Real-Time Weapons Detection System using Computer Vision

Pranav Nale
Department of Information Technology, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, 412115, Maharashtra, India
pranav.nale.btech2019@sitpune.edu.in

Shilpa Gite
Department of Artificial Intelligence and Machine Learning, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, 412115, Maharashtra, India
shilpa.gite@sitpune.edu.in

Deepak Dharrao
Department of Computer Science and Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, 412115, Maharashtra, India
deepakdharrao@gmail.com

2023 Third International Conference on Smart Technologies, Communication and Robotics (STCR) | 979-8-3503-7086-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/STCR59085.2023.10396960

***Abstract*—The growing use of Closed-Circuit Television (CCTV) systems in modern security applications has driven the need for automated surveillance through computer vision. The primary aim is to reduce human intervention while enhancing early threat detection and real-time security assessments. Although advanced surveillance technologies have facilitated monitoring, constant human oversight remains challenging. This has prompted a quest for models capable of identifying unlawful activities with minimal human involvement. Real-time weapon detection, despite advancements in deep learning algorithms and dedicated CCTV cameras, remains a formidable challenge, especially with varying angles and potential obstructions. Existing detection systems are often expensive and require specialized tools, necessitating a cost-effective and reliable alternative that minimizes false positives. This research focuses on creating a secure environment by utilizing real-time resources and deep-learning algorithms for identifying dangerous weapons. Without a predefined dataset for real-time detection, the researchers compiled one from diverse sources, including camera shots, internet images, movie data, YouTube CCTV recordings, and Roboflow Computer Vision Datasets. The proposed weapon detection system employs a hybrid model of Detectron2 and YOLOv7, emphasizing precision and recall in object detection, particularly in challenging conditions like low-light environments. This research contributes to developing an effective, reliable real-time weapon detection system tailored for diverse scenarios.**

Keywords— Weapons, Detection System, YOLOv7, Detectron2, RealTime, Object Detection, Gun Detection, RealTime Object Detection

I. INTRODUCTION

Global crime rates have increased as a result of the increasing use of pistols during gruesome crimes. For a country to grow, the rule of law must be preserved. The prevalence of gun-related criminality is a major concern in many parts of the world [1]. The majority of them are countries where owning a gun is legal. The harm would have been done even if the news they received was false and untrue if it hadn't spread quickly worldwide owing to the media, particularly social media. If a person is in a situation with a weapon, they may become irrational and act violently. This is because people may be brainwashed.

Installing surveillance cameras that can automatically identify firearms and trigger an alarm to notify the operators or security personnel is the solution to the aforementioned dilemma [2, 3]. However, there hasn't been much research done on algorithms for detecting weapons in Real-Time using surveillance cameras, and related studies frequently consider hidden weapon detection, typically utilizing X-rays or millimeter wave pictures using conventional machine learning approaches [4]. Convolutional neural networks (CNN), in particular, have produced ground-breaking breakthroughs in object identification and categorization during the past few years [5]. It has so far produced the best results in common image processing issues, including grouping, detection, and localization. CNN automatically learns features from the data rather than choosing them manually.

The goal of this research is to make object detection models more accurate using well-labelled cutting-edge datasets, enhance the current weapon detection system using Real-Time detection with a combination of video and photodetection, and run a firearm detection analysis on real-time video, as these studies typically perform evaluations on fictitious datasets [6].

II. LITERATURE REVIEW

The modern world is very concerned with security and safety. A nation's ability to attract foreign investment and tourism depends on how safe and secure its environment is. The most recent numbers show increased civilian gun ownership [7]. According to the most recent figures, there are 71.1 million gun owners in India; China, 49.7 million; Pakistan, 43.9 etc. [8], which indicates that risks from firearms are growing internationally. Maintaining the Integrity of the Specifications. Real-time object recognition and categorization became a challenge as a result of significant advancements in the field of CCTV, processing technology, and deep learning models [9]. There has only been a small amount of research in this area, and most of it has been focused on detecting concealed weapons.

