2022 3rd International Conference for Emerging Technology (INCET)
Belgaum, India. May 27-29, 2022

2022 3rd International Conference for Emerging Technology (INCET) | 978-1-6654-9499-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/INCET54531.2022.9824281

Weapon Detection from Surveillance Images using Deep Learning

Anjali Goenka
Department of Computer Science and Engineering
National Institute of Technology, Tiruchirappalli
Tiruchirappalli, India
anjaligoenka25@gmail.com

K. Sitara
Department of Computer Science and Engineering
National Institute of Technology, Tiruchirappalli
Tiruchirappalli, India
sitara@nitt.edu

***Abstract*—Security is the biggest concern in today's world which needs to be addressed to save people from critical threats. We need to detect these threats at the earliest to protect people and take required actions. Security cameras are used almost everywhere now ranging from our home to shopping malls to banks. Currently, not many surveillance cameras have an automatic weapon detection system but with the advancement in technologies, it can be easily equipped. This will help the people in charge concerned to take the appropriate actions and thus prevent crimes. Deep learning techniques are used widely to detect objects as the traditional methods of object detection have their own limitations in certain situations. One such algorithm – Mask RCNN is implemented in this work to detect guns from surveillance video images. Gaussian deblur technique is used to enhance the features of handgun for efficient detection especially in blurred images. The experiment results show that the performance of the model increased with preprocessing.**

***Index Terms*—Convolution Neural Networks, Weapon Detection, Deep Learning, Mask RCNN**

I. Introduction

Everyday more than 500 people are killed all over the world by handheld guns, and more than 1000 people are wounded at their home or workplace. This survey is conducted by the non-profit Gun Violence Archive [1]. Gun violence has far-reaching consequences that go far beyond the lives of those who are killed. Millions of people's lives are changed because of it, and they live in terror of the subsequent mass killing. As a result, there is an increasing social desire for effective technology techniques to reduce violence.

CCTVs or surveillance/security cameras are nowadays located in every public place such as cross roads, schools, parking lots, business places, malls. Because the critical content can be missed and video screens may need continuous monitoring in surveillance, vulnerable circumstances may be missed [2]. Hence, an intelligent surveillance system is vital for detecting security hazards in live video streaming. Deep learning has made a name for itself in the field of machine learning in recent years, particularly in the areas of object recognition, classification, and image segmentation.

The main motivation behind this work is the need for automatic handheld gun detection in the surveillance system as compared to the traditional manual detection system. The traditional methods require continuous monitoring which is not a feasible solution in many scenarios. Various problems are encountered while detecting guns manually. These mainly arise due to the noises in the input image and the quick response required for real time processing of the frames in the videos.

In view of all the above problems and disadvantages of the previous system and methodology, we propose an automatic gun detection system based on deep learning techniques. We propose a pretrained Mask RCNN (Mask Region – based Convolution Neural Network) model [4]. For object instance segmentation, it uses deep CNN (Convolution Neural Network) based learning model. In Mask RCNN, while performing segmentation mask for each instance it can detect the objects of interest in given frame.

The focus of this work is to build an automatic hand gun detection system that can be equipped with the surveillance cameras. Transfer learning technique is used to train our model and various test cases such as blurred images, images with multiple guns are considered from the database[13]. This work can be used in any place where crime activities can take place such as in homicide situations.

The paper is further organized as follows. Section II discusses in detail about the literature survey performed. This helps to conclude Mask RCNN as the suitable algorithm to achieve our objective. Section III discusses about preprocessing steps, along with the proposed work and architecture of the model. Section IV contains the dataset and screenshot of results and all the analysis and evaluation done for understanding the performance of our proposed work. Finally, conclusions and future works are given in Section 5.

II. Literature Survey

The research in handgun detection in images or videos can be broadly classified into two categories –

- Detection using X-ray or millimeter wave images.
- Machine learning based object detection.

Recently deep CNN models are used for detecting weapons. A few works are published in this area.

