FAST National University of Computer and Emerging Sciences

Weapon Detection Using Quantized YOLO v5 and YOLO v8

Group Members:

1. Aahil Halani (20K-0308)
2. Faizan Hakim (20K-0290)

Course Instructor: Dr. Sumaiyah Zahid

Date: 4^{th} May, 2024

Table of Contents

Abstract

Video surveillance is essential for creating a secure and hassle-free environment in all areas of life. It helps identify theft, detect unusual events in crowded locations, and monitor the suspicious behavior of individuals. However, monitoring sun'eillanee eameras manuaDy is quite chaDenging, and thus, fully automated surveillance with smart vide.capturing capabilities is gaining popularity. This approach is a deep learning methodology to remotely monitor unusual actions with accurate information about the location, time of occurrence, and identification of criminals. Detecting criminal conduct in public settings is difficult due to the complexity of real-world scenarios. CCTV cameras can record suspicious incidents in public areas, such as carrying weapons, which helps authorities to take preventive measures to protect citizens. The proposed system employs the state-of-the-art YOLOv8 model for real-time weapon detection, which is faster, more accurate, and better than YOLOvS. To ensure fast performance, the weights of YOL0v8 were quantized. In our experiments, we evaluated the performance of the YOLOv8 and YOL0v5 models for weapon detection. The mean Average Precision (mAP) value achieved using YOLOv8 was 90.1 %, which outperformed the mAP value of 89.1% obtained with YOWvS. Furthermore, by applying weight quantization to the YOL0v8 model, we reduced the inference time by 15% compared to the original YOLOv8 configuration.

Introduction

In recent years, the rise in tragic incidents involving weapons in public spaces has spurred significant concern and urgency for effective preventive measures. Schools, malls, airports, and other public areas have unfortunately witnessed the devastating consequences of such occurrences. In response, there has been a concerted effort to develop and implement advanced weapon detection systems to mitigate these risks.

Traditional approaches to weapon detection often rely on costly and limited sensors, which can be inefficient and fail to provide comprehensive coverage. To overcome these shortcomings, the integration of deep learning techniques, particularly object detection algorithms, has emerged as a promising solution. These algorithms, adept at analyzing real-time videos and images, offer a more effective means of detecting weapons in various environments.

In this research endeavor, we turn our focus to one such advanced algorithm: YOLOv8 (You Only Look Once version 8). Building upon the success of its predecessors, particularly the popular YOLOv5, YOLOv8 represents a significant advancement in weapon detection capabilities. By harnessing the power of deep learning and incorporating innovative features such as anchor-free detection, YOLOv8 demonstrates improved accuracy and efficiency in identifying weapons across diverse scenarios.

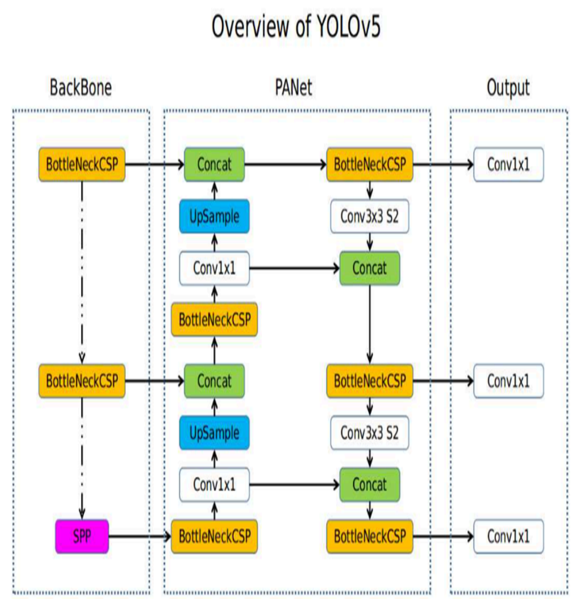
This project aims to evaluate the effectiveness of YOLOv8 in detecting weapons in challenging settings, including crowded areas, low-light conditions, and scenarios involving different types of weapons.
By conducting comprehensive assessments and analyses, we seek to contribute to the ongoing efforts to enhance public safety through cutting-edge technology and innovative methodologies.

