Weapon detection system for surveillance and security

Shehzad Khalid
Dept. of Computer Engineering
Bahria University
Islamabad, Pakistan
shehzad@bahria.edu.pk

Abdullah Waqar
Dept. of Computer Engineering
Bahria University
Islamabad, Pakistan
abdullahwaqar40@gmail.com

Hoor Ul Ain Tahir
Dept. of Computer Engineering
Bahria University
Islamabad, Pakistan
hoorulaintahir@gmail.com

Onome Christopher Edo,
Department of Information Systems,
Auburn University at Montgomery,
Alabama, USA,
oedo@aum.edu

Tenebe, Imokhai Theophilus, PhD
Mineta Transportation Institute,
San Jose State University
California, United States
Yoshearer@gmail.com

2023 International Conference on IT Innovation and Knowledge Discovery (ITIKD) | 978-1-6654-6372-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ITIKD56332.2023.10099733

***Abstract*— Weapon detection is a critical and serious topic in terms of public security and safety, however, it's a difficult and time-consuming operation. Due to the increase in demand for security, safety, and personal property protection, the requirement for deploying video surveillance systems capable of recognizing and interpreting scenes and anomalous occurrences plays an important role in intelligence monitoring. In certain regions of the globe, mass shootings and gun violence are on the increase. Timely detection of the presence of a gun is critical to prevent loss of life and property. Several object detection models are available, which struggle to recognize firearms due to their unique size and form, as well as the varied colors of the background. In this study, we proposed a state-of-the-art system based on a deep learning model YOLO V5 for weapon detection that will be sufficiently resilient in terms of affine, rotation, occlusion, and size. We evaluated the performance of our system on a publicly available dataset and achieved the F1-score of 95.43%. For the detection and segmentation of firearms, we implemented Instance segmentation or Pixel level segmentation which employed Mask-RCNN. The system achieved the detection accuracy (DC) of 90.66% and 88.74% Mean intersection over union (mIoU). The purposed methodology combined different data augmentation and preprocessing methods to improve the accuracy of the proposed weapon detection system.**

Keywords—weapon detection; handguns; Yolov5.

I. INTRODUCTION

Security is a serious concern for the whole world in the modern day, as it is necessary to safeguard sensitive and valuable assets such as a person, house, community, and country. The key objective of this research is to develop a system for monitoring an area's surveillance data. A person carrying a firearm in public is a significant sign of potentially harmful scenarios. Recently, the frequency of situations in which individuals or small groups use weapons to harm or kill people has increased [1].

For monitoring and surveillance tool in the fight against crime Closed Circuit Television (CCTV) are widely used. CCTV's purpose is to monitor the environment to prevent crime and social events. Its applicability is user-dependent, for example, CCTV is used in street surveillance to monitor several activities, including discovering missing persons, detecting drug addiction, and identifying anti-social behavior. Additionally, it is used to collect evidence of a crime and provide it to the appropriate authorities for prosecution. A CCTV system comprises a camera and an unattended operator deployed remotely. A CCTV camera captures and transmits the video to a base station's television screen, where the operator examines it for suspicious activity or evidence collecting. However, the operator's ability to identify suspicious activity is proportionate to his or her attention to each video stream shown on the screen. Due to the low operator-to-screen ratio, the concurrent operation of several video feeds on the same screen, and the operating room's ambient settings, it is difficult for a CCTV operator to watch each video stream activity with total attention at all times. Due to the risk that they would miss detecting any aberrant behavior that might have major security implications, it is critical to detect such anomalies promptly to avert any terrorist action.

Weapons are often used for violent actions rather than self-defense. This issue necessitates the use of contemporary technologies and tactics to survive the consequences of gun violence, just as the great majority of governments use video surveillance systems to watch citizens for signs of terrorism and crime. Some of the advanced nations proposed a variety of approaches to automatically detect firearms in a video sequence. At the moment, the most accurate weapon detection models are built using Deep Convolutional Neural Networks [3-6]. Based on a large quantity of labeled data, these models can learn the unique attributes on their own. We can localize the objects in images by a process called object detection.