It was initially derived from imaging technologies like millimeter waves and utilized for baggage checking and various airport security reasons before being employed for weapon detection. For finding concealed weapons at airports and other secure areas of the body,

979-8-3503-7086-7/23/$31.00 ©2023 IEEE

Sheen et al. introduced the CWD approach, utilizing three-dimensional millimeter (mm) wave imaging technology [10]. X. Zue et al. proposed an alternative CWD method based on a multi-stage decomposition method that fuses color visual images with infrared (IR) pictures [11]. Meanwhile, R. Blum et al. presented a CWD methodology that combines visual and IR or millimeter wave images. This approach incorporates a multiple-resolution mosaic capability, emphasizing the concealed weapon in the target image [12]. These diverse methods highlight the evolving landscape of concealed weapon detection, exploring various imaging technologies and fusion techniques for enhanced accuracy and efficiency.

E. M. Upadhyay proposed an image-fusion-based CWD method. When the scene's picture was present above and beneath exposed areas, they employed IR image and visual effusion to discover hidden weaponry [13]. They used a homomorphic filter taken at various exposure levels, which they then applied to visible and IR images. The current methods achieve high precision by combining different extractors and detectors, either by using simple methods like boundary detection, pattern matching, and easy intensity descriptors or by using trickier methods like cascade classifiers with boosting [14]. Rohith Vajhala published the technique for handgun detection in CCTV systems. For classification, they have combined the backpropagation of artificial neural networks with HOG as a feature extractor [15]. The detection was carried out under various conditions, first with a weapon alone and then with HOG and background subtraction techniques for people before the target object, with a claimed accuracy of 83%.

III. Dataset Construction and Pre-Processing

A. Weapon Dataset Classes

- *Reason for choosing Pistol Class*

We chose the short handheld weapons in the pistol class based on our research and analysis after studying several CCTV films of robberies and shooting incidents. We came to the conclusion that revolvers or pistols were utilized in virtually all of those incidents. Fig. 1 displays a few real-time samples taken from the pistol class dataset.

The dataset for this class consists of image samples of the following weapons:

- ✓ Pistol
- ✓ Revolver
- ✓ Short handheld firearms

Figure 1 Dataset Samples of Pistol Class including Pistol and other short-handled weapons

B. Real-Time Detection Dataset

The binary classification for a real-world scenario is the focus of this study; therefore, two classes were created, with the pistol class including photographs of pistols and revolvers but there exist classes to reduce the confusion while training the model and to decrease the chances for false positives.

C. Data Pre-Processing

The effectiveness of a Machine Learning (ML) model for a given task is influenced by a variety of factors. The representation and quality of the data are crucial at the beginning. It is more difficult to find representation during the training stage if there is a lot of redundant and irrelevant data or noisy data. Processing time for ML problems is significantly slowed down by data preparation and filtering stages. Data cleansing, standardization, processing, extraction, and feature selection are all part of the pre-processing process. The obtained dataset underwent pre-processing to create the final training dataset.

Pre-processing is extremely crucial for better training of a model. Making the dataset the same size or resolution is the first step in pre-processing, which is important

for improved model training. The next step is to apply the mean normalization. Making bounding boxes on these photos, also known as annotation, localization, or labelling, is the next stage. Each image in the data has a bounding box that is labelled. The labelled object's width, height, and value x, and y coordinates are recorded in XML, CSV, or text format. The four primary phases in data preparation are as follows:

- ✓ Image Scaling
- ✓ Data Augmentation
- ✓ Image Labeling
- ✓ Image Filtering using OpenCV
- ✓ RGB to Grayscale
- ✓ Rotation and Perspective

IV. Methodology

A. Object Labelling

This dataset served as the starting point for this project as it is a hybrid of manually scouted class images and a pre-labelled dataset from Roboflow. There were 2693 total photos in this collection, 2080 of which were classified as pistols from the pre-labelled Roboflow dataset, and 613 were manually classified as pistol/short firearms using an offline labelling tool Labelme.

Those 613 classified images were labelled using the dynamic polygonal labelling tool in Labelme, which intricately classifies the subject from the rest of the background, making it better for training and testing in later phases.

