Iraqi Journal of Information and Communications Technology(IJICT)
Vol.7, Issue 1, April 2024
www.ijict.edu.iq
ISSN:2222-758X
e-ISSN: 2789-7362

A MULTI-WEAPON DETECTION USING DEEP LEARNING

Moahaimen Talib [1], **Jamila H. Saud** [2]
[1,2] Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq
moahaimen@gmail.com [1], dr.jameelahharbi@gmail.com [2]
Corresponding Author: **Jamila H. Saud**
Received:30/01/2023; Revised:23/08/2023; Accepted:20/09/2023
DOI:10.31987/ijict.7.1.242

***Abstract*- The escalating usage of light weapons in criminal and terrorist activities has necessitated the development of advanced weapon detection systems. This study centers on the application of the high-speed deep learning detection model, You Only Look Once version 7 (YOLOv7), to address the need for efficient and swift threat identification. The proposed system, trained on a self-curated dataset rich in images of various dangerous light weapons, is designed to recognize and distinguish multiple weapons simultaneously. The unique value of YOLOv7 in object and specifically weapon detection lies in its outstanding speed and accuracy, as demonstrated by certain weapon categories achieving a mean average precision (mAP) of 97%. The system's performance can potentially be further enhanced by augmenting the image dataset for each weapon category. This study, therefore, not only validates the critical importance of YOLOv7 in advancing detection methodologies but also presents a practical solution for bolstering public safety.**

keywords: **Artificial intelligence, Computer Vision, Object Detection, Convolutional Neural Networks, Pattern Recognition**

I. INTRODUCTION

Civilian safety has become a growing concern in recent times due to the rising number of criminal activities and terrorist attacks. In such scenarios, weapons of diverse kinds often pose a significant threat, necessitating efficient detection and identification mechanisms[1]. Even though surveillance cameras have become ubiquitous in various spaces like streets, offices, malls, and banks, accurately identifying weapon types remains a complex task for security personnel. The gravity of the threat presented is determined by the type of weapon used, which calls for a highly accurate detection system [2, 3].

The need for rapid and precise detection systems has given birth to numerous research efforts in weapon detection methodologies. With deep learning at the forefront, the proposed system intends to utilize multiple layers of nonlinear processing units for feature extraction and transformation, thereby learning different data features throughout the, structure of these layers. Deep learning essentially operates on raw data representations [4, 5] for instance, image representations can vary from a vector of pixel intensity values to custom shapes and clusters of features, some of which can provide richer information than others[6]. At the core of these deep learning efforts lies the Convolutional Neural Network (CNN), comprising convolution, pooling, activation function, dropout, fully connected, and classification layers[7].

Despite significant strides in object detection due to advancements in CNN-based systems, challenges persist, such as high false detection rates, substantial time overheads, and increased computational demands. In response to these issues, an innovative solution using the YOLOv7 network model[8] for detecting seven distinct weapon types (AKM, M4, PKS, RPG, PISTOL, KNIFE, MORTAR, FGM-148 Javelin antitank missile, and SNIPER) is proposed. The main contribution focuses on gathering datasets, for each weapon category and applying YOLOv7 to achieve the results in terms of "mAP" (mean precision).

ISSN:2222-758X
e-ISSN: 2789-7362

The research brings forth contributions;

1) The work developed a dataset that is generated automatically and possesses a deal of diversity. This dataset offers an advantage over created datasets in terms of speed and efficiency of generation.
2) The YOLOv7 for weapon detection with an emphasis, on its speed reduced false detection rates, and lower computational requirements is implemented. To enhance the modelâs efficiency, we have also utilized a model size, an efficient inference algorithm, and GPU-based inference.
3) The work introduced training configurations that include an initial learning rate and the use of the Adamax optimizer. These adjustments contribute to an accurate model.
4) To evaluate the modelâs performance, Mean Average Precision (mAP) is used as a metric, for evaluation this metric offers a measure of how the modelers perform.

This paper's structure is as follows: Section II reviews related work in the domain, while Section III details the YOLO network model. Section IV introduces the materials and methods used, followed by Section V, which presents the experimental results. Section VI discusses these findings, and Section VII concludes the study.

