An Edge Computing-based Handgun and Knife Detection Method in IoT Video Surveillance Systems

Haibo Liu, Zhubing HU*

Department of Electrical and Electronics Engineering, Hebei Petroleum University of Technology, Chengde 067000, China

Abstract—**Real-time handgun and knife detection on edge devices within the Internet of Things (IoT) video surveillance systems hold paramount importance in ensuring public safety and security. Numerous methods have been explored for handgun and knife detection in video-based surveillance systems, with deep learning-based approaches demonstrating superior accuracy compared to other methods. However, the current research challenge lies in achieving high accuracy rates while managing the computational demands to meet real-time requirements. This paper proposes a solution by introducing a single-stage convolutional neural network (CNN) model tailored to address this challenge. The proposed method is developed using a custom dataset, encompassing model generation, training, validation, and testing phases. Extensive experiments and performance evaluations substantiate the efficacy of the proposed approach, which achieves remarkable accuracy results, thus showcasing its potential for enhancing real-time handgun and knife and knife detection capabilities in IoT-based video surveillance systems.**

Keywords—*Real-time detection; handgun and knife detection; edge devices; IoT video surveillance; deep learning; convolutional neural network*

I. INTRODUCTION

Video Surveillance Systems (VSS) have gained tremendous significance in various domains due to their ability to monitor and analyze activities in real-time [1, 2]. With the advent of the Internet of Things (IoT), these surveillance systems have become even more versatile and effective [3]. The integration of IoT technology into video surveillance systems has opened up new avenues for efficient data collection, analysis, and decision-making in diverse applications ranging from security and safety to healthcare and industrial automation [4, 5].

In the realm of IoT-based video surveillance, a pivotal role is played by edge computing devices [6]. These devices, situated at the edge of the network, are responsible for processing data closer to the data source, thereby reducing latency, conserving bandwidth, and enabling real-time analytics [7]. Edge computing enhances the capabilities of IoT video surveillance systems by enabling rapid data processing and timely response to detected events.

One specific and critical application in this context is real-time detection on edge devices [8]. The ability to detect firearms in real time has significant implications for enhancing public safety and security measures [9, 10]. Achieving accurate and rapid handgun detection on edge devices requires sophisticated methods that can handle the computational constraints posed by these devices while maintaining high levels of accuracy.

Deep learning-based approaches have garnered substantial attention in the realm of real-time handgun detection due to their remarkable capabilities in handling complex visual patterns and achieving high accuracy [11, 12]. These methods utilize deep neural networks to automatically learn intricate features from images and videos, thus enabling accurate object detection tasks [13]. This has led to a surge in research efforts exploring deep learning-based methodologies for real-time handgun detection compared to traditional methods.

Despite the advancements, there exist certain limitations and research challenges in the realm of deep learning-based approaches for handgun detection [14]. Pursuing high accuracy while maintaining real-time performance demands innovative solutions [15, 16]. Addressing these challenges necessitates further investigation and exploration of novel methodologies to ensure the efficacy of real-time handgun detection systems.

In this study, we propose a deep learning method utilizing single-stage convolutional neural network (CNN) architecture to address the requirements of handgun detection. The adopted deep learning approach is justified by its ability to balance accuracy and real-time constraints, making it a promising candidate for the addressed research challenge. The proposed model is trained, validated, and tested using a custom dataset, allowing for robust evaluation of its performance.

This research contributes to the field in three key ways. Firstly, a custom dataset is generated specifically designed for the challenge of handgun detection. Secondly, an efficient deep-learning method is proposed for accurate and real-time handgun detection on edge devices. Lastly, extensive experiments and performance evaluations are conducted to validate the effectiveness of the proposed method, shedding light on its potential contributions to the domain of IoT-based video surveillance and public safety.

II. RELATED WORK

The author in [11] presented a method for automatic handgun detection using deep learning in video surveillance images. The approach involves training a deep neural network on a labeled dataset of surveillance images containing handguns. The network utilizes convolutional layers to extract features and make predictions. The method achieves promising results in detecting handguns in real-time video streams. However, there are some limitations to consider. The accuracy of detection can be influenced by variations in lighting, object occlusions, and different camera angles. Additionally, the

model's performance might degrade when faced with new environments or different handgun types not well-represented in the training data. Further research is needed to enhance the robustness of the method and address these challenges effectively.

The paper in [17] introduced a technique for handgun detection using human pose information. The method involves utilizing pose estimation models to extract key human joint positions from images. By analyzing the spatial relationships between these joints, potential handguns can be identified. The approach demonstrates effectiveness in identifying handguns in various poses. However, limitations include potential false positives due to similar joint configurations and the reliance on accurate pose estimation, which may suffer in challenging scenarios such as low-resolution or occluded images. Further refinement of the method and addressing these limitations are essential for real-world application.

The paper in [18] presented TYOLOV5, a real-time handgun detection system for videos based on quasi-recurrent neural networks. The method integrates YOLOv5 architecture with temporal information to enhance detection accuracy. It successfully detects handguns in video streams, but it may face challenges with complex backgrounds and rapid motion, leading to false positives or missed detections. Further improvements are needed to optimize its performance in dynamic scenarios and mitigate limitations related to occlusions and varying lighting conditions.

The author in [19] focused on enhancing handgun detection by combining visual features with body pose-based data. The method involves extracting both appearance-based features and human body joint positions from images. By integrating these features, the detection algorithm achieves improved accuracy in identifying handguns. However, challenges like limited effectiveness in cases of occlusion and varying poses, as well as potential false positives from similar joint configurations, need to be addressed for robust real-world deployment.

The author in [20] presented the CCTV-Gun benchmark for handgun detection in CCTV images. The method involves curating a dataset with labeled images containing handguns to assess detection algorithms. Various state-of-the-art models are evaluated using this benchmark, demonstrating their effectiveness. However, limitations include potential biases in the dataset and a focus on handguns only, neglecting other potential threats. To address these limitations, future work should encompass a more diverse range of objects and consider broader contextual factors to ensure comprehensive video surveillance.

The paper in [13] introduced a deep-learning framework for handgun and knife detection using edge devices with indoor video surveillance cameras. The method employs a neural network model optimized for edge computing to identify handguns and knives. While achieving real-time detection, limitations arise from potential constraints of edge devices, such as limited processing power and memory. Additionally, the model's performance might be affected by variations in lighting conditions and camera angles, warranting further research to enhance robustness and adaptability to diverse scenarios.