Figure 2 Polygonal Labelling method in Labelme

B. *Object Recognition*

It is a technique for identifying the actual class or category that an image belongs to by increasing the likelihood just for that particular class. This technique is carried out quickly using CNNs. CNN is frequently used as a backend in cutting-edge Classification and Detection algorithms.

According to Fig. 3, the classification of images and localization of objects fall under the recognition, and combined classification & and localization are used to detect objects. A quick summary of object categorization, localization, and detection is in order

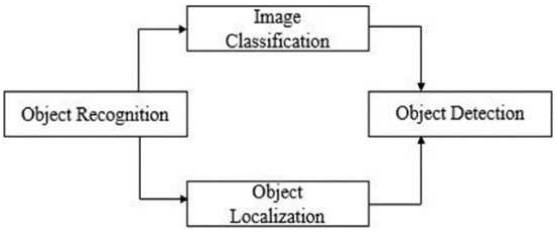

Figure 3 Object Classification and Localization

Image Classification The feature maps are obtained by applying a kernel/filter to the entire picture in the classification model. It then guesses the label based on the likelihood of the extracted characteristic.

Object Localization By providing the height and width that go along with the item's coordinates, this technique produces the precise location of an object within an image. Object Detection The characteristics of the aforementioned algorithms are used in this work. The detection technique provides the class name and the enclosed box's x and y coordinates along with width & and height.

Figure 4 Object Detection using YOLOv7

Figure 5 Instance Segmentation using YOLOv7

In order to produce the box with our chosen threshold, non-max suppression is utilised, as shown in Fig. 4 and 5. The following attributes can be seen in the output:

- ✓ Bounding Box
- ✓ Probability

Object Detection tends to be a very CPU and GPU heavy task. Hence, in the past object detection was highly limited due to a lack of data and poor computing power. As time went on, however, computing power rose, and the world transitioned from CPUs to GPUs (GPU). Originally intended for gaming and enhancing the graphics quality of computers, GPUs are now widely employed for deep learning. Competitions began in ImageNet and comprised around 1000 classes.

C. *Classification and Detection Approach*

Following are the classifier and object detectors are used in this research work.

- ✓ Detectron2
- ✓ YOLOv5
- ✓ YOLOv7
- ✓ Faster RCNN-Inception ResNetV2

Detectron2 One of the most potent deep learning toolboxes for image identification is Detectron2. Instance segmentation, person keypoint detection, panoptic segmentation, object detection, and other activities may be easily switched between because of its versatile architecture [16] [17]. Popular datasets, including COCO, Cityscapes, LVIS, and PascalVOC are supported natively, in addition to other Faster/Mask R-CNN backbone combinations (Resnet + FPN, C4, Dilated-C). Additionally, it offers baselines with pre-trained weights that are ready for usage. The architecture of Detectron2 is shown in figure 6.

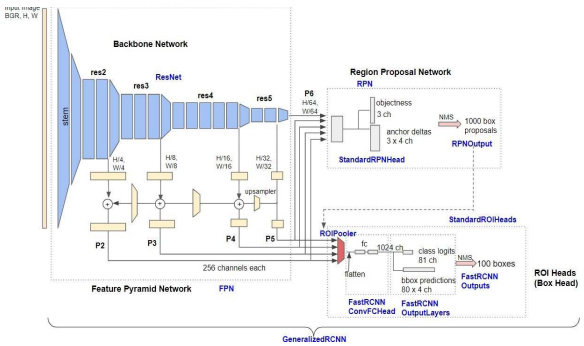

Figure 6 Architecture of Detectron2

The architecture of this network is depicted in the above schematic. It consists of three blocks, namely:

- Backbone Network: Different scales of feature maps are extracted from the input picture.
- Region Proposal Network: From the multi-scale characteristics, it extracts object areas.
- Box Head: In order to acquire precise box positions and classification results, it warps and crops feature maps into a number of fixed-size features by using proposal boxes.

YOLOv7 The newest member of the YOLO (You Only Look Once) family of models is the v7 version. Single-stage object detectors are YOLO models. Image frames in a YOLO model are enhanced by a backbone [18] [19]. These characteristics are merged and blended in the neck, where they

are subsequently transmitted to the network's head. Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be using English units as identifiers in trade, such as "3.5-inch disk drive."