Generally, automated weapon detection system faces various challenges:

- Automatic detection alarm systems need the pistol's precise placement in an image.

- Creating a new dataset is a tiring and time-consuming operation.
- Pistols may be held in a variety of ways with one or two hands, obscuring a considerable portion of the gun.
- Automatic gun detection alarms must be activated in real-time and only when the system is certain that a pistol is present in the scene [7].

These challenges affect the system's overall accuracy and increase the rate of false positives. To overcome this problem, we will apply several augmentation techniques to the dataset.

Fig. 1. Demonstration of automated weapon detection using CCTV camera

II. Related work

This section will highlight the recent approaches towards the problem of weapons detection. Olmos et al. [7] propose a technique for firearm recognition based on deep learning. He formulates this issue as a two-class problem, weapon, and in the background. They created a training database by manually labeling photos from the Internet and employing Faster-RCNN based on VGG16 as a detection model. This model scored a satisfactory accuracy of 91.43 % when trained on a dataset of 6000 photos, but it incorrectly associates the gun with things that may be handled similarly, such as a smartphone, knife pocketbook, card, and bill. Verma et al. [8] suggested an automated gun identification system based on Convolutional Neural Networks for crowded scenes (CNN). Through transfer learning, they employed Deep Convolutional Networks (DCN), a state-of-the-art Faster RCNN model. The experiments used a dataset derived from a portion of the Internet Movie Firearm Database (IMFDb). The system detected and classified three kinds of firearms, shotguns, revolvers, and pistols, with an accuracy of 89.9 cents using Support Vector Machines (SVM).

Luvizon et al. [13] further extended the research by including the human stance and looking to get improved results for identifying movements and activities conducted on the scene. Given that the human stance has sufficient information to assess activity, it is predicted to be beneficial in solving the handgun detection challenge. The author presents a multitask framework for simultaneously estimating 2D and 3D positions from still images and recognizing human actions from video sequences. The approach estimates two-dimensional and three-dimensional poses from a single image or frame sequence. In a unified framework, these stances and visual information are combined to anticipate actions. While body position information is frequently used for classical computer vision issues such as activity and gesture identification, it has not been widely applied to handgun detection.

Elmir et al. [14] proposed a framework that operates in stages. The first phase is picture collection, followed by motion and gun detection. They conducted trials on three distinct models, including the Faster R-CNN, Mobile-Net, and the Convolutional Neural Network CNN. They have validated their approaches and provided a comparison of the above three methods on separate datasets. The training data set comprises 9261 images, 200 of which correspond to the 102 handgun classes. The training data set for the region task has 3000 images. The detection and classification test data set comprises 608 images, of which half them has handguns. They trained the models on a database containing a sample of 420 images. They evaluated the first model using 608 images, the second model using 200 images, and the third model using 420 images. The performance of these models was evaluated, On CNN they achieved 55% accuracy, on Faster R-CNN 80% accuracy was achieved, and 90% accuracy was archived using Mobile-Net. Lai et al. [24] utilize a TensorFlow-based version of Overfeat3, an integrated framework that employs CNNs for classification, detection, and localization, they evaluated their model and achieved 89 % test and 93 % training accuracy from surveillance videos, films, and homemade movies.

Darknet YOLO is a cutting-edge convolutional neural network-based object identification system [10]. It was built first using Darknet, an open-source neural network framework written in CUDA and C [11]. Traditional classifiers find potential areas and identify the intended item using sliding window approaches or selective search. Thus, locations with a high probability of having weapons are identified [10]. YOLO, in contrast to previous approaches, does not reuse a classifier for detection. This method merely examines the image once, the picture is segmented into many sub-regions to do the detection. Five bounding boxes are examined for each sub-region, and the probability of each having an item is determined. YOLO performs detection at a hundred times the speed of Fast R-CNN [10].

