ISSN : 2321-9416

Indo - American Journal of Mechanical Engineering

www.iajme.org

Email : iajme.editor@gmail.com or editor@iajme.org

ISSN 2321–9416 www.iajme.org
Vol 15, Issue 1, Mar 2026

https://iajme.com/10.62645/iajme.2026.v15.i01.pp23-31

Image-Based Weapon Detection Using YOLO v7

[1] **Mrs.M.Rajitha,** [2] **Gaddam Sathwika,**[3] **Suggala Vikas,** [4] **Damera Karthik,** [5] **Arla Suphala,** [6] **Boyini Parmeshwar,**

[1]Assistant Professor, Department of Data Science, Narsimha Reddy Engineering Collage, Maisammaguda(V), Kompally, Telangana.

[2,3,4,5,6] Student, Department of Data Science, Narsimha Reddy Engineering Collage, Maisammaguda(V), Kompally, Telangana.

Abstract—

Despite the tiny size and possible ambiguity of these weapons, the continuing danger of violence involving firearms and knives is a challenge to worldwide public safety and necessitates robust detection technologies. With the goal of improving public safety in both civilian and military settings, this project aims to design a computerized system that can recognize these common weapons. The YOLO v7 and v8 frameworks are used to present a gun and knife detection system, which takes advantage of contemporary deep learning improvements. In order to make monitoring and detection more efficient, our system annotates identified weapons with bounding boxes. The evaluation findings show that the YOLOv7-e6 model is successful; it improves the detection capabilities of knives and weapons and has a much higher mean Average Precision (mAP) than other techniques. Efforts to safeguard communities from the constant danger of armed violence are continuing, and this study adds to those efforts. Things like Yolo v7, object identification, deep learning, and weapons like knives and firearms

I. INTRODUCTION

Numerous public health, psychological, and economic issues are exacerbated by gun violence. Many people lose their lives every year as a result of this kind of violence. Witnessing violence in their communities may have a profound and long-lasting emotional impact on children. Kids' mental and emotional health might take a hit after seeing horrific gun situations. A number of studies have highlighted the prevalence of firearms in criminal offenses such robbery, burglary, rape, and theft. Law enforcement agencies can respond more effectively and reduce the occurrence of such crimes if they are able to spot suspicious activity and act swiftly. [11]. When there is too much going on for humans to keep an eye on everything, automated computer systems may aid by alerting people if anything potentially harmful is occurring. Something suspicious could be happening if individuals are openly carrying weapons like knives or firearms. Informing those monitoring security cameras is a good idea, even if openly carrying guns is legal in certain areas. More and more applications, such as smart transportation systems, are using automated surveillance cameras. Some examples of these methods include traffic monitoring and vehicle identification systems [13]. Since knives and firearms are the most regularly utilized lethal instruments, our primary objective is to utilize computers to detect these weapons. Those in control should be familiar with their appearance while held. A smart surveillance security system capable of detecting weapons, especially firearms and knives, was the focus of this research. Our objective is to do this by identifying a weapon from a taken picture using deep learning and a few computational vision methods. Notably, convolutional neural networks (CNNs) have achieved remarkable achievements in the field of deep learning and machine learning, namely in the areas of object detection and identification, however this is limited to visual data. Recognizing and classifying objects is the first step in any video surveillance software and is essential for object tracking. In order to do this, we trained the "You Only Look Once" (YOLO v7) classification model. A state-of-the-art object recognition classifier, this model operates in real time. To recognize firearms and knives in images, this study details the work on YOLOv7, an item recognition model based on convolutional neural networks (CNNs). The research set out to do three main things: first, see how well the models performed; second, see how the training durations affected the models; and third, see how training the models on a dataset improved the detection quality. The document is structured into five parts. The theoretical groundwork required to comprehend this work is laid forth in the part that follows this introductory one. Section 3 delves into the technique and instruments that were used.

Presented in the fourth are the findings from this investigation. Lastly, the study findings are reviewed in the fifth part.