A. Traditional Methods of Gun Detection

A number of research papers are published as a consequence of research in weapons recognition utilizing 2D X-ray imaging[19]. Nercessian et al. [3] explored the detection of firearms using 2D X-ray images of baggage. The approach characterizes handgun characteristics via edge detection; however, it only works with pistols in a fixed orientation. Similarly, the authors Fliton and Brekon in [5] compared the method with simple descriptors and a more complex RIFT/SIFT solution where the former gives a better performance results. Border/edge detection and pattern matching [6], cascade classifiers with boosting [7] are some other methods which achieve good performance. However, all these methods take time to process and will not be suitable for real time processing as the time taken to identify the guns require 3-5 minutes.

B. Machine Learning Methods

In [8], the author Halima et al. described a method which uses K-means clustering and SVM(Support Vector Machine) as weapon classifier. They discussed an automatic surveillance system for identifying fire weapons in a clustered situation. The first step is to extract SIFT (Scale Invariant Feature Transfom) features from the images and cluster them using K-means clustering. Then, by counting the occurrences of the extracted clusters in each image, a word vocabulary-based histogram is created. This feature is applied to the SVM. Finally, the system classifier determines if a test image contains a weapon or not. This method exhibits high accuracy in images and videos. However, this method failed to detect multiple guns in the same image. The authors in [12] uses K-means clustering algorithm to develop an automatic visual gun detection system. It uses color-based segmentation and Harris interest point detector. This work uses traditional machine learning method and work even for occluded guns in the images. However, the time and space complexity required for this method is more and it gives poor results with illumination changes in the recorded visuals.

C. Advancement to Deep Learning Methods

The work done by authors in [9] is one of the first to apply deep learning models for weapon detection system. The research was carried out on high quality videos which are of high contrast, brightness and better resolution which is normally missing in CCTVs or any other surveillance videos. In this research, the authors found that Faster R-CNN method gave best performance as compared to other deep learning algorithms which was used for analysis. For real time detection system with reduced false positive, alarm will be activated if visuals of gun is present in 5 consecutives frames. This research shows good results even in low quality videos. However, the number of false positives for gun detection in dataset[20] is high.The recall value obtained was less.

For another type of weapon – knives, the authors in [10] developed a weapon detection system which could detect cold steel in real time for surveillance videos. They proposed various algorithms involving classifiers based upon Convolution Networks and region selection techniques. The model was designed for indoor scenarios which uses surveillance videos and it provided excellent results as a real time knife detection system. YOLO version 4 stood out the best among all the other deep learning-based CNN algorithms used in [11]. The authors implemented binary classification as pistol and non-pistol classes and applied the open-source deep learning algorithms such as YOLO version 3, Faster RCNN – Inception ResNetV2, SSD-MobileNet-v1 to analyze. The disadvantage of the system was that it considered white background as a part of the pistol. The author suggested that this is due to training data containing white background in many images. YOLO (You Only Look Once) models are also used for detecting objects in various research. The authors try to improve the models for occluded and small objects and implement the same for traffic signal detection[17][18].

As discussed above, all the research system implemented for weapon detection is slow in nature and cannot be extensively used for real-time detection system. A few methods require supervision of an individual and others require high quality images or videos for good performance. More researchers has to focus in this area of weapon detection where other deep learning CNN algorithms can be analyzed or tuned to increase the system performance.

III. PROPOSED SYSTEM

Based on the literature survey carried out, we can see that deep learning techniques were adapted only recently for weapon detection. We have proposed the deep learning technique – Mask RCNN for handgun detection in this work. First, the images in the dataset are preprocessed using various techniques. To these preprocessed images, we used the Mask RCNN model. This model is pre-trained and is used to create segmentation masks on the RoIs (Regions of Interests) in the input image.

A. Methodology

The proposed architecture of the entire work is shown in Fig. 1.

