Available *online* at: http://ajeeet.ft.unand.ac.id/

Andalas Journal of Electrical and Electronic Engineering Technology

ISSN 2777-0079

Design and Implementation of a Deep Learning-Based Safety Helmet Compliance Detection System Using the Faster R-CNN Method

Palman, Sandy Azizi, Muhammad Ilhamdi Rusydi, Rahmadi Kurnia*

Department of Electrical Engineering, Faculty of Engineering, Universitas Andalas, Padang, 25163, Indonesia

ARTICLE INFORMATION

Received: April 2, 2026
Revised: May 2, 2026
Accepted: May 5, 2026
Available online: June 1, 2026

KEYWORDS

Faster R-CNN, Computer Vision, Personal Protective Equipment

CORRESPONDENCE

E-mail: rusydi@eng.unand.ac.id

A B S T R A C T

Workplace accidents remain one of the major issues in industrial environments and are often caused by low compliance with the use of Personal Protective Equipment (PPE), particularly safety helmets. Manual supervision of PPE usage tends to be inefficient and prone to human error. This study aims to develop an intelligent computer-vision-based system capable of automatically and real-time monitoring helmet compliance. The proposed system employs the Faster Region-Convolutional Neural Network (Faster R-CNN) algorithm to detect and classify workers who are wearing and not wearing helmets. The dataset was obtained from CCTV video recordings in industrial areas, which were converted into image frames for training and testing processes. The experimental results show that the system achieved an accuracy of 90% for helmet-wearing workers and 87% for non-helmet-wearing workers during daytime conditions, and 97% and 91% respectively at night. With an average computation time of 0.1 seconds per frame, the system is capable of real-time detection at up to 10 frames per second. These results indicate that the Faster R-CNN method is effective in detecting PPE compliance and has the potential to be implemented as an automated safety-support system in industrial environments.

INTRODUCTION

According to global data released by the International Labour Organization (ILO), the number of occupational accidents (OA) and occupational diseases (OD) worldwide has reached 430 million cases per year, consisting of 270 million (62.8%) OA cases and 160 million (37.2%) OD cases, resulting in approximately 2.78 million worker fatalities annually. In Indonesia, the number of workers experiencing OA/OD increased each year from 2019 to 2021 [1]. Throughout 2022, PT Semen Padang recorded seven occupational accident cases, comprising four first-aid cases and three severe cases. Therefore, efforts to prevent workplace accidents must continue to be strengthened and enhanced.

As part of PT Semen Padang's efforts to prevent workplace accidents and minimize associated risks, the company has implemented an Occupational Safety and Health Management System (SMK3) integrated into the Semen Padang Management System (SMSP) [2]. One of the key measures for controlling workplace accidents is ensuring the use of personal protective equipment (PPE) in operational areas. However, compliance with PPE requirements—particularly the use of safety helmets—is often neglected by some workers, thereby increasing the risk of workplace accidents [3], [4], [5], [6].

Previous studies have explored the use of computer vision–based detection technologies to identify PPE usage [7], [8], [9], [10]. Convolutional Neural Networks (CNNs) are among the most commonly applied methods for detecting objects in digital images [11], [12], [13], [14]. The Faster R-CNN model has demonstrated strong performance in object detection, particularly in identifying region-based features in images, such as worker faces or safety helmets [15], [16], [17], [18].

Previous research has primarily focused on demonstrating algorithm effectiveness in general or controlled environments, with limited application in complex industrial contexts. Few studies have systematically designed and implemented real-time systems for detecting personal protective equipment (PPE) usage, while simultaneously addressing the accuracy and processing speed requirements critical for industrial applications.

This study proposes a novel approach to monitoring safety helmet compliance in industrial work environments by leveraging an optimized Faster R-CNN algorithm to detect helmet usage among workers at PT Semen Padang Packing Plant (PP) Bengkulu. The system is expected to achieve high detection accuracy and provide automated notifications to management when PPE violations occur, thereby contributing to a reduction in workplace accidents.

