Enhancing Firefighter PPE Compliance Through Deep Learning and Computer Vision

Asmaa Alayed, Razan Talal Alqurashi, Samah Hamoud Alhelali, Asrar Yousef Khadawurdi, Bashayer Fayez Khan
College of Computing, Umm Al-Qura University, Makkah, Saudi Arabia

***Abstract*—Ensuring firefighter safety in high-risk environments requires strict adherence to Personal Protective Equipment (PPE) protocols. This study presents an automated real-time detection system for PPE using deep learning and computer vision techniques, aiming to improve PPE compliance and overall safety monitoring. The research employs advanced object detection models, specifically YOLOv10 and YOLOv11 (You Only Look Once), to identify critical PPE components such as helmets, gloves, boots, and self-contained breathing apparatus (SCBA) units. A custom-annotated dataset of firefighter images was developed to train and evaluate both models using standard performance metrics such as precision, recall, mAP, F1-score, and Intersection over Union (IoU). The results show that YOLOv11 outperformed YOLOv10, achieving a higher mAP@0.5 score of 0.646 compared to 0.586, with improved detection of small and partially occluded objects and a reduction in training time by 11%. YOLOv11 showed improved detection accuracy for small and partially blocked objects and reduced training time by 11%, while maintaining real-time efficiency. The system generates instant alerts when PPE is missing, minimizing reliance on manual monitoring and improving situational awareness in real-time. This research reinforces the role of AI in public safety and AI-powered automation in enhancing critical public safety operations. By integrating deep learning and computer vision into PPE monitoring systems, the study contributes to developing intelligent, responsive solutions aligned with modern safety standards.**

Keywords*—Firefighter safety; Personal Protective Equipment (PPE) Object detection; YOLOv10; YOLOv11; deep learning; computer vision; real-time detection; PPE compliance; AI in public safety

I. Introduction

Firefighting is a highly demanding and dangerous profession where the safety of personnel relies heavily on their ability to wear all necessary personal protective equipment (PPE) correctly. Essential equipment such as helmets, gloves, fire-resistant clothing, boots, and self-contained breathing apparatus (SCBA) shields firefighters from extreme heat, toxic gases, falling debris, and hazardous environments. Ensuring that this gear is worn properly is not only a matter of regulation but also a life-saving necessity.

However, during emergency responses, ensuring complete PPE compliance becomes difficult due to high-pressure conditions, time constraints, and human error. Manual checks are often rushed or incomplete, exposing firefighters to avoidable risks. Although safety standards exist in Saudi Arabia under the General Directorate of Civil Defense [1], ensuring real-time PPE compliance remains a challenge—especially before deployment into active zones.

Recent advances in artificial intelligence and computer vision have created new opportunities for enhancing safety protocols in the firefighting domain. Deep learning models—particularly object detection algorithms like YOLO—have shown impressive real-time performance in detecting and classifying objects across various sectors, including medical applications [2], construction [3], [4], and public safety systems [5]. These technologies offer a promising solution for automatically verifying the presence of PPE using image or video input prior to emergency deployment. This approach reduces reliance on manual inspection, increases accuracy, and enables faster decision-making in critical moments.

Despite existing research in PPE detection across various fields such as construction and industry, a significant research gap remains in applying AI-based systems to the firefighting sector—especially in Saudi Arabia. Most current solutions are limited to laboratory settings or narrow use cases, often lacking generalizability to real-world firefighting scenarios or focusing on only certain PPE components.

To address this gap, this paper presents an AI-driven solution specifically designed for firefighting services in Saudi Arabia. Leveraging YOLOv10 and YOLOv11 models, the system detects missing PPE components from real-time visual inputs, thereby enhancing emergency preparedness and reducing the risk of injury. This work offers a novel application of deep learning in the public safety domain by providing a practical, real-time solution for monitoring PPE compliance.

The remainder of this paper is organized as follows: Section II reviews recent developments in PPE detection using deep learning. Section III describes the data collection process and research methodology. Section IV presents the experimental results and performance evaluation. Section V provides a discussion of the findings, including a comparison with other models across various domains. Finally, Section VI concludes the study and outlines directions for future work.

II. Literature Review

Deep learning has demonstrated strong potential for automating PPE compliance monitoring in high-risk environments such as firefighting, construction, and industrial operations. These techniques aim to minimize human error by enabling real-time visual detection of safety equipment. In the firefighting domain, Sesis et al. [6] applied YOLOv5 with a custom, domain-specific dataset to detect four key PPE components including helmets, gloves, masks, and insulated protective clothing. By using transfer learning, they achieved high detection accuracy, highlighting the model's suitability for emergency response scenarios.

Delhi et al. [3] applied YOLOv3 to construction sites, achieving 96% precision and recall in detecting hard hats and jackets, demonstrating its effectiveness in active work environments. Similarly, Isailovic et al. [7] benchmarked YOLOv5, MobileNetV2-SSD, and Faster R-CNN for detecting head-mounted PPE using a diverse dataset, where YOLOv5 achieved the best results. Wang et al. [4] introduced a real-world construction site dataset (CHV) and compared various YOLO models, showing YOLOv5x had the highest mAP (86.55%), while YOLOv5s demonstrated the fastest inference speed—highlighting the trade-off between accuracy and efficiency. In industrial settings, Chen and Demachi [5] combined OpenPose with YOLOv3 to monitor PPE at the Fukushima nuclear station, achieving high precision and recall for helmets and masks. In healthcare, Loey et al. [2] used YOLOv2 with ResNet-50 to detect medical masks, further illustrating the role of deep learning in public safety applications.

