Proc. of the 2019 IEEE 6th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA 2019)
27-29 August 2019, Kuala Lumpur, Malaysia

Gun detection system using YOLOv3

Arif Warsi[1], Munaisyah Abdullah[2], Mohd Nizam Husen[3], Muhammad Yahya[4]

Sheroz Khan[5], Nasreen Jawaid[6]

[1, 2, 3] *Universiti Kuala Lumpur, Malaysian Institute of Information Technology, Malaysia*
[1]arif.warsi@s.unikl.edu.my, [2]munaisyah@unikl.edu.my,
[4] *Universiti Kuala Lumpur, British Malaysian Institute Malaysia*
[5] *Kulliyah of Engineering, International Islamic University Malaysia*
[6] *Ilma University Karachi, Pakistan*

***Abstract*— Based on current situation around the world, there is major need of automated visual surveillance for security to detect handgun. The objective of this paper is to visually detect the handgun in real time videos. The proposed method is using YOLO-V3 algorithm and comparing the number of false positive and false negative with Faster RCNN algorithm. To improve the result, we have created our own dataset of handguns with all possible angles and merged it with ImageNet dataset. The merged data was trained using YOLO-V3 algorithm. We have used four different videos to validate the results of YOLO-V3 compared to Faster RCNN. The detector performed very well to detect handgun in different scenes with different rotations, scales and shapes. The results showed that YOLO-V3 can be used as an alternative of Faster RCNN. It provides much faster speed, nearly identical accuracy and can be used in a real time environment.**

Keywords—YOLOV3; Handgun detection; False Positive;

I. INTRODUCTION

The number of crimes involving guns violence is increasing in many parts of the world, especially in that places where the possession of guns is legal. [1] Reported incidents caused by guns violence in America in year 201, 2016 and 2017 were 333752, 58908 and 61721 respectively. Another study ranked Malaysia in top 10 countries having highest gun violence in East, South East and South Asia in year 2016 [2]. From these statistics, it can be assumed that violence rate concerning guns are increasing every year becoming a challenge for law enforcement agencies to overwhelm this issue timely.

The solution of this problem is monitoring and early detection using control camera or surveillance systems which can help to prevent this kind of violence and assist policemen or security agents to act timely.

Recently, the area of machine learning mainly in detection and classification of objects and image segmentation has been revolutionized by deep learning. You look only once (YOLO) outperformed other detection algorithm at predict in images what objects they are [3].

In this paper, we present an automatic gun detection system using deep learning mainly YOLOv3 which is compared with the results of [4]. Detecting gun in a scene is very challenging issue because of various subtleties linked with it. The challenges in gun detection is particularly caused by occlusion. Gun to scene and gun to object are the two types of gun occlusion. Real time processing is another main problem in gun detection system that arises during detection. The rest paper is structured as follows: section 2 describes related work, section 3, section 4, section 5.

II. RELATED WORK

Mostly, gun detection research specially emphases on hidden weapon detection and knife detection. Hidden weapon detection is based some techniques of imaging like millimetre wave imaging, infrared imaging used in airport for luggage (containing gun and knife) control applications.

The research work in [5] proposed a visual gun detection system based on Harris interest point and SIFT. They used color based segmentation to select dissimilar object from an image by deploying K-mean cluster algorithm. In [6], researchers used 3D millimetre wave imaging technique to detect the weapon concealed in the body and other hidden location.

In another work [6], researchers deploy a real time gun classifier that detect and classify guns. They also used imageNet dataset for training their model and acquired a mAP of 89% using overfeat-3 algorithm. The research work [7], is based on terahertz human dataset used in deep learning to detect the concealed weapon. The achieved a competitive accuracy compared to other concealed weapon detection system.

[8], used numerous networks like YOLOv2, Sliding window-based CNN, region-based fully convolutional networks (R-FCN), faster region-based CNNS (F-RCNNS) with transfer learning for image classification and detection problems. They showed empirically that fine-tuned CNN features give greater performance than conventional algorithms. The image dataset used in training the model is based on X-rays images.

In [4], the researchers used faster RCNN with VGG-16 based classifier for detecting guns in in YouTube videos and achieved a maximum mean average precision (mAP) of 84.21%. They used ImageNet dataset for training a modelling a handguns detection system.

III. METHODOLOGY

Methodology has been organized as follow:

A. DATASET

We have created our own image dataset containing guns with different position and orientation and merged it with ImageNet dataset. YOLO v3 algorithm has been trained and validated to evaluate our gun detection system for better results. Figure 1 shows some of the examples from the dataset being used for training our model. Figure-1 shows some of the images from ImageNet datasets including the image we merged.

