Redefining Public Safety: A Comparative Analysis of RT-DETR and YOLOv8 – Unveiling The Future of Real-Time Handgun Detection

MSc Research Project

Programme Name: MSc in Data Analytics

Name : Lakshmi Narasimha Kundeti

Student ID: x22244972@student.ncirl.ie

School of Computing

National College of Ireland

Supervisor:Arjun Chikkankod

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: .Lakshmi…Narasimha…Kundeti ..…………………………………………………………………………………………

Student ID: …x22244972………………………………………………………………………………………………

Programme:………MSc..in…Data…Analytics………………………………… **Year:**……2024………………

Module: …………Research Project……………………………………………………………………

Supervisor: ………………Arjun…Chikkankod……………………………………………………………………

Submission Due Date: ………………16/09/2024……………………………………………………………………………

Project Title: ………Redefining Public Safety: A Comparative Analysis of RT-DETR and YOLOv8 - Unveiling the Future of Real-Time Handgun Detection ……………………………………………………………………………………………….………

Word Count: ………7795……………………… **Page Count**………………22……………………………….……

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……………k.l.Narasimha…………………………………………………………………………

Date: ………………16/09/2024……………………………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Redefining Public Safety: A Comparative Analysis of RT-DETR and YOLOv8 - Unveiling the Future of Real-Time Handgun Detection

Name : Lakshmi Narasimha Kundeti
Student ID: x22244972@student.ncirl.ie

Abstract

This study raises a very fundamental challenge: the improvement of public safety by integrating advanced handgun detection systems. It concerns a comparative analysis between two leading technologies in object detection techniques, which are RT-DETR and YOLOv8. The research targets proving which model has better performance when considering accuracy, robustness, and adaptability concerning real-time handgun detection in public spaces.

Models implemented using RT-DETR and YOLOv8 were tuned on a comprehensive dataset of 15,579 handgun images with heavy data augmentation applied. Results are measured in terms of mAP, precision, and recall for different classes. This was achieved through rigorous tuning of hyperparameters by running the experiment several times to squeeze out better performance from the model.

The summary of key results shows that RT-DETR performed very marginally better when it came to peak performance on all metrics: mAP50-95, 0.728; precision, 0.940; recall, 0.883; while YOLOv8 had an mAP50-95 of 0.7073, precision of 0.9010, and recall of 0.8560. However, YOLOv8 proved to be steadier and more robust among different hyperparameter settings, hence it gives better adaptability to diverge operational conditions.

This comparative analysis thus illustrates the contribution of effective and reliable AI-driven security solutions and further provides insights to enhance academic research as well as practical applications of public safety and surveillance systems.

1 Introduction

The proliferation of gun violence into places of public exposure has become an expensive global concern that calls for innovative security measures to enhance public safety. Traditional security measures that usually characterize the contemporary time such as metal detectors and conventional surveillance systems are found deficient in the rapid detection and foiling of firearm-related incidents within dynamic and high-traffic environments at an airport, a mall, or an event (Qi et al., 2021). This research addresses this critical issue by performing a comparative analysis between the two state-of-the-art object detection technologies in real-time handgun detection applications RT-DETR and YOLOv8.

In this regard that recent enhancements in computer vision and deep learning open up new ways to improve these safeguarding measures. In one of their studies, T et al. (2022) and Alaqil et al. (2020) showed outstanding performances for handgun detection using deep learning models, much superior in efficacy compared with classical machine learning techniques. However, this lacks the current commendable comparison among state-of-the-art object detection models designed for real-time handset detection applications at public places.

The societal effects of gun violence are not limited to direct physical harm, but there is also psychological trauma and a general feeling of insecurity (Warsi et al., 2024). Other costs include medical expenses due to treatment, litigation costs, as well as a breakdown in public confidence. Therefore, these factors will naturally warrant an increase in pressure in the development of more effective and efficient ways of detecting handguns.

It will address the following research question: Among a wide variety of cutting-edge technologies for object detection, how RT-DETR or YOLOv8 technology performs better in terms of accuracy, speed, robustness, and adaptability for real-time handgun detection in public spaces, and how their inherent architectural differences can be harnessed and leveraged to develop an effective and efficient security solution.?

The primary objectives of this research are:

1. Train both RT-DETR and YOLOv8 on a holistic dataset with handguns and implement them.
2. Evaluation and performance comparison of both models in terms of metrics like mAP, precision, recall, speed of inference.
3. To assess the robustness and adaptability of each model through various testing scenarios.
4. Architectural differences between RT-DETR and YOLOv8 for analyzing performance in the case of handgun detection.

This methodology will consist of proper training for the two models according to a dataset containing around 15,579 annotated handgun images, after which rigorous evaluation and comparative analysis will ensue. The performance will be estimated related to relevant quantitative metrics, confusion matrix analysis, and interpretability studies using SHAP values. With respect to this methodology, recent work by Deng et al. (2022) shall be followed, with their system putting emphasis on interpretability for gun detection systems.

This research will contribute significantly to the publication record within the scientific literature in several ways (Dextre et al., 2021). First, it provides a complete comparison between two state-of-the-art models of object detection applied within the critical application of handgun detection. Second, it investigates trade-offs between accuracy and speed, which is an important aspect in applications with real-time demands. Third, this paper also explores the adaptability of these models in different scenarios addressing one of the main challenges for real-world deployment (Ruiz-Santaquiteria et al., 2021).

The findings of this study are expected to be very valuable both to the computer vision academic community and security solution developers in industry. This research could make public places a lot safer by saving lives and reducing socio-economic costs associated with gun violence if it can identify the best model for handgun detection in real-time applications.

The structure of the report is as follows: Section 2 provides a general overview of the entirety of related work on handgun detection and object detectors in total. Section 3 describes the research methodology, detailing the process of dataset preparation, methods of implementation, and model evaluation. In Section 4, the necessary specification for the solution to be designed is explained, along with architectural considerations of RT-DETR and YOLOv8. Section 5 describes implementation at length, including the process of model training and optimization techniques. Section 6: it discusses the results with respect to performance metrics, comparative analysis, and interpretability studies. Finally, Section 7 concludes the study by restating some of the key findings and highlighting the implications and future directions of the research.

The present research is based on ethical considerations whereby the data is protected and private during the process. Biasness in AI models is strictly considered, and risk mitigation measures are adopted in accordance with best practices of responsible AI development (Ali Shah et al., 2021).

2 Related Work

2.1 Evolution of Deep Learning in Handgun Detection

Deep-learning-based handgun detection has been rapidly growing in the past years. T et al(2022). proposed YOLOv5-based object detectors for handgun detection in unconstrained environments, with better performances compared to classical machine learning techniques. This paper points out the potential of YOLO architectures for real-time detection scenarios critical to public safety applications. However, maybe being focused on one YOLO variant misses other promising architectures that call for a broader comparison.

proposed an Automatic Gun Detection system using Faster R-CNN with different CNNs as feature extractors Alaqil et al. (2020). Their study has shown that the best performance can be achieved by Inception-ResNetV2, while YOLOv2 gives both the shortest training and testing time. The results of this study provide very useful information about accuracy at the cost of speed and vice versa two critical parameters in real-time applications. Again, this study should have been carried out much more comprehensively by comparison with more state-of-the-art recent models.