While previous studies demonstrate the potential of deep learning in PPE detection, they are largely confined to controlled or domain-specific environments. Therefore, this study seeks to extend his body of work and build upon these foundations by developing a real-time, AI-powered PPE compliance monitoring system specifically tailored to the operational needs of firefighters in Saudi Arabia. Comparison of studies is given in Table I.

TABLE I. COMPARISON OF STUDIES ON PPE DETECTION USING DEEP LEARNING MODELS

Study	Research Title	Focus	Algorithm	Performance	Dataset size and type
Sesis et al. [6]	AI-based Firefighting PPE Detection using YOLOv5	Detection of four Firefighting PPE (helmet, gloves, mask, and insulated protective clothing)	YOLOv5 with transfer learning	mAP@0.5 = 0.834, Precision = 0.914, Recall = 0.735	Custom dataset (342 samples), tailored to firefighting scenarios.
Delhi et al. [3]	Detection of PPE Compliance on Construction Sites Using YOLOv3	Construction site PPE (hard hats, jackets)	YOLOv3, transfer learning	Precision and recall: 96%	Self-collected dataset from video feeds, 2,509 images
Isailovic et al. [7]	The Compliance of Head-mounted Industrial PPE using Deep Learning Object Detectors	Industrial PPE, head-mounted	YOLOv5, MobileNetV2-SSD, Faster R-CNN	YOLOv5 achieved best precision (0.920), recall (0.611)	Public dataset (Roboflow, Pictor PPE), 12,682 images
Wang et al. [4]	Fast PPE Detection for Real Construction Sites Using Deep Learning Approaches	Construction site PPE (helmets, vests)	YOLOv5x (best mAP), YOLOv5s (fastest speed, 52 FPS)	YOLOv5x mAP: 86.55%, faster with YOLOv5s	Public dataset (CHV), 1,330 images, real construction site

III. DATA COLLECTION AND RESEARCH METHOD

A. Dataset Collection

High-quality image data is essential for effectively training deep learning models in firefighter PPE detection. Since no publicly available datasets met the specific requirements for detecting firefighter PPE in operational scenarios, we constructed a custom dataset tailored to this task. The initial internal dataset we collected exhibited limited variability and contained redundant samples, although the image clarity and framing were sufficient. To enhance diversity and address these limitations, we manually collected an additional 574 images from the official social media channels of the Saudi Civil Defense and other publicly available resources. All selected images adhered to the standards defined in the official firefighter PPE guidelines [1], ensuring accurate representation of required equipment.

The final dataset captures firefighters from varied angles, distances, and operational contexts to provide a broad visual representation of real-world scenarios. Further expansion of the dataset is planned to improve representation of diverse environmental conditions—particularly in low-light, crowded, or partially obstructed situations—to enhance the model's generalizability and robustness in real-world deployments.

B. Dataset Annotation

To enable accurate detection of PPE, all images in the constructed dataset were manually annotated by our team using bounding boxes to identify five critical equipment components: boots, fire suit, gloves, helmet, and SCBA. We used the Roboflow platform [8] to facilitate efficient and consistent annotation. Each bounding box precisely localized the corresponding PPE item within the image, allowing the model to recognize multiple equipment types simultaneously. Many images featured multiple equipment components, enabling the model to distinguish between them effectively. This annotated dataset provides a reliable foundation for training and evaluating detection models, ultimately supporting improved firefighter safety and system performance.

Representative annotation scenarios include a fully equipped firefighter (Fig. 1), absence of the SCBA (Fig. 2), and absence of both the SCBA and gloves (Fig. 3). These examples demonstrate the system's capacity to identify compliance gaps prior to deployment, thereby reducing risk and enhancing real-time situational awareness.

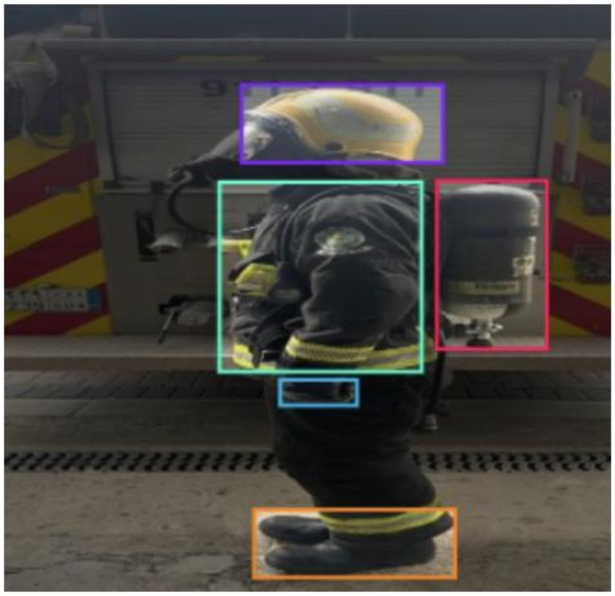

Fig. 1. A firefighter fully equipped with all necessary gear.

Fig. 2. A case of the absence of the SCBA.

Fig. 3. A case of the absence of the SCBA and gloves.

