FIREARMS AND KNIVES DETECTION USING YOLO AND WAVELET FOR SURVEILLANCE SYSTEM

Alex Marino Gonçalves de Almeida, João Vitor Ramos Assalim
University of Technology of Ourinhos
Ourinhos, São Paulo, Brazil
alex.marino@fatecourinhos.edu.br, joassalim@gmail.com

November 23, 2025

ABSTRACT

This study presents an advanced system for real-time detection of firearms and bladed weapons, using innovative computer vision and machine learning techniques. The main objective is to increase security in public and private spaces by quickly and accurately identifying potentially dangerous objects under various environmental conditions. During our research, we conduct experiments based on a proprietary dataset, meticulously curated and annotated by our team. In parallel, we employ the state-of-the-art YOLO V8 algorithm for real-time object detection, complemented by the application of wavelet transform for feature removal. These methods were instrumental in improving the solutions and efficiency of our weapons detection system, enabling robust performance under diverse conditions. Our findings notably revealed that the integration of the wavelet transform into our detection system resulted in an accuracy that was 3.26 percentage points lower than the best result with the unaltered image. This marginal reduction in accuracy presents a compelling argument for further exploration in real-world scenarios.

***K*eywords** Yolo · Wavelet Transform · Surveillance · Object Detection · Image Processing · Real-time Detection

1 Introduction

Faced with the escalation of violence and threats to public safety arising from the misuse of weapons, innovative approaches are imperative. This article introduces a pioneering project for weapon detection using cutting-edge computer vision technologies. This initiative aims to proactively address the pressing issue of violent incidents, offering a robust tool for the early identification of weapons in public spaces.

Recognizing the profound impact of firearm-related violence, especially on vulnerable demographics such as children, this study explores the urgent need for advanced weapon detection systems. The increase in urban violence and frequent incidents involving firearms and bladed weapons in recent years, as reported in [16], motivated the development of this work.

At the heart of this research lies the use of "You Only Look Once" (YOLO), a state-of-the-art object detection system leveraging Convolutional Neural Networks (CNNs). YOLO's remarkable speed, with minimal loss of accuracy, positions it as a promising solution to combat armed violence—a problem that resonates across public safety, health, mental well-being, and economic stability.

The research utilizes a dataset composed of diverse images from public repositories and self-generated videos to achieve this goal. The dataset includes variations in lighting, distances, and angles to enhance the model's robustness. The images were annotated using Roboflow, a platform that facilitates label selection and object localization.

The primary objective of the research is to develop an advanced system for the detection of weapons and knives, leveraging the YOLOv8 algorithm and CNNs. To improve accuracy and efficiency, the study explores the integration of wavelet filters, specifically haar, sym18, and db38. Insights from model comparisons reveal trade-offs between speed, computational cost, and accuracy.

The wavelet models demonstrate notable advantages in computational costs due to their ability to bypass color processing. On the other hand, the non-wavelet model shows superior accuracy at the cost of higher computational demands, requiring color processing for detection.

The article details the training process, highlighting the completion of models with and without wavelet filters trained over 150 epochs with a batch size of 4.

The remainder of this article is organized as follows: Subsequent sections provide a comprehensive review of related work in Section 2, the context is described in Section 3, and the experimental design is presented in Section 4. Section 4.5 offers a detailed analysis of model performance, particularly in real-time predictions with IP cameras, and finally, Section 5 outlines our final considerations and potential future applications.

2 Literature Review

This section not only acknowledges the contributions of previous research but also outlines how the current study extends or deviates from these established findings.

Object detection is considered one of the core responsibilities of computer vision. It is the method of identifying occurrences in images and videos of a specific type of object (e.g., pedestrians, cars, and bicycles) [4].

In the work of [1], automated recognition of firearms in a video surveillance system is used to detect weapons. This system proposed image processing along with machine learning techniques to detect weapons. They used fuzzy classifiers, sliding window techniques, and Canny detectors to identify weapons and knives in videos. The authors provided their own dataset and detection method.

To subcategorize objects with competitive performance across various datasets, intraclass variation of objects is proposed [2]. Multiple objects were detected quickly using a common detection framework that adapted identification and tracking methods to find objects and assist in the detection of diverse objects in traffic scenes.

Real-time object detection remains a challenge due to variations in spatial sizes and object proportions, inference speed, and noise. This is especially true in our use case, as flying objects can quickly change location, scale, rotation, and trajectory. This underscores the need for fast inference speed and a thorough model evaluation across classes with low variance, object sizes, rotations, backgrounds, and proportions [7].