II. Other Works

Researchers have recently begun looking at ways to use images to identify knives and firearms, with the goal of making public spaces safer. They are analyzing the issues and potential solutions. A system for automatic gun identification employing Faster R-CNN architectures, such as ResNet50, Inception-ResNetV2, VGG16, and MobileNetV2, is suggested in the study [1]. It is for mobile apps that MobileNetV2 is selected. The efficiency of the method in identifying weapons is evaluated using YOLOv2. Even though it's sluggish, Faster R-CNN with Inception ResNetV2 has the best mAP at 82%. Although it is the fastest model, VGG16 is not as fast as YOLOv2. To improve surveillance approaches, this research [2] use Convolutional Neural Networks (CNNs). Specifically, it utilizes CNNs to detect fire and pistols in regions that are watched by cameras. In order to immediately notify the authorities, the YOLOv3 algorithm is used to evaluate video frames in real-time and identify any irregularities. A 45 fps detection rate and 89.3%, 82.6%, and 86.5% accuracy on datasets such as IMFDB, UGR, and FireNet are achieved by the model. You may use it both inside and outdoors thanks to its fast detecting rate. Crimes against individuals and property have increased in Thailand due to recent unrest, according to another research [3]. The cops have to constantly monitor the CCTV footage, even though it is used to keep an eye on things. The cops may be able to do more with the aid of a weapons detecting system. Using the TensorFlow Object Detection API, they combed through data from two datasets that identified weapons. Their research indicated that the Faster R-CNN Inception V2 model outperformed the competition when it came to weapon detection, with an impressive high score of 0.540 and average scores of 0.793 and 0.627 across several metrics. Another article that was cited [4] discusses many approaches to locating weapons, including both basic and advanced techniques. When dealing with X-ray and Terhertz images, simple solutions that focus on picture clarity tend to do well. The problem is that individuals employ various collections of images, making it impossible to compare these approaches. Complex approaches demonstrate their efficacy using a shared image collection called Imagenet and use deep learning. Overfeat is one approach that discovers weapons correctly at 89% mAP, whereas Faster RCNN is another that checks 0.2 frames per second and is incredibly quick. Focusing on both deep learning and non-deep learning methods, this research [4] examines a range of algorithms that can detect knives and pistols. Image quality is the foundation of non-deep algorithms, which work well with X-ray and Terahertz pictures. However, comparisons are inaccurate due to the use of diverse datasets. When applied to the Imagenet dataset, deep learning algorithms reveal their accuracy and performance. While Faster RCNN achieves the greatest speed of 0.2 frames per second, the Overfeat approach achieves a more accurate result of 89% mAP. Using data from the Internet Movie Firearm Database, this research [5] examined the effectiveness of the YOLO algorithm in detecting firearms. Police must respond promptly to ensure the safety of the public if they see a person in possession of a firearm in a public place. Operators of security cameras get fatigued and less effective when they are tasked with too much footage. Convolutional neural networks are used by YOLO and similar programs to detect objects in photos; on occasion, these networks outperform humans. Another research [6] used deep learning and transfer learning to create a computer algorithm that could detect common weapons such as handguns and rifles. According to their research, YOLO V3 outperforms both previous versions and conventional computer algorithms. It can run on relatively low-end graphics processing units. Use of this application in surveillance systems has the potential to reduce the occurrence of mass shootings and other violent crimes. Robots that monitor their environments for potential threats might potentially benefit from this technology. Worldwide fear ensued when bombs detonated in Hyderabad, killing dozens of people [7]. Predicting suicide bombers and explosive sites is the job of Image Processing for Conclead Weapon Detection. Uncovered weapons, such as knives, explosives, and firearms, pose a threat to public safety and security. Modern developments in image and sensor technology have made it possible to identify hidden weapons from a distance, which is particularly useful in settings where controlling large crowds is challenging. Although Faster R-CNN is an efficient target identification technique, it does have certain drawbacks, such as using too many negative data. Hard negative sample mining and alternative training are the two strategies that have been suggested. In contrast to alternate training, which uses convolutional layers from both RPN and Fast R—Simulation shows substantial improvements in detection accuracy [8]. Hard negative sample mining is used to collect negative data. Countries with high rates of firearm ownership, whether legal or illegal, also have high rates of crime involving weapons [9]. Human administrators are finding it more difficult to identify any threats due to the proliferation of CCTV cameras. A Haar Cascade Classifier and the OpenCV library are discussed in this work as a means of swiftly locating firearms for the purpose of maintaining public safety. Based on their research, this approach has a 95% success rate in identifying submachine guns. It's a great choice because to its low price and user-friendliness. Combining RPN with Fast R-CNN networks to create Faster R-CNN is an exciting new direction in object identification for R-CNN series [10]. A specific layer known as ROI is connected to it. Pooling for end-to-end object detection. When discussing ResNet101 and PVANET networks, they inquire as to the viability of using Faster R-CNN. Using the deep learning program Caffe, scientists trained many models and discovered that the PVANET network performed the best at correctly detecting items.