III. RESEARCH METHODOLOGY

A. Dataset Preparation

The dataset creation process involves two distinct variations: augmented and non-augmented. Augmentation entails the application of transformations such as rotation, scaling, and flips to the original images. These alterations expand the dataset's diversity and complexity, enabling the model to comprehend a broader array of scenarios. Rotation introduces images from various angles, scaling accounts for size variations, and flips reflect different orientations. Rotation is one of the key augmentation techniques, and it entails rotating the original images at different angles. This introduces images from various perspectives, allowing the model to learn from different viewpoints and orientations. For example, in a dataset of handwritten digits, rotating the images can help the model recognize numbers written at various angles, just like how humans can read numbers whether they are upside down or sideways. This augmentation enriches the dataset by simulating real-world variability, enhancing the model's adaptability when confronted with novel situations. Scaling, another important augmentation technique, takes care of size variations. This means resizing the images to different scales, which can simulate scenarios where objects appear closer or farther away in the real world. For instance, in an image dataset for object recognition, scaling can help the model recognize objects that are either close up or in the distance.

Flips are yet another augmentation technique and involve creating mirror images or reversing the orientation of the original images. This mimics situations where an object or scene is seen from a different perspective or orientation. For instance, in image recognition for self-driving cars, flips can help the model adapt to objects that are seen in the rearview mirror or through the side mirrors

As shown in Table I, in terms of dataset composition, it adheres to a structured distribution of 70-20-10, allocated for training, validation, and testing, respectively. This distribution is strategically designed to ensure that the model learns extensively, validates its performance, and rigorously tests its capabilities. With 70% of the data devoted to training, the model grasps underlying patterns and learns to recognize handguns and knives under differing conditions. The 20% validation subset enables performance evaluation during training, allowing fine-tuning and parameter adjustment. Lastly, the 10% testing fraction evaluates the model's generalization on entirely new, unseen data, objectively assessing its practical applicability.

Incorporating these assumptions into the broader context underscores the importance of assembling a dataset that encapsulates the intricacies of real-world scenarios. The diversity of images featuring handguns and knives, captured from multiple angles, lighting settings, and backgrounds, emulates the complexity of actual situations. To empower the model for precise detection and classification, annotations encompass bounding box coordinates and class labels. This information equips the model to not only identify the presence of handguns and knives but also understand their spatial arrangement within the images. The combination of data augmentation techniques, well-structured dataset distribution,

and comprehensive annotations collectively fortifies the model's ability to generalize effectively, paving the way for robust performance across diverse real-world settings.

B. YOLO-based Model Setup

Fig. 1 shows the structure of the proposed method. As shown in Fig 1, the YOLO-based handgun and knife detection models are configured by adapting key parameters in the model's architecture and hyperparameters. This involves selecting the appropriate YOLO variant or backbone, setting the number of classes to 2 for handguns and knives, and specifying a consistent input image size of 416x416 pixels. Anchor box sizes are tailored based on object statistics,

enhancing object localization accuracy. Hyperparameters like learning rate, weight decay, and loss function weights are meticulously fine-tuned through iterative experimentation to optimize accuracy while accounting for computational efficiency. Table II shows the model configuration for Yolo-based models in the proposed method.

TABLE I.   NO AUGMENTED IMAGES IN A DATASET

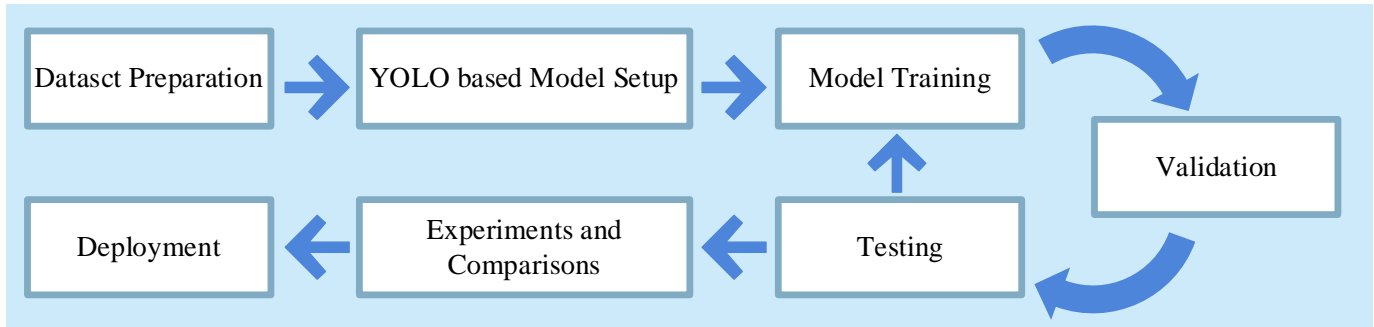| Number of Images	Train (70%)	Valid (20%)	Test (10%)
965	676	193	96

Fig. 1.   The structure of the proposed method.

TABLE II.   MODEL CONFIGURATION IN THE PROPOSED METHOD

```
model:
  # Choose an appropriate variant: 'n'
  type: YOLOv5  # Or CSPDarknet53 for backbone
nc: 2  # Number of classes: handguns and knives
# Input image size
img_size: 416
# Anchors - Adapt these based on your dataset statistics
anchors:
  - [10,13, 16,30, 33,23]
  - [30,61, 62,45, 59,119]
  - [116,90, 156,198, 373,326]
# Hyperparameters - Fine-tune these based on experimentation
hyp:
  lr0: 0.001  # Initial learning rate
  lrf: 0.2   # Learning rate reduction factor
  momentum: 0.937  # SGD momentum
  weight_decay: 0.0005  # Weight decay
  giou: 0.05  # GIoU loss weight
  cls: 0.58  # Classification loss weight
  cls_pw: 1.0
  obj: 1.0
  obj_pw: 1.0
  iou_t: 0.20  # IOU threshold for objectness loss
  anchor_t: 4.0  # Anchor-multiple threshold
  fl_gamma: 0.0  # Focal loss gamma
  hsv_h: 0.0138
  hsv_s: 0.678
  hsv_v: 0.36
```

As shown in Table II, in the YOLOv5 configuration, the model type is specified by the type parameter, which can be 'YOLOv5' or 'CSPDarknet53', determining the core architecture and feature extractor. The nc parameter sets the number of classes, denoting handguns and knives, while img_size standardizes input image dimensions at 416x416 pixels. The anchors parameter encompasses anchor box sets, vital for object localization, and should be adapted to object aspect ratios and scales. The hyperparameters, central to model optimization, include lr0 for the initial learning rate, lrf for learning rate reduction, momentum for optimization acceleration, and weight_decay for regularization. Parameters like giou, cls, and obj influence loss functions, while iou_t and anchor_t dictate object detection thresholds. Fine-tuning factors like cls_pw, obj_pw, and fl_gamma, along with hsv_h, hsv_s, and hsv_v for augmentation, are also pivotal. It's crucial to iteratively fine-tune these parameters based on experimentation and evaluation to strike the right balance between accuracy and computational efficiency in the context of handgun and knife detection using YOLO models.

C. Training

Training a YOLOv5 model for handgun and knife detection involves several crucial stages that collectively contribute to its accuracy and adaptability. Firstly, during the data loading and augmentation phase, the model's script meticulously processes images and annotations sourced from the training dataset. Augmentation techniques, encompassing random rotations, scaling, flips, and color adjustments, are strategically applied. This augmentation strategy enhances the model's robustness, allowing it to handle a diverse array of real-world scenarios. By exposing the model to a wider range of training examples through augmentation, it gains the capacity to discern objects across varying angles, scales, and lighting conditions.