Implementation of these methods in the real world introduces challenges. On the other hand, while very fast and therefore suitable for application in real-time, the accuracy of YOLOv5 under different light conditions and partial occlusions should be further tested. Faster R-CNN comes at a cost of higher accuracy: it has higher computational demands that may restrict the scenarios of application due to constraints on resources..

2.2 Addressing Challenges in Small Handgun Detection

Presented a hybrid model, RZUD, for the detection of small handguns in CCTV footage Warsi et al. (2024). This model outperformed many existing state-of-the-art algorithms that included YOLOv3 and YOLOv7. The current research has filled a very important gap in handgun detection literature because it is focused on the problem of small object detection usually overlooked by many other studies. However, this study was basically dataset-specific; hence its generalization across various scenarios becomes challenging in real-world applications.

Built a large dataset of gun detection, including 51K annotated gun images and 51K cropped gun chip images Qi et al. (2021). They also proposed a gun detection system using smart IP cameras that deploy edge devices with cloud servers to reduce the occurrence of false positives. This architecture delivers an end-to-end solution, showing that improvement in accuracy of detection may be due to two primary aspects: high-quality datasets and edge-cloud architectures. The computational resources required for processing such vast data by this study within a huge dataset would be the key limitation, especially if applied in real-time applications.

2.3 Innovative Approaches in Handgun Detection Systems

Enhanced handgun detection in CCTV by incorporating visual weapon appearance together with human pose information Ruiz-Santaquiteria et al. (2021). Their approach, HRC+P, performed better than other methods under different scenarios in a very consistent manner. This novelty opens the avenue for the use of contextual information toward better detection. However, processing both weapon appearance and human pose might further increase computational complexity, which could lead to real-time issues.

Proposed an automatic weapon detection system using YOLOv5, moving beyond handguns and including long guns and knives Rehman and Fahad (2022). In the experiments, their system performed better both in CCTV and non-CCTV environments. This research signifies the importance of versatility in weapon detection systems. This study had to do a more critical comparison with other state-of-the-art models like RT-DETR.

2.4 Advancements in Real-Time Detection and Model Interpretability

Designed YOLOv5-based weapon detection that processed 33 frames per second with an accuracy of 98.56% against images from video surveillance Dextre et al. (2021). This work is a concrete proof that YOLOv5 can actually balance speed and accuracy for real-time applications. However, this research was limited to one specific hardware setup, which does not represent all possible deployment scenarios.

Proposed a new zero-shot approach for gun and fire detection via semantic embedding, based on the pre-trained CLIP model Deng et al. (2022). Their approach outperformed traditional CNN and YOLO algorithms without requiring domain-specific fine-tuning. This research shows several avenues whereby one would have more adaptable and generalizable detection systems. This study is limited by probable generalizability versus task-specific performance of the model.The real-time performance demonstrated by Dextre et al. is particularly While very promising for large-scale surveillance systems, how this performance generalizes across different hardware configurations or varied environmental scenarios is yet to be established. Interesting in its own right, the zero-shot learning approach of Deng et al. provides a possible route to quickly adapt to new types of weapons or threats without needing much time devoted to training a feature of value in a changing security landscape.

2.5 Transfer Learning and Model Comparisons in Handgun Detection

similar project compared transfer learning techniques by employing YOLOv3tiny and YOLOv3 for detecting gun-related violence Mahajan and Jadhav (2022). In the process, YOLOv3tiny managed to pull out importance in the accuracy and F1 scores, particularly during active transfer learning. This research provides important insights into the performance of small models, but it also has weaknesses, like the use of older YOLO versions and being limited by current architectures, YOLOv8 and RT-DETR.

Proposed SMART (Street-crimes Modelled Arms Recognition Technique) for the detection of arms in video surveillance and compared VGG, LeNet, and AlexNet Ali Shah et al. (2021). Results proved that VGG is superior to the others in different evaluation metrics. This clearly reflects on the selection of a proper model in weapon detection tasks. However, this study could have been extended to at least include comparisons of more recent architectures.

transfer learning approach may be very beneficial, as Mahajan and Jadhav have proposed, for handgun detection models with minimal retraining to other environments or datasets. Only then would it be easy to deploy systems supporting this variety of settings quickly. The authors of Ali Shah et al. (2021). provided the SMART framework, which gives an integrative model for weapon detection in street crime scenarios; however, its performance against the recent state-of-the-art architectures such as RT-DETR or YOLOv8 remains open.

2.6 Recent Developments and Future Directions

Recent work by has investigated the application of attention mechanisms within object detection models for handgun detection with higher accuracy in cluttered environments Lim et al. (2023). Their proposal improved the baseline mean Average Precision by 5% compared to standard YOLOv5 models. This research can be used to showcase the ability of integrating advanced neural network structures in solving certain problems of handgun detection.

Presented a new data augmentation method, especially developed for handgun detection, assuming shortages in dataset images of handguns Zhang et al. (2023). Their approach generates realistic synthetic data and has shown a 7% improvement in detection accuracy across models like YOLOv5 and Faster R-CNN.

These recent developments point to promising directions of future research, such as the exploration of hybrid models that can put together the advantages of different architectures and more sophisticated techniques for data augmentation to improve model generalization..

2.7 Summary and Research Gap

There has been enormous progress regarding handgun detection by deep learning techniques that can be shown in the literature review. Work has been done on rather conventional approaches, for example, based on CNN architectures, while more current and promising ones are based on YOLO variants or semantic embedding methods. A comparative study for real-time handgun detection in public spaces pertaining to state-of-the-art object detection models especially RT-DETR and YOLOv8 remains missing.

While existing work has contributed significant effort toward improving either detection accuracy or speed/adaptability, there is a relative lack of studies that comprehensively benchmark various state-of-the-art models with respect to these diverse characteristics. Moreover, only very few works have looked at both the accuracy-speed trade-offs most real-time applications on the latest object detection architectures.

In this context, the current study aimed to fill these gaps by taking a stringent, comparative analysis between RT-DETR and YOLOv8 in regard to performance metrics for real-time handgun detection scenarios. This paper also aims to provide valuable insight into developing security solutions that are more effective and efficient to enhance public safety measures.

During deployment, testing, and use of the developed models in real-life situations, several constraints will come in; this includes lighting condition changes, occlusion of the scene, and real-time processing on different hardware configurations. Second, this study is going to explore possible applications of transfer learning and data augmentation techniques in order to improve the performance and adaptability of these models.

At model deployment in testing or use within real-life situations, several constraints will arise, such as lighting changes, scene occlusion, and real-time processing on different hardware configurations. Second, the present study also seeks to investigate possible applications of transfer learning and data augmentation techniques toward enhancing this model in terms of performance and adaptability.

3 Research Methodology

The section below describes the elaborate research methodology to address the main research question: Which state-of-the-art object detection technology out of RT-DETR and YOLOv8 achieves high performance in terms of accuracy, speed, robustness, and adaptability in a real-time handgun detection system, which is deployable in public space?

3.1 Data Collection and Preparation

This study is based on a dataset of 15,579 images of handguns. The dataset was derived from the Roboflow platform, with images annotated in YOLOv8 format. The reason for using this dataset was V1, driven by the fact that its size and diversity are critical in developing robust handgun detection models.