C. Dataset Augmentation

In object detection tasks, dataset quality and class balance are critical to model performance. Class imbalance—where certain PPE components are underrepresented—can lead to biased predictions. To address this, we employed data augmentation techniques to expand the dataset and mitigate class imbalance. We experimented with both generic and class-specific augmentation strategies to improve class distribution and data diversity. Despite these efforts, achieving perfect balance was challenging due to the inherent nature of the collected data. Class distributions for each dataset split are summarized below:

- Training Split: 465 images, 2,124 total objects
  - Boots: 314 (14.8%)
  - Fire Suit: 415 (19.5%)
  - Gloves: 283 (13.3%)
  - Helmet: 700 (33.0%)
  - SCBA: 412 (19.4%)
- Validation Split: 53 images, 250 total objects
  - Boots: 39 (15.6%)
  - Fire Suit: 48 (19.2%)
  - Gloves: 34 (13.6%)
  - Helmet: 79 (31.6%)
  - SCBA: 50 (20.0%)
- Test Split: 56 images, 283 total objects
  - Boots: 37 (13.1%)
  - Fire Suit: 59 (20.8%)
  - Gloves: 34 (12.0%)
  - Helmet: 91 (32.2%)
  - SCBA: 62 (21.9%)

Our augmentation pipeline improved data balance and diversity while maintaining high-quality, realistic examples. This enhanced the model's ability to generalize and detect all PPE components reliably across different scenarios.

D. YOLO Algorithms

Deep learning approaches in computer vision have yielded highly promising results across various applications. Among these, the YOLO algorithm has emerged as one of the most efficient and accurate methods for real-time object detection tasks, particularly in critical environments such as emergency response and industrial safety scenarios. YOLO operates by dividing an input image into a grid and performing detection in a single forward pass through a neural network. This design enables it to balance speed and accuracy, making it well-suited for real-time PPE inspection tasks.

YOLO has undergone several iterations over the years, with each version introducing architectural improvements that enhance performance. This study focuses on two of the latest versions—YOLOv10 and YOLOv11—selected for their advanced detection performance, computational efficiency, and applicability to safety-critical tasks [9], [10]. YOLOv10 is a recent advancement in the YOLO series, engineered to deliver high-speed detection with strong accuracy. The model is structured into three major components: a backbone network for feature extraction, a neck for generating multi-scale feature maps, and a detection head for object classification and localization. The architecture includes multiple variants designed to provide a flexible trade-off between inference speed and detection precision. In this research, YOLOv10 was adopted for its capability to reliably detect firefighting PPE—including helmets, gloves, boots, protective suits, and SCBA devices—in real-time. Additionally, transfer learning and data augmentation techniques were employed during training to improve robustness across various environmental conditions [9].

YOLOv11 represents the latest iteration in the YOLO family, incorporating state-of-the-art enhancements in feature extraction, model optimization, and real-time inference. Like its predecessor, YOLOv11 features a backbone, neck, and detection head; however, it introduces a more refined architecture with improved module design to achieve higher mAP while maintaining low latency. The model was selected in this study for its superior balance of accuracy and computational efficiency. YOLOv11 proved particularly effective in detecting multiple PPE components across a range of visual conditions, including occlusion, variable lighting, and different camera angles. These capabilities ensure reliable performance in operational environments where rapid and precise safety inspections are essential [10].

E. Training Methodology

The YOLOv10 and YOLOv11 models were trained using a custom pipeline designed to address class imbalance and enhance detection performance. Class weights were incorporated into the training configuration to ensure underrepresented categories like boots and gloves received higher attention. The training utilized the AdamW optimizer with a weight decay of 0.0005 and was conducted over 200 epochs with a batch size of 16.

A Cosine Learning Rate Scheduler was employed, starting at 0.001 and decaying to 0.000001, with a 15-epoch warmup phase. Data augmentation techniques such as Mosaic, Mixup, Copy-Paste, and geometric/colour transformations were applied to increase data diversity. Additionally, dropout (0.25) was used for regularization. To enhance accuracy and efficiency, strategies like multi-scale training, test-time augmentation (TTA), automatic mixed precision (AMP), and half-precision training were integrated. Early stopping was triggered after 150 epochs without validation improvement, and all results were logged for evaluation.

F. Training Environment

Training deep learning models for object detection requires substantial computational resources, particularly for tasks involving large datasets and complex architectures like YOLOv10 and YOLOv11. In this study, model training was initially conducted using Google Colab's free-tier environment, which offers access to an NVIDIA T4 GPU with 12 GB of VRAM. While suitable for early experimentation, the limited computational capacity resulted in slower training cycles. To overcome these limitations, the training environment was upgraded to Google Colab Pro, which provides access to more advanced GPUs such as the NVIDIA V100 and A100. This upgrade significantly improved the training process, enabling faster iterations, support for larger batch sizes, and greater stability during extended sessions. The combination of free and Pro-tier resources allowed for efficient and cost-effective training, supporting iterative development and comprehensive model evaluation.