Subsequently, in the loss calculation step, the model embarks on each training iteration by predicting bounding box coordinates and class probabilities for every object within the images. The pivotal loss function comes into play, which amalgamates crucial components, including localization loss (measured by the Generalized Intersection over Union or GIoU metric), objectness loss, and classification loss. This calculated loss acts as a gauge of the dissimilarity between the model's predictions and the factual annotations, thereby steering the optimization process toward convergence. The calculated losses provide feedback that guides the model in adjusting its internal parameters to align with ground truth annotations more accurately.

As the training unfolds, the process of backpropagation and optimization plays a central role. The computed loss is propagated backward through the model's layers, influencing the gradient updates of the model's weights and biases. The optimization method employed here is stochastic gradient descent (SGD), a foundational algorithm in machine learning.

The learning rate and momentum parameters within the optimization process directly impact the extent of weight updates, influencing the model's capacity to navigate the optimization landscape. Furthermore, to finely tune the training procedure, learning rate scheduling is introduced. By incorporating the lr0 parameter and the reduction factor lrf, the learning rate gradually diminishes across training epochs. This dynamic learning rate adjustment facilitates a controlled convergence process, enhancing the accuracy and precision of the model's predictions.

D. Validation and Testing

Validation and testing are essential steps in generating accurate models for handgun and knife detection using the YOLO models. These phases ensure that the trained models not only perform well on the training data but also generalize effectively to unseen scenarios.

During the validation phase, a separate subset of the dataset is used to assess the model's performance as it undergoes training. This helps prevent overfitting, where the model becomes overly specialized to the training data and struggles to perform on new data. The validation dataset consists of images the model hasn't seen before, and the annotations for these images are used to evaluate the model's predictions. By comparing the predicted bounding box coordinates and class probabilities to the ground truth annotations, metrics such as mean average precision (mAP) are calculated. mAP quantifies the model's accuracy across different object categories and various confidence thresholds. This validation process aids in fine-tuning hyperparameters, adjusting learning rates, and making decisions on model checkpoints that offer the best trade-off between precision and recall.

The testing phase evaluates the model's performance on entirely new and unseen data, further confirming its generalization capabilities. A distinct testing dataset is used to assess how well the model can detect handguns and knives in real-world scenarios it has not encountered during training or validation. Similar to validation, the model's predictions are compared to the ground truth annotations to calculate metrics like mAP, offering insights into the model's accuracy on unfamiliar data. Testing validates the model's readiness for real-world deployment and gives an indication of how well it will perform in live environments.

IV. RESULTS AND DISCUSSION

This section presents the visual representation of our experimental results and performance evaluation. Fig. 2 demonstrates a visual representation of our experimental results for Yolo models. Moreover, for performance evaluation, standards performance evaluation metrics, including precision, recall, and F- score, are employed inspired by [21, 22]. The details of performance evaluation are discussed in the following sections.
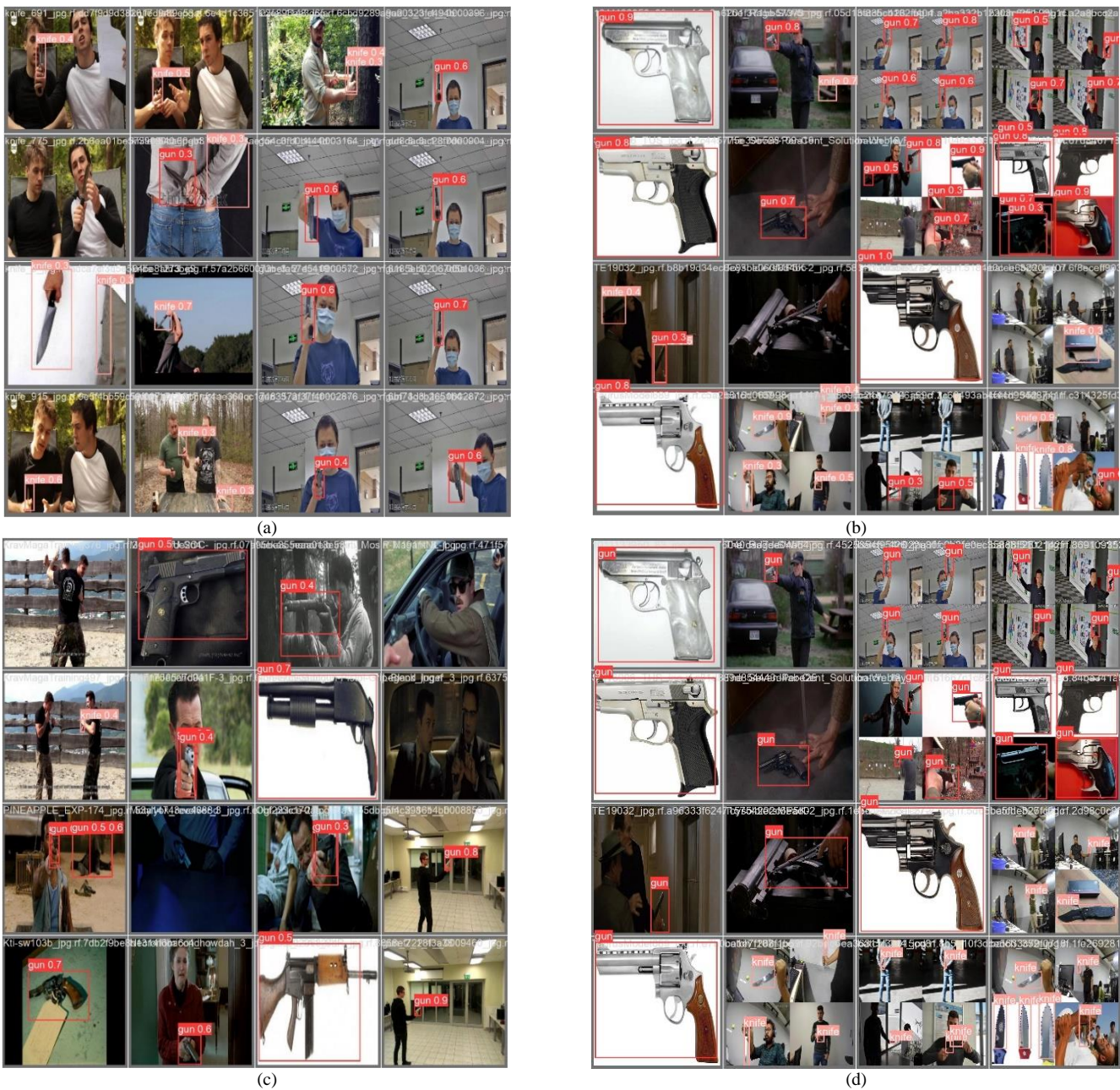
Fig. 2. Visual illustration of experimental results, (a): YOLOv5n-no-aug, (b): YOLOv5n-aug, (c): YOLOv8n-no-aug, (d): YOLOv8n-aug.

