A Deep Learning Approach to Gun Detection in Surveillance System using YOLOv9

Lucy Sumi
Dept. of Computer Science & Engineering
National Institute of Technology Nagaland
Chumukedima, India - 797103
lucysumi866@gmail.com

Shouvik Dey
Dept. of Computer Science & Engineering
National Institute of Technology Nagaland
Chumukedima, India - 797103
send2shouvik@gmail.com

2024 Control Instrumentation System Conference (CISCON) | 979-8-3503-7548-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/CISCON62171.2024.10696152

***Abstract*—The escalating firearm related crimes including open firing, robbery, suicides, mass shootings, homicides, threatening at gun point, etc. has underscored the growing importance of timely detection of weapons. This paper presents a gun detection system based on YOLOv9, a state-of-the-art object detection model not previously utilized in literature for gun detection applications. We conduct a comparative analysis with its predecessor versions of YOLOv9, namely YOLOv5, YOLOv6, YOLOv7 and YOLOv8 to assess their performance in gun detection. Our experiments, conducted on a Gun Movie database, reveal that YOLOv9 consistently outperforms its predecessors, demonstrating that YOLOv9 has a good generalizability in detecting firearms in real-time. Particularly noteworthy are the precision and recall values achieved by YOLOv9, reaching 99.6% and 99%, respectively, demonstrating its superior efficacy in accurately identifying guns.**

***Keywords*—Weapon detection system, Deep learning, CCTV surveillances, You Only Look Once v9 (YOLOv9)**

I. Introduction

Based on the statistics from Pew Research Center, gun-related fatalities in the United States were composed of 54% suicides (26,328) and 43% murders (20,958), with the other instances categorized as accidental (549), involving law enforcement (537), or undetermined (458) [1]. The pandemic era saw a significant surge in gun-related homicides, with a steep 45% rise noted between 2019 and 2021, and a 10% increase in gun-related suicides. These concerning statistics underscore a growing concern about the rise in violence and criminal activities involving weapons.

In recent years, the domain of computer vision has been revolutionized by progress in deep learning techniques, offering promising solutions for detecting and identifying weapons in various settings. Traditional methods of firearm detection often rely on handcrafted features such as Canny edge detectors, Histogram of Oriented Gradients (HOG), Speed-Up Robust Features (SURF), etc. [2], which may struggle to adapt to complex real-world scenarios and variations in firearm appearance. On the contrary, convolutional neural networks (CNNs) and deep learning approaches have exhibited higher efficacy in object detection tasks such as firearm detection. By leveraging CNN architectures of two-stage detectors such as Faster R-CNN (Region-based Convolutional Neural Networks) [3] and one stage detectors like Single Shot Multibox Detector (SSD) [4] and YOLO series (You Only Look Once) [5-7] and, researchers have achieved significant advancements in weapon detection accuracy and speed.

Through this paper, we aim to highlight the potential of deep learning-based method in firearm detection capabilities and mitigating the risks associated with firearm-related crimes. The integration of deep learning technologies holds promise for improving public safety and security in today's increasingly complex and interconnected world.

The key highlights of this paper can be outlined as:

i. We incorporated YOLOv9 [8] in firearm detection which has not been used in existing literature, to the best of the authors knowledge.

ii. A comparative analysis is conducted between YOLOv9 and its predecessors (YOLOv5, YOLOv6, YOLOv7, and YOLOv8); and between YOLOv9-c and YOLOv9-e. This comparative evaluation provides insights into the evolution of the YOLO series and highlights the advancements made in firearm detection capabilities.

The subsequent sections of the paper are structured as follows: A summary of the state-of-the-art-literature is presented in Section 2. Section 3 outlines the proposed work, while Section 4 delves into the experimental analysis and their outcomes discussed. Lastly, Section 5 offers concluding remarks of the study.

II. Related Works

One of the earliest contribution in firearm detection was presented by RK Tiwari et al. [9] that used colour based segmentation aided with K-means clustering, and later exploits SURF technique for locating guns from the segmented images. However, the study did not consider various categories of weaponry and was conducted solely on a minuscule dataset, rendering it impractical for real-world scenarios. Gradually, sliding windows approach was applied with machine learning techniques and image processing to build a robust detector [10]. For instance, M. Grega et al. [11] applied sliding windows across images and used MPEG-7 for extracting features of knives and guns to further classify them using Support Vector Machines (SVM).