Since 2012, researchers have developed a range of object identification methods and architectures, including the R-CNN and other variants [25 – 27]. Joseph Redmon [28] proposed the "YOLO" (You Only Look Once) approach in 2016. In contrast to classic region-based techniques, YOLO is a single-stage technique that performs a single pass over the picture through FCNN, making it relatively efficient as compared to the competitors. YOLOv2 [29] overcomes the problem of relatively lower accuracy due to inaccurate localization by employing batch normalization and better resolution classifiers. YOLOv3 [10] was launched with gradual enhancements. YOLOv4, which stands for Optimal Speed and Accuracy of Object Detection, was introduced in 2020. Two-stage object detection networks include Faster-RCNN and R–FCN, as well as a Region Proposal Network (RPN). This kind of network is slow in detecting an object. A single-stage object detection network, similar to YOLOv3 and Single Shot Detector SSD, is thus suggested. A single forward CNN is used to identify the object's location and class [30].

Mehta at el. [1] used YOLOv3 to detect weapons. They implied two different datasets for the validation of their proposed system. The first dataset namely UGR comprises of total 608 images of which half of them have guns and the remaining are without guns. The other dataset known as IMFDB comprises 6000 images out of which 4000 images are with guns and the remaining 2000 without guns. The YOLOv3 model was trained and tested on both datasets separably. On the UGR dataset, they have reported the F1-score of 90.3% and for IMFDB they have reported 84.5% but they have not applied any preprocessing or data augmentations to further enhance the accuracy of the system. Hashmi et al [12] compared the performance of two versions of YOLO, YOLOv3, and YOLOv4. A new dataset containing 7800 images was obtained from different CCTV footage and Google images and was labeled manually. All the images in the dataset were resized to 448 x 448 x 3 and both models were trained and tested separately on the same dataset. After evaluating the result of YOLOv3, it achieved the F1-score of 77% and YOLOv4 achieved 82%.

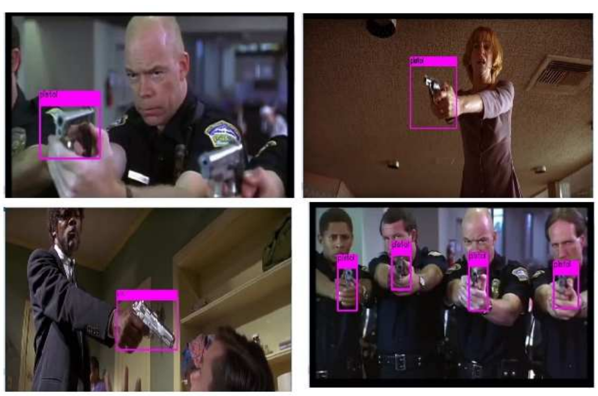

Fig. 2. Weapon detection and localization results obtained by using Yolov3 [15]

Warsi et al. [15] proposed a system for visually detecting the presence of a firearm in real-time video surveillance. The suggested technique employs the YOLO-V3 algorithm and compares the number of false-positive and false-negative predictions to those obtained using the Faster RCNN algorithm. They enhanced the findings by creating their own collection of pistols from all available angles and combining it with the ImageNet dataset. The YOLO-V3 model was used to train and evaluate the dataset. They validated the findings of YOLO-V3 against Faster RCNN using four separate movies. The detector performed poorly at detecting pistols in scenarios with varying forms, sizes, and rotations, with an evaluated F1-score of 75%.

III. Proposed Methodology

A. Dataset

We have employed a publicly available dataset from the University of Granada which contains 2,971 images. We split our dataset into three sets, 70% training set (which contains 2,080 images), 20% for the validation set (which contains 594 images), and lastly 10% for the testing set (containing 297 images). Samples from the dataset are shown in Fig. 3.