III. METHODOLOGY

A. Dataset

If you want your object detection to work, you need a good dataset. Finding high-quality photos for weapon identification remains a challenge, even though the quantity of data has risen. It is difficult to compare research since most of them employ proprietary databases. To improve the performance of weapon identification systems, datasets that mimic genuine surveillance film are used. For training, these models are fed pictures that closely resemble the actual thing. Some studies use picture sets to help collect additional photos, which improves the models' performance. They are also made to function better by creating pictures artificially. From a website named Roboflow, our algorithm retrieved a collection of pictures depicting knives and firearms [17]. This picture has dimensions of 416 × 416 pixels. Out of a total of 400 photos, 360 were used for training the model and 40 for testing purposes.

B. Image Annotation

The program LabelImg is a lightweight and user-friendly option for marking the bounding boxes of items in images. This page provides a basic introduction to LabelImg, when to use it, and how to annotate photographs. The long-term viability of computer vision applications depends on using the right picture annotation software [16].

C. Yolo v7

Following in the footsteps of its forerunners, the YOLO v7 model is architecturally sound. Figure 1 depicts the overall model design. The ultimate prediction is made by the head, which is an FPN. The neck and backbone both use the same formula to extract features. The model's performance was enhanced by including four significant changes.

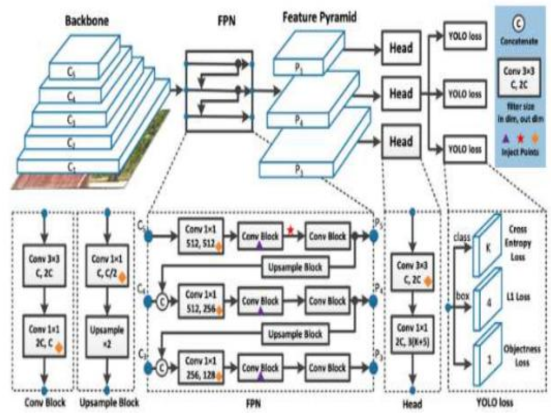

Fig. 1. Yolo v7 Architecture [12]

When it comes to inference speed, the YOLOv7 study is all about making the most of a deep network's backbone layers. In order to investigate the optimal layer efficiency, the writers used cross-stage partial networks (CSPNet). In order to maximize convergence efficiency and effective learning, they focused on controlling the gradient's longest path of change. They settled on an E-ELAN, which stands for Extended Efficient Layer Aggregation Network, as their network architecture. Without modifying the original gradient route, the network's learning capability improves with time. By combining layers and making the network deeper or broader, the YOLOv7 model alters the network's size. This method kept performing at its best over a range of sizes, according to further research. The authors have developed three distinct models: YOLOv7-tiny for edge GPUs, YOLOv7 for standard GPUs, and YOLOv7-W6 for cloud GPUs. They have also used a fundamental model for scaling to create a number of models tailored to certain service needs. They used a method called compound scaling to increase the size of the overall model for YOLOv7, which resulted to YOLOv7-X, after adjusting the size of individual pieces. Their second round of compound scaling yields YOLOv7-E6 and YOLOv7-D6, together referred to as YOLOv7-W6. Furthermore, YOLOv7-E6E is the product of applying the EELAN approach to YOLOv7-E6. YOLOv7-tiny, designed for usage on small computer chips, employs leaky ReLU as its activation function, in contrast to the other variants, which use SiLU [14].

D. Yolo v8

ISSN 2321–9416 www.iajme.org
Vol 15, Issue 1, Mar 2026

Examples of tasks that can be tackled using the latest and greatest YOLO model, YOLOv8, include instance segmentation, object detection, and picture classification. Compared to YOLOv5, YOLOv8 has several improvements and additions to the architecture and development experience. A medium-sized YOLOv8 model achieved a 50.2% mAP, demonstrating the model's strength and accuracy on COCO. In the task-specific domain dataset Roboflow 100, it achieves better results than YOLOv5. YOLOv8 provides a command line interface (CLI) and an easy way to code as a convenience for developers. There is a large and noteworthy community around YOLO, and there are numerous computer vision professionals and resources accessible online to help [15].