Nevertheless, sliding windows is computationally expensive and a time-consuming technique which subsequently led to a shift in trend towards the adoption of CNN for training firearms.
A study by Olmos et al. [12] stands out as one of the earliest contributions to this field using deep learning techniques that incorporated FRCNN. The study meticulously assessed two distinct approaches - region proposal and the sliding - and conclusively demonstrated that the former yielded particularly promising results. Hernández et al. [13] applied FRCNN by introducing a novel binarization based methodology that operates at two levels showcasing a remarkable improvement in the detection of small objects including pistol, knives, smartphone, card, etc. The authors have incorporated deep learning techniques, leading to enhanced robustness,

reduction in false positives and heightened accuracy in comparison to other benchmark multi-class detection models.

An interesting approach was presented by A. Castillo et al. [14] that detects handheld cold steel weapons which resolves the illumination challenge in surveillance videos. This was achieved by using brightness guides process, namely DaCoLT (Darkening and Contrast at Learning and Test).

J. Iqbal et al. [15] presented a distinctive approach to weapon detection, by incorporating orientation information of Object Bounding Boxes (OBB) through the utilization of Axis Aligned Bounding Boxes (AABB) during training. The authors introduced a two-step approach that involves predicting the AABB and OBB consecutively. This method underscores a pivotal role in proposing a unique and effective solution to the challenge of weapon detection from a different angle.

With the rising prominence of YOLO in the domain of object detection, researchers have been exploring innovative ways to enhance firearm detection systems. One notable advancement comes from J. Ruiz et al.[16], who introduced a YOLOv3 based system. This system not only leverages the powerful capabilities of YOLOv3 but also integrates pose estimation information to eliminate false positives and improve the accuracy of weapon detection. Expanding upon this foundation, subsequent studies have delved into the potential of other YOLO variants, including YOLOv4, YOLOv5 and transformer model fused with visual feature extraction [17-19].

Lastly, a recent literature proposed a YOLOv7 based system that focused on the bounding box regression loss of a weapon detection system [20]. This presented a fresh perspective and addressed the accurate localization of the bounding boxes for weapons and showed a substantial result on diverse datasets, demonstrating superior results in comparisons to the existing literature.

To the best of the author's knowledge, there has not been any prior research that uses YOLOv9. Therefore, in this study, YOLOv9 was incorporated for firearm detection which is further explained in next section.

III. YOLOv9 Based Gun Detection System

This study utilized YOLOv9 [8] object detection model as the backbone network for weapon detection system. YOLOv9, an evolution of the YOLO (You Only Look Once) series, excels in terms of speed, efficiency and accuracy in detecting objects within images and video streams. It operates by combining the notion of Programmable Gradient Information (PGI) with the Generalized-ELAN (GELAN) architecture which is based on Efficient Layer Aggregation Network (ELAN) [21]. The Generalized Efficient Layer Aggregation Network or GELAN is a novel architecture that integrates ELAN and CSPNet for gradient path planning and feature aggregation, prioritizing lightweight design, fast inference, and accuracy. GELAN extends ELAN's layer aggregation by allowing any computational blocks, ensuring flexibility.

YOLOv9 model comes in five variants, namely, YOLOv9-n, YOLOv9-s, YOLOv9-m, YOLOv9-c and YOLOv9-e for nano, small, medium, compact and extended respectively. While the former two are lightweight models, the later three are medium to large models, with each varying in number of parameters and performance.

IV. Results and Discussion

A. Dataset detail

For the purpose of this study, we have used Gun Movie database [22] which has been used in the literature [11] comprising of 7 short videos. We have extracted the frames of these videos and used a total of 500 images. The datasets were divided into three sets: 70% for training, 20% for validation, and 10% for testing purposes.

B. Implementation details

The experimental setup involves various hardware and software and tools that are used to facilitate implementation with specific and compatible settings to train and test the models. We trained our model on a workstation with hardware specification: NVIDIA GeForce RTX 2080Ti, AMD Ryzen Thread ripper 2990WX 32-Core processor and 32 GB RAM. Anaconda was installed along with software environment of CUDA 11.2, CUDNN 8.1, TensorFlow-gpu 2.10, Pytorch 11 and python version 3.8 in a windows 11 system. The model has been trained using 300 epochs. The implementation base code of YOLOv9 is available for downloaded from https://github.com/WongKinYiu/yolov9/tree/main. A detailed information on various parameters used during the training process are listed in Table I.