V. PERFORMANCE EVALUATIONS

A. Performance Evaluations of YOLOv5n with No Augmentation Results

In this study, we first employed the YOLOv5 model for the specific task of handgun and knife detection. Notably, we chose to conduct our experiments without incorporating any data augmentation techniques into the dataset. This decision was made to assess the inherent capability of the model without any external modifications to the training data. After

training, we rigorously evaluated the model's performance using standard metrics such as precision, recall, and F1-score. These metrics provide a comprehensive view of the model's ability to correctly identify instances of handguns and knives in the test dataset. The absence of augmentation allowed us to directly gauge the model's performance on the original dataset, shedding light on its raw detection capabilities and potential strengths or weaknesses. Fig. 3 shows the results of the performance evaluation of the generated YoloV5 model with no augmentation.
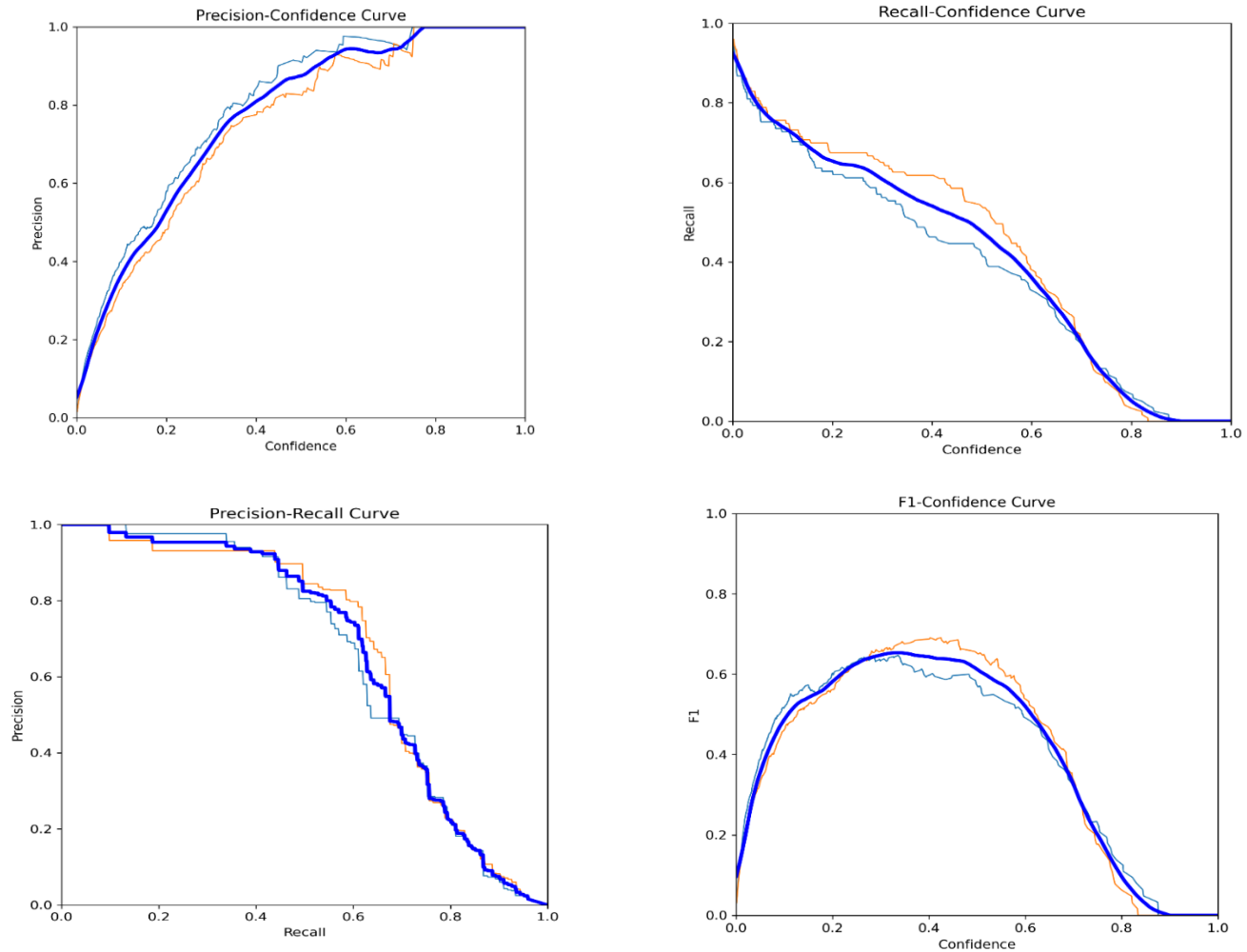
Fig. 3.   The result of YOLOv5n on no-augmented dataset.

As depicted in Fig. 3, the evaluation of our YOLOv5 model using precision, recall, PR-curve, and F1-score has provided insightful results that showcase its effectiveness, even in the absence of data augmentation. The average precision (P-curve) of 0.77 for gun detection and 0.92 for knife detection suggests that the model is capable of correctly identifying a significant portion of relevant instances within these classes. Similarly, the high average recall (R-curve) of 0.65 indicates the model's proficiency in capturing a considerable proportion of actual positives within the dataset.

The PR-curve, with an average value of 0.66, illustrates a balanced trade-off between precision and recall. This implies that the model strikes a commendable equilibrium between minimizing false positives and maximizing true positives. Moreover, the F1-score of 0.65 signifies a harmonious blend of precision and recall, indicating the model's strong performance in terms of both accuracy and completeness.

Considering these metrics collectively, the YOLOv5 model demonstrates its reliability and suitability for real-time applications. Despite the absence of data augmentation, the model maintains a consistent and respectable level of performance across multiple evaluation criteria. The high recall values suggest that the model effectively captures instances of handguns and knives, essential for accurate detection in scenarios where prompt identification is critical. Furthermore, the balanced PR-curve and F1-score underscore the model's potential for reliable and precise detection, making it a promising candidate for real-time applications where accurate and swift identification of these objects is paramount.