Introducing the YOLOv7, the latest addition to the YOLO (You Only Look Once) model family, focusing on single-stage object detection. In YOLO models, image frames undergo enhancement through a backbone [18] [19]. These characteristics are then fused and integrated in the neck before being transmitted to the network's head. Primary units are encouraged to be either SI (MKS) or CGS, with English units permissible as secondary units in parentheses. Noteworthy exceptions include using English units as identifiers in trade, such as "3.5-inch disk drive."

The locations and types of items around which bounding boxes should be created are predicted by YOLO. YOLO conducts post-processing via non-maximum suppression (NMS) to arrive at its final prediction [20].

The industry has seen an increase in the number of YOLO models. As they learn about YOLO and machine Learning, developers can quickly catch up thanks to the compact architecture, while practitioners can power their applications with the bare minimum of hardware thanks to the real-time inference performance. The Architecture of YOLOv7 is shown in Figure 7. YOLOv7 developers sought to advance object identification by creating a network architecture that outperformed competitors in predicting bounding boxes at comparable inference speeds. Achieving this, they made critical adjustments to the YOLO network and training procedures. It's emphasized to avoid combining SI and CGS units, preventing confusion and dimensional imbalances in equations. If mixed units are necessary, the recommendation is to explicitly state the units for each quantity in the equation.

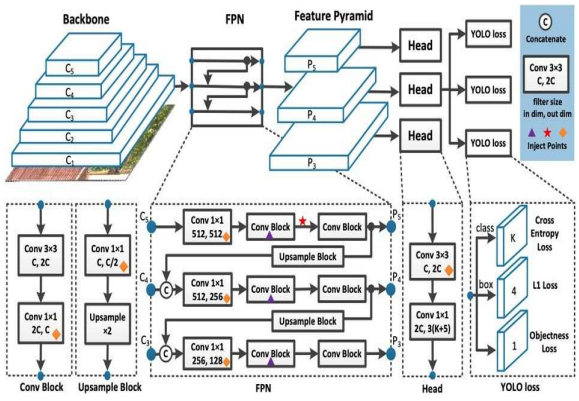

Figure 7 Architecture of YOLOv7

In essence, YOLOv7 represents a significant stride in enhancing the accuracy of bounding box predictions, underscoring the importance of refined network architecture and meticulous training procedures

D. Training Mechanism

The overall approach taken in training and optimization is shown in Fig. 3. Starting with issue definition, locating the necessary dataset, utilising pre-processing techniques, and then training and assessing the dataset are the next steps. Depending on the accuracy of the evaluation, we keep those weights as a classifier; however, if it is inaccurate, the backpropagation procedure and gradient descent technique are used.

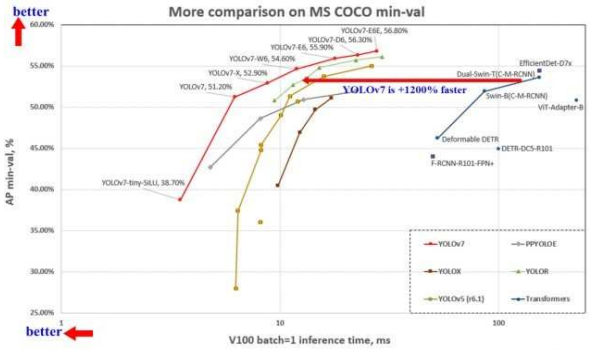

Figure 8 Evaluation of YOLOv7 with its peer networks

E. Confusion Object Inclusion (using YOLOv7)

We have designed the issue to decrease the frequency of false positives and negatives. This weapon class covers all the hand weapons, revolvers, and other firearms, which helps train the model to enhance the accuracy in low-light and unfavourable angles and provides a trustworthy Real-Time solution. This model aims to reduce the confusion between items like mobile phones, metal detectors, selfie sticks, purses, etc., that are sometimes confused with pistol classifications.

V. Data And Results

We have identified firearms in real-time streams that were poor quality, dark, and frame-per-second. Since most previous work focused on recognizing high-quality photos and videos, since such models were developed using good-quality datasets, real-time recognition of low-resolution objects is not achievable. Following model training and model testing on the datasets listed in Table 1, the outcomes are examined.

