Utilizing YOLOv11 for Real-Time Construction PPE Compliance Detection

Ghatfan Emery Razan
Department of Information Technology, Faculty of Computer Science, Universitas Brawijaya, Indonesia
ghatfaneryrazan@student.ub.ac.id

Diva Kurnianingtyas
Department of Informatics Engineering, Faculty of Computer Science, Universitas Brawijaya, Indonesia
divaku@ub.ac.id (corresponding author)

Mirza Hilmi Shodiq
Department of Information Technology, Faculty of Computer Science, Universitas Brawijaya, Indonesia
exquisitemirza@student.ub.ac.id

Simon Fernandes Martua Raja Pandopotan Sitompul
Department of Information Technology, Faculty of Computer Science, Universitas Brawijaya, Indonesia
simonsitompul25@student.ub.ac.id

Azarya Stefanus Lopulalan
Department of Information Technology, Faculty of Computer Science, Universitas Brawijaya, Indonesia
azaryalopulalan@student.ub.ac.id

Kohei Arai
Information Science Department, Saga University, Saga, Indonesia
arai@cc.saga-u.ac.jp

Received: 27 October 2025 | Revised: 5 January 2026 and 4 February 2026 | Accepted: 7 February 2026

 DOI: https://doi.org/10.48084/etasr.15801

ABSTRACT

Personal Protective Equipment (PPE) adherence monitoring remains a persistent challenge in the construction sector. Human oversight has proven ineffective in complex or expansive settings, underscoring the importance of intelligent automation. This study aims to present an object-detection system for real-time PPE monitoring utilizing YOLOv11, the latest version of the You Only Look Once (YOLO) model series. The system categorized compliance and non-compliance into 15 classes, including cover classes and their counterparts. The model was trained on 717 annotated images and evaluated using standard object detection benchmarks, including precision, recall, and mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 0.5 (mAP@0.5). Several data augmentation strategies were used during training to strengthen generalization. During the process, the total precision and mAP@0.5 were 0.810, with the strongest performance in detecting the mask and hardhat categories. Additional testing was conducted on visually degraded images, including blurred, grayscale, and Out-Of-Distribution (OOD) areas, to further assess robustness. Although strong performance was maintained under standard conditions, decreased confidence was acquired when detecting smaller, ambiguous objects. Overall, YOLOv11 performed well for real-time PPE detection under dynamic site conditions, but factors such as class imbalance and susceptibility to visual noise indicated the need for further refinement. Optimizing powerful yet lightweight model variants tailored for edge deployment, as well as sophisticated augmentation strategies and robust domain-adaptation frameworks, should be prioritized in future research to improve real-world applicability.

Keywords-YOLOv11; industry problems; construction site safety; computer vision; industrial PPE detection

I. INTRODUCTION

The construction sector is consistently ranked among the highest in terms of occupational risks and injuries. Based on the study published by the International Labor Organization (ILO), about 2.78 million individuals die because of occupational injuries and diseases annually, of which a considerable percentage is connected to the construction industry [1]. The use of hard hats, safety vests, and other PPE dramatically reduces the likelihood of these incidents. Despite strict industry regulations intended to mitigate workplace risks, compliance with PPE requirements is worryingly low. Examining the issues more deeply underscores an ever-growing need for attention. Recent studies indicate that negligence in PPE use accounts for nearly 70% of all reported construction-related fall accidents [2]. The traditional practice of relying on supervisors to manually ensure compliance with safety protocols has proven ineffective on large or remote construction sites due to limited manpower or oversight.

Many current real-time monitoring systems operate reactively, implementing safety measures after an incident rather than focusing on accident prevention. These observational gaps create an urgent need for intelligent, automated systems that can proactively monitor PPE compliance across the diverse and variable conditions found on work sites [3]. The gaps in the application of deep learning and computer vision technologies have been addressed through various studies. Models such as YOLOv3 and Faster R-CNN have achieved promising results in PPE detection under controlled conditions [4], but the performance declines significantly in real-world environments with poor lighting, obstructions, or low-resolution imagery. A further limitation is that most frameworks can detect only one or two object classes, which limits their suitability for comprehensive safety compliance. This lack of toughness and class diversity constitutes a critical research gap, and failing to address it can lead to preventable injuries. Previous deep learning studies using models such as YOLOv3 and YOLOv5 for PPE detection are constrained. Most of these models are limited to identifying one or two object types and are occasionally tested for real-time performance in cluttered environments. Moreover, limited attention has been given to multi-class compliance by distinguishing proper from improper PPE usage, particularly under adverse visual conditions such as blur or grayscale distortion. The process creates a significant research gap, underscoring the need for a robust system capable of accurate, real-time detection across diverse environments.