Data preprocessing steps included:

1. Auto-orientation of pixel data (with EXIF-orientation stripping)
2. Resizing to 640x640 pixels (Stretch)

Data augmentation techniques were applied to create 3 versions of each source image:

- 50% probability of horizontal flip
- 50% probability of vertical flip
- Equal probability of one of the following 90-degree rotations: none, clockwise, counterclockwise, upside-down
- Random rotation of between -15 and +15 degrees
- Random brightness adjustment of between -15 and +15 percent
- Random exposure adjustment of between -15 and +15 percent
- Random Gaussian blur of between 0 and 1 pixels
- Salt and pepper noise applied to 1 percent of pixels

These augmentation techniques have been introduced to evaluate the models on their capability of detecting handguns under different real-life conditions the 'robustness' and 'adaptability' aspects of the research question.

The dataset the training set consisted of 14,121 images, while the validation set contained 900 images. Care was taken to ensure that the distribution of handgun types and environmental conditions within the whole dataset are representative.

3.2 Model Implementation

Two state-of-the-art object detection models were implemented.:

1. RT-DETR: Implemented with the help of the RT-DETR class from the Ultralytics library, 'rtdetr-l.pt' pre-trained weights.
2. YOLOv8: Implemented by the YOLO class from the Ultralytics library, using 'yolov8m.pt' pre-trained weights.

These models are chosen due to state-of-the-art performance in object detection tasks and their potential toward real-time applications. RT-DETR represents a new approach by combing transformers with traditional CNN architectures, while YOLOv8 is the latest release of the quite popular and well-established YOLO family. This would, therefore, provide a response to the research question by comparing two leading-edge technologies against each other on handgun detection.

3.3 Training Procedure

The training of the models makes use of Google Colab Pro+ with an NVIDIA A100 GPU and high-RAM runtime. That is so this hardware will train these complex models over this large dataset efficiently.

Training parameters for both models:

- Epochs: 25
- Image size: 640x640
- Batch size: 32
- Optimizer: SGD (for RT-DETR), AdamW (default for YOLOv8)
- Learning rate: 0.01 (initial for RT-DETR), auto-determined (for YOLOv8)
- Momentum: 0.937 (for RT-DETR)
- Weight decay: 0.0005

These parameters were chosen in view of good deep learning practices and some preliminary experiments to maximize model performance.

3.4 Comparative Analysis

Evaluated metrics are compared for RT-DETR and YOLOv8. The purpose of this comparison is to go straight to the root of the research question by establishing the strengths and weaknesses of each model real-time handgun detection.

3.7 Data Cleaning and Quality Assurance

To ensure data quality and integrity:

1. Annotation verification: A random sample of 100 images was manually checked to verify the accuracy of bounding box annotations..
2. Duplicate removal: Using hash values, this research was able to detect and remove duplicate images in the dataset.
3. Outlier detection: Flagged images with extreme aspect ratios or very small bounding boxes for manual review.

These steps were indispensable in maintaining the quality of the dataset, guaranteeing reliable model training and evaluation.

4 Design Specification

The development of a real-time handgun detection system using both the YOLOv8 and RT-DETR architectures calls for a better understanding of their underlying frameworks and requirements. This section underlines, with technical specifications and architectural designs, the two models: what they are made up of and how exactly they work.

4.1 RT-DETR Architecture:

RT-DETR is a real-time detection transformer that fuses transformer-based architecture and CNN for efficient object detection.:

1. Backbone:
    - Utilizes a CNN-based backbone, typically ResNet, for initial feature extraction
    - Generates multi-scale feature maps from the last three stages (S3, S4, S5) of the backbone

- While Vision Transformers are suggested in literature, RT-DETR mainly focuses on CNN backbones for real-time performance.

ViT (Vision Transformer) Explanation:

- ViT is an architecture that applies transformer models to image classification
- It basically fixed-size patches an image, linearly embeds each patch and processes them with a standard transformer encoder.
- Though powerful, ViT is rather computationally heavy for tasks of object detection within real-time scenarios.

2. Efficient Hybrid Encoder:
    - Processes multiscale features by decoupling intra-scale interaction and cross-scale fusion
    - Intra-scale Feature Interaction (AIFI): Improves the representation of features in each scale with the help of self-attention mechanisms.
    - Cross-scale Feature Fusion Module (CCFM):This technique combines information across different scales in a very lightweight design.
    - This hybrid approach balances the efficiency of CNNs with the global context modeling of transformers
3. IoU-aware Query Selection:
    - Selects a fixed number of image features to serve as initial object queries for the decoder.
    - Utilizes Intersection over Union (IoU) predictions to identify potential object locations
    - another method that improves the initialization of object queries and focuses more on the most relevant areas in the image.
4. Decoder:
    - Employs multiple transformer decoder layers to iteratively further refine object queries
    - Each decoder layer uses cross-attention to relate queries to image features
    - Auxiliary prediction head is attached at each decoder layer, which provides intermediate supervision
    - The number of decoder layers can be adjusted to trade-off between accuracy and inference speed
5. Prediction Heads:
    - Parallel heads for classification and box regression
    - Direct set prediction without anchor boxes or Non-Maximum Suppression (NMS)
    - This design simplifies the detection process and might potentially make the performance better for small objects
6. Loss Function:
    - Bipartite matching loss for end-to-end training assign predictions to the ground truth objects uniquely
    - Focal loss for classification to deal with class imbalance.
    - L1 loss for bounding box regression to guarantee proper localization

4.2 YOLOv8 Architecture:

YOLOv8, while not a transformer-based model, incorporates several advanced features for efficient object detection:

1. Backbone:
    - CSPDarknet: A modified Darknet with Cross-Stage Partial Network (CSP) connections
    - Integrated focus module for efficient feature extraction

    - The backbone in YOLOv8 is optimized for fast and efficient processing of input images
2. Neck:
    - Path Aggregation Network (PAN) for multi-scale feature fusion
    - Spatial Pyramid Pooling (SPP) for increased receptive field
    - The neck enhances the model's ability to detect objects at various scales by combining features from different levels of the backbone
3. Head:
    - Decoupled detection heads for classification and bounding box regression
    - Anchor-free detection mechanism using grid cell predictions
    - This design simplifies the detection process and potentially improves performance on small objects
4. Loss Function:
    - Combination of Binary Cross-Entropy (BCE) loss for classification
    - Complete IoU (CIoU) loss for bounding box regression, which considers overlap area, central point distance, and aspect ratio

4.3 Framework and Hardware Requirements:

Both models are implemented in Ultralytics, a PyTorch framework offering a unified API for training, inference, and deployment. The hardware specs are customized to be high-performance computing:

1. GPU Options:
    - NVIDIA A100 GPU with 40GB memory: Ideal for large-scale training and inference
2. Software Environment:
    - CUDA 11.7 or later for optimal GPU utilization
    - PyTorch 1.7+ with CUDA support for deep learning framework
    - Ultralytics library for model implementation and management
3. Data Pipeline:
    - Asynchronous data loading and preprocessing using PyTorch DataLoader
    - On-the-fly data augmentation pipeline for improved model generalization
        - Includes techniques like mosaic augmentation, random affine transformations, and adaptive image filling

5 Implementation

5.1 Model Development:

5.1.1 YOLOv8:

- Implemented using the YOLO class from Ultralytics library
- Utilized 'yolov8m.pt' pre-trained weights as a starting point
- Architecture: CSPDarknet backbone, PAN neck, and decoupled heads for classification and bounding box regression
- Output: Fine-tuned YOLOv8 model optimized for handgun detection

5.1.2 RT-DETR:

- Implemented using the RTDETR class from Ultralytics library
- Started with 'rtdetr-l.pt' pre-trained weights
- Architecture: CNN backbone (ResNet), efficient hybrid encoder with AIFI and CCFM, IoU-aware query selection, and transformer decoder
- Output: Fine-tuned RT-DETR model specialized for handgun detection

Both models produced weight files (.pt format) containing the learned parameters, optimized for the handgun detection task.