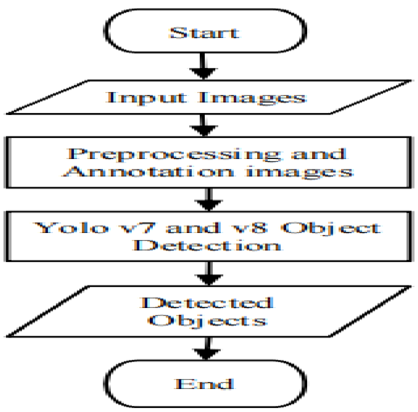

Fig. 2. System Flow Diagram

Figure 2 shows the system flow diagram. Multiple sources provide data into the weapon and knife detection system. These include live video feeds, picture files, and security cameras. The input frames are preprocessed by the system upon receipt, which includes converting them to a YOLOv7 compliant format, scaling the pictures to the input size necessary, and normalizing the pixel values. YOLOv7 is a real-time object identification model that checks every frame for items, including firearms, and then returns the locations of those things, the labels for those objects, and a confidence score for each. In the post-processing steps that follow object recognition, the system checks for bounding boxes with low confidence ratings and, if necessary, uses non-maximum suppression to get rid of overlapping boxes that aren't necessary. Based on the class labels supplied, the algorithm finds bounding boxes that have been identified as knives and weapons by YOLOv7.

E. Evaluation Metrics

Detailed explanations of the assessment techniques used to evaluate the object detection model's efficacy are provided. The accuracy, precision, recall, and mAP at 0.5 IoU were the assessment metrics employed in this research. The mAP ranged from 0.5 to 0.95 IoU. Accuracy is defined as the percentage of true positives relative to the total number of false positives. A visual representation of the model's accuracy in making detections is provided by the figure. Recognition of tiny items requires a high degree of accuracy due to the fact that these things are notoriously difficult to see and distinguish from their surroundings. With recall, we can see how well the model can spot all the real items in the picture, no matter how little. It checks the model's accuracy in identifying genuine things in the picture by counting how many of them it gets right. To put it simply, it reveals if the model is able to correctly identify the minute things in the picture. At a given IoU threshold, the mAP determines the model's average accuracy for all object classes. Whether the anticipated and ground-truth bounding boxes are identical is determined by this criterion.

IV. EVALUATION RESULTS

This study's training and testing datasets were supplied in YOLO format. Each photo in the test and train folders had a corresponding.txt file that included all the information about the bounding boxes and classes of knives and firearms. This preprocessing step might affect the model's performance; for example, the accuracy and recall of the model could be affected by the quality of the dataset's bounding boxes, which in turn affect the model's output. Figures 2, 3, 4, 5, and 6 show the recall, mean average precision, and accuracy of each model. The time required to train each model is shown in Table 1.

When it comes to this dataset, the Yolo v7-e6 model takes the most time to train while the Yolo v8 model is the quickest. Yolo v7-e6 model, on the other hand, achieves the highest system accuracy in this dataset. With an accuracy score of 0.68, the bounding box of the identified knife is shown in Figure 3. Even still, the bounding boxes usually include the knife and some of the hand, suggesting that the model may be associating the knife with the hand positions of the people in the pictures.

ISSN 2321–9416 www.iajme.org
Vol 15, Issue 1, Mar 2026

Fig. 3. Detected Knife with Bounding Box

Figure 4 shows that the model successfully identified every pistol in the scene.The models function well on the whole, but sometimes they may provide an inaccurate result, such as when they try to draw a bounding box around a hand or arm or when two people are carrying identical things.

Fig. 4. Detected Gun with Bounding Box

See how the YOLOv7 e6 models identified some objects in a few pictures in Figure 5. The models perform well for the most part, although they do make a few errors, most notably when they attempt to draw a box around a hand or arm, or when someone is carrying something like. An example of a failed detection system is shown in the image in the second row and second column.