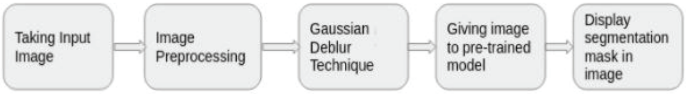

Taking Input Image
Image Preprocessing
Gaussian Deblur Technique
Giving image to pre-trained model
Display segmentation mask in image

Fig. 1: Block Diagram of the proposed system

In the above system, we take an input image and apply different preprocessing techniques – such as resize, flip, shift etc. Then we perform Gaussian Deblur technique for image sharpening. The resultant image is passed to the trained model where it gives bounding boxes around the guns, confidence value and the segmentation masks.

B. Preprocessing Steps

Images contain various properties or features such as size, dimensions, color, resolution, etc. Some images are in RGB, and some are in Grayscale. Some images are movie shots, and some are images from surveillance videos. Considering all these variations, data preprocessing is performed on the images will be input to the model, which is pre-trained on the COCO dataset[4].

The training model is fed with images of the same size, i.e., 200x200 pixels. It is the smallest size possible that can be assigned to generate images of the same size. This helps to reduce the memory required and thus the time needed to process the image. A value smaller than this will cause the image quality to degrade and impact the performance. After this, data augmentation is performed on the training dataset. To minimize the noise in the images, a Gaussian deblur filter is applied to deblur an image using a Gaussian function. Equation-1 represents the Gaussian kernel in 2-dimensional form. It's like a nonuniform low-pass filter that keeps the spatial frequency low while decreasing visual noise and minor characteristics.

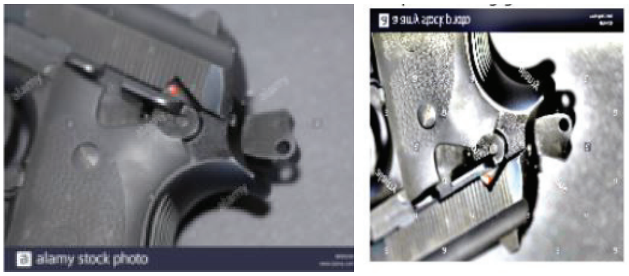
$$G_{2_D}(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-x^2-y^2}{2\sigma^2}} \tag{1}$$

where σ signifies the standard deviation of the distribution; and x and y are the location indices. The value of σ controls the variance around a mean value of the Gaussian distribution. This value will then further determine the extent of blurring/deblurring effect around a pixel.

We use the function as given in equation 2 to convolve the image using Gaussian filter. One way to add two images is to mix them so that the two appear together.

$$F(x) = \alpha * f_1(x) + \beta * f_0(x) + \gamma \tag{2}$$

Different values of α, β and γ can be assigned to the images. This allows the image to be transparent or translucent depending on the added weight. For getting deblurred image, we set 4, –4 and 128 for α, β and γ respectively by trial and error approach. F(X) denotes the image obtained by the preprocessing step. Source image is given as input in $f_1(x)$ and the blurred version of the image is given as input in $f_0(x)$. This helps to enhance the features of the image and make it deblurred.



(a) Original Image

(b) Resultant image after preprocessing

Fig. 2: Preprocessing Technique - a) Original Image b) Resultant image after preprocessing

The original image is blended with the blurred image of the input image. The output image obtained has clear features compared to the original image. Figure 2(a) and 2(b) corresponds to the images before and after preprocessing techniques respectively.

C. Proposed Model - Mask RCNN Architecture

Mask R-CNN model is an advancement of the Faster R-CNN model which is divided into two parts - object detection and semantic segmentation. Using instance segmentation, it can perform pixel level identification on object. By using CNN feature generator, it can extract the required features from the given input image. On the extracted features, it applies Region Proposal Network(RPN) to create the region proposal map. RoI feature maps are covered with boundary boxes and are wrapped into fixed dimension using pooling process. Pooled boxes are transferred to fully connected CNN to extract the classification and boundary box prediction. Figure 3 shows the the architecture of the Mask RCNN architecture[14].