TABLE I. Parameter Settings for Training

Parameters	Values	Parameters	Values
Batch size	16	Initial learning rate	0.01
Workers	2	Final learning rate	0.0001
Image size	640*640	Training mode	Single GPU
Epochs	300	Momentum	0.937
Optimizer	SGD	Weight decay	0.0005

C. Experimental ananlysis

This section offers a comparative analysis of state-of-the-art YOLO detectors as recently published, specifically, YOLOv5, YOLOv6, YOLOv7, YOLOv8 and YOLOv9. Table II provides a detailed outcome of the same.

TABLE II. Performance Evaluation of Yolov9 vs. Existing Baselines (YOLOv5, YOLOv6, YOLOv7 and YOLOv8)

Model	GFLOP	#Para. (M)	P	R	mAP @.50	mAP @.5:95
YOLOv5	16.5	7.2	98.2	97.2	97.5	45.8
YOLOv6	44.2	18.5	97.2	98	98	45.8
YOLOv7	104.7	37.1	96	97	95.6	37.6
YOLOv8	28.4	11.1	96.9	96	96.4	46.2
YOLOv9	236.6	50.6	**99.6**	**99**	**99.2**	**48.7**

As seen in the table, YOLOv9 outperforms the other versions of YOLO with a precision of 99.6% and recall 99%. The achieved mAP@.5 and mAP@.5:95 are 99.2% and 48.7%, respectively. The achieved results are attributed to the incorporation of PGI and GELAN, which not only prevents loss of data during gradient updates by capturing and retaining crucial information from the input data, but also boosts the learning capabilities of the model throughout the detection process. This results in more accurate and robust detections in real-world scenarios. Information loss may lead to instances where detection accuracy is compromised, occurrence of false positives and missed detections. The results depicted in the table are also illustrated in Fig. 1.

Fig. 1. Comparative performance evaluation of YOLOv5, YOLOv6, YOLOv7, YOLOv8 and YOLOv9.

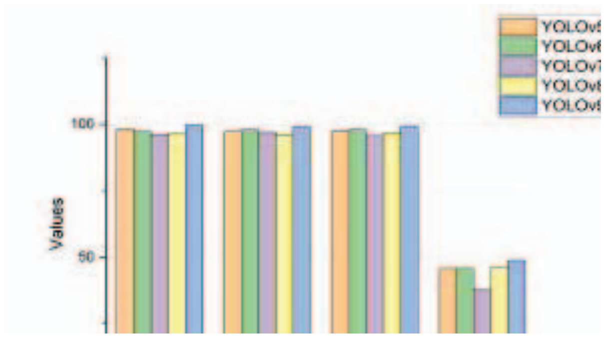


Fig. 1. Comparative performance evaluation of YOLOv5, YOLOv6, YOLOv7, YOLOv8 and YOLOv9.

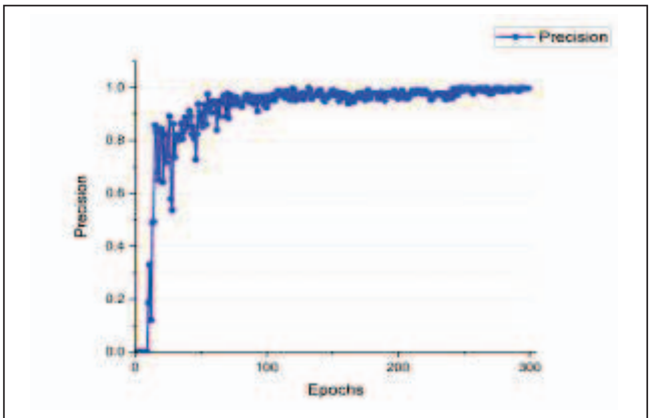
The graphs illustrating precision, recall and mAP values of YOLOv9 throughout the during training process are presented in Fig. 2.

During the drafting of this paper, the YOLOv9-n, YOLOv9-s and YOLOv9-m models were not available for analysing from the authors' official sources. Therefore, we have trained our dataset on YOLOv9-c and YOLOv9-e variant. In Table III, we present a comparison between the performance of YOLOv9-c and YOLOv9-e.