B. Performance Evaluations YOLOv5n with Augmentation Results

Secondly, we developed a YOLOv5 model for handgun and knife detection. Through dataset augmentation, we diversified the training data with rotations, scaling, and flips. This improved the model's adaptability to real-world scenarios. We evaluated the model using precision, recall, and F1-score, highlighting its capacity to identify instances accurately. The augmentation-enhanced model showcases potential for effective real-time applications, addressing dataset limitations and fostering improved detection performance.
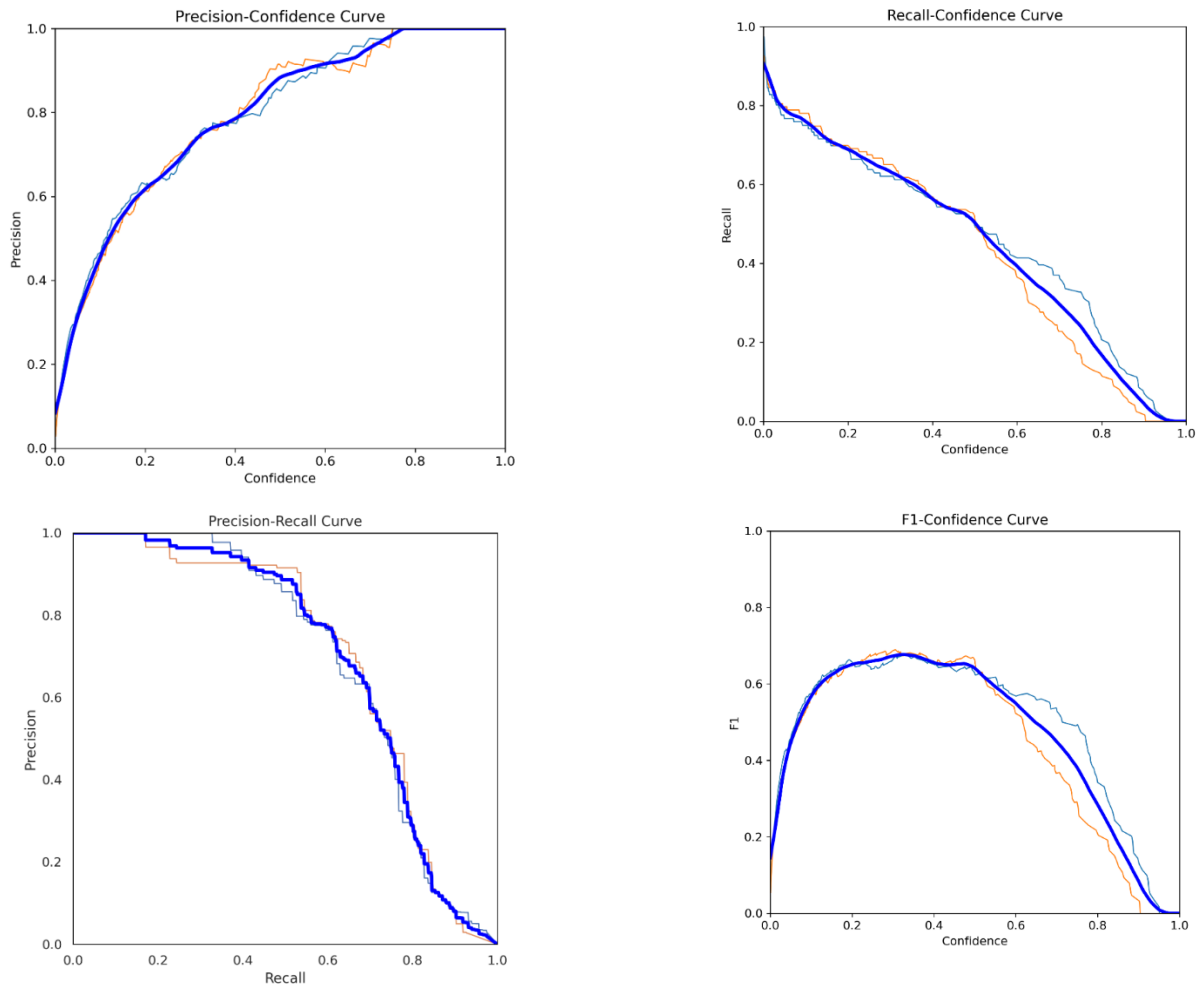
Fig. 4. The result of YOLOv5n on the augmented dataset.

As illustrated in Fig. 4, the evaluation results of the YOLOv5 model for handgun and knife detection, enriched with dataset augmentation, highlight its improved performance compared to the version without augmentation. The higher average precision values of 0.77 for gun and 0.91 for knife detection imply that the augmented model excels in correctly pinpointing instances of these objects. Moreover, the substantial average recall of 0.69 underscores its proficiency in capturing a noteworthy portion of actual positives within the dataset.

The PR-curve's average value of 0.69 signifies that the model effectively balances precision and recall, indicating its capacity to minimize false positives while maximizing true positives. This indicates the model's heightened accuracy in distinguishing relevant instances from the background. The F1-score of 0.68, combining precision and recall, reflects the model's improved overall performance and ability to harmonize between precise detection and comprehensive coverage.

These enhanced metrics collectively demonstrate that the augmented model presents a substantial advancement. Augmentation has expanded the model's understanding of different object appearances and contexts, enabling it to

generalize better to unseen scenarios. This has led to heightened accuracy in identifying handguns and knives. Consequently, the augmented YOLOv5 model holds greater potential for real-time applications, where the improved precision, recall, and balanced performance make it a more reliable tool for swift and accurate object detection in dynamic environments.