The CNN structure used here is Feature Pyramid Network(FPN). It is a feature extractor created with accuracy and speed for such pyramid concepts. It provides many multi-scale feature maps with higher useful content for object detection than the normal feature pyramid FPN network is capable of detecting very small objects. In our experiment, we deployed a single backbone using ResNet 101.

We extract numerous feature map layers and send them to a Region Proposal Network(RPN) for object detection using the FPN as a feature detector. RPN performs 3 × 3 convolutions on the feature maps before performing a separate 1 × 1 convolution for class predictions and boundary box regression. RPN head is made up of 3 × 3 and 1 × 1 convolutional layers. On all feature maps, the same head is used.

IV. EXPERIMENTAL RESULTS

A. Dataset

Handgun Dataset was obtained from Soft Computing and Intelligent Information Systems [13]. Dataset contains images with guns, no guns, multiple images of guns from movie screenshots, movie posters, and even surveillance videos. 2844 images were annotated using VGG Image Annotator(VIA). These annotated images were used for object detection in Mask RCNN. The resolution of the images varies from 300x150 to 500x250 pixels. Figure 4 shows some of the images from the dataset such as image from surveillance video as in Fig. 4(c), image with a blurred gun as in Fig. 4(b) and a high quality image of a gun as in Fig. 4(a).

B. Experimental Evaluation and Analysis

We have trained our proposed model using 80% of the original dataset and tested it against the remaining 20%. Results obtained were measured with various performance factors such as Precision, Accuracy and Loss values[14].

The precision ratio is calculated by dividing the True Positives (TPs) with the sum of TPs and False Positives (FPs). This will describe the performance of the proposed system by correctly identifying images containing guns out of all the images that are having guns.

$$Precision = \frac{TPs}{TPs + FPs} \tag{3}$$

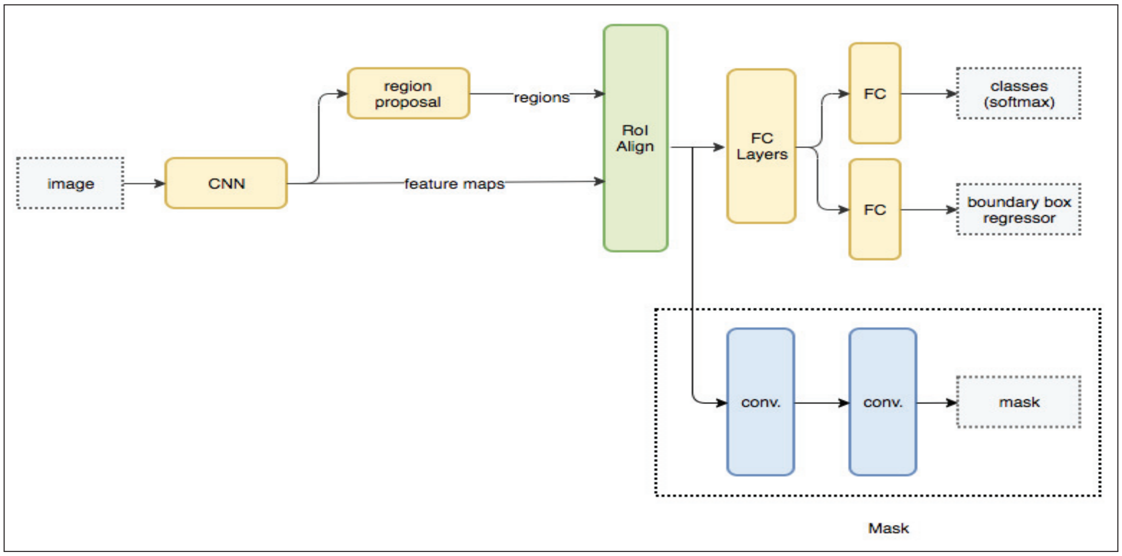
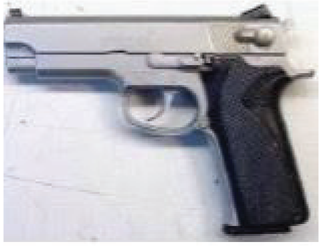
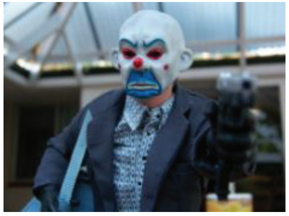