II. RELATED WORK

The domain of weapon detection has been an active research field due to its pivotal role in maintaining security and safety. Research in this area can be primarily used Deep Learning models due to their superiority over traditional algorithms which demonstrated limited success due to the diversity of weapons and their varying appearances in different contexts.

The advent of deep learning ushered in a new era for weapon detection. Deep learning models, and more specifically, CNNs, have shown remarkable progress in this field. They can learn hierarchical features automatically, making them more robust and adaptable to varying conditions.

Tabik and Herrera in 2017 [9] delved into the realm of automatic handgun detection by employing two methodologies: region proposal and sliding window combined with a histogram of gradients, all under the Faster Regional CNN (R-CNN) framework. For their model training, they harnessed a novel dataset named Pistol Detection and introduced a unique assessment metric for movie detection systems, dubbed Activation Time for Alarms per Interval. Their research achieved an accuracy rate of 84.21%. However, their methodology's inherent limitation lies in the use of the Faster R-CNN framework, which can be computationally intensive. Additionally, the reliance on separate region proposal mechanisms could introduce inefficiencies. In contrast, our model, which utilizes YOLOv7, offers a more integrated and streamlined approach to object detection. YOLOv7's architecture ensures rapid processing speeds and a more holistic accuracy, eliminating the need for separate region proposal systems and hence, outperforming the methods presented by Tabik and Herrera.

Fernandez-Carrobles, Denizz, and Maroto in 2019[10] pioneered in the domain of object detection with a primary aim to detect firearms and knives in real-time videos. Their method centered on the Faster R-CNN approach, integrating the powers of CNN and RPN, ultimately harnessing the SqueezeNet architecture to achieve an accuracy of 85.45%. Nonetheless, the research faced evident challenges. The Faster R-CNN model, despite its innovativeness, was resource-intensive, leading to slower inference speeds. In contrast, our YOLOv7-based model offers a solution that's both computationally efficient and

ISSN:2222-758X
e-ISSN: 2789-7362

swift. Unlike the Faster R-CNN, YOLOv7 identifies and classifies objects in one forward pass, significantly enhancing the detection speed and overall performance, and making our approach superior in real-time scenarios.

Verma and Dhillon in 2019[11] spearheaded a technique for automatic gun recognition in crowded scenarios using the Faster R-CNN model. Their strategy incorporated transfer learning to evolve a Deep Convolutional Network (DCN), an offshoot of Faster Region-based CNN. They leveraged the Internet Movies Firearms Database (IMFDB) and another reference database for model assessment, obtaining a commendable 93% accuracy. However, the model grappled with inherent limitations. Like its Faster R-CNN predecessors, it necessitated immense computational resources, constraining real-time performance and adaptability on devices with limited processing capabilities. Contrasting this, our YOLOv7-driven model addresses such setbacks. With YOLOv7's streamlined architecture, it accomplishes simultaneous detection and classification in a singular pass, promising swifter detection times and more versatile applications, emphasizing its enhanced viability in practical scenarios.

Gelana and Yadav 2022[12] unveiled a methodology rooted in using a CNN classifier via a sliding window approach. The classifier training involved techniques of background removal and edge detection, using high-definition ClosedCircuit Television (CCTV) footage for both training and testing datasets. Their reported accuracy stood at a notable 93.84%. However, their method has distinct shortcomings. Primarily, the reliance on high-quality CCTV footage limits the algorithm's broader applicability, assuming the consistent availability of such pristine data. In reality, often surveillance videos can be of lower quality or grainy, which could compromise this model's effectiveness. Moreover, the chosen sliding window technique brings inherent issues, such as struggles with varying object dimensions, potential escalation in false positives from repeated detections, and the computational heft it demands, posing difficulties for real-time implementations or devices with restricted processing capabilities. In contrast, our YOLOv7-centric model effectively bypasses these challenges, affirming its greater suitability for diverse practical applications.

Dwivedi, Singh, and Kushwaha in 2022[13] presented a CNN-based weapon detection technique, focusing on data generation for visual identification. Their model, underpinned by the trained weights of the Visual Geometry Group-16 (VGG-16) network, demonstrated an impressive 97% accuracy for knives and 99% for pistols after training with diverse weapon and non-weapon imagery. However, its primary limitation was the specific focus on only knives and handguns, reducing its efficacy in scenarios with varied weapons. Moreover, the dependency on the VGG-16 network meant considerable computational demands, potentially impacting real-time applications and limiting deployment on less powerful devices. Contrarily, Our Model offers broader weapon detection capabilities without being restricted to specific types. Additionally, The Yolov7 architecture allows for more efficient processing, making it apt for real-time scenarios and devices with varying computational strengths.