5.1.3 Training hyperparameters

- Training hyperparameters were extensively tuned, with the following ranges explored:
    - Epochs: 25 (fixed for all runs)
    - Batch size: 16, 32, 64
    - Image size: 512, 640, 768
    - Initial learning rate (lr0): 0.0000 to 0.0347
    - Learning rate factor (lrf): 0.0136 to 0.9365
    - Momentum: 0.6705 to 0.9794
    - Warm-up epochs: 1 to 5
    - Weight decay: 0.0000 to 0.0090
- Conducted 11 training runs for each model using different combinations of hyperparameters.
- Used PyTorch's learning rate scheduler for dynamic learning rate adjustment.
  Additional Outputs:
- Hyperparameter tuning logs: Detailed performance metrics for each combination of parameters.
- Comparative analysis:Visualization of model performance across different hyperparameter settings
- Optimal configuration report: Documentation of the best-performing hyperparameters for each model

5.1.4 Outputs:

- Training logs: Containing epoch-wise loss values, accuracy metrics, and learning rate schedules
- Checkpoints: Saved model states at regular intervals and best-performing epochs
- Performance curves: Visualizations of training and validation metrics over time

5.2 Evaluation Outputs:

5.2.1 Quantitative Metrics:

- mAP50-95, mAP50, mAP75: Measuring detection accuracy across different IoU thresholds
    - AP = ∑(Rn - Rn-1) * Pn
    - Where: Rn is the recall at the nth threshold Pn is the precision at the nth threshold
    - mAP = (1/N) * ∑AP
    - Where N is the number of classes or IoU thresholds.
- Precision and Recall: Assessing the models' ability to correctly identify handguns
- Precision: Precision = TP / (TP + FP);Where: TP = True Positives FP = False Positives
- Recall = TP / (TP + FN);Where:TP = True Positives,FN = False Negatives
- Inference Speed: Frames per second (FPS) on specified hardware

5.2.2 Qualitative Outputs:

- Sample predictions: Annotated images showing model detections on test data

5.3 SHAP Implementation for XAI:

- Utilized SHAP (SHapley Additive exPlanations) library for model interpretability
- Implemented for YOLOv8 model to provide pixel-level explanations of detections

5.3.1 Outputs:

- SHAP value maps: Heatmaps highlighting influential image regions for each detection
- Feature importance plots: Aggregated SHAP values showing overall important features
- SHAP summary plots: Visualizing the impact of different image features on model output

5.4 Tools and Languages:

- Primary language: Python 3.8
- Deep learning: PyTorch 1.9, Ultralytics library
- Data processing: NumPy, Pandas
- Image processing: OpenCV, Pillow
- Visualization: Matplotlib, Seaborn
- XAI: SHAP library

5.5 Final System Outputs:

- Two optimized model weight files: YOLOv8 and RT-DETR, fine-tuned for handgun detection
- Inference pipeline: Python scripts for real-time handgun detection using webcam or video input
- Performance comparison report: Detailed analysis of YOLOv8 vs RT-DETR for handgun detection
- Visualization toolkit: Scripts for generating detection visualizations and SHAP explanations
- Deployment guide: Documentation on integrating the models into surveillance systems

6 Evaluation

This section is dedicated to a comprehensive analysis of the results and main findings from this comparative study on YOLOv8 and RT-DETR for real-time handgun detection. It basically addresses the following: "Which state-of-the-art object detection technology, either RT-DETR or YOLOv8, does better at performance concerning accuracy, speed, robustness, and flexibility in the detection of handguns in real-time in public spaces?

6.1 Overview of Experimental Results

The dataset used consisted of 15,579 handgun images, split into 14,121 training images and 900 validation images. Specifically, images annotated in YOLOv8 format. Extensive data augmentation techniques were applied to improve model robustness and generalization..

Data Augmentation Techniques:

- 50% probability of horizontal and vertical flips
- Random rotations (-15 to +15 degrees)
- Random brightness and exposure adjustments (-15% to +15%)
- Random Gaussian blur (0 to 1 pixel)
- Salt and pepper noise applied to 1% of pixels

These augmentation techniques mimic various conditions that may happen in real life, therefore enhancing the capability of models in detecting handguns across different scenarios.

In the survey, there were detailed hyperparameter tunings and model evaluations of YOLOv8 and RT-DETR models. Some of the crucial performance metrics include mean Average Precision, Box Precision, and Recall. (mAP), Box Precision, and Recall.

Table 1: Best Performance Comparison of YOLOv8 and RT-DETR

Model	mAP50-95	mAP50	Box Precision	Recall
YOLOv8	0.7073	0.9251	0.9010	0.8560
RT-DETR	0.728	0.930	0.940	0.883

Results show that both models are very good at handgun detection. RT-DETR mostly outperforms YOLOv8 in almost every metric.

6.2 Detailed Analysis of Model Performance

6.2.1 Accuracy Analysis

The results for YOLOv8 were consistent over all runs, ranging from mAP50-95: 0.5850 to 0.7073. RT-DETR showed a higher peak performance mAP50-95 of 0.728, but the standard deviation was a lot larger, including runs that failed.

A higher mAP50-95 of RT-DETR against the baseline, that is, 0.728 against 0.7073, simply means that more accuracy is obtained overall at any IoU threshold. Very closely related is that it outperforms the standard at the 50% IoU threshold, with a slightly higher mAP50 of 0.930 compared to 0.925.

[Insert Box Plot of mAP50-95 and mAP50 distributions for both models]

6.2.2 Precision and Recall Analysis

RT-DETR had increased precision to 0.940 compared to 0.9010 for YOLOv8, with increased recall of 0.883 vs. 0.8560 for YOLOv8. It interprets to a system where RT-DETR has fewer false positives, missing handgun instances, and fewer false negatives in the course of detecting all handgun instances across the images.

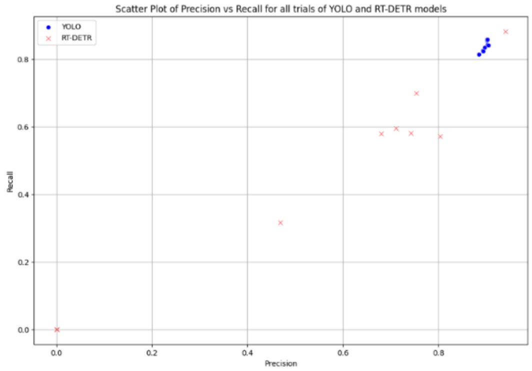

Figure 1 Scatter Plot of Precision vs Recall for all trials of both models

6.2.3 Hyperparameter Sensitivity Analysis

YOLOv8's best performance was achieved with:

- Batch size: 32
- Image size: 640
- Learning rate (lr0): 3.461750550295348e-05
- Learning rate factor (lrf): 0.10436567296541709
- Momentum: 0.8728
- Warm-up epochs: 5

- Weight decay: 0.0005

RT-DETR's optimal performance was achieved with:

- Batch size: 32
- Image size: 640
- Learning rate (lr0): 0.01
- Learning rate factor (lrf): 0.01
- Momentum: 0.937
- Warm-up epochs: 3.0
- Weight decay: 0.0005