G. Evaluation Metrics

To accurately evaluate the performance of the proposed PPE detection models for firefighters, a set of fundamental object detection metrics was utilized. These metrics were chosen to reflect the model's effectiveness in identifying and classifying multiple PPE components, including small and partially obscured items such as gloves or mask straps, which are critical in firefighting environments. The core evaluation indicators include Average Precision (AP) and mAP. These values are derived by comparing the model's predictions against the actual annotated labels in the dataset. AP is calculated as the area under the precision-recall (PR) curve for a given class, whereas mAP represents the average of AP across all detected PPE categories, as shown in Eq. (1) and Eq. (2).

$$AP = \int_0^1 p(r)\, dr \tag{1}$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{2}$$

Using the prediction results, three additional classification metrics were calculated:

- Precision, which quantifies the ratio of correctly predicted positive cases to the total predicted positives:

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

- Recall, which measures the ratio of correctly predicted positives to all actual positives:

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

- F1-Score, the harmonic mean of precision and recall:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

These metrics—Eq. (3), Eq. (4), and Eq. (5)—are critical in understanding the trade-off between missing detections and generating false alarms, particularly in safety-critical scenarios such as firefighting. Furthermore, IoU was used to evaluate the spatial accuracy of predicted bounding boxes. IoU calculates the overlap between the predicted and ground-truth boxes relative to their union, as expressed in Eq. (6):

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{6}$$

This metric is essential for assessing how well the model localizes PPE elements in complex visual scenes.

By leveraging this set of metrics, the evaluation framework ensures a comprehensive, balanced, and realistic analysis of model performance under the challenges posed by real-world firefighting environments.

IV. Results

A. Mean Average Precision and Model Size

YOLOv11 demonstrated a notable improvement in mAP compared to YOLOv10, making it more effective for detecting firefighter PPE in real-world conditions. Specifically, YOLOv11 achieved an mAP@0.5 of 64.6%, outperforming YOLOv10, which reached 58.6%, as shown in Fig. 4 and Fig. 5. This indicates YOLOv11's stronger ability to accurately detect PPE elements such as helmets, gloves, and SCBA gear. Additionally, despite YOLOv11 having a slightly larger architecture with 20 million parameters compared to YOLOv10’s 16.5 million, its efficiency in detection accuracy justifies the additional model size, especially for safety-critical applications. YOLOv11 maintained a practical balance between model complexity and performance, which positions it as a more reliable choice for deployment in environments requiring high detection accuracy and consistency.

Fig. 4. YOLOv10: mAP@0.5 and mAP@0.5:0.95 over 200 training epochs.

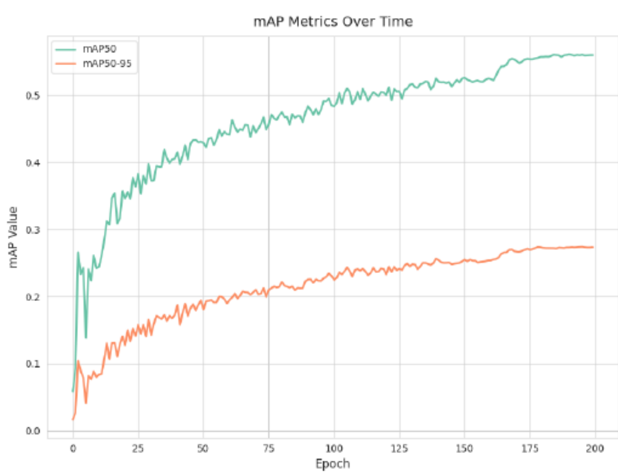

Fig. 5. YOLOv11: mAP@0.5 and mAP@0.5:0.95 over 200 training epochs.

Furthermore, YOLOv11 showed consistent improvement in detecting underrepresented PPE classes, particularly gloves and boots. This can be attributed to the refined data augmentation strategies and class-weighted training techniques, which addressed the earlier dataset imbalance. The better mAP results across all categories affirm the model's enhanced generalization ability in real-world, unstructured firefighting scenarios. YOLOv11 also demonstrated greater resilience under challenging visual conditions, such as varying lighting, partial occlusions, and dynamic backgrounds. Its performance remained stable in scenarios with overlapping gear or unconventional viewing angles, where YOLOv10 occasionally faltered. The enhanced feature extraction pipeline in YOLOv11 likely contributed to this robustness.

In practical deployment simulations, YOLOv11 also exhibited lower false negative rates, particularly in safety-critical gear detection. This reliability is crucial when monitoring firefighter readiness, as missing even one critical item (e.g., SCBA) can be life-threatening. The trade-off in model size for higher accuracy is justified in operational settings where real-time decisions depend on detection outcomes. Lastly, the precision-recall tradeoff and confidence calibration of YOLOv11 indicate that it is better suited for integration into alert systems. It provides more consistent outputs that can be used as triggers for decision-support tools or automated gatekeeping mechanisms in firefighting stations.

Collectively, these results underline that while both models are viable, YOLOv11 offers tangible advantages that support its deployment in real-world safety monitoring systems.

B. F1-Confidence, Precision-Confidence, PR Curve, and Recall-Confidence

These curves provide deeper insight into the reliability of both models across different confidence thresholds.

1) F1-Confidence: YOLOv11 consistently maintains a higher F1-score across all confidence levels. At the default threshold (0.25), it reached 0.74, compared to YOLOv10’s 0.69. This indicates a better balance between precision and recall, especially in more uncertain detections.

2) Precision-Confidence: YOLOv11 sustains over 80\% precision until the threshold drops below 0.18, while YOLOv10’s precision begins to degrade earlier, around 0.24. This shows YOLOv11’s robustness in maintaining accuracy even with lower confidence predictions.

