2022 Fifth College of Science International Conference of Recent Trends in Information Technology (CSCTIT) | 979-8-3503-3492-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/CSCTIT56299.2022.10145737

A Multi-Weapon Detection Using Synthetic Dataset and Yolov5

Moahaimen Talib Abdullah
Computer Science Dept.,College of Science,
Mustansiriyah University –
Baghdad, Iraq
moahaimen@uomustansiriyah.edu.iq

.Jamila Harbi ALameri
Computer Science Dept.,College of Science,
Mustansiriyah University
Baghdad, Iraq
Dr.jameelahharbi@uomustansiriyah.edu.iq

***Abstract*—With increasing terrorist and criminal activities around the world, detecting several types of weapons at the same time has become crucial. Especially most of these groups are used lightweight weapons which are easily lifted and run with them. For that reason, finding an intelligent system capable of discovering these types of weapons with high quality and speed becomes of utmost necessity. This paper focuses on utilizing the deep learning networks technique (namely YOLOV5) to detect various weapons with high accuracy and speed. YOLOV5 is a member of compound-scaled object detection, which is commonly trained on the Microsoft COCO dataset. However, the proposed system was trained on the self-collected dataset, which consists of sixty thousand images of different types of guns used in Iraq. Thus, the proposed system has the ability to detect different types of weapons at the same time. Furthermore, the system can achieve high performance by increasing the number of trained images for every kind of weapon. The results showed a high mAP reach of 97% of some types of weapons.**

KEYWORDS—Weapon Detection;Cnn,Yolov5;deep learning; Anti-Terrorism

I. INTRODUCTION

Multiple Detection of weapons is so crucial these days as of the great danger around the world while many attacks happen yearly, relatedtoterrorists or criminal activities such as bank robbery [1]. Since the suspected individuals may carry different types of weapons, it is very important to detect each weapon and its type to understand the type of danger that dealing with, and enable the security forces to respond according to the level of danger the type of weapon presents, although there are millions of cameras spreading in streets, offices, moles, and banks or other institutes, it is still difficult to detect or identify the weapons even for the most trained security personnel especially if he is watching many monitor screens[2],[3].

It is crucial become a high requirement to build a system that can detect and identify the type of weapon at high speed and accuracy.in order to send an alarm to security personnel to make them respond to the danger. using deep learning, feature extraction, and conversion, it employs multiple layers of nonlinear processing units [4]. The deep learning framework is built on the learning of multiple levels of data features. Deep learning is based on learning from the main data representation [5]. A vector of density values per pixel or features (such as clusters of edges) and custom shapes are used to represent an image, with some expressing the data better than others [6]. The Convolutional Neural Network (CNN), which comprises convolution, pooling, activation function, dropout, fully connected, and classification layers [7], represents the core architecture of the deep learning idea.

Recently, Convolutional neural network (CNN) based object detection techniques have made significant improvements in object detection fields. The effect is not sufficient given the persistently high false detection rate, high time overhead, and high computational overhead.

As a result, we use the YOLOv5 (You Only Look Once) [8] lightweight network model as the basis for our work which proposes a new model to detect seven different weapon types (AKM, M4, PKS, RPG, PISTOL, KNIFE, andSNIPER)the contribution was in collecting dataset which of 7 classes of weapons, and finding the highest “mAP, accuracy” Results in using yolov5x6 model.

The rest of the paper consists of section II related work, section III preprocessing, section IV yolo network model, Section V materials and methods, section VI experimental results, section VII discussion, section VII conclusion

II. RELATED WORKS

M. Milagro Fernandez-Carrobles, Oscar Deniz, and Fernando Maroto [9], proposed a method of Gun and knife detection based on Faster R-CNN for video surveillance, the main objective of his work is the development of an object detector that efficiently locates guns and knives in real-time video. For that purpose, an approach based on deep learning techniques and more especially through the Faster R-CNN methodology will be adopted. The object detection approach uses internally a CNN and a Regional Proposal Network (RPN) for the classification and location processes respectively, SqueezeNet architecture achieving an 85.45% accuracy.