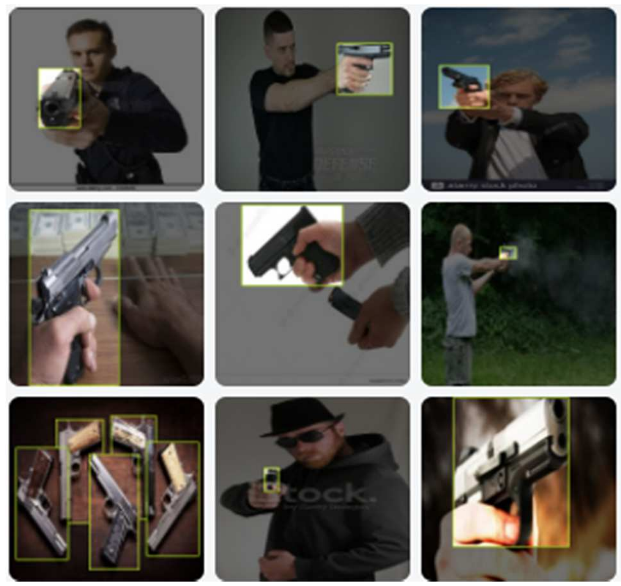

Fig. 3. Samples from the dataset

B. Preprocessing techniques applied to the datasets

The images in our dataset were pre-processed prior to using it to train our proposed deep-learning-based model. These approaches are mostly employed to generate standard image representation with consistent size, orientation, and color space. TABLE I presents the techniques applied to the images belonging to the dataset.

TABLE I. PREPROCESSING TECHNIQUES APPLIED TO THE IMAGES

Preprocessing type	Preprocessing
Auto-Orient	Applied
Resize	Stretch to 416x416
Grayscale	Applied

C. Augmentation techniques applied to the datasets

The problem of labeled data scarcity has been addressed through data augmentation. Using image fusion, geometric alteration, and other approaches for data augmentation, new images were generated which was required to train the deep learning model. This caters to the various affine transformations that we expect a weapon to undergo in different images based on the location and distance of the camera from the weapon as well as the way the weapon is placed in the scene or held by a person. It has been observed through empirical evaluation that the trained model using augmented training data, after applying augmentation techniques as presented in TABLE II, yielded better accuracies as compared to the original training data. Data augmentation makes the training results more resilient, improves prediction accuracy, accelerates convergence, and reduces training process time. This enhances the accuracy and consistency of the training.

TABLE II. AUGMENTATIONS APPLIED TO THE IMAGES FROM THE DATASET

Augmentation type	**Ranges**
Crop	0 - 20% Zoom
Flip	Vertical, Horizontal
Noise	Up to 5% of pixels
90° Rotate	Counter-Clockwise and Clockwise
Rotation	Between -15° and +15°
Cutout	3 boxes with 10% of image size each
Blur	Up to 10 pixels
Crop	0 - 20% Zoom
Noise	Up to 5% of pixels

We managed to obtain 7,131 images after the application of the augmentation step from which 6,240 images were used for training, 594 images for validation, and 297 images for the test set. A sample of the dataset after applying the above-mentioned augmentations is shown in Fig. 4 below.

Fig. 4. Sample of the dataset after applying the augmentations

D. *Proposed method for weapon detection*

We have used a state-of-the-art object detection algorithm known as YOLO V5 "You Only Look Once". As with every other single-stage object detector, it is composed of three major components: Backbone, Neck, and Head. The internal workings and structure of the model are shown in Fig. 5.

1) Backbone: The Backbone is mostly utilized to extract important features from an image. In YOLO v5, for extracting important properties from an image Cross-stage Partial Networks (CSP) serve as the backbone. The CSPNet is utilized in YOLO v5 because it eliminates computational bottlenecks by uniformly spreading computation throughout all convolutional layers, the objective is to maximize the usage rate of each calculation unit. Compressing feature maps during the feature pyramid construction process benefits in lowering memory costs, the object detector reduces memory use by 75%.

2) Neck: The Neck is mostly used to generate feature pyramids; it aids in the model's generalization when objects are scaled. It assists in identifying the same thing in a variety of sizes and scales. The Feature pyramids aid the model in performing effectively on unseen data. In our YOLO v5, we employ a Path Aggregation Network (PANet) to generate feature pyramids for the neck.

3) Head: The head is mainly utilized to complete the process of detection. It creates final output vectors with objectless scores, class probabilities, and bounding boxes by applying anchor boxes to features.