This study develops a YOLOv11-based real-time detection system for 15 classes of PPE to address the gap, comprising both compliant and non-compliant categories. The objectives are to analyze performance across standard and complex scenarios, assess robustness to distortions, and determine optimization thresholds. This study is expected to advance theory in multi-class safety detection and to contribute a practical, scalable Artificial Intelligence (AI) surveillance solution to improve safety in high-risk sectors, such as construction.

II. BACKGROUND

The development of intelligent safety monitoring systems is closely tied to advances in artificial neural networks, deep learning, and object detection technologies. These systems form the foundation for reliable, real-time compliance monitoring in high-risk domains, such as construction. Deep learning originates in Artificial Neural Networks (ANNs), inspired by the human brain, which consists of interconnected, weighted nodes (neurons) arranged in layers. Common architectures in deep learning include the Multilayer Perceptron (MLP), trained through supervised learning, where connection weights are iteratively adjusted. Although simple ANNs can handle basic recognition, complex pattern recognition requires deeper models with multiple hidden layers to extract features hierarchically [5, 6]. Deep learning, a branch of machine learning, uses multi-layered neural networks for automated feature extraction and complex decision-making, eliminating the need for manual feature engineering. This capability is crucial for visual recognition, such as detecting safety equipment, where subtle details are critical. Convolutional Neural Networks (CNNs) are foundational to deep learning for image analysis because they effectively capture spatial hierarchies. Moreover, this field has experienced rapid growth driven by large-scale datasets, high-performance hardware, including GPUs and TPUs, and advances in algorithms [7, 8].

Deep learning plays a crucial role in visual perception systems, and object detection, a core application of deep learning, classifies and localizes objects using bounding boxes. Compared to earlier methods such as Haar Cascades and SIFT, which rely on less accurate handcrafted features, modern frameworks, including Faster R-CNN, Single Shot Detectors (SSDs), and YOLO, offer superior accuracy through end-to-end processing, enabling real-time applications in industrial safety [9, 10]. Among these methods, YOLO achieves exceptional efficiency by analyzing the entire image in a single forward pass, partitioning it into a grid to predict bounding boxes and class probabilities simultaneously, which supports high inference speed and global context awareness.

Recent versions, including YOLOv5 and YOLOv11, further improve performance in cluttered backgrounds by incorporating attention layers, stronger backbones, and powerful data augmentation [11, 12]. The YOLO framework's flexibility has been demonstrated across diverse domains, including unconventional ones. For example, in environmental monitoring, a study by [13] found that while YOLOv6 achieved higher mAP and faster inference for detecting plastic waste, YOLOv7 provided greater robustness against visual noise, making it more reliable in cluttered environments, despite a lower mAP of 0.512. In the healthcare sector, another study [14] showed that both YOLOv8 and MobileNet-v2 achieved perfect scores of 1.0 for non-invasive dehydration analysis, and that YOLOv8 showed superior real-time classification without errors, whereas MobileNet-v2 experienced some misclassifications.

The evolutionary balancing of speed, accuracy, and robustness has driven architectural innovations, leading to YOLOv11, which is utilized in this study. The YOLOv11 architecture in this study offers significant improvements over

earlier versions. The Cross-stage Feature Fusion (C2f) module improves context understanding, the Spatial Pyramid Feature Fusion (SPFF) improves small-object detection, and the Bidirectional Feature Pyramid Network (BiFPN) enables adaptive feature fusion. These modifications improve performance in detecting objects of various sizes and those that are complex or sealed, making YOLOv11 well-suited for challenging environments such as construction sites. These technological developments strike a crucial balance between inference speed and accuracy, making the technology ideal for advanced, intelligent PPE monitoring systems, particularly in high-risk areas that require immediate response. Regarding construction site safety, YOLO-based and other deep learning frameworks are foundational to the development of automated detection systems. By enabling real-time identification of safety breaches, these frameworks enable timely preventive actions. The growing availability of annotated PPE datasets and live surveillance feeds empowers solutions to bridge the gap between established safety policies and on-site implementation.