Gyanendra K. Verma, and Anamika Dhillon [10], proposed a technique of automatic gun recognition from crowded scenes using Faster R-CNN Deep Learning, The Deep

Convolutional Network DCN, a state-of-the-art Faster Region-based CNN model, through transfer learning. The Internet Movie Firearms Database (IMFDB), a reference gun database, was used in the work to assess gun detection. For detecting the visual handheld gun, favorable system performance was achieved. Additionally, it shows how the CNN model increases classification accuracy when compared to a large number of training photos, which is especially useful in situations when a generous liberal is frequently unavailable, this work achieved 93% accuracy.

Neelam Dwivedi, Dushyant Kumar Singh, and Dharmender Singh Kushwaha [11] presented a method of providing a novel approach based on CNN for recognizing visual weaponry. he used data creation for visual weapon detection. The weights of the pre-trained Visual Geometry Group-16 (VGG-16) network are used to initiate this suggested CNN model. By exposing this CNN model to a wide variety of weapons (knives and handguns exclusively) and non-weapon pictures during training, these weights are further adjusted, this work achieved 97% accuracy for knives and 99% accuracy for handguns.

Olmos et al., SihamTabik and Francisco Herrera [12] examined two methods for automatic handgun detection: region proposal and sliding window with a histogram of gradients (with Faster R-CNN). They utilized a recently produced dataset to train both models (Pistol Detection). For evaluating detection models employed for automated detection technologies for movies, they created a metric (Activation Time for Alarms per Interval), this work achieved a precision of n 84,21%.

Gelana and Yadav [13] used the sliding window method to apply a CNN classifier that has been trained using background removal and edge detection. The training and testing datasets are constructed using high-quality CCTV videos, which is not always practical and restricts the use of the algorithm even though they show 93.84% accuracy.

Briefly, the above researchers attempted to classify firearms weapons, knives, and other types of weapons, but no research attempted to detect and recognize different types of weapons such as M4, and AKM or heavy weapons such as RPG using YOLOV5. This study proposed a model that detects and recognizes several types of weapons and our proposed model using YOLOV5 has more accuracy and more Frame Per Second [FPS]

III. YOLO NETWORK MODEL

Joseph Redmon introduced the real-time object identification system YOLO (You Only Look Once) in 2016. When he first began building the YOLO algorithm, another author, Alexey Bochkovskiy, released an article on it when it didn't make much progress. Subsequently, a succession of YOLO appeared, leading to YOLOV2, YOLOV3, and finally YOLOV4. On May 30, 2020, the Ultralytics LLC team released YOLOV5 [18] in the midst of YOLOV4. YOLOV5 switched from the darknet to Pytorch and achieved 140 FPS in Tesla P100, compared to YOLOV4's 50 FPS. The advantages and architecture of YOLOV5 are virtually identical to those of Yolo V4. However, Yolov5 is easier to train and use for object detection than Yolov4.YOLO's objective is to identify objects more quickly than conventional CNNs while maintaining accuracy [14]. It does this by just looking at the image once and treating object detection as a single task. The pipeline shrinks the input, uses a single CNN to process it, then thresholds the output based on the model's level of confidence [1]. To carry out the detection, YOLO divides the image into several sub-regions and assigns five anchor boxes to each. The zone with the highest likelihood is chosen [14] once the probability of a certain object is computed.

We utilize the most recent version of the YOLO family, YOLO-v5, in this work. Bounding box estimation and object identification were initially combined in one end-to-end differentiable network by the object recognition model YOLO. The setting in which YOLO is created and maintained is the darknet. The YOLO-v5 model, which is the first to be created using the PyTorch framework as opposed to earlier YOLO models [16], is lighter and easier to use than earlier YOLO iterations. A clever CNN is the foundation of YOLO-real-time v5's object identification system. The bounding boxes and probability for each zone are determined by this method after it divides the image into regions. These bounding boxes are weighted using the predicted probabilities [17]. The method "you only look once" at the image since it only requires one forward propagation pass through the neural network to conclude. After a non-max suppression (which ensures that the object detection algorithm only recognizes each object once), it then outputs known objects together with the bounding boxes.