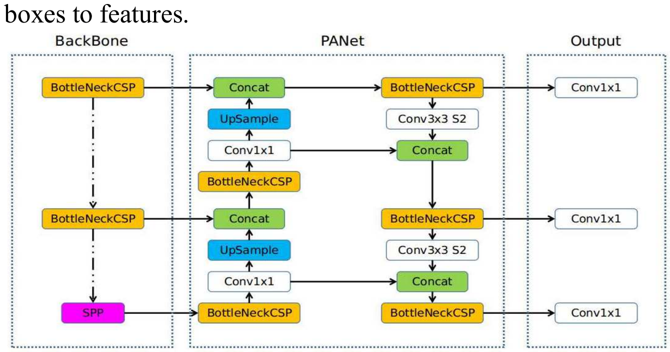

Fig. 5. Internal workings of YOLO V5

E. *Proposed method for instance level segmentation*

Pixel-level segmentation is the process by which you assign a class to each pixel in an image. Firstly, we acquired the dataset and labeled the images using the VGG Image annotator and then we trained our model using the Mask RCNN algorithm. Mask R-CNN employs the two-step technique, with the first stage that uses Region Proposal Network (RPN). In addition to predicting the box offset and class in the second stage, Mask R-CNN also produces a binary mask for each RoI.

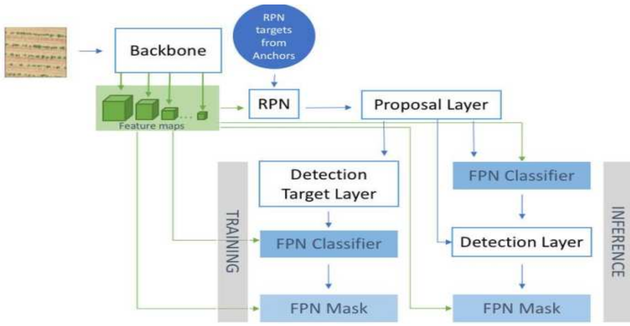

Fig. 6. Architecture of Mask R-CNN

Fig. 6 shows the architecture of mask RCNN which has been used for instance level segmentation in the proposed method. The RGB source image is sent to the Backbone network, which is responsible for feature extraction at varying sizes. This Backbone is a ResNet-101 with prior training. Each block in the ResNet design generates a feature map, with the resultant collection serving as an input to other blocks: the Feature Pyramid Network (FPN), as well as the Region Proposal Network (RPN). The RPN generates scores for each anchor based on the chance that it would be categorized as negative, positive, or neutral and the proposal Layer starts by retaining the highest scores to choose the best anchors. The next step is the Detection Target Layer on the training route. This is not a network layer, but rather an additional filtering step for the Regions of Interest (ROIs) generated by the proposal Layer. Nonetheless, this layer utilizes the ground truth boxes to calculate the overlap with the ROIs and sets them to true if IoU > 0.5 and false otherwise.

IV. Results

A. Performance Matrix

We used the dataset of 2,971 images which became 7,131 after applying augmentation from which 297 images were included in the test set to evaluate our model performance. The proposed algorithm, YOLO V5 was trained using 6,240 images. The performance of the proposed system is shown in TABLE III.

TABLE III. Confusion Matrix Obtained from Proposed Region-Based Segmentation

	Actual YES	**Actual NO**
Predicted YES	282	16
Predicted NO	11	0

As shown in TABLE IV below, we obtain an average value of precision is 94.63%, an average value of recall is 96.25%, and an average value of accuracy is 95.43%.

TABLE IV. Experimental Results of Region Segmentation Approach

Precision	**Recall**	**F1-score**
0.9463	0.9625	0.9543

This means that the model's performance is excellent. YOLO v5 outperforms other algorithms such as EfficientNet, Faster RCNN, and SSD in terms of latency. The YOLO is frequently utilized in real-world applications because of its higher inference speed. The delay of YOLOv5 is 24 milliseconds. The Qualitative results of region-based segmentation of weapons based on the proposed approach are also presented in Fig. 7.

Fig. 7. Results of the proposed model

B. Comparison of Proposed Approach with Competitors

The comparison of F-1 scores obtained from different experiments concludes that YOLO V5 outperforms other approaches that include YOLO V3, YOLO V4, and Faster RCNN. The performance of the proposed method was compared with existing approaches mentioned in the literature review, shown in TABLE V. The acquired results demonstrate the higher performance of the proposed method. Several approaches used deep learning, while others employed image-processing techniques. The suggested model employs a convolutional neural network, which is more efficient and accurate in contrast to earlier models. After developing this model, we can simply use real-time classification outcomes in the real world.