3) PR Curve: YOLOv11 achieved a noticeably higher area under the curve (AUC = 0.59) compared to YOLOv10 (AUC = 0.51). It maintains higher precision across a broader range of recall values, indicating more stable performance as the model tries to capture all ground truth objects, as shown in Fig. 6 and Fig. 7.

4) Recall-Confidence: YOLOv11 demonstrated higher recall values across confidence thresholds, especially for harder-to-detect classes like gloves. It achieved 0.37 recall for gloves at the 0.25 threshold, compared to YOLOv10’s 0.29.

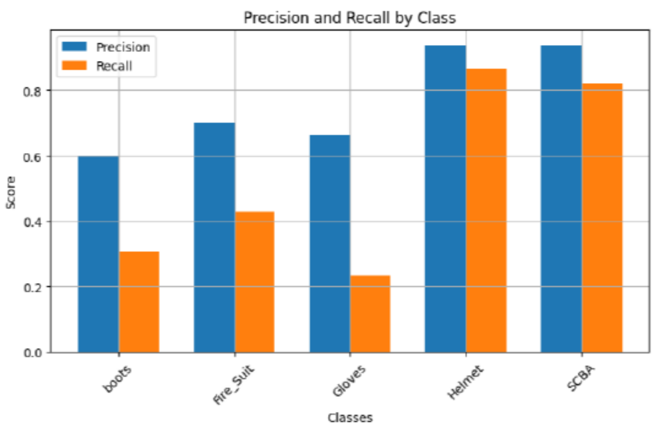

Fig. 6. YOLOv10 Precision and Recall by class.

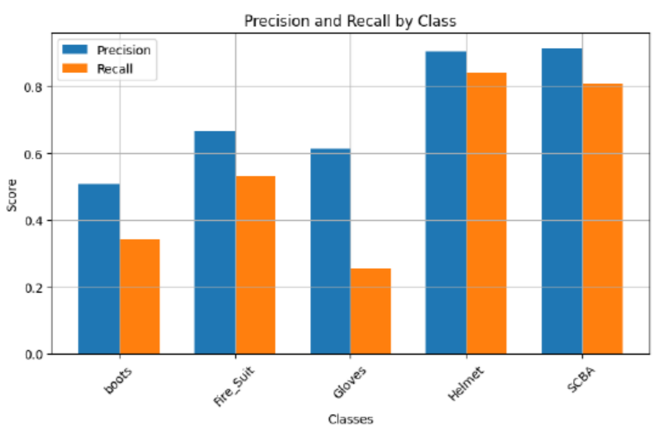

Fig. 7. YOLOv11 Precision and Recall by class.

C. IoU Performance Analysis

IoU was used to assess the spatial alignment between predicted and ground-truth bounding boxes across different PPE classes. The following breakdown highlights the performance of YOLOv10 and YOLOv11 in detecting key equipment components:

1) YOLOv10 IoU Performance:

a) Helmet: YOLOv10 achieved a mean IoU of 0.681, with 88.5% of predictions exceeding the 0.5 IoU threshold. This reflects reliable bounding box alignment for helmets, which are relatively easier to detect due to their distinct features.

b) Gloves: The mean IoU for gloves was 0.220, with only 28.1% of predictions above the 0.5 threshold. This highlights the model's struggle with accurately localizing smaller and less distinct objects like gloves.

As shown in Fig. 8, YOLOv10 demonstrated solid performance for prominent gear like helmets but had significant difficulty with smaller items such as gloves.

2) YOLOv11 IoU Performance:

a) Helmet: YOLOv11 recorded a mean IoU of 0.700, with 90% of predictions exceeding the 0.5 IoU threshold. This demonstrates improved spatial accuracy for helmet detection compared to YOLOv10.

b) Gloves: The mean IoU for gloves improved to 0.350, with better alignment accuracy. This reflects YOLOv11's enhanced capability in detecting and localizing smaller objects more precisely than YOLOv10.

Fig. 9 illustrates the improved bounding box accuracy of YOLOv11, particularly for challenging classes such as gloves, showing clear gains over its predecessor.

D. Speed

Speed is critical in real-time PPE detection systems for firefighting, where decisions must be made instantaneously. While YOLOv11 offers better accuracy, it operates slightly slower than YOLOv10, achieving 0.04 seconds per frame compared to YOLOv10's 0.033 seconds. This minor speed trade-off is acceptable given the improved detection precision and recall, which outweighs the marginal delay in processing.