Fig. 5. Results For Knife and Gun with Precision Value

A virtually uniform learning curve is shown by the fact that all performance measurements become consistent after 70 epochs (Fig. 6). Additionally, these measurements show that the model is learning and improving its performance consistently across the training epochs.

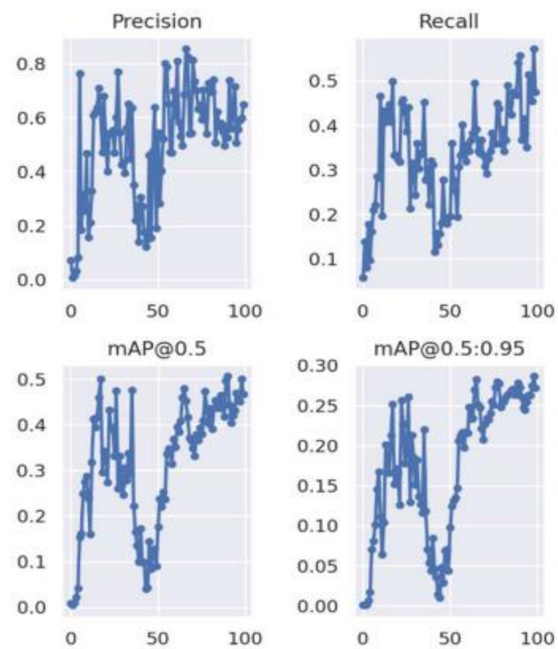

Fig. 6. Yolo v7 Validation Metrics

The measures in Fig. 7 seem to level out at around 90 epochs, which indicates that the model's learning rate has likely plateaued. A lack of high-quality or sufficient training data might be at blame.

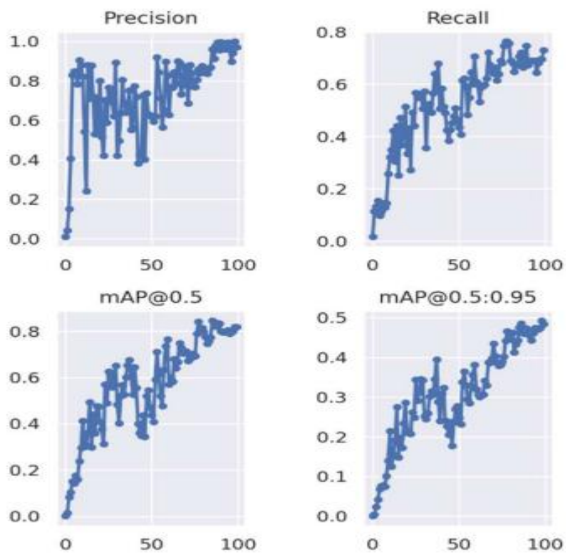

Fig. 7. Yolo v7-x Validation Metrics

Figure 8 shows that all performance metrics stabilize at around 60 epochs, which means that the learning curve is always smooth. Also, the metrics keep going up as the training epochs proceed, which means the model is learning and becoming better all the time.

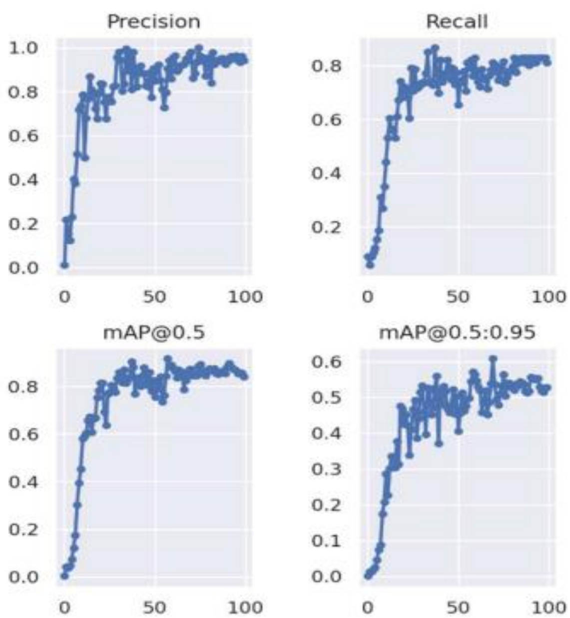

Fig. 8. Yolo v7-w6 Validation Metrics

With very small changes, as seen in Fig. 9, the metrics remain rather steady throughout the training epochs, indicating that the model is converging, especially at epoch 48. Even though the issue is inherently complicated, the model nonetheless manages to achieve peak performance and provide satisfying solutions.