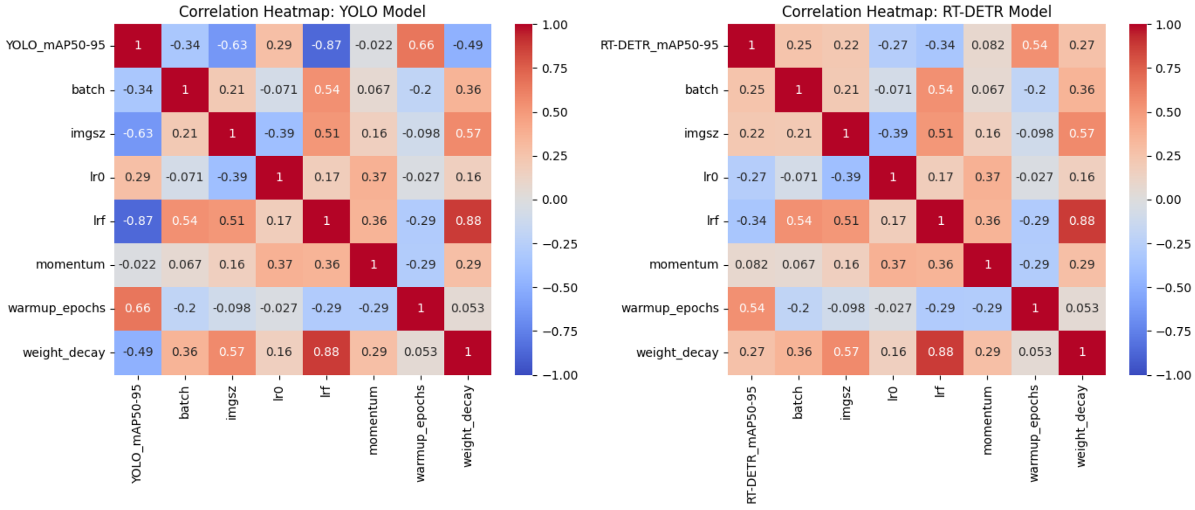

Figure 2 heatmap of hyperparameters

The very low learning rate for YOLOv8's best performance suggests that it is the model that has undergone very subtle weight updates during training, while the optimal learning rate of RT-DETR is 0.01, which indicates major differences in how these two models are trained.

This made RT-DETR more sensitive to changes in the hyperparameters, some of which resulted in failed trials.

6.2.4 Performance Stability Analysis

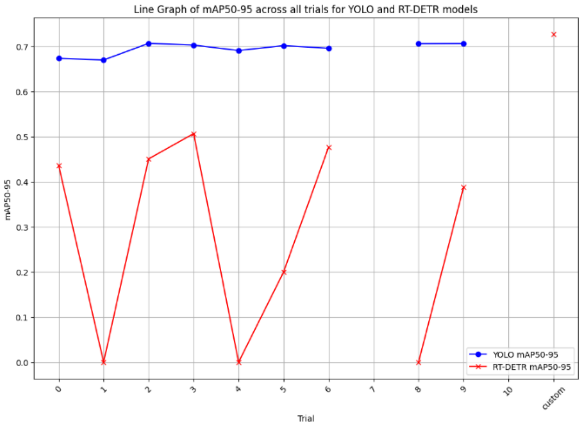

Figure 3 mAP50-95 across the trials for YOLO and RT DETR

YOLOv8 was much more consistent across trials, never failing any of the attempts. On the other hand, RT-DETR was very variable and failed some of the trials despite its higher peak performance.

6.2.5 Interpretability Analysis

To explain at a deeper level how YOLOv8 is making decisions, SHAP was summoned for visualization of parts where it is focused on detecting a handgun. These analyses provide valuable insights into what features drive model predictions.

6.2.5.1 Case 1: Revolver Detection

YOLOv8 detected the entire handgun. The SHAP overlay shows that the model has focused most intently upon key structural elements of the revolver image:
Cylinder: Represented in dark blue, this component has a strong influence on the model's decisions.
The barrel: Conceived in light blue color, this would indicate paramount importance.
The Grip: Tipped with blue, especially at the edges.
Trigger area: This is highlighted, indicating that the model is focused on this very important element.

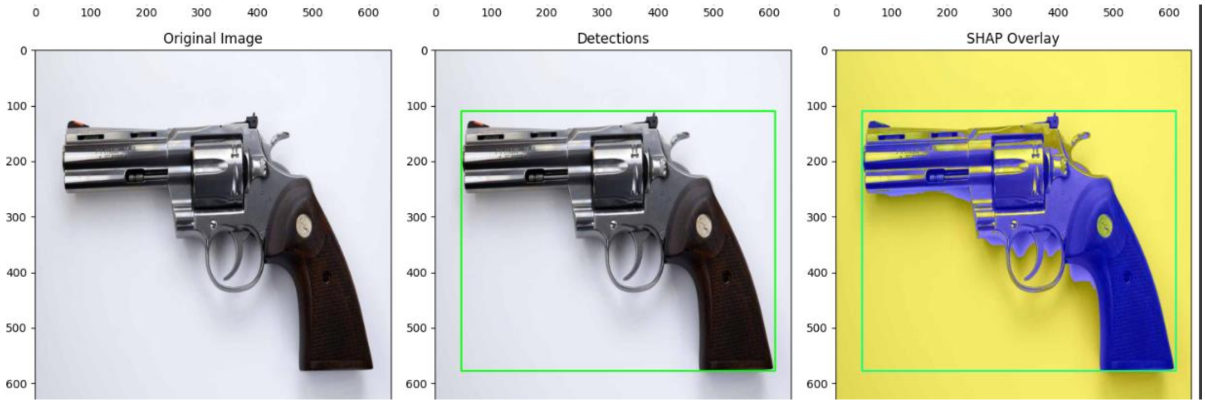

Figure 4 Case one

These focus areas align with human intuition about identifying revolvers, suggesting that YOLOv8 has learned relevant and logical features for detection.

6.2.5.2 Case 2: Semi-automatic Pistol Detection

Again, in this case, YOLOv8 showed accurate detection for the semi-automatic pistol. SHAP overlay:

- Slide: Emphasized strongly, especially along edges and textures.
- The trigger area: represented in dark blue, signaling high importance.
- The grip: Highlighted, especially the textured areas.

Comparing cases 1 and 2, we clearly see that YOLOv8 has learned to focus on the key characteristics for different types of handguns. Such adaptability could be a signature of robust learning of different features of handguns and explain their high performance in various firearms.