III. YOLO NETWORK MODEL

The application of YOLOv7 for weapon detection signifies a noteworthy leap in both speed and accuracy, pushing the boundaries of what is feasible in real-time surveillance and threat mitigation. Its adaptability to various scenarios and robustness in handling a range of image qualities highlight its superiority and potential as a reliable solution for weapon detection tasks.

ISSN:2222-758X
e-ISSN: 2789-7362

Deep convolutional neural networks with only one stage Both YOLOv3 and YOLOv4 have produced positive results for object detection. A number of network architectures are utilized by YOLOv5 and by enhancing YOLOv4 with two different kinds of Cross Stage Partial "CSP" modules, YOLOv5 is much more suited to object detection and recognition in terms of detection precision and computational complexity. Therefore, this paper proposes a method for multi-weapon detection using Yolov] the latest in the YOLO lineage[14]. YOLOv7, a PyTorch-based object detection model, exhibits improved user-friendliness and reduced complexity compared to its predecessors [15]. Furthermore, it has an embedded bounding box prediction and object recognition within an end-to-end differentiable network[16], facilitating real-time object detection via a smart CNN[17].

Our research focuses on the application of the YOLOv7 model in object detection technology, which is currently the latest member of the YOLO lineage[18]. Developed with the aim of optimizing bounding box estimations and object identification, YOLO ingeniously integrates these functions into a single end-to-end differentiable network[19]. YOLOv7 further refines this model, leveraging the user-friendly PyTorch framework to construct a streamlined model that incorporates a potent CNN for real-time object detection.

YOLOv7's superiority over other real-time object detectors is well-established. This model efficiently subdivides images into regions, subsequently computing the probability and bounding boxes for each of these segments. Each bounding box's weight correlates with the predicted probabilities, following the YOLO [20]principle of performing a single forward pass for object detection Fig. 1 shows the supremacy of Yolov7 in detection and speed over previous versions of YOLO and other object detection models.

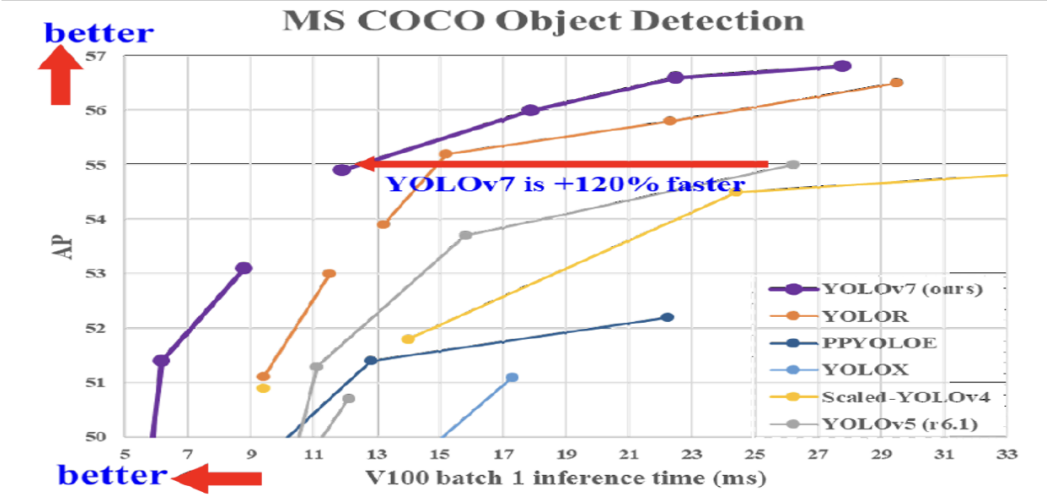

Figure 1: Performance comparison of Yolov7 against other Yolo versions

Central to the architecture of YOLOv7 is the idea of controlling the shortest longest gradient path, a strategy that enables a deeper network to effectively learn and converge. The YOLOv7 model introduces the concept of Extended Efficient Layer Aggregation (E-ELAN), a key computational strategy that forms the backbone of its structure.

