https://doi.org/10.34836/pk.2026.325.7

Weapons detection in video surveillance systems using deep neural networks

2LT Maciej Płoński
Central Forensic Laboratory of the Police
ORCID: 0009-0005-8847-0372
e-mail: maciej.plonski@policja.gov.pl

Abstract

The purpose of this paper is to conduct an empirical assessment of the usefulness of selected deep neural network architectures in detection of weapons in CCTV images. The analysis compared single-stage YOLOv5 detectors and two-stage Faster R-CNN detectors, each trained under identical conditions on a limited dataset. The experiment used a dataset from the Kaggle platform comprising 4,000 images of firearms and knives. The models were trained with transfer learning on COCO pre-trained weights, and their performance was assessed in terms of mAP, precision, recall and frames per second (FPS). The results obtained indicate that the YOLO family offers a favourable trade-off between accuracy and speed, enabling video stream processing in near real time.

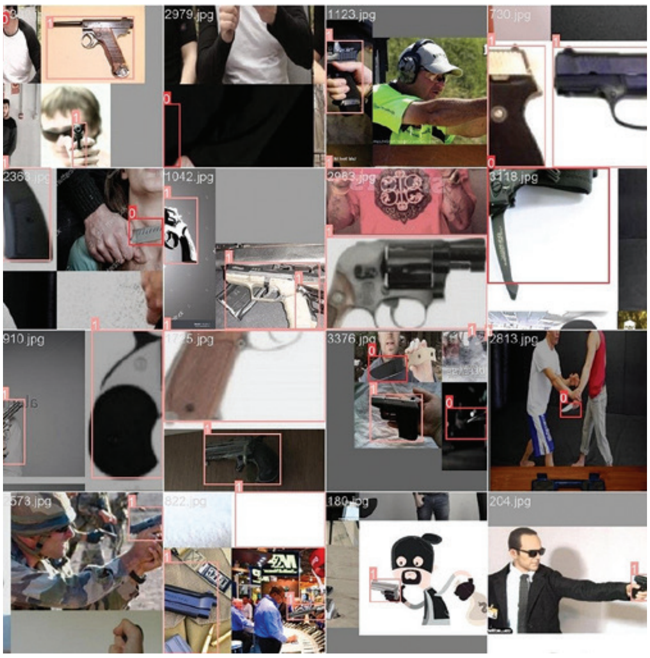
Keywords: artificial intelligence, deep neural networks, weapons detection, YOLO, Faster R-CNN, CCTV monitoring

Introduction

The growing number of CCTV cameras in urban and industrial environments means that it is no longer practicable for operators to analyse all video streams manually. Estimates by market analysts suggest that the global number of such devices has exceeded one billion in recent years, with a significant proportion being installed in public spaces and critical infrastructure (Torregrosa-Domínguez et al., 2024). In this context, the task of detecting dangerous situations, such as the appearance of a weapon, can be automated by means of deep learning methods. Real-time solutions are of particular practical importance, as they enable triggering responses such as operator alert, access block or patrol call before an incident escalates. Traditional monitoring observation is vitiate by vigilance decrement. As noted by Velastin et al. (2006), even trained operators cease to be able to spot objects effectively after around 20 minutes of continuous observation (Torregrosa-Domínguez et al., 2024).

Studies into object detection have preserved a distinction between two solution families: fast single-stage models, such as YOLO and SSD, and more accurate, albeit slower, two-stage models, including Faster R-CNN and Mask R-CNN (Reswara et al., 2023; Yusro et al., 2023). Single-stage models achieve a FPS rates, which makes them well-suited for alarm applications. Two-stage models may offer greater classification precision at the expense of speed. However, Polish-language literature lacks a systematic description of the effectiveness of these approaches to weapons detection in CCTV monitoring scenarios, and an assessment of the benefits of transfer learning in the case of small, specialised datasets. An earlier comparison was carried out in a master's thesis by M. Płoński (2025), in which YOLOv5 and Faster R-CNN were compared on a shared dataset of images.

The aim of this study is to determine which of the two commonly used approaches, i.e. the single-stage (YOLO) or two-stage (Faster R-CNN) models, better meets the requirements for weapons detection in surveillance with limited training data, and to estimate the impact of transfer learning on performance.



Phot. 1. Examples of object location and classification

Literature review

Weapons detection is a special case of small-object detection. Pistols, revolvers and knives usually occupy a small part of the frame; they are often partially obscured and appear under a variety of lighting conditions. Studies indicate that the low resolution of urban CCTV cameras installed at great distances limits the practical effectiveness of detection and requires the use of data augmentation and scale-invariant models (Torregrosa-Domínguez et al., 2024). Salazar et al. (2020) noted that a in a camera feed a handgun may occupy around 3% of the frame. More recent implementations based on YOLOv8 are designed to operate in live video surveillance mode, which underscores the need for effective real-time detection of small objects.