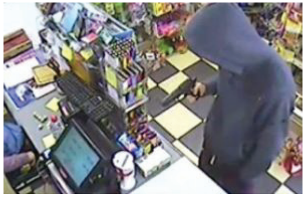
Fig. 3: Mask RCNN Architecture



(a) High quality image and gun with close look



(b) Image with blurred gun



(c) Surveillance Camera Frame

Fig. 4: Examples from Dataset

Recall ratio is calculated by dividing the TPs with the sum of TPs and False Negatives (FNs). This means that all the images which contained guns in the database will be identified as having guns.

$$Recall = \frac{TPs}{TPs + FNs} \tag{4}$$

Accuracy is calculated as the ratio of correct predictions with the total number of images in the database.

$$Accuracy = \frac{TPs + TNs}{TPs + FPs + FNs + TNs} \tag{5}$$

F1 value is calculated by taking the harmonic mean of the precision rate and recall rate. Formula for calculating F1-measure is as follows:

$$F1 = \frac{2 * Recall * Precision}{Precision + Recall} \tag{6}$$

C. Results

Out of 2844 images, 2355 images are correctly marked whereas 489 images are incorrectly marked. Thus, our model has a precision of 88.45% and has a recall of 81%. Accuracy of the system is 82.76% and its F measure is 84.69%. Amount of memory required to store the dataset decreases after reducing the image size. This improved the space complexity of the model. Also, the model takes less time to extract the features of the image due to preprocessing steps on the blurred image. Hence the time complexity of the model and its overall performance improves. Table I shows the comparison results of the proposed Mask-RCNN model with and without applying Gaussian deblur feature on dataset[13].

From the result analysis in Table I, we can see that precision value has improved by 5%. This is due to reduced number of false positives obtained from Mask R-CNN model compared to the Faster R-CNN method. The accuracy of Mask R-CNN model is lesser due to the segmentation mask which masks each pixel of the object. In Faster RCNN [9], background of the image inside the bounding box is considered during training. This results in more accuracy and an increased number of false positives. In Mask RCNN, pixels of the gun are considered. Mask RCNN tends to mask extra pixels (refer Fig. 7) or leave out some pixels depending on the shape of the gun resulting in lower accuracy. To improve this, we should include more guns of different shapes and sizes such as machine guns, assault rifles, snipers and shotguns in the dataset.

TABLE I: Evaluation of the input for Mask RCNN model

Model-Feature	Recall (%)	Precision (%)	F1 Score (%)	Accuracy (%)
Mask RCNN (With Gaussian deblur feature)	81.25	88.45	84.69	82.76
Mask RCNN (Without Gaussian deblur feature)	80.55	82.92	81.72	80.82
Faster RCNN - selective search method with VGG-16 based classifier[9]	100	84.21	91.43	90.62

In surveillance videos, images contain lots of noise, and detecting metal guns in black and white images becomes more difficult. Moreover, the images are blurred due to out of focus and low-quality recording. Hence, the Mask RCNN model was unable to detect guns in such images. For this we applied an addweighted function available in OpenCV library, where the original image is preprocessed by inverse filter using Gaussian filter to give a deblurred image which consists of intensified edges of the handgun. After addition of this feature, overall accuracy of the model is increased by approx. 2%.

Table II shows the Mask Loss and Class Loss before and after applying Gaussian deblur feature. Though the values obtained are very less, the model gives a comparatively high RPN loss value. These results also show that detecting guns is difficult compared to other weapons such as knives as various objects especially which are black in colour and have metallic effects can easily be confused as handguns.