The outcomes for various approaches are assessed as stated in the methodology section. Because pistols and revolvers were utilized in 97% of the robbery incidents, our key issue statement is real-time detection. As a consequence, various outcomes for the YOLOv7 technique have been assessed here.

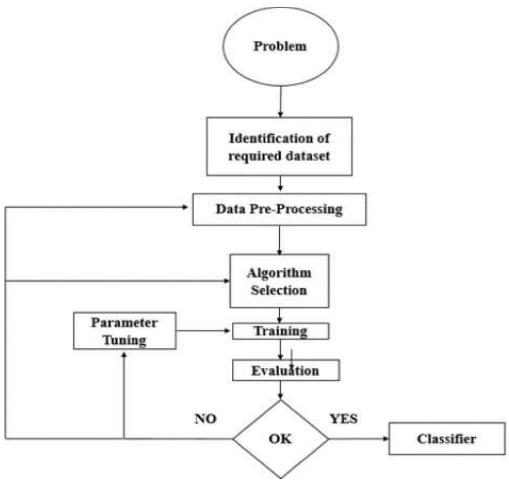

Figure 9 Training Flow Diagram

A. Dataset Experimentation Results

For the highest-performing model, mean average precision (mAP), along with the traditional metrics of F1-score and

frames per second, were used to compare the performance of various models. These terms are derived using equations 1, 2, and 3 below. The precision-to-recall ratio is the F1 score.

$$Precision = TP/TP + FP \quad (1)$$

$$Recall = TP/TP + FN \quad (2)$$

$$F1 - Score = 2 * Precision * Recall/Precision + Recall \quad (3)$$

The model of our method that performs the best overall is YOLOv7. Fig. 10 displays the YOLOv7 performance graph for loss and mean average precision (mAP) on a validation dataset.

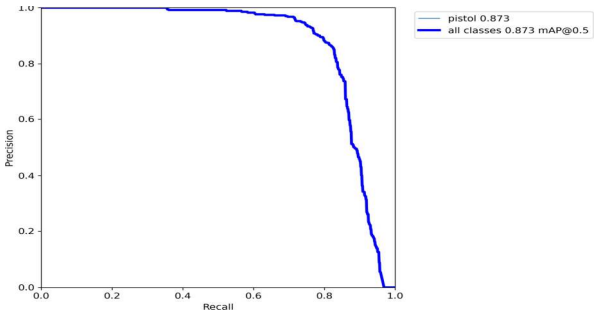

Figure 10 Precision & Recall Metrics

We can observe how smoothly the model loss curve converges to the optimal level and how exactly it does so, producing a very strong loss score of 0.84 and a mAP of 91.73%. The average precision values for the relevant class are averaged to provide the mAP.

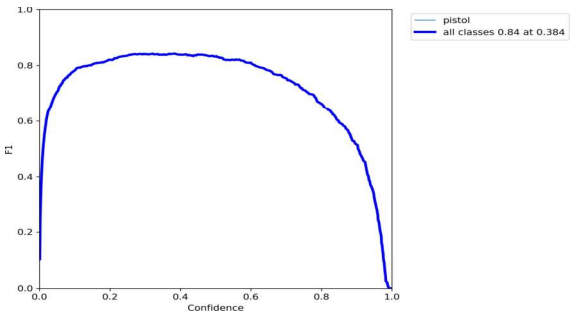

Figure 11 F1 Score & Confidence Metrics

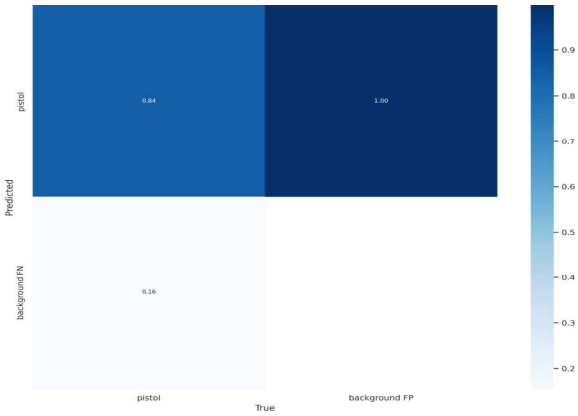

Figure 12 Confusion Matrix for trained dataset

B. Detection Results - Pistol Class in Images

In this section, we have provided the output images showing detection of the pistol. figure 13 shows Pistol image detection in blurred images, and figure 14 shows Pistol image detection in varied lighting conditions.