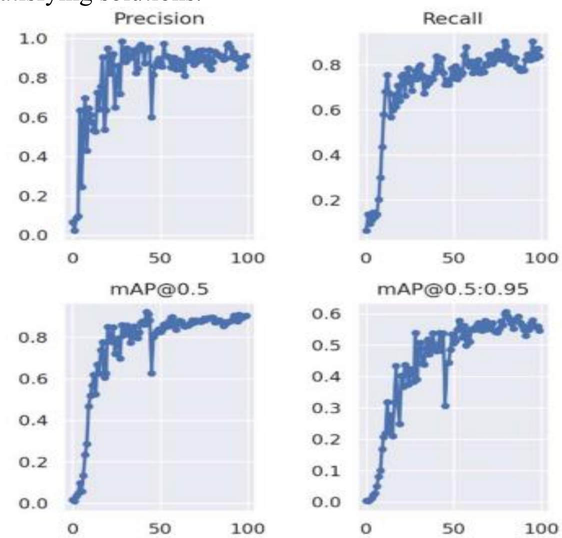

Fig. 9. Yolo v7-e6 Validation Metrics

Nevertheless, the results exhibit a significant amount of variation over epochs, suggesting a learning tendency that is not entirely regular. Notwithstanding these variances, all measures show steady growth across the training epochs, which points to the model's continuous learning and development. Fig. 10 shows that the model is converging, but the large swings between epochs raise the question of whether this is happening again.

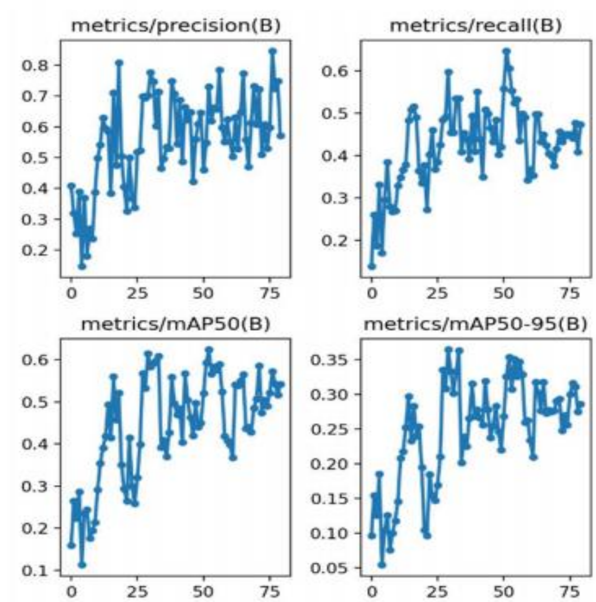

Fig. 10. Yolo v8n Validation Metrics

The time needed to train YOLO models for 100 epochs is shown in Table 1. Training takes the longest at 1.239 hours for YOLO v7-e6, and the shortest at 0.447 hours for YOLO v8n.

TABLE I. TRAINING DURATION OF YOLO V7 MODELS

Model	Training Time for 100 Epoch
Yolo v7	0.651 hours
Yolo v7-x	0.950 hours
Yolo v7-w6	0.752 hours
Yolo v7-e6	1.239 hours
Yolo v8n	0.447 hours

At IoU thresholds of 0.5 and 0.5 to 0.95, Table 2 shows the recall, accuracy, and mean Average accuracy (mAP) scores for several YOLO models. If the model has a high recall value, it means it can recognize more relevant items with little chance of missing any. The opposite is true for lower precision values, which indicate less accuracy in accurately recognizing the objects that are detected. Overall, Yolo v7-e6 outperforms the other model in this case, as it obtains a higher mAP at 0.5 IoU and a higher mAP over an IoU value range of 0.5 to 0.95.