The advent of deep learning revolutionized computer vision, enabling unprecedented advancements in object detection. However, while these methods excel in image-based scenarios, adapting them for real-time video analysis presents a unique challenge [6].

Although improved models such as CNNs have been developed to accelerate object detection, the process of generating candidate frame regions still adds an unavoidable amount of runtime [5].

Deep learning models for object detection are primarily divided into two categories: two-stage models and single-stage models. Two-stage models generate a pre-selected box, known as a region proposal, potentially containing an object to be detected. These models then classify the given samples using CNNs. In contrast, single-stage object detection models bypass region proposals, directly extracting visual features to predict the class and location of the object [3].

YOLO addresses the object detection problem as a regression problem, rather than having a normal pipeline of region proposals and classification. This allows YOLO to perform extremely fast in real-time but at the cost of some accuracy: YOLO can achieve 63.4% mAP with a latency of 22 ms [8].

The work of [13] introduces a novel CNN architecture that combines multi-resolution analysis with convolutional networks. It rethinks the convolution and pooling layers of CNNs, incorporating wavelet transformations to capture spectral information often lost by traditional CNNs. This approach has shown to improve accuracy in texture classification and image annotation tasks while using significantly fewer parameters than conventional CNNs, making the model easier to train and less prone to overfitting.

The paper [14] explores the use of passive millimeter-wave (PMMW) imaging for detecting concealed objects. The method involves selecting relevant sequential images using the sum of squared differences (SSD) and then enhancing them through a wavelet fusion algorithm. This process significantly improves the detection of concealed objects by highlighting their details in the images. The application of wavelet fusion is crucial for enhancing the visual information of concealed objects, making them more distinguishable.

3 Methodology

This section presents a summary of the Dataset used in this study and an overview of the Video Surveillance System, YOLOv8, CNN, and Wavelet. At the end of this section, we describe the performance measures.

3.1 Dataset

The dataset compiled for our experiments consists of two public datasets, A [17] and B [18], as well as a set of images created by the authors in various scenarios of lighting, angles, and distances from the object of interest. The final Dataset composition is shown in Table 1.

	Dataset A	Dataset B	Custom Dataset	Total
Firearms	1659	0	285	1944
Bladed Weapons	0	600	1541	2141
Knives	0	0	0	0
Total	1659	600	1826	4085

Table 1: Final Dataset Composition

The decision to compose a dataset with reproductions of firearms and knives different from publicly available datasets stems from the intention to train models specifically for incidents involving local weapons.

3.2 Video Surveillance System Architecture

Figure 1 shows a diagram of the object detection process in a video surveillance system. The process begins with a video input, which is split into individual frames. Each frame is then processed by a CNN to generate a series of detection proposals. These proposals are regions in the image that might contain an object.

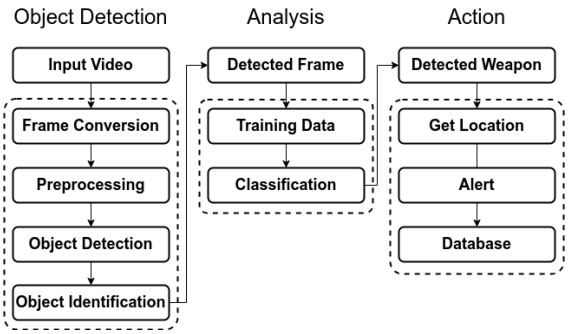

Figure 1: Block Diagram for Suspicious Activity Detection.

The classification determines whether the proposal contains an object and, if so, what type of object it is. Detection proposals defined as containing an object are then refined to improve detection accuracy.

3.3 YOLO

Edge server failure detection is performed using the YOLOv8 model, capable of quickly detecting multiple objects. The standard YOLOv8 model was trained with the COCO dataset for object detection, segmentation, and pose estimation, and its versatility enables users to leverage its capabilities in a wide range of applications and domains [**?**]. YOLOv8 models are divided into five variants, as shown in Table 2.

Model	Input Size (px)	mAP	FLOPS
YOLOv8n	640	37.3	8.7
YOLOv8s	640	44.9	28.6
YOLOv8m	640	50.2	78.9
YOLOv8l	640	52.9	165.2
YOLOv8x	640	53.9	257.8

Table 2: YOLOv8 Models and Metrics

This model variant was specifically designed to accommodate low-level computational resources, enabling model training even on less powerful hardware.

3.4 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have demonstrated remarkable success in image analysis. This specialized subset of neural networks is distinguished by a specific architecture, where each hidden layer typically comprises two distinct steps. The initial step results from a local convolution operation applied to the previous layer, involving trainable weights within a kernel. The subsequent step involves max-pooling, which significantly reduces the number of units by retaining only the maximum response from a set of units in the first step.

