronics and Communications in Japan, 102(11), 25-32.

[24] H. Mousavi, S. Mohammadi, A. Perina, R. Chellali, and V. Murino, “Analyzing tracklets for the detection of abnormal crowd behavior,” in Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision, pp. 148–155, IEEE, Waikoloa, HI, USA, January 2015.

[25] Damashek, A., & Doherty, J. (2015). Detecting guns using parametric edge matching. Tech. Rep.

[26] Glowacz, A., Kmieć, M., & Dziech, A. (2015). Visual detection of knives in security applications using active appearance models. Multimedia Tools and Applications, 74, 4253-4267.

[27] Alqubaa, A., & Tian, G. Y. (2012). Weapon detection and classification based on time-frequency analysis of electromagnetic transient images. International Journal of Advances in Systems and Measurements, 3(3).

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

[29] Tiwari, R. K., & Verma, G. K. (2015, January). A computer vision based framework for visual gun detection using SURF. In 2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO) (pp. 1-5). IEEE.

[30] Alajrami, E., Tabash, H., Singer, Y., & El Astal, M. T. (2019, October). On using AI-based human identification in improving surveillance system efficiency. In 2019 International Conference on Promising Electronic Technologies (ICPET) (pp. 91-95). IEEE.

[31] Landa, J., Jun, C., & Jun, M. (2017, January). Implementation of a Remote Real-Time Surveillance Security System for Intruder Detection. In 2017 9th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) (pp. 102-105). IEEE.

[32] Sivakumar, P. (2021, July). Real Time Crime Detection Using Deep Learning Algorithm. In 2021 International Conference on System, Computation, Automation and Networking (ICSCAN) (pp. 1-5). IEEE.

[33] Patel, K., & Patel, M. (2021, July). Smart surveillance system using deep learning and RaspberryPi. In 2021 8th International Conference on Smart Computing and Communications (ICSCC) (pp. 246-251). IEEE.

[34] Arthi, R., Ahuja, J., Kumar, S., Thakur, P., & Sharma, T. (2021). Small object detection from video and classification using deep learning. In Advances in Systems, Control and Automations: Select Proceedings of ETAEERE 2020 (pp. 101-107). Springer Singapore.

[35] ain, H., Vikram, A., Kashyap, A., & Jain, A. (2020, July). Weapon detection using artificial intelligence and deep learning for security applications. In 2020 International conference on electronics and sustainable communication systems (ICESC) (pp. 193-198). IEEE.