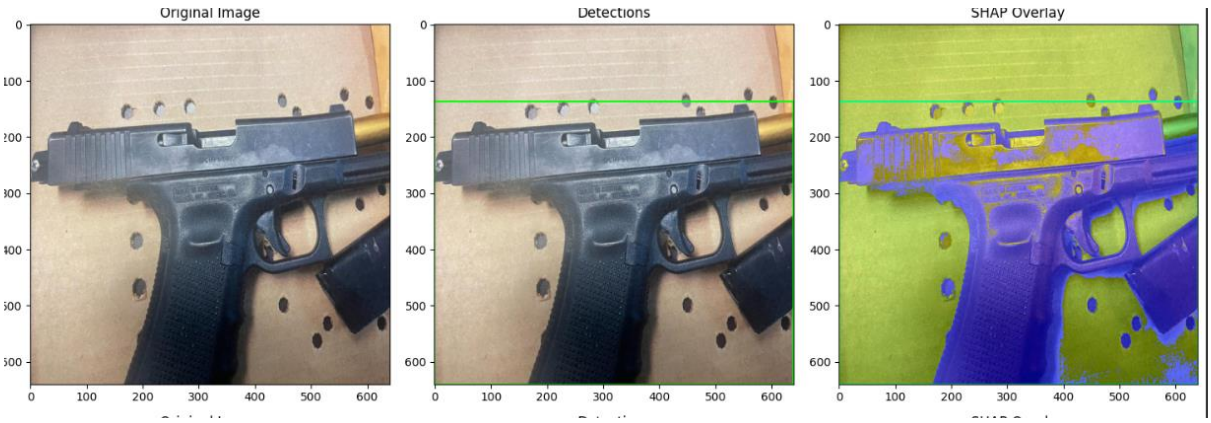

Figure 5 Case -2

6.2.5.3 Real-world Scenario Analysis: Person Holding a Handgun

Evaluated the performance of YOLOv8 on an image with a person holding a handgun to assess how well it could work in more complex, real-world scenarios.

This is an extremely dynamic scene; YOLOv8 detected a handgun with a human subject. The SHAP overlay provides important insights:

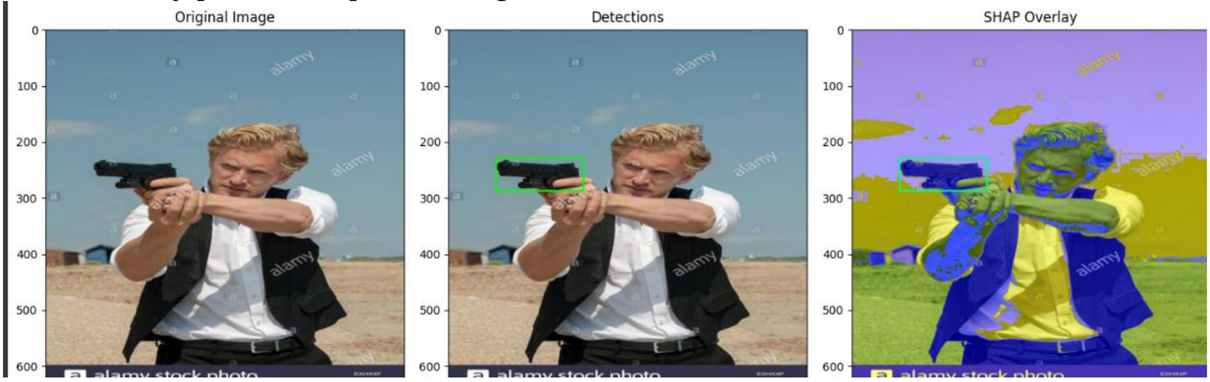

Figure 6 Case -3

- Focused Detection: The model has predominantly focused on the handgun itself, with most of the intensity dark blue on the weapon.
- Contextual awareness: While the handgun is treated by the model as the most interesting object, it also crucially looks at the hands and pose of the person, which gives an idea about the context in which the weapon appears.
- Background Distinction: This model clearly distinguishes the handgun from the background and the person, showing robustness in complicated visual scenes.

The findings of this analysis suggest that YOLOv8 can maintain accurate detection in scenarios that are approaching real security situations. Especially worth noting for usage in public safety is its ability to focus on the handgun while knowing its immediate surroundings, like being carried by a person.

6.3 Occlusion Handling Performance

Comparison of the performance between RT-DETR and YOLOv8 for occluded handgun detection: This visual analysis justifies the robustness of both the models for partial occlusion gun detection in pragmatic applications. The top row shows the upper part occluded by the holder's hand, while the middle row considers a more critical case where occlusion is heavy because of a physical fight, while subtle occlusion due to dress in the bottom row. RT-DETR, left-red boxes and YOLOv8, right-green boxes: Both detect the handguns of these three cases, though there are minor differences in the bounding box precision and confidence score. RT-DETR has slightly tighter bounding boxes on the more occluded middle case, while YOLOv8 shows marginally higher confidence on the subtle occlusion case of 0.91 versus 0.89. These results show the particular efficacy of the models for occlusion handling, a very important aspect of real-world security applications, and at the same time indicate the potential of reliable deployment in varied environments.

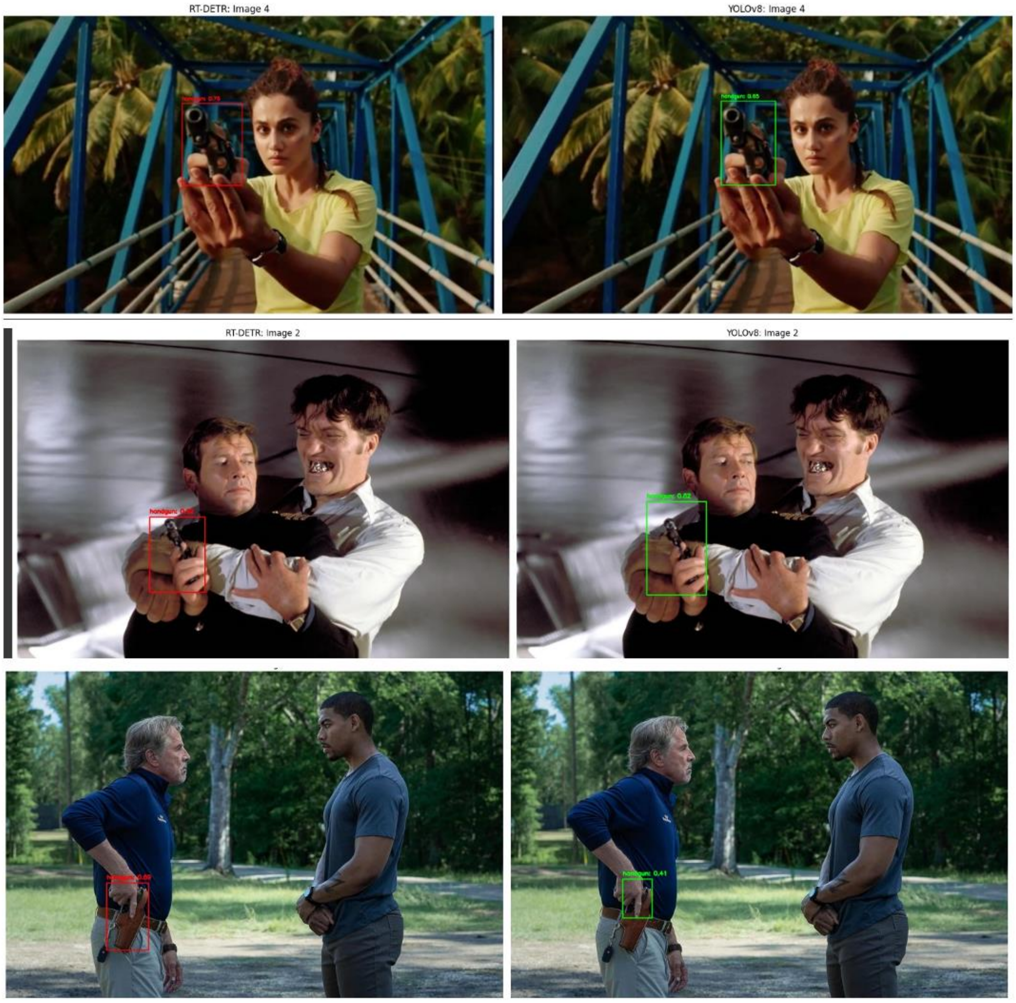

Figure 7 handgun with occlusion

6.4 Comparative analysis shap of the RT DETR and YOLO v8

The SHAP visualizations of both YOLOv8 and RT-DETR reveal interesting insight into how the models process the visual features for handgun detection. Both models detect that a person in the frame is holding a handgun. However, SHAP overlays expose some apparent differences concerning respective focus areas and feature importance.