ISSN:2222-758X
e-ISSN: 2789-7362

The idea of Efficient Layer Aggregation (ELAN), which the E-ELAN extends, has emerged from the pursuit of improved layer aggregation techniques in YOLO models. ELAN and E-ELAN are essentially computational units aimed at enhancing feature extraction capacity, maintaining a low computational cost, and facilitating a healthy gradient flow during back-propagation Fig. 2 shows Extended efficient layer aggregation networks. The proposed EELAN does not change the gradient transmission path of the original architecture at all but uses group convolution to increase the cardinality of the added features, and combine the features of different groups in a shuffle and merge cardinality manner. This way of operation can enhance the features learned by different feature maps and improve the use of parameters and calculations.

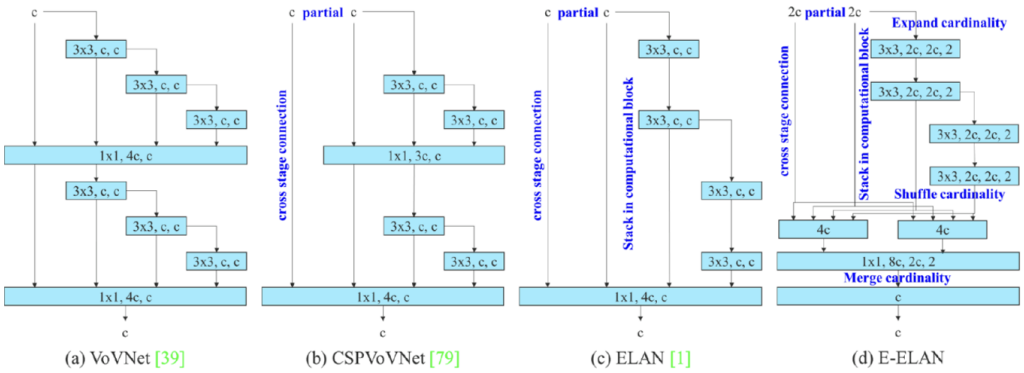

Figure 2: networks with extended efficient layer aggregation

ELAN, in its design, applies the principle of 'cross-stage partial connections.' This structure allows layers from different stages (or depths) in the model to share information, thereby enhancing feature learning. The E-ELAN extends this idea by introducing 'expand', 'shuffle', and 'merge' techniques, which further strengthen the representational capacity of the network.

The authors of YOLOv7 have significantly pushed the boundaries of efficient layer aggregation with the introduction of E-ELAN. E-ELAN, with its advanced layer aggregation strategy, continually enhances the network's learning ability, supporting the development of models at different scales to cater to varied inference speed requirements. Importantly, while E-ELAN modifies the design of the computational blocks, it keeps the transition layer's design intact.

YOLOv7 also employs model scaling, a technique that involves adjusting certain attributes to create models of varying scales, to cater to different needs related to inference speeds. Group convolution, employed in E-ELAN, expands the channel and cardinality of computational blocks, causing an increase in the output width of a computational block due to the model's sizing based on concatenation. This necessitates the consideration of specific adjustments when scaling a model using concatenation-based methods.

In their pursuit of redefining the state of the art in object detection, the authors of YOLOv7 have made significant strides, such as introducing an auxiliary head for enhanced training, implementing model scaling techniques, and incorporating re-parameterization planning. All these features combine to produce a model that is robust, accurate, and effective in real-time

ISSN:2222-758X
e-ISSN: 2789-7362

object detection.

IV. PROPOSED METHOD

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

1) Image preprocessing: (preprocessing steps (resizing all images into unified size 416×416, auto-Orientation, auto-adjust contrast))
2) Image augmentation
3) Processing images
4) Building the weapon dataset
5) Training using the Yolov7 model
6) Detection of weapons by inputting a real-time video.

The pipeline of the work may be divided into three main categories Generating Dataset, Training Phase, and Detecting Phase, as shown in Fig. 3.