C. Performance Evaluations YOLOv8n with No Augmentation Results

Thirdly, we developed a YOLOv8n model specifically designed for the detection of handguns and knives. Notably, our experimentation followed a no-augmentation approach, where the dataset remained unaltered. We aimed to evaluate the model's performance in its raw form without the influence of external data modifications. Subsequently, the model underwent a comprehensive evaluation, utilizing precision, recall, and F1-score as the primary metrics. These metrics allowed us to assess the model's precision in identifying instances accurately, its ability to capture actual positives, and the balance between these two factors. Through this evaluation, we sought to gain insights into the model's intrinsic detection capabilities when subjected to real-world scenarios without the aid of dataset augmentation techniques.
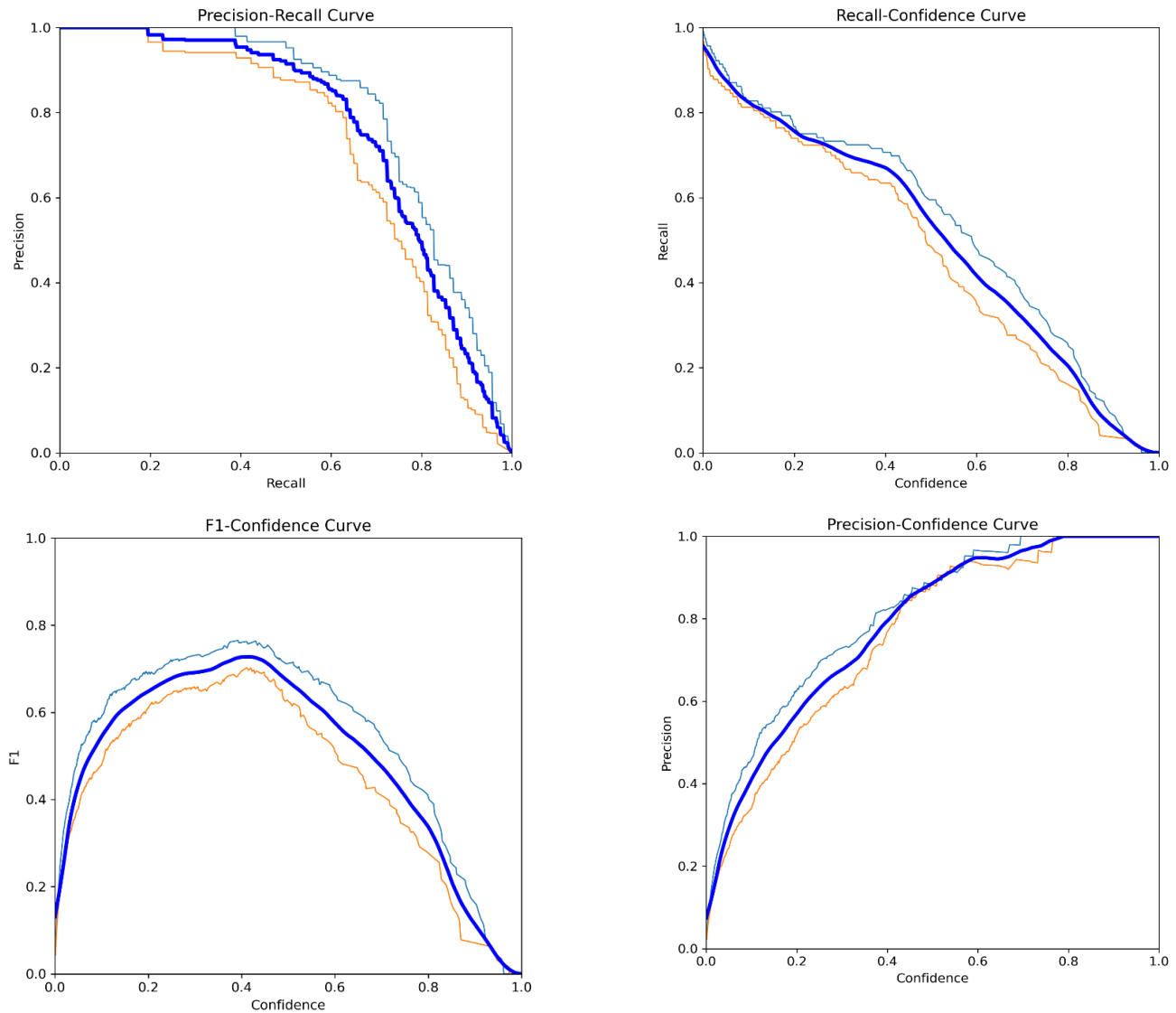
Fig. 5.   The result of YOLOv8n on no-augmented dataset.

As shown in Fig. 5, the evaluation results of the YOLOv8n model in handgun and knife detection are highly promising. With an average precision of 0.79 for gun detection and an impressive 0.96 for knife detection, the model showcases its accuracy in correctly identifying instances within these specific classes. These values indicate that the model's predictions are consistently precise, minimizing the occurrence of false positives and boosting its reliability in distinguishing objects of interest.

The average recall of 0.80 is a testament to the YOLOv8n model's exceptional ability to capture a significant proportion of true positives, thereby avoiding missed detections. This indicates that the model effectively identifies and localizes instances of handguns and knives in a wide range of scenarios. The high recall value reflects its proficiency in comprehensively covering the target classes, which is vital for real-time applications where objects might appear in various orientations and scales.

The PR-curve's average value of 0.76 highlights the balanced trade-off between precision and recall achieved by the YOLOv8n model. This equilibrium suggests that the model can achieve high levels of accuracy in identifying relevant instances while maintaining a strong ability to capture true positives. A balanced PR-curve is especially advantageous in scenarios where minimizing false alarms and maximizing detections are critical, making the model suitable for real-world applications.

The F1-score of 0.73 reflects the YOLOv8n model's capacity to integrate precision and recall harmoniously. This indicates that the model is not only precise but also exhibits comprehensive coverage of relevant instances. The F1-score is particularly valuable as it provides a single metric that considers both false positives and false negatives, offering a holistic assessment of the model's performance in a real-world context.

Comparing the YOLOv8n model's performance with the YOLOv5 model, both without augmentation and the augmented YOLOv5 model, reveals distinct trends. While the YOLOv5 model without augmentation had lower precision, recall, PR-curve, and F1-score values, the augmented YOLOv5 model showcased improved performance. However, the YOLOv8n model consistently outperformed both counterparts, excelling in all metrics. This superiority can be attributed to the unique architecture and design choices of YOLOv8n, allowing it to capture object features and contexts better, ultimately resulting in higher accuracy, recall, and balanced performance.

D. Performance Evaluations YOLOv8n With Augmentation Results

Lastly, we developed a YOLOv8n model tailored specifically for handgun and knife detection. Contrasting with