Figure 13 Pistol image detection in blurred image

Figure 14 Pistol image detection in varied lighting conditions

C. Detection Results - Pistol Class in Video

In this section, we have presented output result of pistol detection from videos frames. Figure 15 shows Pistol detection in pre-fed video and Figure 16 shows Pistol detection in CCTV footage.

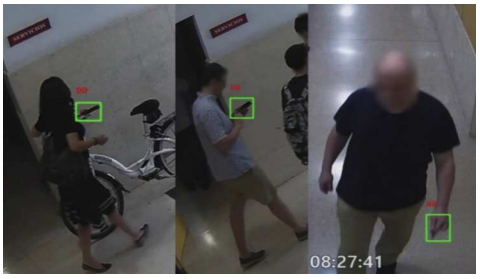

Figure 15 Pistol detection in pre-fed video

Figure 16 Pistol detection in CCTV footage

D. Detection Results - Pistol Class in Real-Time

In this section we have presented output result of Pistol detection in Real-Time. We have taken two out in two different direction as shown in figure 17 Pistol detection in Real- Time (Right Angle) and figure 18 Pistol detection in Real-Time (Left Angle) We may infer from earlier trials that

the idea made a difference in the modern weapon detection systems.

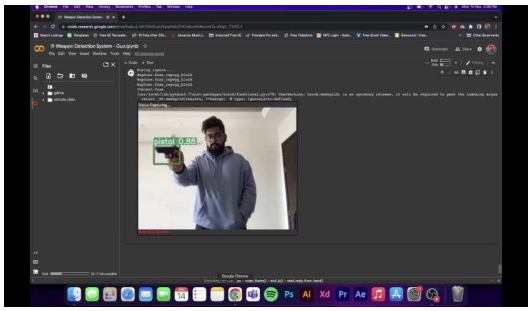

Figure 17 Pistol detection in Real-Time (Right Angle)

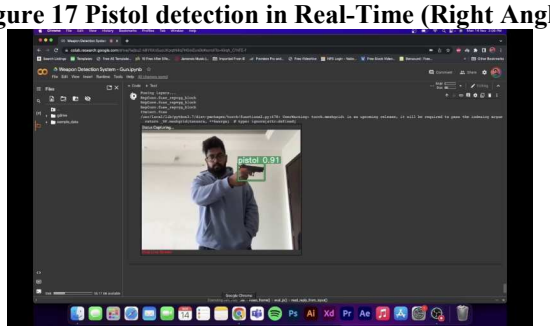

Figure 18 Pistol detection in Real-Time (Left Angle)

E. Discussion

We may infer from earlier trials that the idea made a difference in modern weapon detection systems.

Fig. 10, 11 and 12 show the various results and metrics of experimentation on the dataset, namely, precision, recall, and F1-score for evaluation.

Figures 13, 14, 15, and 16 show the detection of various firearms in pre-fed footage in low-light and unfavorable camera angles.

For the real-time situation, the object detection model YOLOv7 performed admirably in terms of speed and detection precision from Fig. 17 and 18.

Findings indicate that the optimum approach is to initially train in synthetic pictures before training in actual photos for fine-tuning.

VI. CONCLUSION & FUTURE WORK

This study proposes an improved real-time automatic weapon detection system for monitoring and command applications, addressing challenges related to distance-dependent accuracy. The research aims to enhance security, promoting economic benefits by attracting security-conscious investors and visitors. Object detection algorithms utilizing Region of Interest (ROI) outperformed those without, with the YOLOv7 model, trained on a new database, demonstrating exceptional results. It achieved a mean average precision (mAP) of 87.3%, an F1-score of 91%, and a confidence score of nearly 98%, surpassing previous real-time studies.