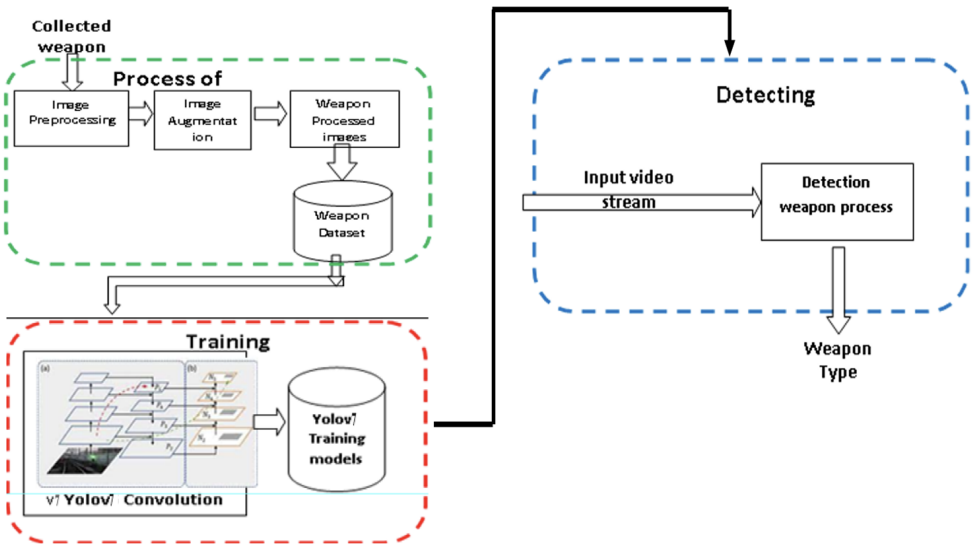

Figure 3: The pipeline of the system

V. EXPERIMENTAL SETUP

This section describes the materials and methods used in the proposed system. The study utilized a synthetic dataset that was automatically constructed, as there isn't a standard weapon dataset readily available. This technique employs a Python script to provide cropped images of objects of interest with a background, and the same script generates annotations. Using this method, the images that feature realistic backdrops and objects but are not actual photographs is created. An automated approach allows us to construct datasets much more quickly than manual methods. For example, 1000 synthetic images

and annotations can be generated in less than an hour, which is significantly faster than manually annotating 1000 distinct images as shown in Fig. 4 . The objects of interest include (akm, m4, pks, rpg, pistol, knife, mortar, fgmen-148 javelin antitank missile, sniper). Masks and cropped images of these items are used in various configurations. background images are simply various pictures downloaded from the internet. To add complexity to the background, cropped images and masks of different objects as background noise is used, including cars, chairs, guitars, and other items. The total number of YOLOv7formatted, labeled, and annotated images, with an equal number of weapons in each, is 60,000 . Pytorch can be configured with two distinct processors to accommodate GPUs or central processing units (CPUs). GPUs typically

perform better than CPUs when running Pytorch programs. For this study, the YOLOv7 model was trained on Google Colab Pro+ using an Nvidia V100 GPU and 58 GB of RAM, Fig. 4 shows how the dataset was built.

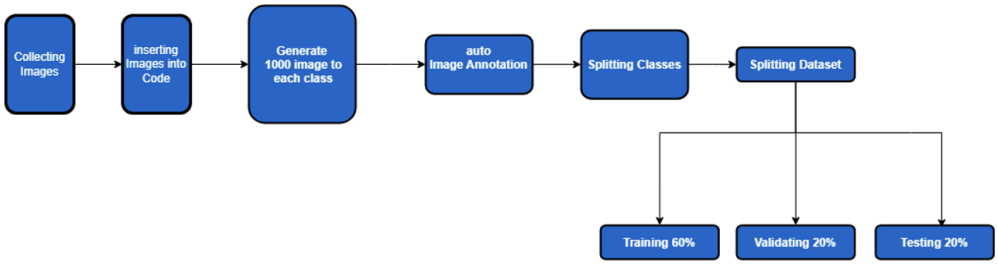

Figure 4: Dataset Generation

VI. EXPERIMENTAL RESULTS AND DISCUSSION

Our study involves the analysis of nine distinct types of weapons, the dataset for which is divided into three sections: 60% for training, 20% for validation, and 20% for testing. The training dataset includes 40000 images, while the testing and validation datasets comprise 20000 images each this shown in Table I.