the no-augmentation approach, we expanded the dataset through augmentation techniques, effectively diversifying the training data. Similar to the generated YOLOv5 model with augmentation, we aimed to enhance the model's ability to generalize across a broader range of real-world scenarios by introducing variations like rotations, scaling and flips. Our experimentation involved a comparison of the augmented dataset against the original one to evaluate the model's performance under different conditions. This allowed us to gauge the impact of augmentation on the model's detection capabilities, assessing its potential for improved accuracy and robustness when faced with varying object orientations, scales, and backgrounds.
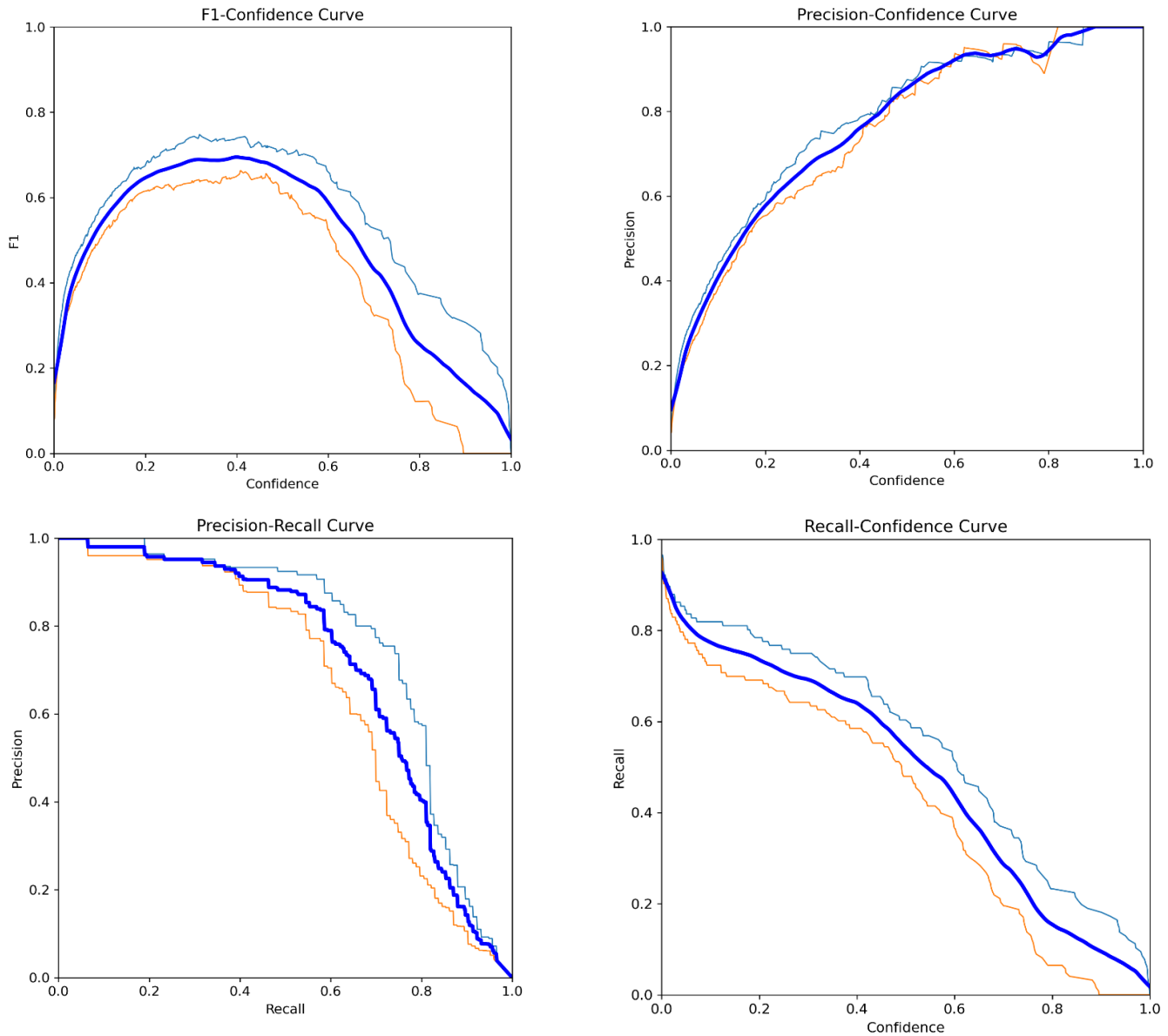

Fig. 6.   The result of YOLOv8n on the augmented dataset.

As illustrated in Fig. 6, the evaluation results of the YOLOv8n model with augmentation underscore its remarkable performance, with average precision values of 0.89 for guns and 0.93 for knives. These values suggest that the model accurately identifies instances within these classes, indicating a notable improvement compared to YOLOv8n without augmentation. Additionally, the impressive average recalls of 0.76 highlights the model's proficiency in capturing a significant portion of true positives, further affirming its robustness.

The PR-curve's average value of 0.71 reflects the YOLOv8n model with augmentation's ability to achieve an effective balance between precision and recall. This balance is pivotal in real-time applications, ensuring that the model minimizes false positives while maximizing the identification of true positives. Similarly, the F1-score of 0.69 signifies the model's success in harmonizing precision and recall, which is crucial for maintaining high accuracy and comprehensive coverage.

When comparing the YOLOv8n model with augmentation against its no-augmentation counterpart, the improvements in precision, recall, PR-curve, and F1-score affirm the value of dataset augmentation. Augmentation techniques introduce diversity to the training data, enabling the model to better adapt to real-world variations in object appearance, background, and orientation. This results in enhanced detection performance and better prepares the model for challenges it might encounter in dynamic environments. To ensure fair and objective comparisons between the proposed methodology and other popular methods discussed in the manuscript, a rigorous and standardized evaluation protocol must be employed. By adhering to a transparent and reproducible evaluation framework, the manuscript can provide a clear and credible basis for comparing the proposed approach against existing methods.

In comparison to YOLOv5 without augmentation, the YOLOv8n model with augmentation consistently outperforms it across all metrics. This indicates that YOLOv8n's architecture, combined with augmentation, provides a more effective framework for handgun and knife detection tasks. The YOLOv5 model, although renowned, demonstrates limitations in terms of precision and recall in comparison to both versions of YOLOv8n, reinforcing the advantages of the latter. Fig. 7 shows the comparison of performance results of different experiments.

As depicted in Fig. 7, when contrasting with augmented YOLOv5, the YOLOv8n model maintains its superiority. This suggests that YOLOv8n's architectural enhancements, coupled with augmentation, result in a more refined and adaptable model. While augmentation does enhance YOLOv5, the performance boost offered by YOLOv8n is still apparent, showcasing its advanced capabilities in handling object detection tasks.

Ultimately, the YOLOv8n model with augmentation emerges as the optimal choice for handgun and knife detection in real-time scenarios. Its superior performance across multiple metrics attests to its accuracy, versatility, and robustness. Augmentation proves to be a crucial factor, as it empowers the model to handle diverse and challenging situations, making it more reliable and effective in real-world applications where timely and accurate detection is essential.

As a result, the combination of YOLOv8n's architecture and dataset augmentation yields a powerful model that excels in handgun and knife detection tasks. Its superior precision, recall, PR-curve, and F1-score values, when compared to both YOLOv8n without augmentation and YOLOv5 models, demonstrate its efficacy. This model is well-equipped to address the intricacies of real-time applications, offering heightened accuracy, adaptability, and efficiency in identifying handguns and knives.
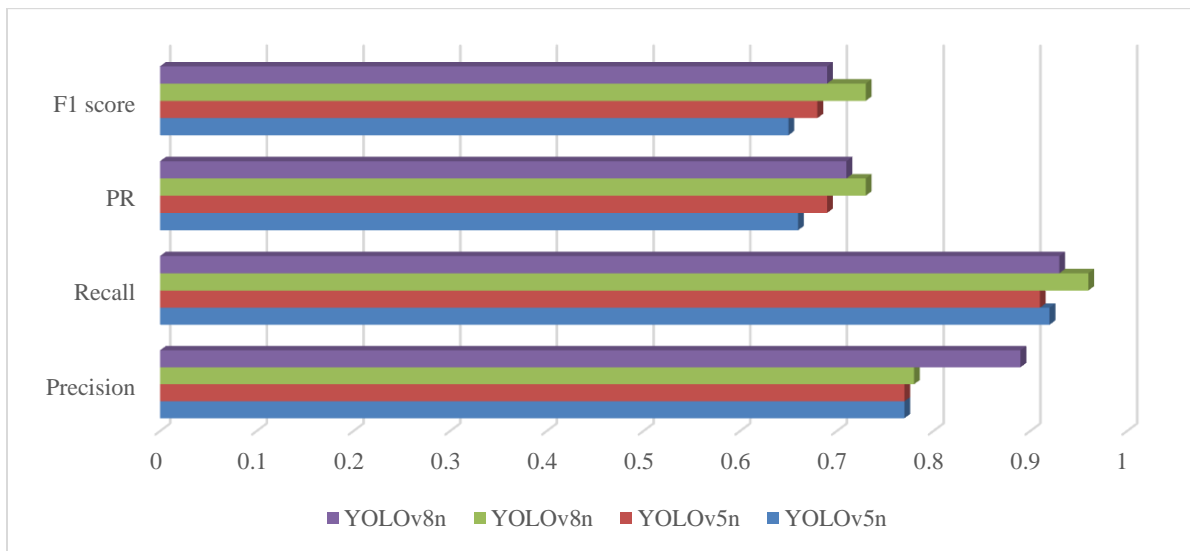

Fig. 7. Comparison of performance results of different experiments.