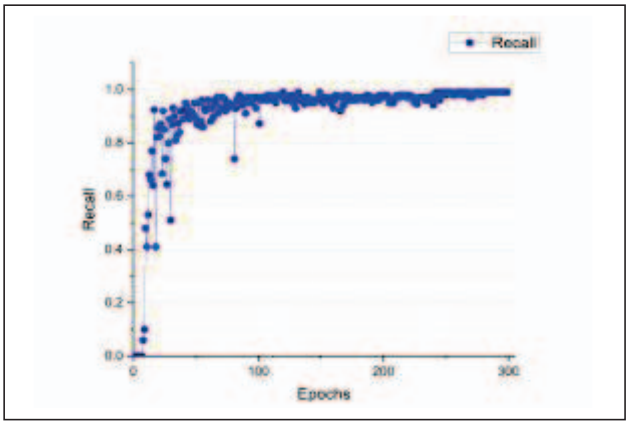
TABLE III. COMPARISION BETWEEN YOLOV9-C AND YOLOV9-E

Model	GFLOP	#Para. (M)	P	R	mAP @.50	mAP @.5:95
YOLOv9-c	236.6	50.6	**99.6**	**99**	**99.2**	**48.7**
YOLOv9-e	240.7	68.5	95.8	95	94.4	45.7

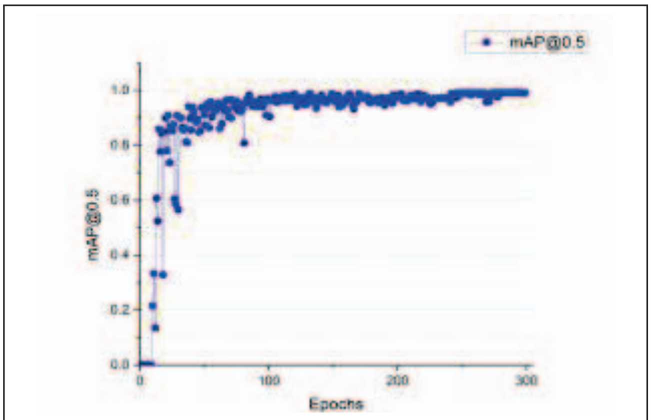
Here, YOLOv9-c outperforms YOLOv9-e due to the risk of overfitting with larger models like YOLOv9-e, which could occur when the dataset lacks sufficient information. This overfitting can lead to a lack of generalization ability on new, unseen data. Consequently, simpler models or those with fewer parameters are often preferred in such situations, as they are less prone to overfitting. The results of the same are displayed in Fig. 3 and the results of detection are displayed in Fig. 4.



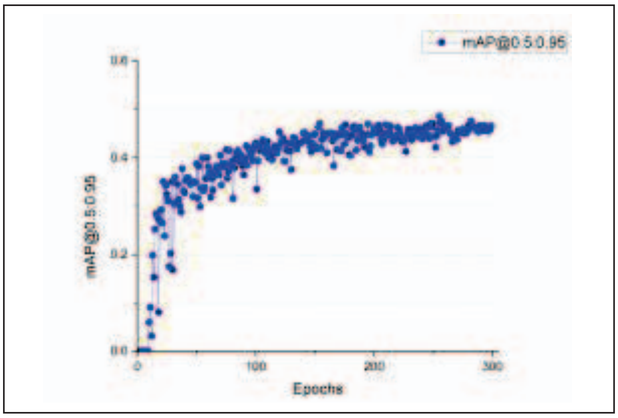
(a)



(b)

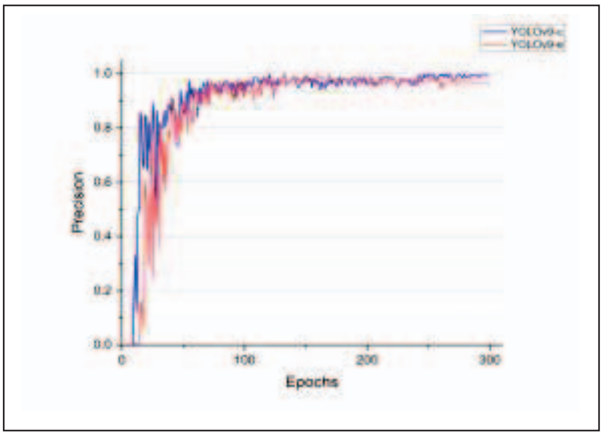


(b)