TABLE I
Dataset for Weapons

Training	Testing	Validating
40000	20000	20000

The training process for the YOLOv7 model utilizes input images with a resolution of 416×416 pixels, formatted in RGB, and spans over 50 epochs. With this setup, high precision is achieved, recall, and mAP values for various weapon classes. For instance, the precision for the pistol, akm, and $\mathrm{m4}$ classes were $0.991, 0.992$, and 0.987 , respectively. The recall values for these classes were $0.923, 0.949$, and 0.938 , respectively, and the mAP values were $0.948, 0.958$, and 0.956 . In totality, the overall mAP was 95%, signifying a high degree of accuracy in comparison with similar studies, The results in the Dataset, as shown in Table II.

This work focus on evaluating object detection accuracy using metrics like mean Average Precision (mAP), Recall (R), and Precision (P). True Positive (TP) is when positive samples are correctly identified; False Positive (FP) is when

ISSN:2222-758X
e-ISSN: 2789-7362

TABLE II
Results of Training Synthetic dataset

Weapon Class	**P**	**R**	**mAP**
Pistol	0.991	0.923	0.948
AKM	0.992	0.949	0.958
M4	0.987	0.938	0.956
Javlin	0.977	0.921	0.949
Mortar	0.971	0.933	0.952
RPG	0.992	0.925	0.952
PKS	0.988	0.913	0.951
Sniper Drugnove	0.988	0.922	0.952
Knife	0.992	0.922	0.945

negative samples are mistakenly marked as positive; False Negative (FN) is when positive samples are incorrectly marked as negative. Additionally, $\mathrm{P(K)}$ represents precision at the Kth detection, while $\Delta \mathrm{R(K)}$ denotes the change in recall from the $(\mathrm{K}-1)$ to the Kth detection, as show in eq. 1 and 2 and 3 :

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$mAP = \frac{1}{C} \sum_{K=1}^{N} P(K) \Delta R(K) \tag{3}$$

Another metric considered is the Intersection over Union (IoU)[21], which determines the similarity between the predicted and actual bounding boxes. The accuracy $(P(k))$ and recall $(R(k))$ are ascertained based on the number of object categories (C), the number of loU thresholds (N), and the loU threshold (k).

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

During the 50 epochs of the YOLOv7 model training, learning started from the second epoch onward, offering an ideal learning curve for our study. The metrics used - accuracy, precision, and mAP - increased significantly with each epoch.

Our distinctive contribution to the literature is the use of the YOLOv7 model to counteract terrorist attacks by creating a unique dataset encompassing nearly all weapon types used in assaults. The dataset is applied in two stages: firstly, the model is trained using the training and validation datasets; secondly, the developed model is evaluated using a test dataset it has never encountered, yielding an impressive classification success rate Fig. 5 mAP and precision and recall for each weapon classshows each class results.