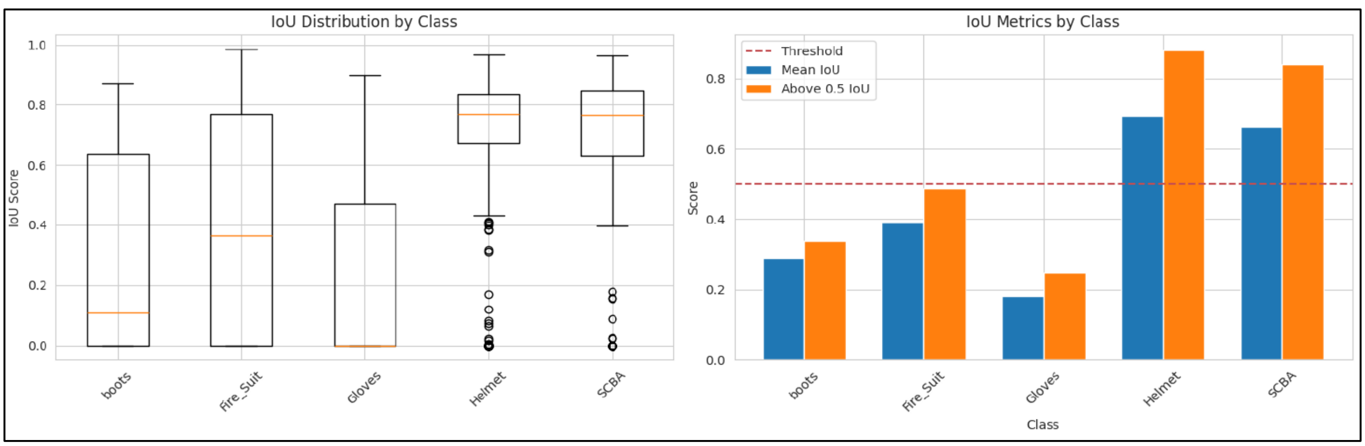

Fig. 8. IoU metrics for YOLOv10.

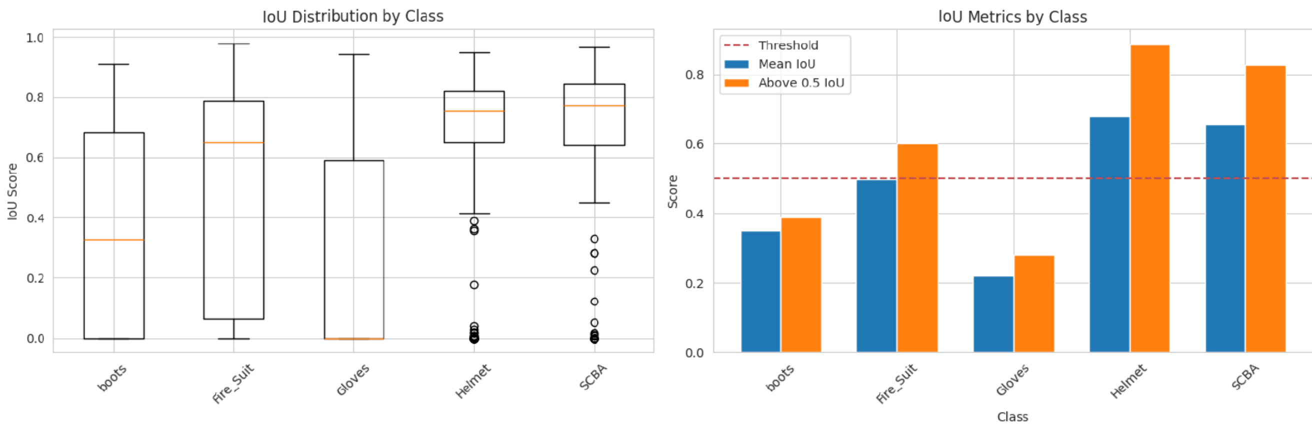

Fig. 9. IoU Metrics for YOLOv11.

V. Discussion

The results confirm that YOLOv11 significantly outperforms YOLOv10 in most evaluation metrics relevant to real-time PPE detection. This performance boost is particularly critical when working with safety-critical gear such as SCBA and gloves, which are harder to detect due to occlusion or small size. The increase in mAP, F1-score, and IoU, especially for underrepresented classes, reflects the effectiveness of class-weighted training strategies and augmentation pipelines tailored to firefighter environments. YOLOv11 also demonstrated greater consistency in handling complex backgrounds and overlapping equipment, suggesting its stronger generalization to real-world firefighting scenarios.

The detailed analysis reveals that while both models provide valuable performance, YOLOv11's architectural enhancements lead to improved detection of small and occluded PPE items. The incorporation of advanced augmentation and training strategies further boosted model robustness. However, increasing model complexity slightly impacts inference speed, a common trade-off in deep learning applications. As shown in Fig. 10 and Fig. 11, YOLOv11 was able to correctly recognize all critical equipment components, including the fire suit, gloves, boots, and helmet—demonstrating its superior reliability in safety-critical scenarios. In contrast, YOLOv10 failed to detect the fire suit, which is crucial for ensuring firefighter readiness and personal safety in hazardous conditions.

Moreover, the visual results affirm that YOLOv11 can identify all PPE components with high accuracy, making it suitable for integration into real-time alert systems or automated PPE compliance gates. This could reduce manual inspection errors and accelerate firefighter readiness verification.

Despite its slightly lower speed, YOLOv11's advantages in accuracy, localization, and class coverage outweigh the performance trade-off, making it more practical for deployment in Saudi Arabia's civil defense operations, where precision and speed are both mission-critical.

The proposed system, based on YOLOv10 and YOLOv11, is specifically optimized for image-based data in the visual modality, particularly RGB images captured in real-world firefighting environments. The models are best suited for high-resolution still images or video frames where PPE components are visible under variable lighting, motion blur, partial occlusions, or cluttered backgrounds. Due to their architecture, YOLO models excel in object detection tasks on spatial data with well-defined bounding box annotations. The system performs reliably on moderate-sized datasets (e.g., ~600–1,000 images), particularly when enhanced through data augmentation to account for class imbalance and variability in pose, angle, and context.