The newest offering from YOLO, the YOLOv5 network, provides the benefits of high detection accuracy, quick detection speed, and lightweight features. In YOLOv5, there are primarily 4 models: the benchmark model YOLOv5l, the extended model YOLOv5x, and the preset simplified versions YOLOv5s and YOLOv5m. They differ primarily in the number of feature extraction modules and convolution kernels present in different regions of the network, as well as in the size and quantity of model parameters in this work the model Yolov5x6 was used [19].

The input, backbone, neck, and head networks make up the YOLOv5 network topology [20], as shown in Table I. The input terminal uses adaptive anchor, adaptive image scaling, and mosaic data augmentation, among other features. A CNN called the Backbone network [21] collects several fine-grained images to create image features. The focus module, CONV-BN-Leaky ReLU (CBL) module, CSP1 X module, and other modules make up the majority of its components. The Neck network, which is mostly used to construct FPN and PAN, is a collection of feature aggregation layers comprising mixed and integrated image information. The CBL module, Upsample module, CSP2 X module, and other modules make up most of it. The Head terminal recognizes the bounding box's loss function as GIoU_ Loss.

TABLE I : THE STRUCTURE OF THE YOLOv5 NETWORK.

YOLOv5	Features
Input	Mosaic data augmentation, adaptive anchor, adaptive image scaling
Backbone	Focus, CBL, 3×CSP1_X, SPP
Neck	CBL, 5×CSP2_X, Upsample, Concat, FPN+PAN
Head	GIoU_Loss

IV. PREPROCESSING.

After collecting images from google, the images were preprocessed by selecting the weapon image and the mask of each image, then applying python code changing the size to 416*416 pixels for all images, and auto-Orientation, auto-adjust contrast, after the preprocessing, other python code applied to mask the augmentation for each image so multiplying the number of images, to enhance the training, the training phase is done using YOLOV5X6 model. the training is done by 32 batch size, 50 epochs, with a learning rate equal to 0.01

V. MATERIALS AND METHODS

In this section, we explain the proposed system contains materials and methods that are used.

Since there is no standard weapon dataset available, the study used a Synthetic Dataset that is automatically generated. This method uses a python script to add a backdrop to cropped images of interesting objects. Using the same script, annotations are produced. With this method, we produce images that aren't quite real photographs but have completely realistic backgrounds and objects. When compared to a human procedure, an automated process allows us to build datasets significantly faster. For instance, it can be completed in less than an hour with 1000 synthetic photographs and comments. Compared to manually annotating 1000 different photographs, it is considerably faster.

AKM, M4, PKS, RPG, PISTOL, KNIFE, and SNIPER are the objects of interest. Cropped images and masks of these objects in various positions.Backdrop images (simply various images downloaded from the internet).Cropped images and masks of various items (such as vehicles, chairs, guitars, etc.) will be utilized as background noise to add complexity to the background.

60000 images are the total number of creating images, which all have an equal number of weapons in them, labeled and annotated in YOLOV5 format.

The algorithm used in this work consist of the following steps:

- Image preprocessing:(preprocessingsteps(resizing all images into unified size 416*416,auto-Orientation, auto-adjust contrast)
- Image augmentation
- Processing images
- Building the weapon dataset
- Training using the Yolov5x6 model
- Detection of weapons by inputting a real-time video.

To accommodate graphics processing units (GPUs) or central processing units (CPUs),Pytorch can be configured with two separate processors. On average, Pytorch programs perform better on GPUs than on CPUs. The model in this study the Yolov5x6 model training was on Google Colab pro+ with v-100 Nvidia GPU, and 58 Gb Ram

VI. EXPERIMENTAL RESULTS

Seven different types of weapons are the subject of the study's experiments, and the dataset used to train the network is split into three groups: 60% for training, 20% for testing, and 20% for validation. as shown in Table II

TABLE II DATASET FOR WEAPONS

Training	Testing	Validating
40000	20000	20000

The Yolov5x6 model was trained by 50 epochs, the model's input image had a resolution of 416x416 pixels and is formatted in RGB. While training the Yolov5 began to learn from the second epoch of 50 epochs of the Yolov5x6 model, which is excellent for the work of study it gave great learning through epochs, and with each epoch.