ISSN:2222-758X
e-ISSN: 2789-7362

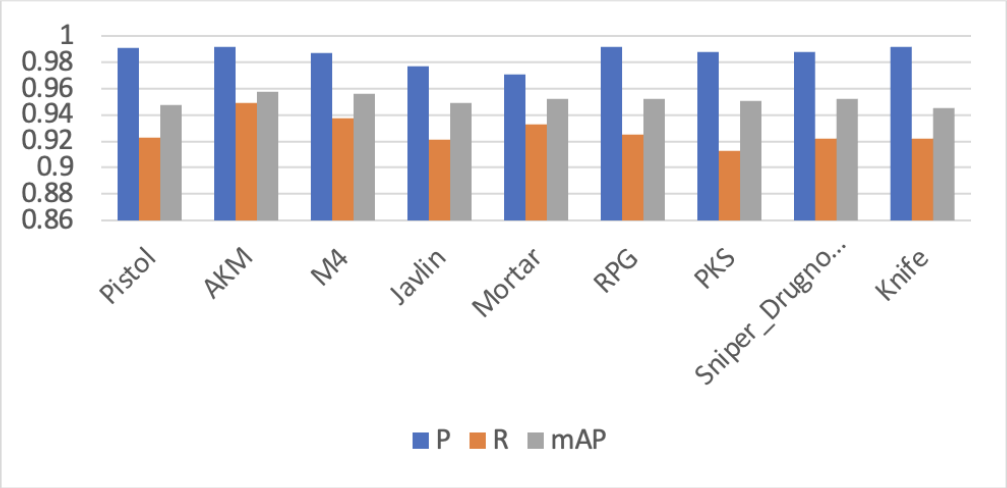

Figure 5: mAP and precision and recall for each weapon class

The proposed model's efficacy has been evaluated using weapon images both on individuals and in isolation, as well as real-time videos to assess the model's performance in real-world scenarios. Compared to previous works using YOLOv7, our study achieved superior results, with a total precision of 99%, a total recall of 93%, and a total mAP of 95% on various backgrounds containing multiple objects. Given YOLOv7's speed and precision in weapon recognition, our work is pivotal for high-speed weapons detection, with our model being deployable on machines with average computational requirements Fig. 6 shows a sample of weapons type.

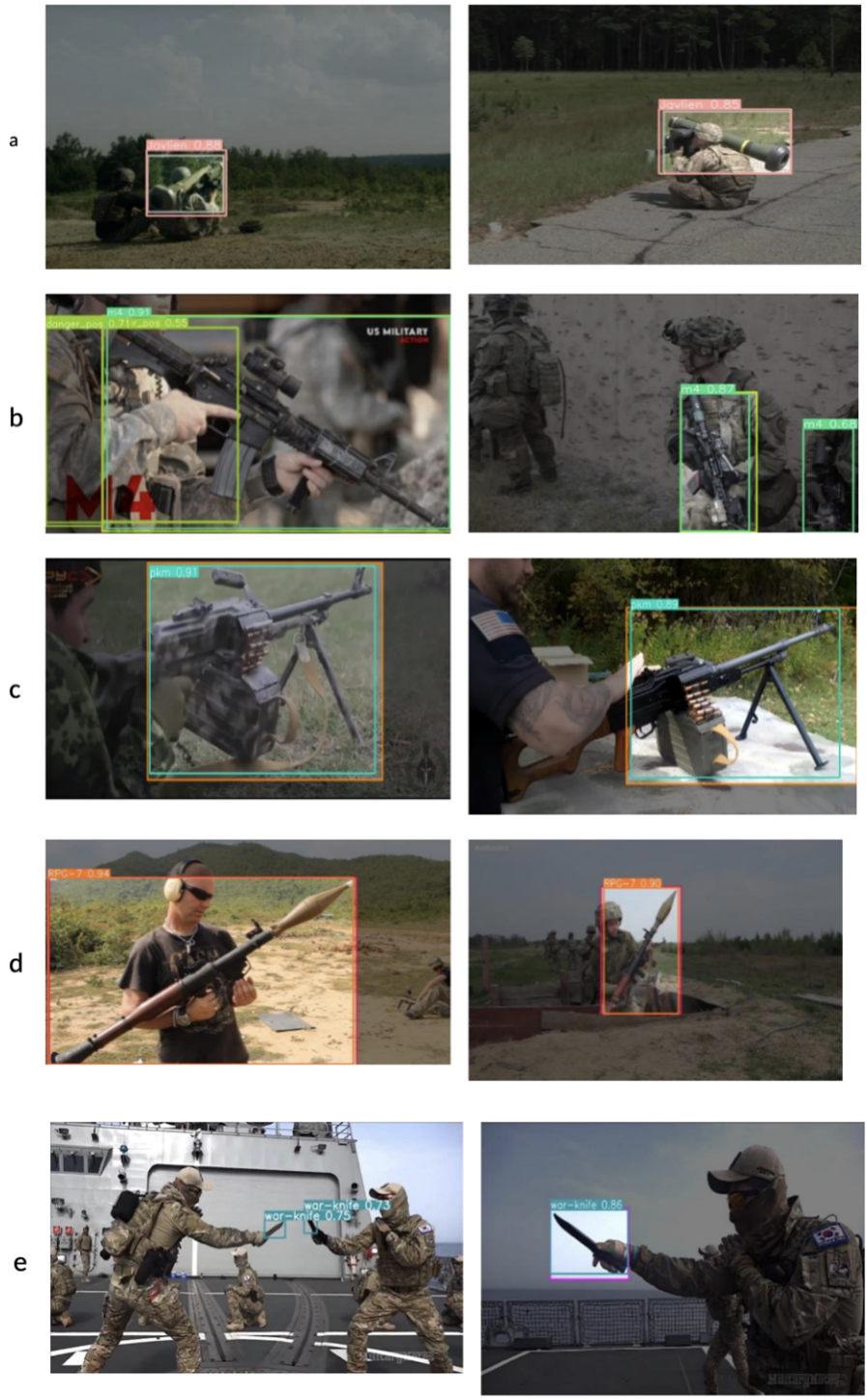

Figure 6: Samples of Weapons Types Recognize by the Model (a) FGM Javlin (b) M4 (c) PKM (d) RPG (e) War-Knife