B. YOLO-V3 ALGORITHM

In this work, we have used transfer learning for training YOLOv3 model for gun detection and used the weight trained on ImageNet by YOLOv3 team instead of starting from zero.

YOLOv3 is an object detection algorithm widely used for real time processing. Input image is divided into $M \times M$ grids. A single object is then predicted by this grid cell.

Logistic regression is used to predict an object scores for each bounding box by YOLOv3 and changes the method to compute the cost function. If a ground truth object is early overlapped by a bounding box prior more than others, the resultant object score should be 1. No cost is experienced for other prior with overlap greater than predefined threshold 0.5.

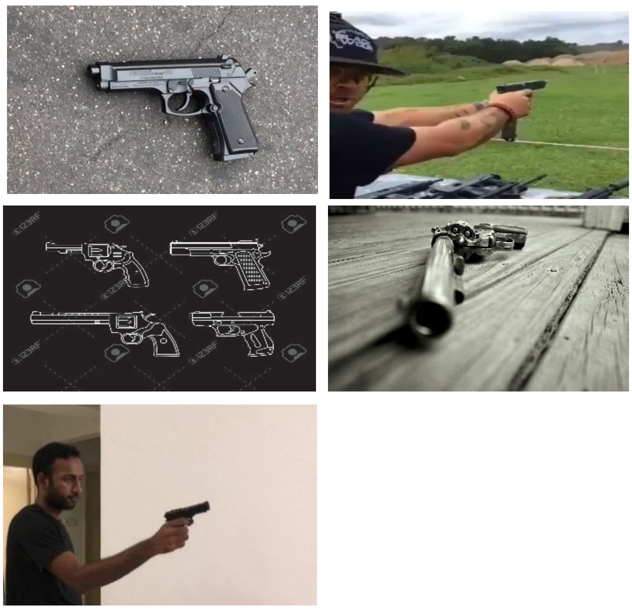

Figure 1: images used for training YOLOv3 model.

Only a single boundary box prior is associated with each ground truth object. If bounding box prior is not allocated, no classification and localization lost incurs, only confidence loss on objects incurs. We compute the loss by using the following equations.

$$l_x = \sigma(t_x) + c_x$$

$$l_y = \sigma(t_y) + c_y$$

$$l_w = p_w e^{t_w}$$

$$l_h = p_h e^{t_h}$$

$$Pr(\text{Object}) * \text{IOU}(\text{b}, \text{object}) = \sigma(t_o)$$

Where l_x and l_y are loss function and l_y and l_h are generated from sigma function with the range 0 and 1. The network calculates 5 coordinates for each bounding box t_x, t_y, t_w, t_h and t_o.

Note that loss function only corrects the classification error in the same grid cell.

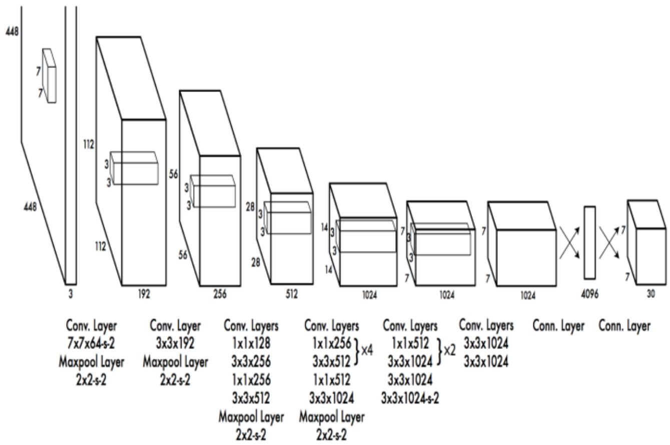

Figure 2: shows the block diagram of YOLO-V3 algorithm

IV. RESULTS AND DISCUSSION

Detecting handguns in videos not only focus on gun detection but also emphasis on minimizing the number of false positives and providing real time detection.

We have estimated YOLO-V3 based classification model considering a single class to specify the presence of handgun. YOLO has been used by researchers for detecting different objects related to their interest. However, this paper is to check how the YOLOv3 neural network behave when there is occluded background in detecting the handgun. We selected four videos as benchmark from the research work [4]. They have used Faster RCNN with VGG16 algorithm for detecting handguns in different videos. We used theirs ImageNet dataset and merged with our dataset. The dataset has been trained on YOLOv3.