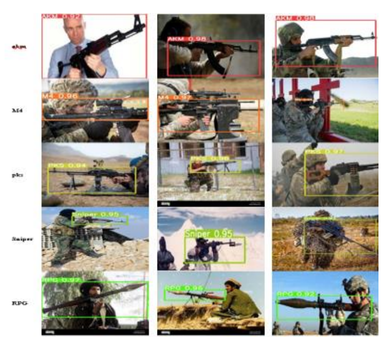

Fig. 1 Samples of Weapons Types Recognize by the Model

This work achieved the best results, the total precision was 99% and the Total Recall was 93% and the Total mAP is 95% which is very high comparing other works using Yolov5 on different backgrounds with many other objects.

This article also primarily assesses the recall and precision of item detection. The definitions of precision, recall, and mean average precision (mAP):

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$mAP = \frac{1}{C}\sum_{k=i}^{N} P(k)\Delta R(k) \quad (3)$$

k is the IoU threshold, Intersection over Union (IoU) [23] is a standard for object detection that identifies how closely the expected and real bounding boxes resemble one another. T. The equation can be used to explain IoU,whereIoU, a normalized index, has a value range of [0, 1]..

$$\mathrm{IoU} = \frac{\mathrm{area}(\mathrm{box}(\mathrm{Pred}) \cap \mathrm{box}(\mathrm{Truth}))}{\mathrm{area}(\mathrm{box}(\mathrm{Pred}) \cup \mathrm{box}(\mathrm{Truth}))} \quad (4)$$

P(k) is the precision, and R(k) is the recall, where C is the number of object categories, N is the quantity of IoU thresholds, k is the IoU threshold, and so forth.

VII. DISCUSSION

By creating a dataset that differs from other research in the literature and includes nearly all of the weaponry that terrorists employ in their assaults on people, the Yolov5x6 model is used in this work to prevent terrorist attacks. This will be utilized in autonomous security control, which can identify and detect the type of weapon with fast speed and high accuracy, using the AKM, M4, RPG, PKS, Dragunov Sniper, Pistol, and Knife. Training, validation, and test datasets are created using the data utilized in the dataset for the generated model's training procedure. The proposed model is trained using the training and validation dataset in the first stage. A reasonable rate of classification success is attained when the model trained in the secondstage is tested using a test dataset that the model is unable to observe. To evaluate how well the proposed model performs in terms of availability and categorization in the actual world, our proposed model has also been tested using images of several classes of weapons discovered on humans and weapons alone and it is tested on real-time videos.

. The importance of this work is very high-speed detection weapons due to the use of Yolov5 which is the fastest object detection algorithm, and also the accuracy of detecting weapons. The proposed model of this work can be used on computers with medium computer specifications

VIII. CONCLUSIONS

By improving the task efficacy and efficiency of security forces in security applications, the model put out in this work will help close many security gaps. In this way, the study's initial value is increased by the fact that the findings can be applied straight to new applications. The study's findings are anticipated to inspire and inform subsequent research, particularly with reference to autonomous security units.

In future work, the types of weapons will be increased and the future will make it possible to make this model on drones to detect the terrorist gangs from above and specify all types they use to attack civilians

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection." [Online]. Available: http://pjreddie.com/yolo/

[2] Z. K. Abbas and A. A. Al-Ani, "Anomaly detection in surveillance videos based on H265 and deep learning," International Journal of Advanced Technology and Engineering Exploration, vol. 9, no. 92, pp. 910–922, 2022, doi: 10.19101/IJATEE.2021.875907.

[3] C. Asia-Pacific Signal and Information Processing Association. Annual Summit and Conference (2019 : Lanzhou Shi, Asia-Pacific Signal and Information Processing Association, IEEE Signal Processing Society, and Institute of Electrical and Electronics Engineers, 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC).