TABLE V. Comparison of Proposed Approach with Existing Deep Learning Approach

Sr. no	**Paper**	**Technique / Classifier**	**F1-score/ Accuracy (%)**
1	Verma et al. [8]	SVM and DCN	89.9
2	Olmos et al. [7]	Faster R-CNN	91.43
3	Mehta et al. [1]	YOLOv3	90.3
			84.5
4	Hashmi et al. [12]	YOLOv3	77
		YOLOv4	82
5	Elmir et al. [14]	CNN	55
		Fast R-CNN	80
		Mobile-SDD	90
6	Warsi et al. [15]	YOLOv3	75
7	**Proposed**	**YOLOv5**	**95.43**

A graphical representation of the comparison of the F1-score of the proposed method against the existing deep learning based approaches mentioned in the literature is shown in Fig. 8. We obtained an average value of precision is 94.24%, and an average value of recall is 95.97% as shown in TABLE VI, whereas TABLE VII shows the accuracy of detection and as well as the Mean Intersection over union mIoU for the accuracy of the mask. We have achieved a detection accuracy of 90.66% and mIoU of 88.74%. The qualitative results of pixel-level segmentation using mask RCNN are shown in Fig. 9.

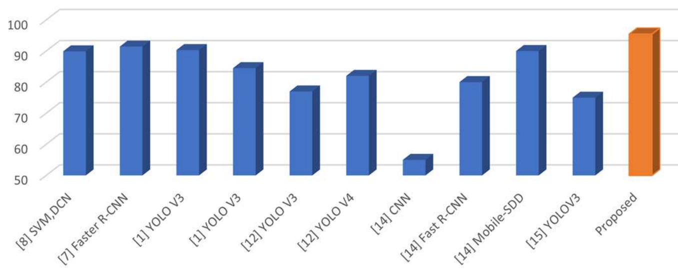

Fig. 8. F1-score of the proposed system against the existing methods

TABLE VI. PRECISION AND RECALL FOR MASK R-CNN

Model	Precision	Recall
Mask RCNN	0.9424	0.9597

TABLE VII. TACCURACY AND INTERSECTION OVER THE UNION OF MASK R-CNN

Model	Detection Accuracy (DA)	Mean Intersection Over Union (IOU)
Mask RCNN	90.66%	88.74%

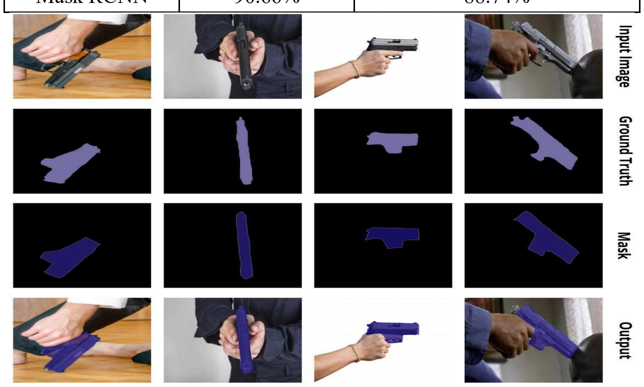

Fig. 9. Results obtained by our Mask R-CNN model

V. CONCLUSION

The purpose of this research is to provide an efficient real-time weapon detection deep learning model with a high accuracy metric. We evaluated several deep-learning algorithms and methodologies for the early identification of firearms. According to our research, deep learning algorithms have shown outstanding results in terms of both speed and accuracy. This technology will aid police departments in different areas to detect weapons automatically and respond to them in a swift time to prevent harmful incidents.

REFERENCES

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

[2] Tiwari, R.K. and Verma, G.K., "A computer vision based framework for visual gun detection using harris interest point detector" *Procedia Computer Science,* vol. 54, pp.703-712, 2015.