TABLE II: Comparison Table with Loss values difference before and after applying Gaussian feature

Approach used for MASK RCNN: ResNet101(COCO)	After Gaussian Filter - Deblur	Before Gaussian Filter - Deblur
Validation Mask Loss	0.1909	0.2795
Validation Class Loss	0.0384	0.0786

Figure 5 shows the training loss during last epoch. 100 iterations are performed during a single epoch. From Fig 5, we can see that for the initial 40 iterations, training loss increased and then decreased.

Some sample images and the output obtained are shown in Fig. 6, Fig. 7 and Fig. 8.

Test Case 1: For high quality images and close look, guns are detected with high accuracy and reduced false positives. Figure 6 shows the output with an occluded gun.

Test Case 2: For images with multiple guns, confidence value of detection is high. The model detects all the guns and no other object is marked as gun. Thus, no false negative in Fig. 7.

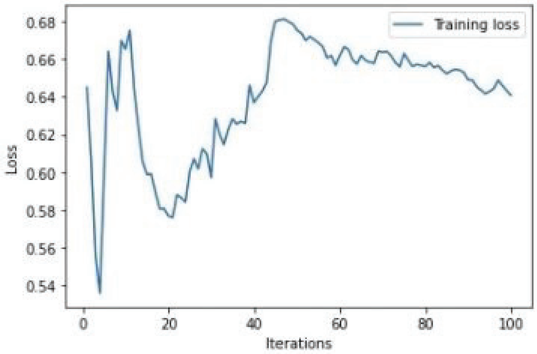


Fig. 5: Training loss curve across 100 iterations.

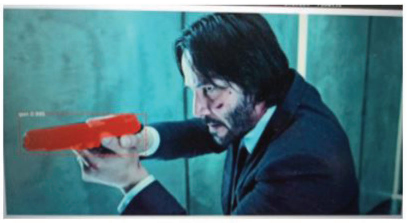


Fig. 6: Testing Result 1 - Output image with occluded gun

Test Case 3: For a black and white image, it is observed from Fig. 8 that the weapon is detected with good confidence value.

Test Case 4: For an image with blurred gun (ref Fig. 4(b)), it can be observed from Fig. 9 that the model detects the gun with high confidence value.

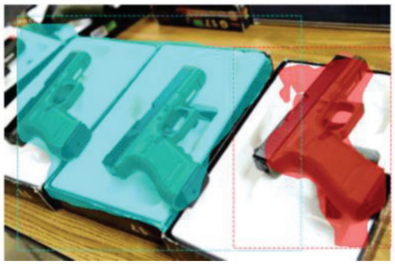


Fig. 7: Testing result 2 - Output image with multiple guns

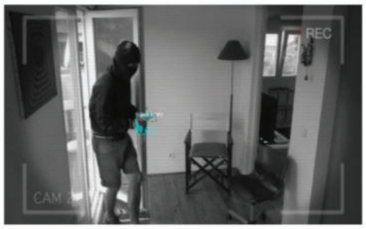


Fig. 8: Testing result 3 - Output image with GrayScale

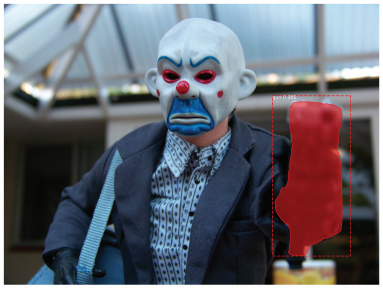


Fig. 9: Testing result 4 - Output image with blurred gun

V. CONCLUSION

We have developed and implemented a portable gun detection technique for an effective video surveillance system. The system combines a variety of preprocessing techniques with Mask RCNN to provide the best results. Our method detects the presence of multiple weapons in real-time and is unaffected by changes in size, scale, rotation, affine, and even partial occlusion. By adopting this technique, the system performance is improved, and real-time processing requirements such as space and time complexity can be reduced. In the future, models such as YOLOv4 can be applied [15] [16], and even different categories of guns can be included in the dataset to improve the performance for real-world applications.