The researchers prioritized real-time weapon detection with minimized false positives and negatives, utilizing a new training database and the latest deep learning model. Future work focuses on further reducing false positives and negatives, possibly expanding to more classes. The study suggests integrating object identifiers with movement and 3-D position approximation for enhanced recall and accuracy. Recommendations include limiting the identification of common items and triggering alarms only for successive frames with identified weapons, contributing to ongoing efforts for a more accurate and reliable real-time weapon detection system.

REFERENCES

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

[2] Olmos, R., Tabik, S., Herrera, F.: Automatic handgun detection alarm in videos using deep learning. Neurocomputing 275, 66–72 (2018)

[3] Xiao, Z., Lu, X., Yan, J., Wu, L., Ren, L.: Automatic detection of concealed pistols using passive millimeter wave imaging. Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST) pp. 1–4 (2015)

[4] González, J.L.S., Zaccaro, C., Álvarez García, J.A., Morillo, L.M.S., Caparrini, F.S.: Real- time gun detection in CCTV: An open problem. Neural Netw 132, 297–308 (2020)

[5] de Azevedo Kanehisa, R.F., de Almeida Neto, A.: Firearm Detection using Convolutional Neural Networks. In: ICAART (2). pp. 707–714 (2019)

[6] Yadav, P., Gupta, N., Sharma, P.K.: A comprehensive study towards high-level approaches for weapon detection using classical machine learning and deep learning methods. Expert Systems with Applications p. 118698 (2022)

[7] Karp, A.: Estimating global civilian-held firearms numbers (2018)

[8] Reid, A.J.: The gun problem (2022)

[9] Darker, I.T., Kuo, P., Yang, M.Y., Blechko, A., Grecos, C., Makris, D.: Automation of the CCTV-mediated detection of individuals illegally carrying firearms: Combining psycholog- ical and technological approaches. Proc. SPIE 7341 (2009)

[10] Sheen, D.M., Mcmakin, D.L., Hall, T.E.: Three-dimensional millimeter-wave imaging for concealed weapon detection. IEEE Trans. Microw. Theory Techn 49(9), 1581–1592 (2001)

[11] Xue, Z., Blum, R.S., Li, Y.: Fusion of visual and IR images for concealed weapon detec- tion. In: Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002.(IEEE Cat. No. 02EX5997). vol. 2, pp. 1198–1205 (2002)

[12] Blum, R., Xue, Z., Liu, Z., Forsyth, D.S.: Multisensor concealed weapon detection by using a multiresolution mosaic approach. In: IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004. vol. 7, pp. 4597–4601 (2004)

[13] Upadhyay, E.M., Rana, N.K.: Exposure fusion for concealed weapon detection. Proc. 2nd Int. Conf. Devices Circuits Syst. (ICDCS) pp. 1–6 (2014)

[14] Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324 (1998)

[15] Vajhala, R., Maddineni, R., Yeruva, P.R.: Weapon detection in surveillance camera images (2016)

[16] Pradhan, A., Niaz, S.Y., Pradhan, M.P., Pradhan, R.: DETECTION AND RECOGNITION OF TEXTS FEATURES FROM A TOPOGRAPHIC MAP USING DEEP LEARNING. Suranaree Journal of Science & Technology 29(5) (2022)

[17] Hung, C.P., Choi, J., Gutstein, S.M., Jaswa, M.S., Rexwinkle, J.T.: Soldier-led adaptation of autonomous agents (SLA3). In: Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III. vol. 11746, pp. 743–754 (2021)

[18] Doan, T.S., Nguyen, T.K.T., Vo, T.A.: Weapon Detection with YOLO Model Version 5, 7, 8 (2023)

[19] Kumar, S., Kumar, C.: Deep Learning based Target detection and Recognition using YOLO V5 algorithms from UAVs surveillance feeds. In: 2023 International Conference for Advancement in Technology (ICONAT). pp. 1–5 (2023)

Li, P., Che, C.: SeMo-YOLO: a multiscale object detection network in satellite remote sensing images. In: 2021 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (2021)