Single-stage and two-stage models for object detection differ in their processing strategies. Single-stage detectors, such as YOLOv5 and YOLOv8, formulate a task as a joint regression and classification within a single network, which enables FPS rates to be achieved with acceptable accuracy (Reswara et al., 2023; Yusro et al., 2023). Two-stage detectors, e.g. Faster R-CNN, first generate proposals for regions of interest, and then classify them and refine boundaries, which usually improves the accuracy of recognition but increases the cost of computation. Comparative literature consistently concludes that YOLO is better suited for applications that require speed, whereas Faster R-CNN may have the edge in scenarios where minimising false alarms is crucial (Reswara et al., 2023; Aboyomi, 2023; Yusro et al., 2023).

Since 2012, the ImageNet dataset, which was used in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) between 2012 and 2017, has played a key role in the development of object detection methods. It comprises over 14 million images assigned to a thousand classes and forms the basis for training modern convolutional networks (Deng et al., 2009). Since 2023, several datasets designed for training weapon detectors have been available on the Kaggle platform, including the Weapon Detection Dataset for YOLOv5, which contains 4,000 images of pistols and knives, as well as more recent multi-class datasets. These datasets are relatively small, which justifies the use of transfer learning. The use of COCO pre-trained weights has improved the performance of the YOLOv5 model by approximately 9 percentage points mAP@0.5 compared to training from scratch. Kunchala (2024) reported similar results for YOLOv8.

Methodology

The experiment was based on the Weapon Detection Dataset for YOLOv5, comprising 4,000 images, including 2,000 examples of firearms and 2,000 examples of knives. The dataset is publicly available on the Kaggle platform. The images were manually annotated in the YOLO format, i.e. using the coordinates of a bounding box and an object class label. The data was divided into a training set (80%), a validation set (10%) and a test set (10%). Scene separation between subsets was ensured to prevent information leakage and overestimation of results.

In order to improve the model's generalisation, standard data transformations were applied: random scaling, brightness adjustment, rotation and horizontal mirroring. In addition, the Mosaic technique was used in some of the tests, which improved YOLOv5's ability to recognise small objects. The augmentations used artificially increase the diversity of training samples, which is particularly important when a dataset is limited in size.

Two object detection models were selected for comparison: YOLOv5 in its small variant, representing lightweight, single-stage detectors, and Faster R-CNN with the ResNet-50 backbone, which is representative of the classic, two-stage approaches commonly described in the literature (Reswara et al., 2023; Vijayakumar et al., 2023). In both cases, transfer learning was applied using COCO pre-trained weights. The training process was carried out until the mAP value stabilised on the validation set, with the minimum number of epochs set at 100 for YOLOv5 and 50 for Faster R-CNN.

The performance of the detectors under comparison was assessed with the use of four metrics. First, mAP@0.5 as the primary indicator of object detection precision. Second, mAP@[0.5:0.95], i.e. the mean average precision calculated for IoU thresholds ranging from 0.5 to 0.95 in 0.05 increments. Third, precision and recall, which describe the quality of the 'object vs. non-object' classification. Fourth, the number of frames processed per second (FPS) as a measure of the model's suitability for real-time applications.

Results

The key results of the experiment are summarised below (Table 1). Both YOLOv5 and Faster R-CNN achieved results that fell within the ranges reported in the most recent comparative studies on these architectures (Reswara et al., 2023; Yusro et al., 2023).

Tab. 1. Comparison of the effectiveness and efficiency of the YOLOv5

Model	mAP@0.5 (%)	mAP@[0.5:0.95] (%)	Precision (%)	FPS
YOLOv5	71	36–40	90	139
Faster R-CNN	68	38–42	93	9–10

The results obtained are consistent with the observations reported in the literature. YOLOv5 achieves a significantly higher frame rate, at 139 FPS, than Faster R-CNN, at around 9 to 10 FPS, whilst maintaining high accuracy in object localisation. With the development of the YOLO architecture, the mAP achieved for simple classes, including weapons, is comparable to the results of two-stage models. On the other hand, Faster R-CNN offers higher precision at the expense of speed. The consistency of these relationships is confirmed by comparative studies, e.g. Yusro et al. (2023), who reported a higher overall accuracy for YOLOv5 (89.12 per cent) compared with Faster R-CNN (83.92 per cent) in an identical test environment.

Discussion

The results obtained clearly point to various application scenarios for the detection architectures under consideration.

Real-time monitoring

In scenarios where the priority is the shortest possible response time, e.g. CCTV monitoring of public spaces, single-stage detectors are recommended. YOLOv5, and YOLOv8 in more recent implementations, delivers very high throughput whilst maintaining sufficient detection accuracy, as confirmed by both the results of this study and the literature (Yusro et al., 2023; Ahmed and Das, 2025). Ahmed and Das (2025) point out that YOLOv5 is a strong candidate for real-time applications, offering a favourable trade-off between speed and precision.