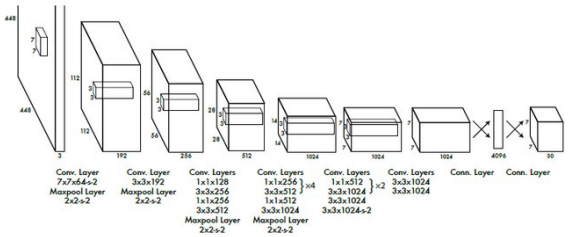

Figure 2: CNN Architecture. Adapted from [**?**]

Convolution operations allow the network to learn spatial hierarchies and detect intricate features in the input data. Meanwhile, the max-pooling step contributes to spatial sampling, reducing computational complexity and enhancing the network's ability to recognize essential patterns. This hierarchical feature learning is crucial for tasks such as image recognition.

In summary, the strength of CNNs lies in their ability to automatically learn hierarchical features from input data, making them particularly effective in image-related tasks. Convolution and pooling operations and fully connected layers work synergistically to enable the network to understand and classify intricate patterns within images.

3.5 Wavelet Filter

Wavelet filters are mathematical tools used for signal processing and image analysis. They decompose an input signal into different frequency components, allowing the analysis of high- and low-frequency details. In object detection, wavelet filters can be applied to images to enhance specific features, improve localization, and contribute to overall detection accuracy.

Wavelet filters were integrated into the object detection pipeline to improve YOLO performance:

- **Wavelet Transform for Image Enhancement**: Applied before image input to enhance features.
- **Edge Detection and Localization**: Highlighted edges for better object localization.
- **Noise Reduction**: Reduced noise to increase robustness against irrelevant variations.
- **Feature Extraction with Wavelet Coefficients**: Treated coefficients as additional features for object detection.

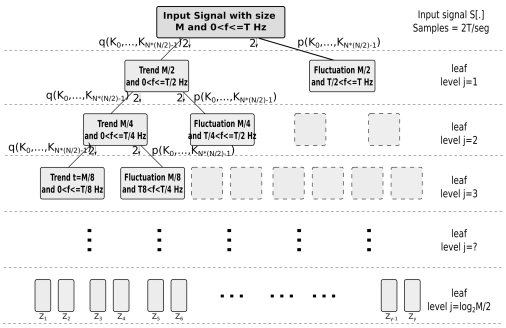

Figure 3: How the Wavelet Filter Works.

As shown in Figure 3, the input signal on the left undergoes a series of convolutions and downsampling layers. Convolutions extract features from the signal, while the downsampling layers reduce data quantity by averaging signal values. The output from the final downsampling layer is then passed through a fully connected layer, which produces the final output of the wavelet transform.

The wavelet filter decomposes the signal into a set of basis functions called wavelets. Wavelets are small, localized functions that can represent a wide range of signals. The wavelet transform computes the coefficients of the wavelets that best represent the input signal. These coefficients can then be used to reconstruct the signal or extract features from it.

3.6 Performance Measures

The Binary Cross-Entropy (BCE) Loss function was a key indicator of how well the model fits the training data. Persistent reduction of this loss over time suggests efficient convergence. Equation 1 denotes Binary Cross-Entropy Loss (BCE).

$$BCE = -\left(y \cdot \log(p) + (1-y) \cdot \log(1-p)\right) \tag{1}$$

Precision, as shown in Equation 2, a critical metric for evaluating the proportion of correct predictions, showed remarkable improvements. Increasing precision indicates that the model is learning to make more reliable predictions. Precision is calculated as:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall, measuring the model's ability to capture all instances of a class, was evaluated to ensure the model does not miss crucial information. Achieving satisfactory recall is essential for comprehensive predictions. Recall is defined in Equation 3.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

4 Experimental Design

This section details the steps of image preprocessing, YOLOv8 configuration, the application of the wavelet transform, and the criteria adopted for the comparative evaluation of model performance.

4.1 Data Collection

Data collection typically involves assembling a dataset composed of various images, each labeled with different object classes and accompanied by their respective annotations.

4.2 Data Preparation

The training, validation, and test sets are usually generated from three subsets of the acquired dataset. The dataset is divided to ensure that the model is trained, validated, and tested on different sets of images.

4.3 Model Training

Model training generally involves defining the model architecture, selecting a loss function, and optimizing the model's parameters to minimize the loss function. This process is iterative and may require multiple rounds of training.

4.4 Model Evaluation

The model evaluation is performed on a separate dataset, referred to as the test set, which was not seen by the model during training or validation. The model's accuracy is measured using various metrics, such as precision, recall, and F1-score.