REFERENCES

[1] Gun Violence Archive, Apr. 2021, [online] Available: https://www.gunviolencearchive.org/

[2] G. F. Shidik, E. Noersasongko, A. Nugraha, P. N. Andono, J. Jumanto and E. J, ”A systematic review of intelligence video surveillance: Trends techniques frameworks and datasets”, IEEE Access, vol. 7, pp. 457-473, 2019.

[3] S. Nercessian, K. Panetta, S. Agaian, ”Automatic detection of potential threat objects in X-ray luggage scan images”, in: Proceedings of the IEEE Conference on Technologies for Homeland Security, pp. 504–509, 2008.

[4] K. He, G. Gkioxari, P. Dollár and R. Girshick, ”Mask R-CNN”, Proc. IEEE Int. Conf. Comput. Vis., pp. 2980-2988, Apr. 2017.

[5] G. Flitton, T.P. Breckon, N. Megherbi, ”A comparison of 3d interest point descriptors with application to airport baggage object detection in complex ct imagery”, Pattern Recogn. 46 (9) 2420–2436, 2013.

[6] R. Gesick, C. Saritac, C.-C. Hung, ”Automatic image analysis process for the detection of concealed weapons”, in: Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies, ACM, p. 20, 2009.

[7] Z. Xiao, X. Lu, J. Yan, L. Wu, L. Ren, ”Automatic detection of concealed pistols using passive millimetre-wave imaging”, in: 2015 IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, pp. 1–4, 2015.

[8] N.B. Halima, O. Hosam, ”Bag of words-based surveillance system using support vector machines”, Int. J. Secur. Appl. 10 (4), pp. 331–346, 2016.

[9] R. Olmos , S. Tabik , F. Herrera , ”Automatic handgun detection alarm in videos using deep learning”, Neurocomputing 275, pp. 66–72, 2018.

[10] Alberto Castillo, Siham Tabik, Francisco Pérez, Roberto Olmos, Francisco Herrera, ”Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning”, Neurocomputing, Vol 330, pp 151-161, 2019.

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

[12] R. K. Tiwari and G. K. Verma, ”A computer vision based framework for visual gun detection using SURF”, Proc. Int. Conf. Electr. Electron. Signals Commun. Optim. (EESCO), pp. 1-5, Jan. 2015.

[13] Andalusian Research Institute in Data Science and Computational Intelligence - https://dasci.es/transferencia/open-data/24705/

[14] Gonzalez, Sebastian & Arellano, Claudia & Tapia Farias, Juan, ”DeepBlueBerry: Quantification of Blueberries in the Wild Using Instance Segmentation”, IEEE Access, PP. 1-1, 2019.

[15] A. A. Ahmed and M. Echi, ”Hawk-Eye: An AI-Powered Threat Detector for Intelligent Surveillance Cameras”, in IEEE Access, vol. 9, pp. 63283-63293, 2021.

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

[17] Y. Li, S. Li, H. Du, L. Chen, D. Zhang and Y. Li, ”YOLO-ACN: Focusing on Small Target and Occluded Object Detection”, in IEEE Access, vol. 8, pp. 227288-227303, 2020.

[18] C. Dewi, R. -C. Chen, Y. -T. Liu, X. Jiang and K. D. Hartomo, ”Yolo V4 for Advanced Traffic Sign Recognition With Synthetic Training Data Generated by Various GAN”, in IEEE Access, vol. 9, pp. 97228-97242, 2021.

[19] M. Singh, S. Singh, ”Image Segmentation Optimisation for X-Ray Images of Airline Luggage”, IEEE CIFHSPS, pp. 10-17, 2004.

[20] Dataset Link: https://github.com/SihamTabik/Pistol-Detection-in-Videos