III. METHODOLOGY

A deep learning method using the YOLOv11 object detection framework (specifically YOLOv11-n) was implemented to build a real-time PPE compliance monitoring system, given the construction site's numerous challenges. Therefore, the workflow was modified to be resilient to these conditions, as shown in Figure 1. This study used a deep learning method based on the YOLOv11-n object detection architecture to develop a real-time PPE compliance monitoring system. The workflow was designed to ensure robustness under diverse environmental conditions common in construction sites.

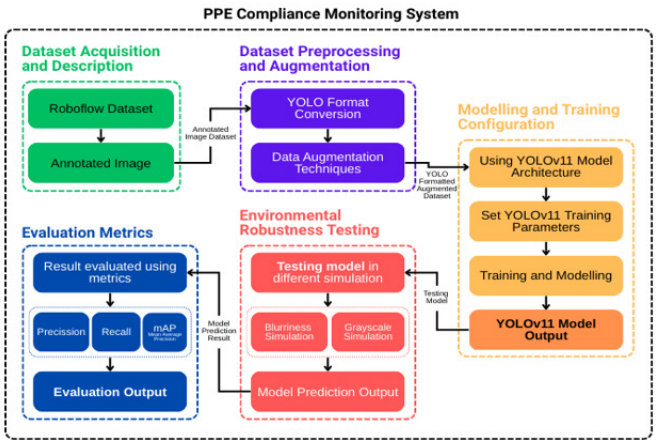

Fig. 1. Methodology followed in this study.

A. Dataset Acquisition and Description

Following the data collection protocols established in the reference study using YOLOv11-s [15], the model developed in this study used the YOLOv11-n architecture and the dataset from the standard data source, Roboflow. As the reference study operated on a dataset of 2,092 images for PPE detection, this study uses a similar dataset with an enlarged scope by using the PPE Dataset from Roboflow, which is licensed under CC by 4.0 [16], that has a total of 4,196 images comprising PPE, heavy machinery, and hazardous construction zones. The images were captured from multiple perspectives, including horizontal viewpoints and surveillance camera angles, to ensure diversity and representativeness in the training data. The dataset features supervised annotations across 15 classes, ranging from compliance indicators (e.g., hardhat, excavator, and safety boots) to noncompliance indicators (e.g., No hardhat and No safety boots).

B. Dataset Preprocessing and Augmentation

The images were converted to YOLO format, including bounding boxes and class labels. The dataset was randomly split into training, validation, and test sets at 70%, 20%, and 10%, respectively. Data augmentation methods, including flipping, rotation, brightness adjustment, and scaling, were applied to improve generalization. Augmentation was prioritized for underrepresented classes to address class imbalance and improve model robustness.

C. Model Architecture and Training Configuration

Training was performed in the YOLOv11-n framework, where images were resized to a square 640×640 pixels and processed in batches of 16 across a 100-epoch run. The YOLO configuration file was used for multi-class mappings, along with other model hyperparameters. A cosine learning rate was used, initialized at 0.01 to govern the optimization trajectory, and Stochastic Gradient Descent (SGD) with momentum then drove the parameter updates.

D. Evaluation Metrics

The model performance was evaluated using the conventional object-detection metrics of precision, recall, and mAP at an IoU threshold of 0.5. Each metric represents a different aspect of detection performance and class-level prediction capability. The mAP was computed as the mean of the per-class precision values derived from the precision-recall curves.

E. Environmental Robustness Testing

Studies often discussed robustness, but the supporting evidence was limited to controlled lab lighting conditions. In practice, the team generated alternative data slices to adapt the architecture to real-world lighting conditions. One slice, simulating common challenges in daily photography, applied a progressive Gaussian blur to the test images, gradually eroding edges and forcing the model to balance motion artifacts against fine detail. A second slice converted all images to grayscale, removing color information and revealing hidden dependencies on hue, testing whether shape alone could support accurate predictions. A final batch introduced unconventional orientations, occlusions, and extreme angles relating to those captured by the street camera after heavy rain. These perspective distortions required the model to adapt, reflecting the unpredictable viewpoints encountered in field deployments. Concurrently, corner-case injections directly address a limitation noted in recent reviews: strong algorithms frequently fail under predictable, real-world conditions.