Background

The evolution of object detection algorithms has been instrumental in advancing the field of computer vision, particularly in applications such as weapon detection in public spaces. Two prominent architectures that have significantly contributed to this progress are YOLOv8 and YOLOv5, both iterations of the You Only Look Once (YOLO) model. Understanding the architectural nuances of these models is crucial for comprehending their effectiveness in real-world scenarios.

YOLOv5 Architecture:

The YOLOv5 architecture represents a culmination of advancements in object detection, leveraging the core principles of the YOLO model while incorporating novel optimizations. Comprising three primary components – the backbone, neck, and output – YOLOv5 demonstrates a streamlined approach to real-time object detection.



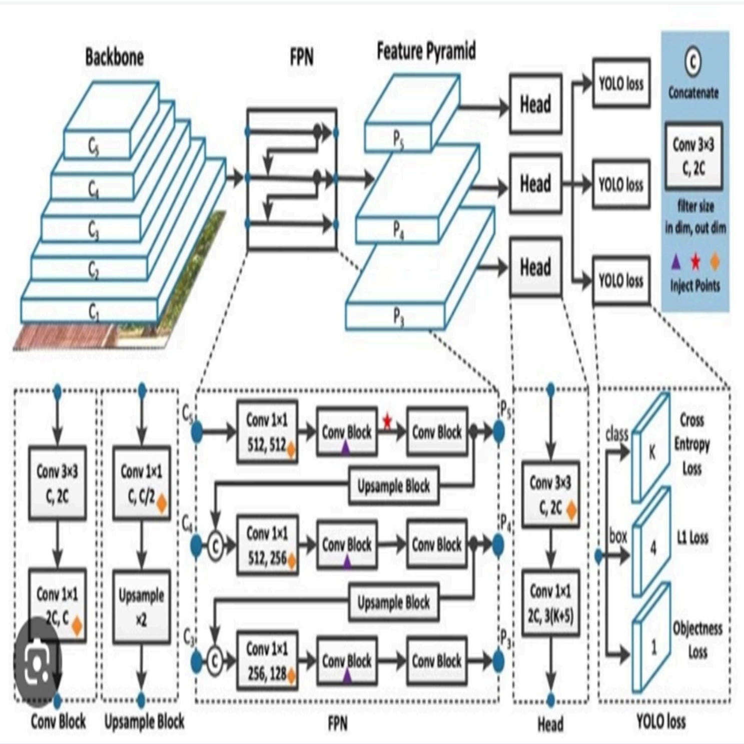
At its core, YOLOv5 employs a convolutional neural network (CNN) backbone, which serves as the foundation for feature extraction. This backbone is responsible for processing input images and extracting high-level features that are essential for accurate detection. Notably, YOLOv5 introduces several algorithmic optimizations within the backbone, enhancing its efficiency and performance.

The neck of the YOLOv5 architecture facilitates feature fusion and refinement, enabling the model to capture contextual information and improve the accuracy of object detection. Through techniques such as adaptive anchor frame calculation and mosaic data augmentation, YOLOv5 adapts dynamically to varying datasets, thereby enhancing its robustness and versatility.

Finally, the output component of YOLOv5 generates bounding boxes and class predictions for detected objects. By employing advanced algorithms for post-processing and inference, YOLOv5 achieves real-time performance while maintaining high detection accuracy.

YOLOv8 Architecture:

Building upon the foundation laid by YOLOv5, YOLOv8 introduces several enhancements aimed at further improving detection accuracy and efficiency. One of the key changes in YOLOv8 is the introduction of the C2f block, which replaces the C3 block used in previous iterations. This architectural modification enhances feature extraction capabilities, enabling the model to capture more discriminative features from input images.



Additionally, YOLOv8 introduces anchor-free detection, a significant departure from traditional anchor-based approaches. With anchor-free detection, the model predicts the center of objects directly, eliminating the need for manually specifying anchor boxes. This approach offers greater flexibility and efficiency, leading to improved performance, especially in scenarios with varying object scales and aspect ratios.

Moreover, YOLOv8 incorporates user-friendly features such as a command-line interface (CLI) and a dedicated GitHub repository, streamlining the deployment and usage of the model in practical applications. By combining these advancements with its superior accuracy and performance, YOLOv8 represents a state-of-the-art solution for weapon detection in public areas.