[4] V. Kaya, S. Tuncer, and A. Baran, "Detection and classification of different weapon types using deep learning," *Applied Sciences (Switzerland)*, vol. 11, no. 16, Aug. 2021, doi: 10.3390/app11167535.

[5] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–27, 2009, doi: 10.1561/2200000006.

[6] H. A. Song, B. K. Kim, T. L. Xuan, and S. Y. Lee, "Hierarchical feature extraction by multi-layer non-negative matrix factorization network for classification task," *Neurocomputing*, vol. 165, pp. 63–74, Oct. 2015, doi: 10.1016/j.neucom.2014.08.095.

[7] S. Masood, U. Ahsan, F. Munawwar, D. R. Rizvi, and M. Ahmed, "Scene Recognition from Image Using Convolutional Neural Network," in *Procedia Computer Science*, 2020, vol. 167, pp. 1005–1012. doi: 10.1016/j.procs.2020.03.400.

[8] L. Jiang, H. Liu, H. Zhu, and G. Zhang, "Improved YOLO v5 with balanced feature pyramid and attention module for traffic sign detection," *MATEC Web of Conferences*, vol. 355, p. 03023, 2022, doi: 10.1051/matecconf/202235503023.

[9] M. Milagro Fernandez-Carrobles, O. Deniz, and F. Maroto, "Gun and knife detection based on Faster R-CNN for video surveillance." [Online]. Available: http://visilab.etsii.uclm.es

[10] G. K. Verma and A. Dhillon, "A Handheld Gun Detection using Faster R-CNN Deep Learning," in *ACM International Conference Proceeding Series*, Nov. 2017, pp. 84–88. doi: 10.1145/3154979.3154988.

[11] N. Dwivedi, D. K. Singh, and D. S. Kushwaha, "Employing data generation for visual weapon identification using Convolutional Neural Networks," *Multimed Syst*, vol. 28, no. 1, pp. 347–360, 2022.

[12] R. Olmos, S. Tabik, and F. Herrera, "Automatic handgun detection alarm in videos using deep learning," *Neurocomputing*, vol. 275, pp. 66–72, Jan. 2018, doi: 10.1016/j.neucom.2017.05.012.

[13] S. Gupta and A. Mahajan, "Weapon Detection in Video Surveillance using Computer Vision Techniques," 2021. [Online]. Available: http://ijesc.org/

[14] R. F. de Azevedo Kanehisa and A. de Almeida Neto, "Firearm Detection using Convolutional Neural Networks.," in *ICAART (2)*, 2019, pp. 707–714.

[15] "Glenn Jocher. Yolov5 in PyTorch. https://github.com/ultralytics/yolov5, 06 2020. (undefined 28/4/202110:16)."

[16] R. Couturier, H. N. Noura, O. Salman, and A. Sider, "A deep learning object detection method for an efficient clusters initialization," *arXiv preprint arXiv:2104.13634*, 2021.

[17] B. Yan, P. Fan, X. Lei, Z. Liu, and F. Yang, "A real-time apple targets detection method for picking robot based on improved YOLOv5," *Remote Sens (Basel)*, vol. 13, no. 9, May 2021, doi: 10.3390/rs13091619.

[18] A. C. G, K. Krishnan, and K. S. Angel Viji Associate Professor, "Multiple Object Tracking using Deep Learning with YOLO V5." [Online]. Available: www.ijert.org

[19] Q. Song *et al.*, "Object Detection Method for Grasping Robot Based on Improved YOLOv5," *Micromachines (Basel)*, vol. 12, no. 11, p. 1273, 2021.

[20] Y. Lecun, L. Eon Bottou, Y. Bengio, and P. H. Abstract|, "Gradient-Based Learning Applied to Document Recognition."

[21] J. Yu, Y. Jiang, Z. Wang, Z. Cao, and T. Huang, "Unitbox: An advanced object detection network," in *Proceedings of the 24th ACM international conference on Multimedia*, 2016, pp. 516–520.