IV. RESULTS AND DISCUSSION

A. Confusion Matrix Analysis

The confusion matrix in Figure 2 showed a clear performance disparity across the 15 object categories. The diagonal entries indicated exceptional accuracy for classes such as earmuffs, with a perfect true-positive score of 1.00, followed closely by strapped, overall suit, and no safety boots at 0.93, 0.90, and 0.88, respectively. The results demonstrated the model's robust ability to detect larger or more distinctively shaped safety features. However, classes such as no earmuff and no overall suit showed significant performance degradation, with true-positive rates dropping to 0.05 and 0.50, respectively. The matrix revealed a critical issue with background confusion, particularly for the no earmuff class, where 93% of instances were incorrectly predicted as background. This pattern of high false negatives suggested that the detector frequently missed subtle or small-scale features, a known challenge in multi-class detection pipelines [17]. Despite these specific weaknesses, the majority of critical safety classes met actionable detection standards, confirming the viability of the YOLOv11 framework for baseline monitoring. The model also indicated the need for targeted data augmentation for underperforming minority classes.

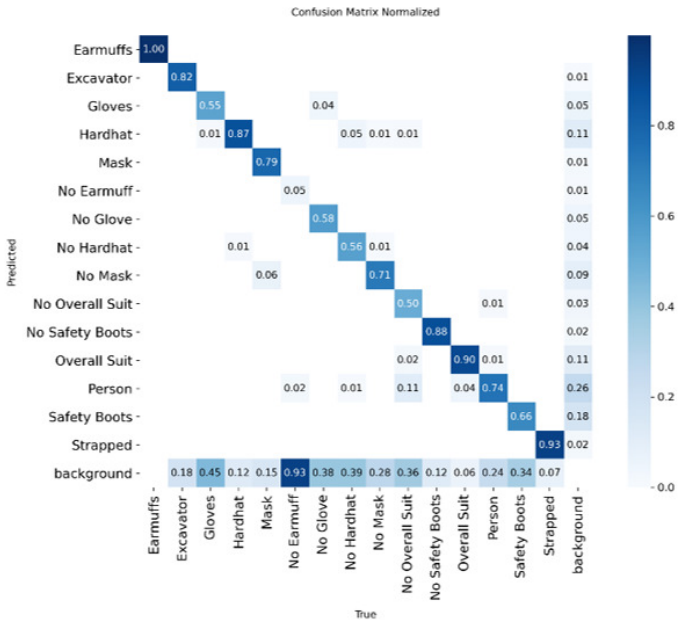

Fig. 2. Confusion matrix.

The confusion matrix in Figure 2 further quantified the model's classification performance by providing a direct breakdown of correct and incorrect predictions. The exceptional diagonal values for earmuffs and strapped, at 1.00 and 0.93, confirmed that the algorithm had effectively learned the distinct visual features of these safety items. Consequently, the matrix showed a severe localization failure for the no-earmuff condition, recording only 0.05 true positives and misclassifying 93% of the time as background. A similar trend was observed for gloves, with 45% background confusion.

B. Confidence-Based Precision and Recall Evaluation

Figures 3-5 show the relationship between the model confidence scores and performance on precision and recall metrics. The precision-confidence curve (Figure 3) indicates that the earmuffs and overall suit classes maintained near-perfect precision across thresholds. No earmuff class showed inconsistent behavior or instability, indicating difficulty establishing reliable detection confidence. Complementing this outcome, the recall-confidence curve (Figure 4) shows that classes such as earmuffs and excavators achieved high recall even at varying confidence levels. However, recall for no earmuff was critically low, starting below 0.4 and worsening as confidence increased. These results indicate that the model's visual features were not reliably learned for distinct objects, such as a hardhat or a strapped.

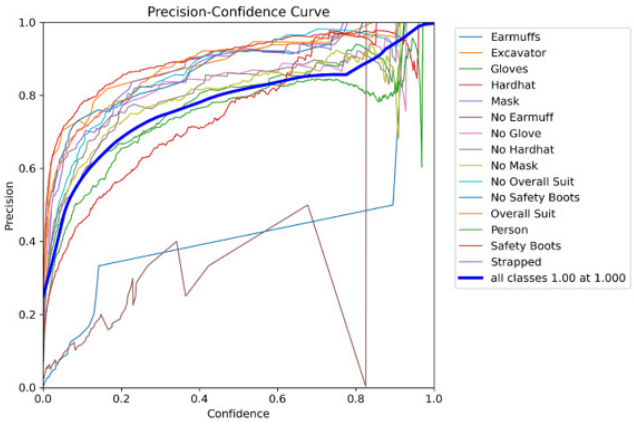

Fig. 3. YOLOv11 Precision-Confidence curve.