ISSN:2222-758X
e-ISSN: 2789-7362

VII. Conclusion

This study set out to determine the improving the task efficacy and efficiency of security forces in security applications, the model proposed in this work will help close various security gaps. This increases the initial value of the findings because they can be used in new contexts. The study's findings are anticipated to inform and direct future research, notably in autonomous security systems.

Funding

None

ACKNOLEDGEMENT

The author would like to thank the reviewers for their valuable contribution in the publication of this paper.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, âYou Only Look Once: Unified, Real-Time Object Detection,â in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779â788. doi: 10.1109/CVPR.2016.91.

[2] Z. K. Abbas and A. A. Al-Ani, âDETECTION OF ANOMALOUS EVENTS BASED ON DEEP LEARNING-BILSTM,â *Iraqi Journal of Information and Communication Technology*, vol. 5, no. 3, Dec. 2022, pp. 34â42, doi: 10.31987/ijict.5.3.207.

[3] B. S. Mahmmed and A. I. Majeed, âFACE DETECTION AND RECOGNITION USING GOOGLE-NET ARCHITECTURE,â *Iraqi Journal of Information and Communication Technology*, vol. 6, no. 1, Apr. 2023, pp. 66â79, doi: 10.31987/ijict.6.1.228.

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

[5] Y. Bengio, âLearning deep architectures for AI,â *Foundations and Trends in Machine Learning*, vol. 2, no. 1, 2009, pp. 1â27, doi: 10.1561/2200000006.

[6] H. A. Song, B. K. Kim, T. L. Xuan, and S. Y. Lee, âHierarchical feature extraction by multi-layer non-negative matrix factorization network for classification task,â *Neurocomputing*, vol. 165, Oct. 2015, pp. 63â74, doi: 10.1016/j.neucom.2014.08.095.

[7] S. Masood, U. Ahsan, F. Munawwar, D. R. Rizvi, and M. Ahmed, âScene Recognition from Image Using Convolutional Neural Network,â in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 1005â1012. doi: 10.1016/j.procs.2020.03.400.

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

[9] R. Olmos, S. Tabik, and F. Herrera, âAutomatic handgun detection alarm in videos using deep learning,â *Neurocomputing*, vol. 275, Jan. 2018, pp. 66â72, doi: 10.1016/j.neucom.2017.05.012.

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

[11] G. K. Verma and A. Dhillon, âA Handheld Gun Detection using Faster R-CNN Deep Learning,â in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Nov. 2017, pp. 84â88. doi: 10.1145/3154979.3154988.

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

[13] N. Dwivedi, D. K. Singh, and D. S. Kushwaha, âEmploying data generation for visual weapon identification using Convolutional Neural Networks,â *Multimedia Systems*, vol. 28, no. 1, 2022, pp. 347â360.

[14] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, âYOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,â in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7464â7475.

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

[16] Y. Lecun, E. Bottou, Y. Bengio, and P. Haffner, âGradient-Based Learning Applied to Document Recognition,â 1998.

[17] R. F. de Azevedo Kanehisa and A. de Almeida Neto, âFirearm Detection using Convolutional Neural Networks,â in *ICAART (2)*, 2019, pp. 707â714.

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

[19] Q. Song et al., âObject Detection Method for Grasping Robot Based on Improved YOLOv5,â *Micromachines*, vol. 12, no. 11, 2021, p. 1273.

[20] B. Yan, P. Fan, X. Lei, Z. Liu, and F. Yang, âA real-time apple targets detection method for picking robot based on improved YOLOv5,â *Remote Sensing*, vol. 13, no. 9, May 2021, doi: 10.3390/rs13091619.

[21] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, âGeneralized intersection over union: A metric and a loss for bounding box regression,â in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Jun. 2019, pp. 658â666. doi: 10.1109/CVPR.2019.00075.