[3] Krizhevsky, A., Sutskever, I. and Hinton, G.E., "Imagenet classification with deep convolutional neural networks" *Advances in neural information processing systems,* vol. 25, pp.1097-1105, 2012.

[4] Zhang, Q., Yang, L.T., Chen, Z. and Li, P., "A survey on deep learning for big data" *Information Fusion,* vol. 42, pp.146-157, 2018.

[5] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. and Berg, A.C., "Imagenet large scale visual recognition challenge" *International journal of computer vision,* vol. 115(3), pp.211-252, 2015

[6] Zhao, Z.Q., Zheng, P., Xu, S.T. and Wu, X., "Object detection with deep learning: A review" *IEEE transactions on neural networks and learning systems,* vol. 30(11), pp.3212-3232, 2019.

[7] Olmos, R., Tabik, S. and Herrera, F., "Automatic handgun detection alarm in videos using deep learning" *Neurocomputing,* vol. 275, pp.66-72, 2018.

[8] Verma, G.K. and Dhillon, A., "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, November 2017.

[9] Grega, M., Matiolański, A., Guzik, P. and Leszczuk, M., "Automated detection of firearms and knives in a CCTV image" *Sensors,* vol. 16(1), p.47, 2016.

[10] Redmon, J. and Farhadi, A., "Yolov3: An incremental improvement" *arXiv preprint arXiv:1804.02767,* 2018.

[11] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., "You only look once: Unified, real-time object detection" in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788, 2016.

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

[13] Luvizon, D.C., Picard, D. and Tabia, H., "2d/3d pose estimation and action recognition using multitask deep learning" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5137-5146, 2018.

[14] Elmir, Y., Laouar, S.A. and Hamdaoui, L., "Deep Learning for Automatic Detection of Handguns in Video Sequences" in *JERI*, April 2019.

[15] Warsi, A., Abdullah, M., Husen, M.N., Yahya, M., Khan, S. and Jawaid, N., "Gun detection system using YOLOv3" in *IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*: IEEE, pp. 1-4, August 2019.

[16] Tiwari, Rohit Kumar, and Gyanendra K. Verma. "A computer vision-based framework for visual gun detection using harris interest point detector." *Procedia Computer Science*, vol. 54, pp. 703-712, 2015.

[17] González, J.L.S., Zaccaro, C., Álvarez-García, J.A., Morillo, L.M.S. and Caparrini, F.S., "Real-time gun detection in CCTV: An open problem" *Neural networks*, vol. 132, pp.297-308, 2020.

[18] Akcay, S., Kundegorski, M.E., Willcocks, C.G. and Breckon, T.P., "Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery" *IEEE transactions on information forensics and security,* vol. 13(9), pp.2203-2215, 2018.

[19] Egiazarov, A., Mavroeidis, V., Zennaro, F.M. and Vishi, K., "Firearm detection and segmentation using an ensemble of semantic neural networks" in *European Intelligence and Security Informatics Conference (EISIC)*: IEEE, pp. 70-77, November 2019.

[20] Buckchash, H. and Raman, B., "A robust object detector: application to detection of visual knives" in *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*: IEEE, pp. 633-638, July 2017.

[21] Castillo, A., Tabik, S., Pérez, F., Olmos, R. and Herrera, F., "Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning" *Neurocomputing,* vol. 330, pp.151-161, 2019.

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

[23] Singleton, M., Taylor, B., Taylor, J. and Liu, Q., "Gun identification using TensorFlow" in *International Conference on Machine Learning and Intelligent Communications*, Cham: Springer, pp. 3-12, July 2018.

[24] Lai, J. and Maples, S., "Developing a real-time gun detection classifier" *Course: CS231n, Stanford University,* July 2018.

[25] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 580–587, 2014.

[26] R. Girshick, "Fast r-CNN," in *Proceedings of the IEEE International Conference on Computer Vision,* pp. 1440–1448, 2015.

[27] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems,* pp. 91–99, 2015.

[28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 779–788, 2016.

[29] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 7263–7271, 2017.

[30] Li, P. and Zhao, W., "Image fire detection algorithms based on convolutional neural networks" Case Studies in Thermal Engineering, vol. 19, p.1006, 2020.