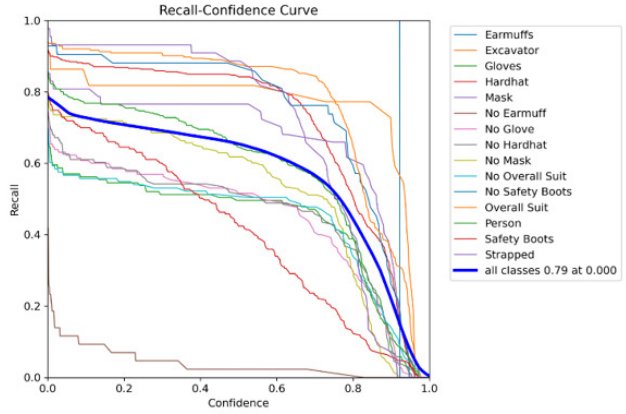

Fig. 4. YOLOv11 Recall-Confidence curve.

C. Precision-Recall Curve and mAP

The precision-recall curves shown in Figure 5 served as a definitive report card for the detection model. When the plots clustered near the upper right corner, as observed for earmuffs, strapped, and overall suit, with equal values of 0.995, 0.918, and 0.911, it showed both high accuracy and strong recall. The situation was less favorable for no earmuffs and gloves, with 0.055 and 0.549, respectively, and the curves remained

reduced, suggesting that the training samples were possibly insufficient or that the features were subtle enough to be indistinguishable from the background. A mAP at IoU 0.5 of 0.713 (thick blur trace) followed the performance reported for YOLOv5 and YOLOv7 in recent factory-focused studies [18], indicating that YOLOv11 could match or exceed effectiveness.

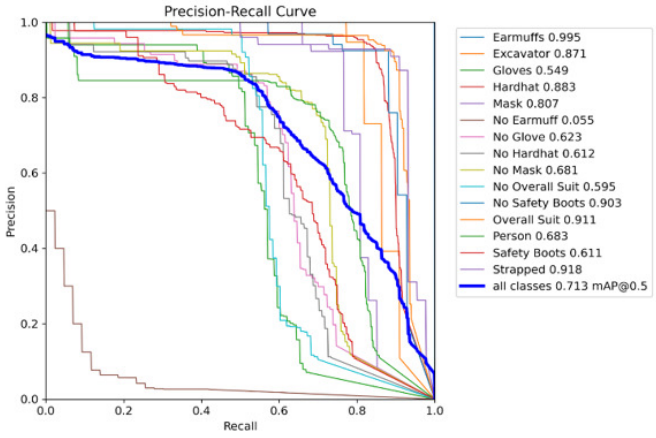

Fig. 5. YOLOv11 Precision-Recall curve.

D. Quantitative Summary of Detection Metrics

Table I presents a snapshot of YOLOv11's overall output, with global precision near 0.91, global recall around 0.67, and mAP at IoU 0.50 of 0.810. The balance between accuracy and practical coverage indicated the framework's effectiveness in real-world, uncontrolled video environments. Examining the per-class numbers revealed significant blind spots in the analysis. Recall for no mask was only 0.40, and the vehicle reached 0.45, indicating that subtle cues were often missed. Consequently, any large-scale deployment required a safer, better-balanced dataset to address these deficiencies. Comparing mAP with recent papers showed a clear dip, a difference that was easily attributed to the smaller training corpus. [19] had explained the role of data volume in shaping model generalization, and these results validated that point. The architectural enhancements in the YOLOv11-C2f modules, the SPFF layer, and the embedded blocks pushed performance beyond previous YOLO iterations on the same limited dataset, regardless of the process.

E. Class-Wise Evaluation Summary

As shown in Table I, the model achieved strong performance across most classes, with global precision of approximately 0.91, recall of nearly 0.67, and mAP@0.5 of 0.810. These values indicated that YOLOv11 achieved a balance between accuracy and detection coverage in real-world applications.

F. Comparative Analysis with the Baseline Study

The proposed YOLOv11-n architecture was benchmarked against the YOLOv11-s model used by [15] to evaluate trade-offs between computational capacity and detection performance (Table II). While the reference study used the larger YOLOv11-s architecture, which benefits from deeper feature extraction layers and achieves a superior mAP@0.5 of 0.810 and a precision of approximately 0.91, this performance advantage must be contextualized by the significant disparity in task complexity. The reference model was limited to detecting a restricted set of broad categories (e.g., humans, helmets, high-visibility vests, gloves, and boots), whereas this study expanded the scope to 15 distinct classes, necessitating the resolution of fine-grained distinctions for small-scale objects such as earmuffs and strapped. Critically, despite deploying the significantly lighter YOLOv11-n (nano) architecture optimized for edge efficiency, the proposed model achieved a global recall of 0.669, effectively matching the reference study's approximately 0.67. This indicates that with robust hyperparameter tuning, the nano model is as effective as the heavier small variant at locating hazards, even though it yields slightly lower precision (0.799) due to increased classification difficulty. Consequently, this study demonstrates that YOLOv11-n offers a scientifically viable trade-off, prioritizing computational efficiency and granular coverage over raw precision for resource-constrained environments.