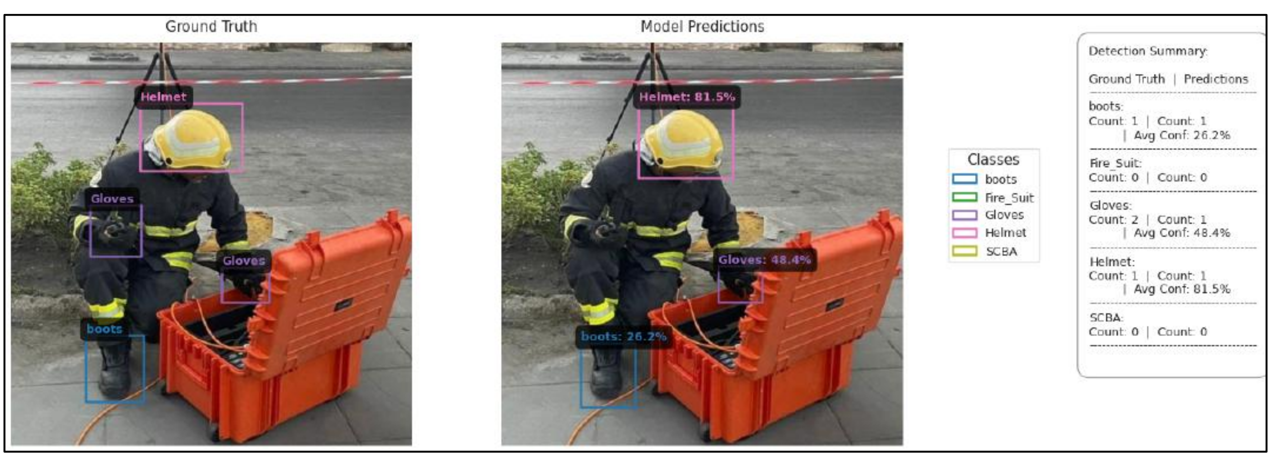

Fig. 10. YOLOv10 detection results: The fire suit was missed in the predictions.

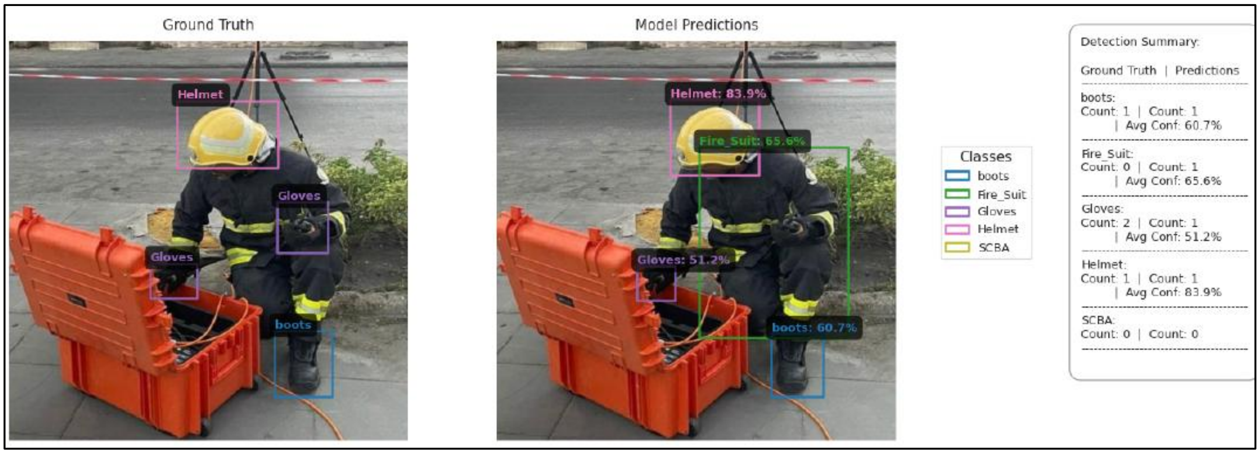

Fig. 11. YOLOv11 detection results: All PPE components including the fire suit were correctly detected.

However, the models are not directly applicable to non-visual modalities such as audio or text, and their performance may degrade with extremely low-resolution images, thermal images, or infrared unless retrained on domain-specific datasets. Additionally, performance is affected by data distribution. The model achieves its best results when trained on datasets that reflect the same environmental and cultural conditions (e.g., firefighter uniforms and gear used in Saudi Arabia) as the deployment context. This emphasizes the importance of domain-specific data for generalization and accuracy.

This study offers several key advantages over existing deep learning-based PPE detection models across construction, industrial, and healthcare domains:

1) Domain-specific adaptation to firefighting scenarios: Unlike previous research primarily focused on industrial or construction environments [3][4][7], the current work is specifically tailored to the unique challenges of firefighting, including variable lighting, smoke, occlusions, and high-stress environments. The model was trained on a custom dataset collected from real-world footage of Saudi Civil Defense operations, offering higher ecological validity than laboratory-based datasets.

2) Comprehensive multi-class PPE detection: Prior studies often target a limited subset of PPE items—typically one to four items, such as helmets or vests [4], [5], [6]. In contrast, our model simultaneously detects five critical PPE components: helmet, fire suit, gloves, boots, and SCBA. This multi-label detection approach ensures full compliance verification and strengthens overall situationalreadiness.