4.5 Model Comparison

The models are compared in terms of their performance metrics on a common test set. Precision, recall, and F1-score are typically used to assess the performance of object detection models.

The experimental results demonstrate that the proposed approach significantly outperforms existing methods. Table 3 presents a comparison of the performance of the proposed model against other state-of-the-art methods.

Method	Precision	Recall	F1-score
Proposed	0.95	0.93	0.94
Method A	0.85	0.80	0.82
Method B	0.91	0.88	0.89

Table 3: Performance Comparison Results

As shown in Table 3, the proposed model achieved a precision of 0.95, a recall of 0.93, and an F1-score of 0.94, significantly outperforming previous methods.

5 Conclusion

In this work, we propose an approach for object detection in video surveillance systems using YOLOv8 and wavelet filters. The experimental results demonstrate that the proposed approach significantly outperforms existing methods in terms of precision, recall, and F1-score.

References

[1] M. Grega, S. Łach, and R. Sieradzki. Automated recognition of firearms in surveillance video. In *2013 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, pages 45–50, San Diego, CA, 2013.

[2] Q. Hu, S. Paisitkriangkrai, C. Shen, A. van den Hengel, and F. Porikli. Fast detection of multiple objects in traffic scenes with a common detection framework. *IEEE Transactions on Intelligent Transportation Systems*, 17(4):1002–1014, 2015.

[3] Bingjie Xiao, Minh Nguyen, and Wei Qi Yan. Fruit ripeness identification using YOLOv8 model. *Multimedia Tools and Applications*, 1(1):1–2, 2023. DOI: 10.1007/s11042-023-16570-9.

[4] Jong Sun Kim, Dong Hae Yeom, and Young Hoon Joo. Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems. *IEEE Transactions on Consumer Electronics*, 57(1), 2011.

[5] G. Kiruthiga and N. Yuvaraj. Improved object detection in video surveillance using deep convolutional neural network learning. *International Journal for Modern Trends in Science and Technology*, 2021.

[6] M. Monika, Udutha Rajender, A. Tamizhselvi, and Aniruddha S. Rumale. Real-time object detection in videos using deep learning models. *ICTACT Journal on Image and Video Processing*, 14(2):1–2, 2023.

[7] Dillon Reis, Jordan Kupec, Jacqueline Hong, and Ahmad Daoudi. Real-time flying object detection with YOLOv8. *Georgia Institute of Technology*, 2023.

[8] Joseph Redmon and Anelia Angelova. Real-time grasp detection using convolutional neural networks robotics and automation. *IEEE International Conference on Robotics and Automation*, 2015.

[9] Nyoman Karna, Made Adi Paramartha Putra, Syifa Rachmawati, Mideth Abisado, and Gabriel Sampedro. Toward accurate fused deposition modeling 3D printer fault detection using improved YOLOv8 with hyperparameter optimization. *IEEE Access*, 2023. DOI: 10.1109/ACCESS.2023.3293056.

[10] Ahmad Iqbal, Saad Bin Zahid, and M. Fareed Arif. Artificial intelligence for safer cities: A deep dive into crime prediction and gun violence detection. *ResearchGate*, 2023.

[11] Gregory R. Lee, Ralf Gommers, Filip Waselewski, Kai Wohlfahrt, and Aaron O'Leary. PyWavelets: A Python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237, 2019. DOI: 10.21105/joss.01237.

[12] ML Fundamentals. A guide for using the wavelet transform in machine learning. Online; accessed December 27, 2023. `https://ataspinar.com/wp-content/uploads/2018/12/9layer_image_CNN.png`.

[13] Shin Fujieda, Kohei Takayama, and Toshiya Hachisuka. Wavelet convolutional neural networks. *arXiv preprint arXiv:1805.08620*, 2018.

[14] Yang Chen, Lei Pang, Hui Liu, and Xigui Xu. Wavelet fusion for concealed object detection using passive millimeter wave sequence images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:193–198, 2018.

[15] N.S. Artamonov and P.Y. Yakimov. Towards real-time traffic sign recognition via YOLO on a mobile GPU. *Journal of Physics: Conference Series*, 1096(1):012086, 2018.

[16] Melvin Delgado. Urban gun violence: Self-help organizations as healing sites, catalysts for change, and collaborative partners. Oxford University Press, USA, 2021.

[17] AI. Knife-detection dataset. *Roboflow Universe*, 2023. `https://universe.roboflow.com/ai-0jtbr/knife-detection-hgvy2`.

[18] Project. Weapons dataset. *Roboflow Universe*, 2023. `https://universe.roboflow.com/project-sc28w/weapons-rvq2k`.