In the subsequent sections of this report, we delve deeper into the architectural intricacies of YOLOv8 and YOLOv5, analyzing their respective components and functionalities. Through comprehensive evaluations and comparisons, we aim to elucidate the strengths and weaknesses of each model, thereby informing future developments in the field of object detection and public safety.

Experiment Setup

Dataset

The Roboflow platform simplifies computer vision development by facilitating the deployment of custom datasets and employing model training techniques. It offers access to both public datasets and allows users to contribute their own. In a recent study, 2986 weapon photos from sources like Google and YouTube were collected via Instagram and annotated using Roboflow. The platform seamlessly integrates datasets into YOLO models, with both labels and photos included. A training set comprising 80% of the data and a validation set comprising 20% were created from the dataset for model training and validation purposes, respectively.

Yolo V5

YOLO (You Only Look Once) is a popular object detection algorithm that divides input images into a grid of cells and predicts bounding boxes and class probabilities for objects within each cell. It resizes the input image, divides it into a grid, uses predefined anchor boxes to detect objects of varying sizes, predicts class probabilities and bounding box coordinates for each object, applies non-maximum suppression to remove duplicate detections, and outputs a list of class probabilities and bounding boxes for identified objects.

Yolo V8

While the primary objective of this study is to evaluate the performance of YOLOv8 and YOLOv5, their operational principles remain largely similar. The key advancements in YOLOv8, such as the adoption of anchor-free detection, a refined loss function, and enhancements to internal convolutional blocks, distinguish it from previous versions and contribute to its improved mean Average Precision (mAP) values. However, apart from these changes, the overall methodology and steps involved in both versions of YOLO remain consistent with previous iterations.

Quantization

Quantization reduces memory requirements in neural networks by converting parameters from 32-bit floats to smaller representations like 8-bit integers, shrinking model size by a factor of 4 and improving power efficiency by reducing processing and memory access. It lowers network latency by utilizing integer data types for faster operations compared to floating-point. Although it may impact network accuracy by altering information representation, research suggests the loss is often negligible compared to gains in latency reduction, memory savings, and energy efficiency. The extent of accuracy impact depends on precision loss, network architecture, and training techniques.

Metrics:

To determine performance of the models, we have used the Mean Average Precision (mAP) which can be determined by the following equation:

$$mAP = \frac{1}{|O|} \sum_{c \in O} AP(c) \quad (1) \qquad AP(c) = \frac{TP(c)}{TP(c) + FP(c)} \quad (2)$$

where O is set if all objects in the data set. The value of $AP(c) = 1$ would represent perfect detection while the value 0 will represent no detection being made.

Results

Table 1 demonstrates an accuracy improvement from 89.1% with YOLOv5 to 90.1% with YOLOv8, showcasing the algorithm's evolving precision.

Version	Precision	Recall	mAP50	mAP50-95	Inference Time
Yolo V5	92.4%	84.2%	89.1%	67.4%	8.7ms per image
Yolo V8	92.6%	81.7%	90.1%	72.7%	9.0ms per image

Weight quantization in YOLOv8 converts 32-bit float values to 8-bit integers, simplifying the architecture. While there's a slight accuracy decrease, the inference time notably reduces by 15%, from 9ms to 7.6ms. Comparing the quantized YOLOv8 to the unquantized YOLOv5, the inference time decreases by about 12.5%, with minimal accuracy impact.

Version	Precision	Recall	mAP50	mAP50-95	Inference Time
Yolo V5 (without quantization)	92.6%	81.7%	90.1%	72.7%	9.0ms per image
Yolo V8 (without quantization)	91.6%	81.4%	88.1%	70.2%	7.6ms per image

Conclusion

To conclude, the importance of efficiently detecting weapons for public safety cannot be overstated. YOLO has emerged as a promising solution, with iterations like YOLOv5 and YOLOv8 showing significant potential but facing challenges in accuracy and efficiency. To address these issues, weight quantization was employed, resulting in a streamlined architecture capable of achieving high accuracy in real-time weapon detection. With an impressive inference time of just 7.6ms and a mean Average Precision (mAP) value of 88.1%, this approach presents a viable solution for enhancing public security. Moving forward, continued refinement and optimization of such algorithms will be crucial for ensuring the safety and well-being of communities worldwide.