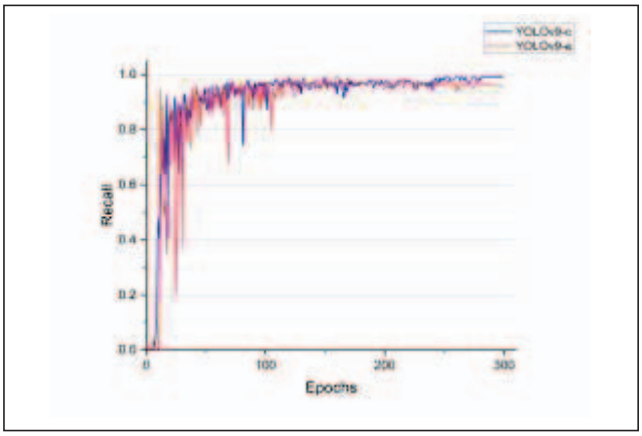


(d)

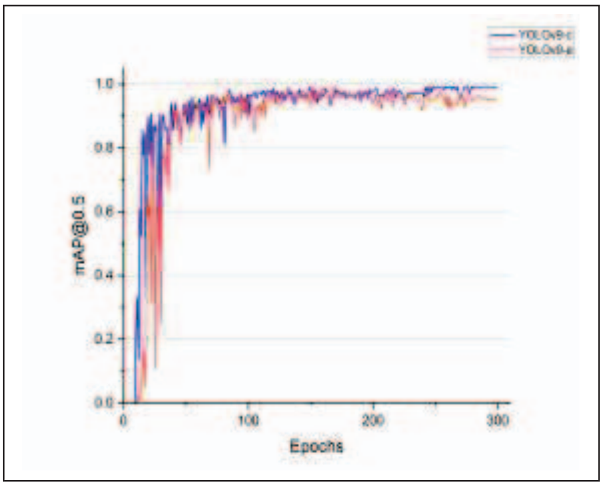
Fig. 2 Graphs depicting performance evaluation of various metrics (a) precision, (b) recall, (c) mAP@.5 and (d) mAP.5:95 achieved by YOLOv9.



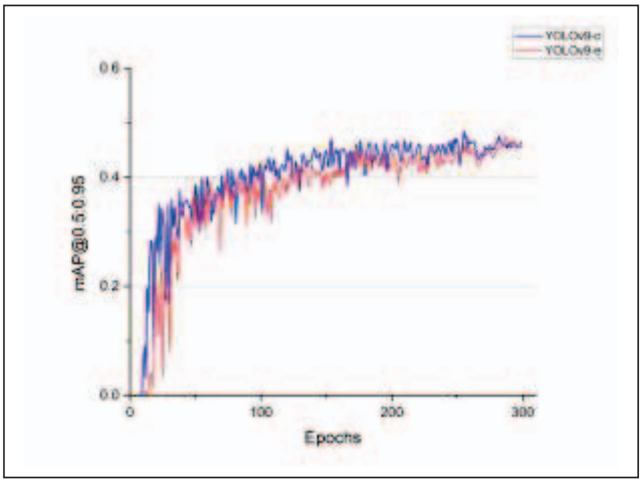
(a)



(b)



(b)



(d)

Fig. 3 Graphs depicting performance comparision between YOLov9-c and YOLov9-e

Fig. 4 Examples of detection predictions obtained by YOLOv9 on test data.

V. Conclusions and Future Works

This paper introduces a gun detection system employs the state-of-the-art YOLOv9 detector. Our study involved a comparative analysis with its predecessor versions, including YOLOv5, YOLOv6, YOLOv7, and YOLOv8, to evaluate performance and efficacy. Through experimentation, we have demonstrated the superiority of YOLOv9 in terms of precision, recall, and mean Average Precision (mAP), showcasing its enhanced capabilities in accurately detecting firearms within images and video streams. The results underscore the importance of leveraging advanced object detection models like YOLOv9 for enhancing security measures and public safety initiatives. Future research could explore the integration of 3D-based and rotating bounding boxes into gun detection systems, enhancing their spatial understanding and adaptability. Moreover, extending this study to address challenges posed by low-quality datasets presents a promising avenue for investigation. Overcoming complexities associated with such data could significantly bolster the robustness and practicality of gun detection systems.