3) Improved detection for small and occluded items: YOLOv11 achieved significantly higher mAP, F1-score, and IoU, especially for underrepresented and smaller items like gloves and SCBA units, compared to YOLOv10 and prior benchmarks such as MobileNetV2-SSD and Faster R-CNN [7]. Our class-weighted training and targeted augmentation pipeline led to a 59% improvement in glove recall over YOLOv10.

4) Real-time monitoring capability with alert integration: While existing models may offer high accuracy, many lack real-time alerting or decision-support integration. Our proposed solution maintains an inference speed under 0.04 seconds per frame, enabling real-time detection and compatibility with automated gatekeeping or alarm systems at fire stations.

5) Environmental hazard awareness and long-term safety enhancements: Beyond immediate PPE detection, this study acknowledges long-term firefighter health risks, including exposure to harmful substances like PFAS (per- and polyfluoroalkyl substances), which are associated with cancer and immune system disorders. Our future vision includes integrating environmental hazard awareness to create a more holistic safety system that enhances both short-term operations and long-term health outcomes.

6) First deployment-ready system for Saudi civil defense: This work constitutes the first firefighter-specific PPE compliance monitoring system designed for potential deployment within Saudi Arabia. It establishes a foundation for future research and development in regional safety automation and may serve as a replicable framework for other high-risk sectors.

VI. Conclusion and Future Work

This study represents a notable advancement in enhancing firefighter safety by developing an AI-based system for PPE compliance. As the first implementation of its kind in Saudi Arabia, the proposed model integrates deep learning techniques with real-time computer vision to detect missing or improperly worn protective gear with high precision. This capability enables timely intervention in hazardous environments, potentially preventing serious injuries or fatalities.

By leveraging advanced YOLO architectures trained on a domain-specific dataset, the system demonstrates robust performance and reliability, setting a new benchmark for automated safety enforcement in high-risk operational contexts. The findings underscore the feasibility and effectiveness of applying artificial intelligence to critical occupational safety challenges, particularly in sectors where real-time monitoring and rapid decision-making are essential.

Beyond its technical contributions, this research illustrates the broader implications of AI-driven automation in promoting safer work environments. It highlights how intelligent systems can augment human oversight and institutional safety protocols, contributing to a more proactive and data-informed approach to risk management. As such, the study provides a foundation for future exploration of similar technologies across other high-risk domains, including industrial operations, emergency response, and construction, where adherence to safety standards is equally vital.

- Future enhancements will focus on expanding dataset diversity through advanced augmentation methods and synthetic data generation, particularly to improve detection of underrepresented PPE classes such as gloves and boots. Model performance can be further optimized by leveraging transfer learning with pre-trained networks and incorporating adaptive training techniques for advanced detection use cases. Additionally, optimizing the system for deployment on edge devices by adopting lightweight architectures will facilitate broader scalability and real-time performance in operational settings.
- To extend system functionality, an automated timing feature will be implemented to monitor the duration firefighters take to fully equip themselves, providing valuable insights into response readiness. Finally, the system will be integrated with safety management platforms and wearable technologies to support comprehensive, real-time monitoring and decision-making in critical firefighting environments.

REFERENCES

[1] General Directorate of Civil Defense, Saudi Arabia, "Personal Protective Equipment Guide," https://998.gov.sa, 2017, accessed: 2024-10-14.

[2] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "Fighting against covid-19: A novel deep learning model based on yolo-v2 with resnet-50 for medical face mask detection," Sustainable Cities and Society, vol. 65, p. 102600, 2020.

[3] V. S. K. Delhi, R. Sankarlal, and A. Thomas, "Detection of personal protective equipment (ppe) compliance on construction site using computer vision based deep learning techniques," Frontiers in Built Environment, vol. 6, p. 136, 2020.

[4] Z. Wang, Y. Wu, L. Yang, A. Thirunavukarasu, C. Evison, and Y. Zhao, "Fast personal protective equipment detection for real construction sites using deep learning approaches," Sensors, vol. 21, no. 10, p. 3478, 2021.

[5] S. Chen and K. Demachi, "A vision-based approach for ensuring proper use of personal protective equipment (ppe) in decommissioning of fukushima daiichi nuclear power station," Applied Sciences, vol. 10, no. 15, p. 5129, 2020.

[6] A. Sesis, I. Siniosoglou, Y. Spyridis, G. Efstathopoulos, T. Lagkas, V. Argyriou, and P. Sarigiannidis, "A robust deep learning architecture for firefighter ppes detection," IEEE Transactions on Industrial Informatics, 2022.

[7] V. Isailovic, A. Peulic, M. Djapan, M. Savkovic, and A. M. Vukicevic, "The compliance of head-mounted industrial ppe by using deep learning object detectors," Scientific Reports, vol. 12, no. 1, p. 16347, 2022.

[8] "Roboflow – Documentation," https://docs.roboflow.com/, 2024, accessed: 2025-01-27.

[9] "Ultralytics YOLOv10 Documentation," https://docs.ultralytics.com/models/yolov10/, 2024, accessed: 2025-01-27.

[10] "Ultralytics YOLOv11 Documentation," https://docs.ultralytics.com/models/yolo11/, 2024, accessed: 2025-01-27.