TABLE II. COMPARISON OF YOLO V7 AND V8 MODELS

Model	Precision	Recall	mAP@.5	mAP@.5:.95
Yolo v7	0.852	0.617	0.674	0.33
Yolo v7-x	0.97	0.731	0.818	0.485
Yolo v7-w6	0.94	0.812	0.841	0.528
Yolo v7-e6	0.911	0.837	0.903	0.546
Yolo v8n	0.702	0.597	0.614	0.365

CONCLUSION AND FUTURE WORK

Various methods using various criteria were discussed in this study in order to locate knives and firearms. Then, we compared their performance to that of other existing approaches using metrics such as recall, accuracy, and mAP). The YOLO architecture has shown promising results in a number of object recognition tasks, and it is famous for its fast speed and high performance when it comes to model selection. The significance of identifying tiny pistols and knives in this study led to the comparison of many YOLO model versions, namely, YOLOv7, YOLOv7-x, YOLOv7-w6, YOLOv7-e6, and YOLOv8n. After extensive research and comparison, the YOLOv7-e6 model was shown to be the most effective for object identification, despite having an excessively lengthy training time. Finally, YOLOv8's significant improvements and new features are shown when compared to YOLOv7. YOLOv8's superior real-time object detection performance is a result of its reduced latency, increased accuracy, and absence of anchors. In upcoming projects including deep learning, computer vision, neural networks, and image processing, this model will be used for various datasets and real-time systems in order to get exceptional item recognition performance.

REFERENCES

1. R. M. Alaqil et al., "Automatic gun detection from images using faster R-CNN," 2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH), Nov. 2020. doi:10.1109/smart-tech49988.2020.00045
2. P. Mehta, A. Kumar, and S. Bhattacharjee, "Fire and gun violence based anomaly detection system using Deep Neural Networks," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Jul. 2020. doi:10.1109/icesc48915.2020.9155625
3. N. Hnoohom et al., "Weapon detection using faster R-CNN inception V2 for a CCTV surveillance system," 2021 25th International Computer Science and Engineering Conference (ICSEC), Nov. 2021. doi:10.1109/icsec53205.2021.9684649I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4.A. Warsi, M. Abdullah, M. N. Husen, and M. Yahya, "Automatic handgun and Knife Detection Algorithms: A Review," 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM), Jan. 2020. doi:10.1109/imcom48794.2020.9001725
5. R. Kanehisa and A. Neto, "Firearm detection using convolutional neural networks," Proceedings of the 11th International Conference on Agents and Artificial Intelligence, 2019. doi:10.5220/0007397707070714

6. S. Narejo, B. Pandey, D. Esenarro vargas, C. Rodriguez, and M. R. Anjum, "Weapon detection using Yolo V3 for smart surveillance system," Mathematical Problems in Engineering, vol. 2021, pp. 1–9, May 2021. doi:10.1155/2021/9975700
7. B. Y. Gopinath, V. S. Krishna, and G. Srilatha, "Concealed Weapon Detection Using Image Processing," International Journal of Electronics & Communication Technology, vol. 5, no. Spl-3, pp. 13– 17, Mar. 2014.
8. Y. Liu, "An improved faster R-CNN for object detection," 2018 11th International Symposium on Computational Intelligence and Design (ISCID), Dec. 2018. doi:10.1109/iscid.2018.10128 Few
9. A. Jain, Aishwarya, and G. Garg, "Gun detection with model and type recognition using Haar cascade classifier," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Aug. 2020. doi:10.1109/icssit48917.2020.9214211
10. B. Liu, W. Zhao, and Q. Sun, "Study of object detection based on faster R-CNN," 2017 Chinese Automation Congress (CAC), Oct. 2017. doi:10.1109/cac.2017.8243900
11. T. Santos, H. Oliveira, and A. Cunha, "Systematic review on weapon detection in surveillance footage through deep learning," Computer Science Review, vol. 51, p. 100612, Feb. 2024. doi:10.1016/j.cosrev.2023.100612
12. https://blog.roboflow.com/yolov7-breakdown/
13. M. Grega, A. Matiolański, P. Guzik, and M. Leszczuk, "Automated detection of firearms and knives in a CCTV image," Sensors, vol. 16, no. 1, p. 47, Jan. 2016. doi:10.3390/s16010047
14. https://viso.ai/deep-learning/yolov7-guide/
15. https://blog.roboflow.com/whats-new-in-yolov8/
16. https://viso.ai/computer-vision/labelimg-for-image-annotation/
17. https://universe.roboflow.com/123-klcxw/guns-and-knives icl65/dataset/1