References

[1] J. Gramlich, (2023, Apr. 26). What the data says about gun deaths in the U.S. [online]. Available: https://www.pewresearch.org/short-reads/2023/04/26/what-the-data-says-about-gun-deaths-in-the-u-s/

[2] H. Bay, T. Tuytelaars and L.V. Gool, "Surf: Speeded up robust features," in *Comput. Vision 9th Eur. Conf. Comput. Vis. (ECCV-2006),* Berlin, Heidelberg, 2006, pp. 404-417.

[3] S. Ren, K. He, R. Girshick and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Adv. in Neural Inf. Process. Syst. 28*, 2015.

[4] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu and A.C. Berg, "Ssd: Single shot multibox detector," in *Comput. Vision– ECCV 2016: 14th European Conference*, Amsterdam, The Netherlands, 2016, pp. 21-37.

[5] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. of the IEEE conf. Comput Vis. and Pattern Recognit.*, 2016, pp. 779-788.

[6] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng Y, L. Li, et al., "YOLOv6: A single-stage object detection framework for industrial applications," 2022. arXiv preprint arXiv:2209.02976.

[7] C. Y. Wang, A. Bochkovskiy and H.Y. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. of the IEEE/CVF Conf. Comput. Vis. and Pattern Recognit.*, 2023, pp. 7464-7475.

[8] C. Y. Wang, I. H. Yeh and H.Y. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information", 2023. arXiv preprint arXiv:2402.13616.

[9] R. K. Tiwari and G. K. Verma, "A computer vision based framework for visual gun detection using SURF," in *Intl. Conf Elect., Electron, Signals, Commun. and Optimim. (EESCO)*, 2015, pp. 1-5.

[10] L. Sumi and S. Dey, "Gun Detection System for Surveillance Cameras Using HOG-Assisted KNN Classifier." in *Int. Conf. Big Data, Mach. Learn., and Appl.,* Singapore, 2021, pp. 221-233.

[11] M. Grega, A. Matiolański, P. Guzik and M. Leszczuk, "Automated detection of firearms and knives in a CCTV image," *Sensors*, vol. 16, issue 1, pp. 47, Jan. 2016.

[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.

[13] F.P. Hernández, S. Tabik, A. Lamas, R. Olmos, H. Fujita and F. Herrera, "Object detection binary classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillances," *Knowl. based Syst.*, Apr. 2020.

[14] A. Castillo, S. Tabik, F. Pérez, R. Olmos and F. Herrera, "Brightness guided pre-processing for automatic cold steel weapon detection in surveillance videos with deep learning," *Neurocomputing*, vol. 330, pp. 151-161, 2019.

[15] J. Iqbal, M. A. Munir, A. Mahmood, A. R. li and M. Ali, "Leveraging orientation for weakly supervised object detection with application to firearm localization," *Neurocomputing*, vol. 440, pp. 310-320, Feb. 2021.

[16] J. R. Santaquiteria, A. V. Mata, N. Vallez, G. Bueno, J. A. Garcia and O. Deniz, "Handgun detection using combined human pose and weapon appearance," *IEEE Access*, vol. 9, pp. 123815-26, Sept. 2021.

[17] M. T. Bhatti, M. G. Khan, M. Aslam and M. J. Fiaz, "Weapon detection in real-time cctv videos using deep learning*", IEEE Access*, vol. 9, pp. 34366-82, Feb. 2021.

[18] J. R. Santaquiteria, A. Mata, N. Vallez, O. Deniz and G. Bueno, "Improving handgun detection through a combination of visual features and body pose-based data," *Pattern Recognit.*, vol. 136, pp. 109252, Apr. 2023.

[19] L. Sumi and S. Dey, "YOLOv5-based weapon detection systems with data augmentation," *Intl. J. Comput. and Appl.*, vol. 45, issue 4, pp. 288-296, Apr. 2023.

[20] L. Sumi and S. Dey, "Improved Bounding Box Regression Loss for Weapon Detection Systems using Deep Learning", *Int. J. of Inf. and Technol.*, pp. 1-17, May. 2024.

[21] C.Y Wang, H. Y. Mark and I. H Yeh, "Designing network design strategies through gradient path analysis," *J. of Inf. Sci. and Eng. (JISE)*, vol. 39, issue, pp. 975–995, Nov. 2022.

[22] Gun Movie Database. Available: https://kt.agh.edu.pl/grega/guns/