The detection results are examined frame by frame in the videos during the experiments and measured a detection as true positive if the gun and bounding box overlapping is more than 50%. If the human eye recognizes the gun is considered as ground truth. Table 1 depicts the True positive #TP, False Positive #FP and Ground Truth #GT.

Based on the videos used in benchmark, YOLO shows high number of true positive in video 1, 3, 6 and less false positive in videos 1 and 6 as shown in table 1.

Table 1: validating the results of YOLOv3 with Benchmark [4] for handgun detection

Video Id	frames	#TP	#GT_P	#FP	Precision	Recall	F1 Measure	Technique
1	516	73	220	1	98.64%	33.18%	**66.36%**	Ours
	393	60	162	8	88.24%	37.04%	52.17%	Benchmark
2	692	276	959	41	87.06%	28.77%	**43.26%**	Ours
	627	467	778	11	98.7%	60.03%	74.36%	Benchmark
3	523	33	65	46	41.77%	50.76%	**45.833%**	Ours
	441	25	58	15	62.5%	43.1%	51.02%	Benchmark
6	267	249	402	8	96.51%	61.94%	**75%**	Ours
	212	141	290	30	82.46%	48.62%	61.17%	Benchmark

The detector has achieved a good balance in terms of precision and F1 measure in video 1 and 6 Figure 3 shows the YOLOv3 detector has low false positive in all the benchmark videos.

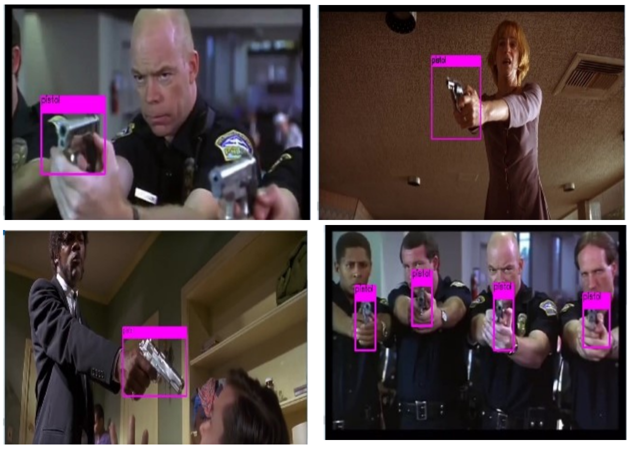

Figure 3: an example of accurate handguns detected in four different frames

False negative can be used to explain low recall in the videos frame which is due to the detection of handguns in occluded environment or when the handgun is fastly moved.

The false negatives strongly rely on the frame quality and when the handgun is visible clearly.

V. Conclusion

In this paper, we evaluated the performance of the YOLOv3 based detector on four different videos. The objective was to minimize the false positive using YOLOv3 algorithm. YOLOv3 based model has been trained with a dataset containing ImageNet and our own customized dataset. It is clear from the results that YOLOv3 has a good detection performance even in low quality videos as than faster RCNN. The advantage of YOLOv3 over Faster RCNN is its speed. The processing speed of YOLOv3 is 45 frames per seconds while Faster RCNN has 8 frame per seconds. Two out of four videos showed better accuracy than Faster RCNN algorithm.

References

[1] Gun Violence Archive, “Past Summary Ledgers : Gun Violence Archive,” [Online].

[2] Muhammad Amin, B., K. Mohammad Rahim, and M. S. Geshina Ayu. "A trend analysis of violent crimes in Malaysia." *Health* 5, no. 2 (2014): 41-56.

[3] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali 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.

[4] Olmos, Roberto, Siham Tabik, and Francisco Herrera. "Automatic handgun detection alarm in videos using deep learning." *Neurocomputing* 275 (2018): 66-72.

[5] Tiwari, Rohit Kumar, and Gyanendra K. Verma. "A computer vision-based framework for visual gun detection

using harris interest point detector." *Procedia Computer Science* 54 (2015): 703-712.
[6] Arceda, V. Machaca, K. Fernández Fabián, and Juan Carlos Gutíerrez. "Real time violence detection in video." (2016): 6-7.
[7] Akcay, Samet, Mikolaj E. Kundegorski, Chris G. Willcocks, and Toby P. Breckon. "Using Deep Convolutional Neural Network Architectures for Object Classification and Detection Within X-Ray Baggage Security Imagery." *IEEE Transactions on Information Forensics and Security* 13, no. 9 (2018): 2203-2215.