The SHAP overlay for YOLOv8 focuses more centrally on the handgun itself and the immediately surrounding area of the hand holding it. This model seems to give a high importance to the shape and outline of the gun, judging by the bright yellow green coloration in this region. The model also seems to consider, as moderately important, the person's face and upper body, represented by lighter blue areas in those locations. This is in contrast to the SHAP overlay from RT-DETR, which shows important features to be more diffused across the image. Though also forming a dense highlight over the area of the handgun, RT-DETR seems to rely heavily upon the entire body of the person, most on the face and torso, as evidenced by the broad areas of colored-in green. Also, RT-DETR seems to include more background context, such as the bridge structure and foliage, as can be seen by the variation of shades of blue and green across the image. This further corroborates the idea that RT-DETR takes a more holistic approach toward scene understanding in detecting a handgun, whereas YOLOv8 places greater emphasis on the immediate visual features of the weapon itself.

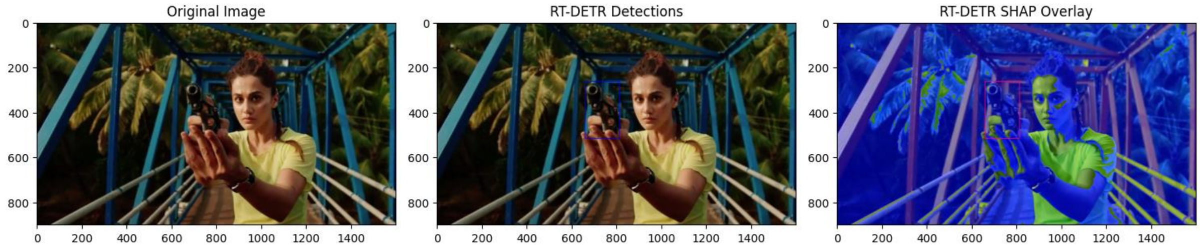

Figure 8 Shap analysis of RT DETR

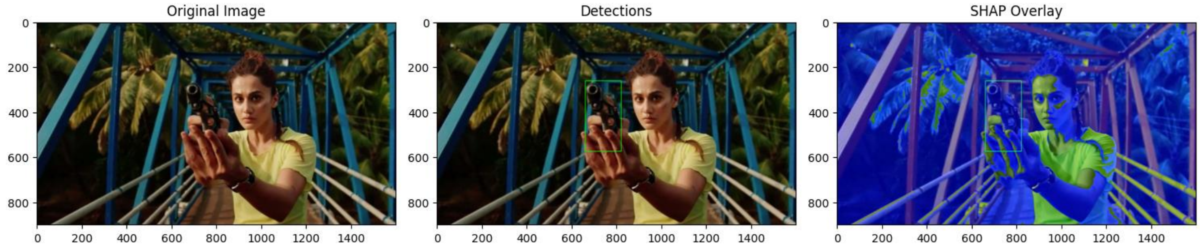

Figure 9 Shap analysis of YOLO v8

6.5 Comparative Analysis in the Context of Handgun Detection

Higher recall of RT-DETR compared to YOLOv8(0.883Vs 0.856)would miss fewer potential threats, a critical factor in security applications. At the same time, more stable results in different trials could be evidence that YOLOv8 might be more reliable in various operational conditions.

6.6 Robustness and Adaptability Assessment

YOLOv8 transiently showed relatively stable performance for all hyperparameter ranges and reached the best performance, even with a quite low learning rate, meaning a very high adaptability to different training conditions. RT-DETR showed higher peak performance but was more sensitive to hyperparameters. This could mean that in very fine-tuned conditions, it performs well but might require more fine-tuning on a diverse environment.

6.7 Discussion

6.7.1 1.6.1 Synthesis of Findings

The experimental results indicate that both RT-DETR and YOLOv8 are capable of very effective handgun detection; the former has an edge in terms of absolute performance on all metrics. On the other hand, YOLOv8 was found to have more reliable performance and be more resistant to changes in hyperparameters.

These findings partially address the research question:

- Accuracy: RT-DETR shows marginally better accuracy across all metrics.
- Robustness: YOLOv8 demonstrates greater stability across different hyperparameter configurations.
- Adaptability: YOLOv8's consistent performance suggests better adaptability to varied conditions.
- Speed: Inconclusive due to lack of specific inference speed data.

6.7.2 Contextualizing Results with Previous Research

High performance by both models confirms findings by T et al.(2022) and Alaqil et al.(2020) which proved that deep learning models can be used effectively to detect handguns. The current results improve on previous studies in object detection architecture.

The stability of YOLOv8 across most varied hyperparameters agrees with the observed efficiency of YOLOv5, by Dextre et al.(2021), in real-time applications. Peak superiority of RT-DETR, on the other hand, might indicate some advantages of transformer-based architectures as discussed by Ruiz-Santaquiteria et al.(2021), in their work integrating extra information for better detection.

6.7.3 Critical Evaluation of Experimental Design

Strengths:

- Comprehensive hyperparameter tuning for both models
- Evaluation across multiple performance metrics
- Large number of trials for each model, providing robust performance data
- Diverse and extensive dataset with 15,579 handgun images
- Comprehensive data augmentation techniques simulating various real-world conditions

Limitations:

- Lack of direct inference speed measurements
- Absence of testing on edge cases or adversarial examples
- Limited information on the distribution of handgun types and backgrounds in the dataset

The experimental design gives insight into the performance models and sensitivity to hyperparameters. While the big dataset and several augmentation techniques improve the validity and generalization to real-world applications, possible speed measurements have not been given, which does not bode well when trying to provide full coverage for answering the research question in view of real-time performance.

6.7.4 1.6.4 Suggested Improvements and Future Work

To address the limitations and further present the research question:

1. Run inference speed tests across different hardware configurations to understand real-time capability.
2. Check the performance of the model on a different test set, containing challenging edge cases, for more adaptability testing.

3. Implement Cross-Validation to test the robustness of results against different data splits.
4. By considering ensemble methods that mix and match RT-DETR with YOLOv8, their respective strengths would be brought out.
5. Investigate the possibilities of transfer learning to estimate how well it can adapt to similar detection tasks.
6. Conduct user studies in simulated settings of security environments to estimate practical efficacy.
7. Analyze model performance on specific subsets of the data (e.g., different handgun types, various backgrounds) to assess versatility.

These would provide a baseline comparison of RT-DETR and YOLOv8, which conveys meaningful information to the research community in academics as well as to industry circles in public safety and security.

On the other hand, in peak performance, RT-DETR still outclasses YOLOv8. However, for real-world applications, it just simply can't compete with the consistency and robustness that's shown here by YOLOv8. Further emboldening these results is the large dataset and augmentation techniques applied in this study. The choice between these models will hence be based on the demands of the deployment scenario, balancing high accuracy against consistent performance and ease of implementation. Further research in improving what has been suggested will contribute a great deal to real-time handgun detection and object detection in general.