Frame-by-frame/tracking analysis

In a post-incident analysis of video footage, where a delay is acceptable, the priority is to ensure the greatest possible accuracy in locating the object and to minimise false alarms. In such circumstances, it may be appropriate to use two-stage models, which provide more refined predictions and generate fewer false positives (Ahmed and Das, 2025). Despite Faster R-CNN's lower processing speed, a delay of around several seconds in forensic analysis scenarios remains acceptable, provided it results in higher detection precision.

Data quality and limitations

It is worth noting that the study was based on relatively homogeneous, correctly labelled images. In urban surveillance environments, low resolution, poor lighting and partial obscurity of the weapon can significantly reduce the effectiveness of detectors. Torregrosa-Domínguez et al. (2024) indicate that limitations arising from image quality require additional measures, such as increasing the model's input resolution, applying segmentation or fusing results from multiple detectors, in order to maintain high performance in real-world environments.

The implementation of automatic weapon detection systems should be carried out in accordance with the applicable legislation. In the Polish context, regulations – including the Police Act and the GDPR – require that an automated system should only generate an indication of a potential threat, whilst the decision to intervene remains with the operator. These systems therefore serve a supporting role and enhance the effectiveness of prevention and evidence-based analysis, without replacing human involvement.

Conclusions

The study confirms that deep neural networks enable the effective detection of weapons in CCTV footage by means of transfer learning, achieving accuracy of over 70% mAP@0.5 despite the limited size of the training dataset. In the experiment presented, YOLOv5 achieved 71% mAP@0.5, which confirms the usefulness of this class of models in the detection of small objects.

The results indicate that the YOLOv5 architecture offers significantly higher computational throughput than Faster R-CNN – approximately 139 FPS compared with approximately 10 FPS, respectively – with a relatively minor loss of classification accuracy. These characteristics make YOLO family models more suitable for real-time operational applications in public safety systems.

At the same time, Faster R-CNN remains a valuable tool in the analysis of evidence, where the acceptable delay is offset by greater localisation precision and fewer false alarms. In scenarios where the priority is to minimise false positives and maximise precision, a two-stage detector remains useful and can complement single-stage solutions (Ahmed and Das, 2025).

Literature

1. Aboyomi, D. D. (2023). YOLO vs. SSD vs. Faster R-CNN: A comparative study on object detection. Information Technology and Engineering Journal, 8(2), 45–54.
2. Ahmed, I., & Das, R. (2025). Comparative analysis of YOLO and Faster R-CNN models for detecting traffic objects. International Journal of Ad-

vanced Computer Science and Applications, 16(3). https://doi.org/10.14569/IJACSA.2025.0160342
3. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. [In:] 2009 IEEE Conference on Computer Vision and Pattern Recognition (pp. 248–255). IEEE. https://doi.org/10.1109/CVPR.2009.5206848
4. Kaggle. (2023). Weapon detection dataset for YOLOv5. https://www.kaggle.com/datasets/balraj98/weapon-detection-object-detection
5. Kunchala, H. K. R. (2024). Real-time gun detection in video streams using YOLOv8 (master's thesis, California State University). California State University ScholarWorks.
6. Płoński, M. (2025). Analiza porównawcza architektur YOLO i Faster R-CNN w detekcji broni z wykorzystaniem transfer learningu (master's thesis). Akademia Policji w Szczytnie.
7. Reswara, E., Sari, R., & Widodo, A. (2023). Comparison of object detection algorithm using YOLO vs Faster R-CNN. [In:] Proceedings of the 2023 International Conference on Big Data Technologies (ICBDT). ACM. https://doi.org/10.1145/3627377.3627443
8. Salazar González, J. L., Zaccaro, C., Álvarez-García, J. A., Soria-Morillo, L. M., & Sancho Caparrini, F. (2020). Real-time gun detection in CCTV: An open problem. Neural Networks, 132, 297–308.
9. Torregrosa-Domínguez, Á., Álvarez-García, J. A., Salazar-González, J. L., & Soria-Morillo, L. M. (2024). Effective strategies for enhancing real-time weapons detection in industry. Applied Sciences, 14(18), 8198. https://doi.org/10.3390/app14188198
10. Velastin, S. A., Boghossian, B. A., & Vicencio-Silva, M. A. (2006). A motion-based image processing system for detecting potentially dangerous situations in underground railway stations. Transportation Research Part C: Emerging Technologies, 14(2), 96–113.
11. Vijayakumar, K. P., Pradeep, K., Balasundaram, A., & Dhande, A. (2023). R-CNN and YOLOv4 based deep learning model for intelligent detection of weaponries in real-time video. Mathematical Biosciences and Engineering, 20(12), 21611–21625. https://doi.org/10.3934/mbe.2023956
12. Yusro, M. M., Ali, R., & Hitam, M. S. (2023). Comparison of Faster R-CNN and YOLOv5 for overlapping object detection. Baghdad Science Journal, 20(3), 893–903. https://doi.org/10.21123/bsj.2022.7243

Legal acts

13. Act of 6 April 1990 on the Police (Journal of Laws of 1990, item 179).
14. Act of 10 May 2018 on the protection of personal data (Journal of Laws of 2018, item 1000).