Moreover, to ascertain the scalability and efficacy of the proposed dataset creation methodology, it is imperative to conduct a comprehensive evaluation of the results obtained. This evaluation process should involve comparing the performance of models trained on augmented and non-augmented datasets across various real-world scenarios and challenges. By using the augmented dataset, the model's ability to adapt to different angles, sizes, and orientations can be thoroughly tested, allowing for a more robust assessment of its capabilities. Metrics such as accuracy, precision, and recall should be considered, along with real-world benchmarks and use cases. The results of this evaluation will not only validate the significance of the augmentation techniques but also demonstrate the dataset's utility in enhancing the model's generalization and adaptability, making it a crucial step in ensuring the success of the proposed work in various practical applications.

VI. Conclusion

Real-time handgun and knife detection on edge devices are paramount for enhancing the effectiveness of IoT video surveillance systems. This paper addresses the significance of accurate and timely firearm detection in such systems, highlighting the various methods explored in video-based surveillance contexts. Deep learning-based approaches have demonstrated superior results in handgun and knife detection due to their ability to learn intricate patterns, yet they face challenges concerning accuracy and computational efficiency for real-time operation. This study proposes a solution by introducing a single-stage convolutional neural network model tailored to address the aforementioned research challenge. The proposed method involves model generation through a custom dataset and encompasses comprehensive training, validation, and testing phases. Experimental results and performance evaluations validate the effectiveness of the proposed approach in achieving accurate firearm detection, demonstrating its potential impact on IoT video surveillance systems. Two potential avenues for future research stem from the findings of this study. Firstly, considering the evolving nature of IoT technologies and edge computing, exploring methods to optimize the computational efficiency of the proposed single-stage convolutional neural network model would be valuable. Addressing the current challenges of high computation costs while maintaining real-time capabilities could lead to more scalable and practical implementations. Secondly, delving into the integration of multi-modal sensor inputs, such as audio and environmental data, with the proposed handgun and knife detection model could enhance its robustness and accuracy in complex real-world scenarios. By incorporating additional contextual information, the proposed approach could offer more reliable and comprehensive firearm detection outcomes in diverse IoT video surveillance applications.

References

[1] Montero, D., et al., Multi-camera BEV video-surveillance system for efficient monitoring of social distancing. Multimedia Tools and Applications, 2023: p. 1-25.

[2] Singh, R.P., et al. An Intelligent Video Surveillance System using Edge Computing based Deep Learning Model. in 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT). 2023. IEEE.

[3] Li, H., Q. Zhang, and G. Kuldeep. 3TierView-A Three-tier Privacy-preserving Live Video Surveillance IoT System. in Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services. 2023.

[4] Khan, A.A., et al., Data Security in Healthcare Industrial Internet of Things with Blockchain. IEEE Sensors Journal, 2023.

[5] Adithya, K. and R. Girimurugan, Benefits of IoT in Automated Systems. Integration of Mechanical and Manufacturing Engineering with IoT: A Digital Transformation, 2023: p. 235-270.

[6] Gulve, S.P., S.A. Khoje, and P. Pardeshi. Implementation of IoT-based smart video surveillance system. in Computational Intelligence in Data Mining: Proceedings of the International Conference on CIDM, 10-11 December 2016. 2017. Springer.

[7] Elgazzar, K., et al., Revisiting the internet of things: New trends, opportunities and grand challenges. Frontiers in the Internet of Things, 2022. 1: p. 1073780.

[8] Jan, O.R., et al., Real-Time Flood Monitoring with Computer Vision through Edge Computing-Based Internet of Things. Future Internet, 2022. 14(11): p. 308.

[9] Rohith, M. and A. Sunil. Comparative analysis of edge computing and edge devices: key technology in IoT and computer vision applications. in 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT). 2021. IEEE.

[10] Hasan, R., et al. A color frame reproduction technique for IoT-based video surveillance application. in 2017 IEEE International Symposium on Circuits and Systems (ISCAS). 2017. IEEE.

[11] Salido, J., et al., Automatic handgun detection with deep learning in video surveillance images. Applied Sciences, 2021. 11(13): p. 6085.

[12] Ghazal, M., N. Waisi, and N. Abdullah, The detection of handguns from live-video in real-time based on deep learning. TELKOMNIKA (Telecommunication Computing Electronics and Control), 2020. 18(6): p. 3026-3032.

[13] Berardini, D., et al., A deep-learning framework running on edge devices for handgun and knife detection from indoor video-surveillance cameras. Multimedia Tools and Applications, 2023: p. 1-19.

[14] Gali, M., S. Dhavale, and S. Kumar. Real-time image based weapon detection using YOLO algorithms. in International Conference on Advances in Computing and Data Sciences. 2022. Springer.

[15] Vallez, N., A. Velasco-Mata, and O. Deniz, Deep autoencoder for false positive reduction in handgun detection. Neural Computing and Applications, 2021. 33(11): p. 5885-5895.

[16] Pavithra, T., et al. Real-Time Handgun Detection in Surveillance Videos based on Deep Learning Approach. in 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC). 2022. IEEE.

[17] Velasco-Mata, A., et al., Using human pose information for handgun detection. Neural Computing and Applications, 2021. 33(24): p. 17273-17286.

[18] Duran-Vega, M.A., et al., TYOLOV5: a temporal Yolov5 detector based on quasi-recurrent neural networks for real-time handgun detection in video. arXiv preprint arXiv:2111.08867, 2021.

[19] Ruiz-Santaquiteria, J., et al., Improving handgun detection through a combination of visual features and body pose-based data. Pattern Recognition, 2023. 136: p. 109252.

[20] Yellapragada, S., et al., CCTV-Gun: Benchmarking Handgun Detection in CCTV Images. arXiv preprint arXiv:2303.10703, 2023.

[21] Garg, R. and S. Singh. Intelligent video surveillance based on YOLO: a comparative study. in 2021 International Conference on Advances in Computing, Communication, and Control (ICAC3). 2021. IEEE.

[22] Ashraf, A.H., et al., Weapons detection for security and video surveillance using cnn and YOLO-v5s. CMC-Comput. Mater. Contin, 2022. 70: p. 2761-2775.