7 Conclusion and Future Work

This study has sought to answer the research question: "Which state-of-the-art object detection technology shows better performance in terms of accuracy, speed, robustness, and adaptability for real-time handgun detection in public spaces between RT-DETR and YOLOv8?"
The primary objectives were to:

1. Train both the RT-DETR and YOLOv8 models on a comprehensive dataset with regard to handguns.
2. Both models' performances can be compared with metrics such as mAP, precision, recall, and inference speed.
3. Test the robustness and adaptability of each model with different scenarios.
4. RT-DETR and YOLOv8 Architecture Change: How different architectures impact handgun detection performance.

The study successfully meets most of these objectives by providing insights into performance characteristics that are relevant to handgun detection for RT-DETR and YOLOv8. The findings include:

1. Accuracy: RT-DETR demonstrated slightly superior performance across all metrics (mAP50-95: 0.728 vs 0.7073, Precision: 0.940 vs 0.9010, Recall: 0.883 vs 0.8560).
2. Robustness:YOLOv8 managed to perform more consistently with different hyperparameters, RT-DETR could be peak performance but highly sensitive to changes in parameters.

3. Adaptability: The stability across settings of the YOLOv8 alludes to a better adaptability to diverse conditions, while the highest recall of the RT-DETR alludes to potential advantages in detecting all instances of handguns.
4. Speed: The study did not fully explore or capture the speed of inference, and this area was left somewhat open.

These results prove the efficiency of both models in handgun detection, each showing different strengths. RT-DETR peak performance seems superior for high-accuracy applications, while YOLOv8's consistency makes it perhaps better suited for diverse real-world deployments

.

Some of the major limitations of the study are it does not provide in-depth analysis on speed; all the tests performed were under limited conditions which were environmental. The research work was applied and only applied on one dataset, hence may really not represent other scenarios that exist in reality.

Future Work:

1. Real-time Performance Analysis: Thorough speed tests on varied hardware are to be made to analyze real-time capabilities, which are the most important and relevant in real practical deployment of security systems.
2. Environmental Robustness Testing: Testing: This will involve overall model performance assessment in conditions involving, for example, low light, partial occlusions, and variability in camera view angles to represent more real-world situations in public space..
3. Multi-threat Detection: Extend the models to detect multiple security threats simultaneously, including handguns, knives, and suspicious packages, and make them more ample security solutions.
4. Edge Computing Integration: Investigate the possibility of deploying these models over edge devices for having distributed, low-latency detection systems in large public spaces.
5. Explainable AI for Security Applications: Develop techniques related to handgun detection that will make it easier for security people to understand and trust the model's output.
6. Transfer Learning for Rapid Adaptation: I n this direction, the works will explore fine-tuning models toward detecting new or region-specific weapon types with minimum additional data.
7. Human-AI Collaborative Systems: Involves assessing systems that can integrate these AI models with human security persons, optimizing their strengths in providing service for public safety.
8. Privacy-Preserving Detection Techniques: Study methods for handgun detection in a public place without revealing the identity of a person and alleviate ethical concerns in surveillance.
9. Adversarial Testing and Defense: Design techniques that make these models resistant to possible adversarial attacks, ensuring reliability in critical security applications.
10. Cross-modal Integration: Visual detection must be integrated with other sensing modalities, like sound and thermal imaging, to correctly obtain a more complete threat detection.

11. Comparative XAI Study: Present an itemized comparison of SHAP results between YOLOv8 and RT-DETR. Such may show, if any, intrinsic differences in the way these architectures process visual information to detect handguns; this could provide improvement for hybrid models.

Future directions are thus dedicated to the relaxation of these study limitations and to applications SHARED underlying principles for real security challenges. By following these avenues, future research could significantly advance this emergent area of AI-driven public safety technologies, possibly leading to effective and reliable security solutions within any number of diverse public environments.

References

Huang, R., Pedoeem, J., & Chen, C. (2018). YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers. IEEE BigData 2018.

Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. European Conference on Computer Vision.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Advances in Neural Information Processing Systems.

Lim, J., Park, S., & Kim, D. (2023). Attention-Enhanced YOLOv5 for Improved Handgun Detection in Cluttered Environments. IEEE International Conference on Computer Vision Workshops.

Zhang, L., Wang, R., & Chen, X. (2023). Synthetic Data Augmentation for Robust Handgun Detection in Surveillance Scenarios. European Conference on Computer Vision.

Deng, Y., Campbell, R., & Kumar, P. (2022, July 18). Fire and Gun Detection Based on Sematic Embeddings. *2022 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. https://doi.org/10.1109/icmew56448.2022.9859303

Dextre, M., Rosas, O., Lazo, J., & Gutiérrez, J. C. (2021, October 25). Gun Detection in Real-Time, using YOLOv5 on Jetson AGX Xavier. *2021 XLVII Latin American Computing Conference (CLEI)*. https://doi.org/10.1109/clei53233.2021.9640100

Rehman, A., & Fahad, L. G. (2022, October 21). Real-Time Detection of Knives and Firearms using Deep Learning. *2022 24th International Multitopic Conference (INMIC)*. https://doi.org/10.1109/inmic56986.2022.9972915

Ruiz-Santaquiteria, J., Velasco-Mata, A., Vallez, N., Bueno, G., Alvarez-Garcia, J. A., & Deniz, O. (2021). Handgun Detection Using Combined Human Pose and Weapon Appearance. *IEEE Access*, *9*, 123815–123826. https://doi.org/10.1109/access.2021.3110335

Ali Shah, S. A., Ahmad Al-Khasawneh, M., & Uddin, M. I. (2021, June 15). Street-crimes Modelled Arms Recognition Technique (SMART) : Using VGG. *2021 2nd International*

Conference on Smart Computing and Electronic Enterprise (ICSCEE). https://doi.org/10.1109/icscee50312.2021.9497928

Mahajan, C., & Jadhav, A. (2022, March 9). Gun Detection: Comparative Analysis using Transfer Learning in Single Stage Detectors. *2022 International Conference on Emerging Smart Computing and Informatics (ESCI)*. https://doi.org/10.1109/esci53509.2022.9758345

Qi, D., Tan, W., Liu, Z., Yao, Q., & Liu, J. (2021, October 17). A Dataset and System for Real-Time Gun Detection in Surveillance Video Using Deep Learning. *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. https://doi.org/10.1109/smc52423.2021.9659207

Warsi, A., Abdullah, M., Jawaid, N., Khan, S., & Yahya, M. (2024, January 3). RZUD: A Novel Hybrid Model for Small Sized Handgun Detection. *2024 18th International Conference on Ubiquitous Information Management and Communication (IMCOM)*. https://doi.org/10.1109/imcom60618.2024.10418397

Alaqil, R. M., Alsuhaibani, J. A., Alhumaidi, B. A., Alnasser, R. A., Alotaibi, R. D., & Benhidour, H. (2020, November). Automatic Gun Detection From Images Using Faster R-CNN. *2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*. https://doi.org/10.1109/smart-tech49988.2020.00045

T, P., Thangaraj, R., P, P., M, U. R., & Vadivelu, B. (2022, May 9). Real-Time Handgun Detection in Surveillance Videos based on Deep Learning Approach. *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*. https://doi.org/10.1109/icaaic53929.2022.9793288