TABLE I. MODEL EVALUATION METRICS

Class	Precision	Recall	mAP@0.5
All	0.799	0.669	0.713
Earmuffs	0.399	1	0.995
Excavator	0.893	0.818	0.871
Gloves	0.814	0.508	0.549
Hardhat	0.92	0.85	0.883
Mask	0.911	0.766	0.807
No earmuff	0.344	0.023	0.055
No glove	0.854	0.531	0.623
No hardhat	0.854	0.538	0.612
No mask	0.819	0.635	0.681
No overall suit	0.921	0.513	0.595
No safety boots	0.907	0.881	0.903
Overall suit	0.922	0.89	0.911
Person	0.781	0.685	0.683
Safety boots	0.738	0.488	0.611
Strapped	0.909	0.909	0.918

TABLE II. PERFORMANCE COMPARISON WITH THE REFERENCE STUDY

Metric	This study (proposed)	Reference study [15]
Model architecture	YOLOv11-n (nano)	YOLOv11-s (small)
Dataset size	4,196 images	2,092 images
Object classes	15 classes	5 classes
Global precision	0.799	~0.91
Global recall	~0.669	~0.67
mAP @0.5	0.713	0.81

G. In-Dataset Evaluation

The model showed strong baseline performance on in-distribution test images (Figure 6), with the excavator achieving 0.92 and the overall suit achieving 0.78, even at night. On the other hand, detection for Person reduced from 0.81 to 0.63. This followed known CNN limitations, in which spatial information was lost in deeper layers [20]. Misclassifications on occluded or non-standard items highlighted challenges with edge cases.

Fig. 6. Result of dataset evaluation.

H. Out-of-Distribution Evaluation

On out-of-distribution data, the model generalized well across distinct classes, such as an excavator, achieving 0.95. Precision for smaller PPE was highly unstable, while visible hardhat instances reached 0.91 (Figure 7). Ambiguous items, such as masks and gloves, achieved critically low confidence scores of 0.30 and 0.36, respectively, indicating significant limitations in contextual interpretation [21].

Fig. 7. Result of out-of-dataset evaluation.

I. Robustness Testing – Blurriness and Grayscale

The YOLOv11 system was subjected to progressive blurring and grayscale conversion tests, simulating common suboptimal conditions encountered on construction sites to evaluate its real-world robustness [22]. The results in Figures 8–12 showed significant variations in performance across these scenarios.

1) Non-Blurred Scenario

Under optimal conditions, the model demonstrates robust baseline detection, achieving high confidence scores for distinct PPE classes, specifically 0.86 for hardhat and 0.89 for person (Figure 8). The system effectively handles multi-person scenes with minimal false positives, validating the YOLOv11s architecture's ability to extract sharp features when edge definitions are clear. The results are also consistent with YOLOv5 and YOLOv7 benchmarks [23].

Fig. 8. Result of the non-blurries scenario.

2) Low Blurriness Scenario

A 20% blurring level causes a negligible decrease in performance. While confidence scores for smaller objects regressed slightly, such as the hardhat detections dipping marginally from 0.86 to 0.82, the model still produced valid bounding boxes for all primary subjects (Figure 9). This confirms the network's resilience to mild signal degradation, retaining recall effectively despite minor loss of high-frequency detail [24].

3) Mid Blurriness Scenario

At 50% blur, the model's performance exhibits a critical inflection point. Confidence scores for person plummeted from 0.80 to 0.38, and smaller objects like gloves were completely lost (Figure 10). Localization stability degraded significantly, with bounding boxes becoming loose or failing to encompass the entire object, signaling a breakdown in feature extraction for non-rigid bodies [24].

Fig. 9. Result of the low-blurries scenario.

Fig. 10. Result of mid-blurries scenario.

4) *High Blurriness Scenario*

High blur by 70% induced catastrophic failure across all minority classes. The model failed to detect hardhats and safety vests entirely in multiple frames, and person detection confidence collapsed to 0.30, often approaching the rejection threshold. The total loss of texture information left the model blind to small-scale PPE, resulting in near-zero recall for safety compliance items [24].

5) *Grayscale Scenario*

The grayscale evaluation shows a strong reliance on color features for certain classes. While person detection remained relatively stable, with a confidence level approaching 0.79, color-dependent classes suffered. Next, hardhat confidence dropped from 0.86 to 0.76, and discriminative classes, such as safety vests, which are typically high-visibility orange or yellow, showed reduced separation from the background. This confirms that while shape features are sufficient for general object detection, chromatic information is essential for high-confidence PPE classification [25].

Fig. 11. Result of high-blurries scenario.

Fig. 12. Result of grayscale scenario.

V. CONCLUSION

In conclusion, this study introduced a real-time PPE monitoring framework based on the YOLOv11 architecture, achieving a mean Average Precision (mAP) of approximately 0.810 at an Intersection over Union (IoU) threshold of 0.5, with precision consistently exceeding 0.9 across major classes. A

significant innovation was its multi-axis compliance grading, which included categories such as No Hardhat to provide specific, contextual feedback beyond simple presence/absence detection. Despite promising performance and strong generalization in stress tests, the model showed limitations. Robustness was reduced in cluttered scenes and when processing blurred or grayscale images. The model was prone to occasional false positives, and class imbalance adversely affected recall and mAP. As the system could significantly improve occupational safety by automating compliance checks, future work should focus on compiling larger datasets or developing domain-adaptation methods. The study directions included deploying lightweight model variants on edge devices and investigating the incorporation of ensemble or transformer architectures to improve reliability in real-world field conditions.

ACKNOWLEDGMENT

This study is supported by funding assistance for publishing reputable international scientific journals for lecturers by the Faculty of Computer Science, Universitas Brawijaya (Decision No: 03917/UN10.F1501/B/KU/2025).

REFERENCES

[1] *COVID-19 and the world of work. Updated estimates and analysis*. ILO Monitor, 2020.

[2] R. Sehsah, A.-H. El-Gilany, and A. M. Ibrahim, "Personal protective equipment (PPE) use and its relation to accidents among construction workers," *La Medicina del Lavoro*, vol. 111, no. 4, pp. 285–295, 2020, https://doi.org/10.23749/mdl.v111i4.9398.

[3] P. R. C. Abordo *et al.*, "Smart surveillance system using ESP32 and camera-based motion detection with IM technology," *International Journal of Research Studies in Educational Technology*, vol. 8, no. 2, pp. 63-74, July 2024, https://doi.org/10.5861/ijrset.2024.8012.

[4] G. L. Hung, M. S. B. Sahimi, H. Samma, T. A. Almohamad, and B. Lahasan, "Faster R-CNN Deep Learning Model for Pedestrian Detection from Drone Images," *SN Computer Science*, vol. 1, no. 2, Apr. 2020, Art. no. 116, https://doi.org/10.1007/s42979-020-00125-y.

[5] J. H. Yousif and M. J. Yousif, "Critical Review of Neural Network Generations and Models Design." Computer Science and Mathematics, Nov. 09, 2023, https://doi.org/10.20944/preprints202309.1149.v2.

[6] J. Zhang and N. Bai, "Augmentation Embedded Deep Convolutional Neural Network for Predominant Instrument Recognition," *Applied Sciences*, vol. 13, no. 18, Sept. 2023, Art. no. 10189, https://doi.org/10.3390/app131810189.

[7] M. Hasanat, W. Khan, N. Minallah, N. Aziz, and A.-U.-R. Durrani, "Performance evaluation of transfer learning based deep convolutional neural network with limited fused spectro-temporal data for land cover classification," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 6, pp. 6882–6890, Dec. 2023, https://doi.org/10.11591/ijece.v13i6.pp6882-6890.

[8] O. T. Lisungu and K. Ogudo, "Unleash the Power of Deep Neural Networks and Transfer Learning for Enhanced Face Recognition," *International Conference on Artificial Intelligence and its Applications*, pp. 87–94, Dec. 2023.

[9] L. Du, "Object Detectors in Autonomous Vehicles: Analysis of Deep Learning Techniques," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 10, pp. 217-224, Oct. 2023, https://doi.org/10.14569/IJACSA.2023.0141024.

[10] L. Yu and S. Liu, "A Single-Stage Deep Learning-based Approach for Real-Time License Plate Recognition in Smart Parking System," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 9, pp. 1142-1150, Oct. 2023, https://doi.org/10.14569/IJACSA.2023.01409119.

[11] W. Wang, Y. F. Sun, W. Gao, W. Xu, Y. Zhang, and D. Huang, "Quantitative detection algorithm for deep-sea megabenthic organisms based on improved YOLOv5," *Frontiers in Marine Science*, vol. 11, Feb. 2024, Art. no. 1301024, https://doi.org/10.3389/fmars.2024.1301024.

[12] Z. Zhou, "Detection and Counting Method of Pigs Based on YOLOV5_Plus: A Combination of YOLOV5 and Attention Mechanism," Mathematical Problems in Engineering, vol. 2022, no. 1, Jan. 2022, Art. no. 7078670, https://doi.org/10.1155/2022/7078670.

[13] N. L. Kirana, D. Kurnianingtyas, and Indriati, "A Deep Learning Approach to Plastic Bottle Waste Detection on the Water Surface using YOLOv6 and YOLOv7," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18623–18630, Dec. 2024, https://doi.org/10.48084/etasr.8592.

[14] N. Daud, A. I. Widjanarko, and D. Kurnianingtyas, "Leveraging AI Models for Enhanced Urine Analysis: Evaluating YOLOv8 and MobileNet-v2 in Dehydration Detection Post-COVID-19," in *2024 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, Nov. 2024, pp. 571–578, https://doi.org/10.1109/COMNETSAT63286.2024.10862742.

[15] S. Sivanraj, D. N. L. S. Uduwage, and M. Tripathi, "Real-time safety detection on construction sites using a vision-language and NLP-based model," Advanced Engineering Informatics, vol. 69, Jan. 2026, Art. no. 103889, https://doi.org/10.1016/j.aei.2025.103889.

[16] "Personal Protective Equipment Dataset," *Roboflow*. https://universe.roboflow.com/training-dataset-ta/personal-protective-equipment-9clx4.

[17] E. Ongko and H. Hartono, "Hybrid approach redefinition-multi class with resampling and feature selection for multi-class imbalance with overlapping and noise," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 3, pp. 1718–1728, June 2021, https://doi.org/10.11591/eei.v10i3.3057.

[18] N. 'Izzaty Mohd Yusof, A. Sophian, H. F. Mohd Zaki, A. A. Bawono, A. H. Embong, and A. Ashraf, "Assessing the performance of YOLOv5, YOLOv6, and YOLOv7 in road defect detection and classification: a comparative study," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 1, pp. 350–360, Feb. 2024, https://doi.org/10.11591/eei.v13i1.6317.

[19] A. Xiong, R. W. Proctor, W. Yang, and N. Li, "Embedding Training Within Warnings Improves Skills of Identifying Phishing Webpages," *Human Factors*, vol. 61, no. 4, pp. 577–595, June 2019, https://doi.org/10.1177/0018720818810942.

[20] X. Jing, X. Liu, and B. Liu, "Composite Backbone Small Object Detection Based on Context and Multi-Scale Information with Attention Mechanism," *Mathematics*, vol. 12, no. 5, Feb. 2024, Art. no. 622, https://doi.org/10.3390/math12050622.

[21] N. D. Nath and A. H. Behzadan, "Deep Convolutional Networks for Construction Object Detection Under Different Visual Conditions," *Frontiers in Built Environment*, vol. 6, Aug. 2020, Art. no. 97, https://doi.org/10.3389/fbuil.2020.00097.

[22] A. H. Zaray, A. Hasan, S. Johari, P. A. Hashmat, and K. N. Jha, "Client and contractor perspectives on attributes affecting construction quality in a war-affected region," *Engineering, Construction and Architectural Management*, vol. 30, no. 10, pp. 4762–4781, July 2022, https://doi.org/10.1108/ECAM-01-2022-0059.

[23] C. Dewi, A. P. S. Chen, and H. J. Christanto, "Recognizing Similar Musical Instruments with YOLO Models," *Big Data and Cognitive Computing*, vol. 7, no. 2, May 2023, Art. no. 94, https://doi.org/10.3390/bdcc7020094.

[24] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness," presented at the *International Conference on Learning Representations*, 2019.

[25] E. Guney, H. Altin, A. E. Asci, O. U. Bayilmis, and C. Bayilmis, "YOLO-Based Personal Protective Equipment Monitoring System for Workplace Safety," *JITSI : Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 5, no. 2, pp. 77–85, June 2024, https://doi.org/10.62527/jitsi.5.2.238.