METHODS

The steps carried out in this research follow the flowchart shown in Figure 1. In the problem identification stage, information is collected from the field regarding obstacles in monitoring compliance with the use of personal protective equipment (PPE)

https://doi.org/10.25077/ajeeet.v6i1.218

in the work area. A literature review is also conducted using journal articles and previously published studies. In the system design stage, a helmet detection system is developed for field work areas, capable of classifying workers who are wearing helmets and those who are not, as well as providing notifications when non-compliance is detected.

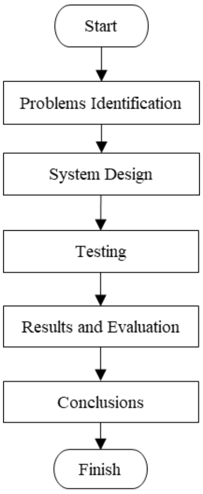

Figure 1. Flowchart of research steps

The next stage is system testing. The designed system is evaluated to assess its performance. The testing process involves observing the model during execution and measuring its performance based on the applied method. All data obtained during the testing phase are recorded. Finally, the results and analysis stage is conducted to examine the overall system performance. Improvements are then implemented to enhance detection accuracy and speed. Model refinement is achieved by adjusting the input size, increasing the amount of training data, and modifying the Faster R-CNN architecture.

System Design

The system to be developed will utilize CCTV cameras as input for the machine learning/deep learning model [19], [20], [21]. The overall system design to be implemented is illustrated in Figure 2.

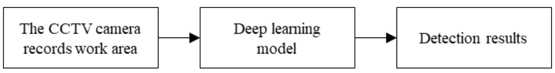

Figure 2. System block diagram

Objects such as helmets and heads without helmets are detected using a deep learning model.

Data Collection and Annotation

Data collection was carried out by recording employees in several rooms and field areas at PT Semen Padang PP Bengkulu. The recordings were captured as videos at 24 frames per second. Each recorded video was sampled every 10 seconds [22], [23]. Video collection was conducted over several days, covering different locations and varying numbers of employees. The CCTV camera installation layout is shown in Figure 3 and 4. The CCTV cameras were installed with a spacing of 4.5 meters between each unit. The distance between each camera and the outermost pole was 9.2 meters. The cameras used were high-resolution devices with a frame size of 1920 × 1080 pixels and an average frame rate of 24 frames per second.

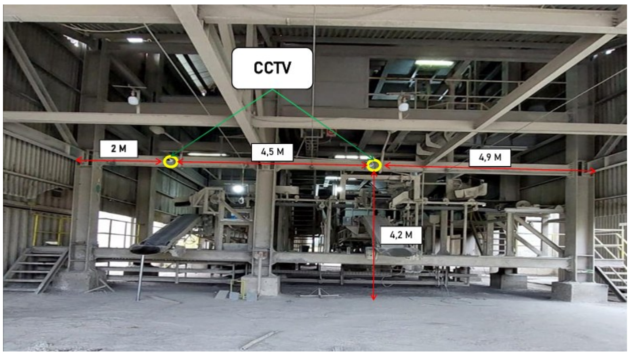

Figure 3. Front-view camera installation position.

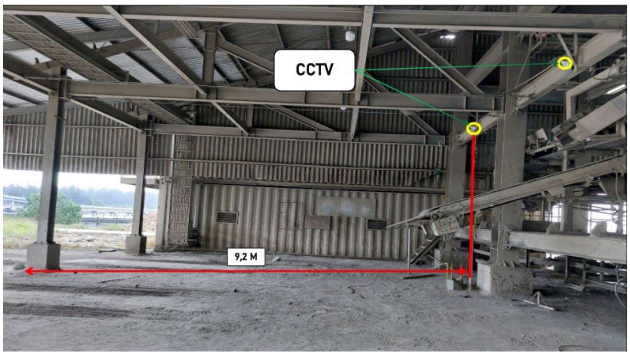

Figure 4. Side-view camera installation position

Data Preprocessing

The preprocessing step applied in this study was the resizing process [24]. The original input image size of 1920×1080 pixels was converted to 640×640 pixels. Using the original 1920×1080 resolution requires substantial computational resources, making the training process difficult to perform efficiently [25]. The 640×640 resolution was selected to facilitate the training process in terms of both speed and required computing resources [26], [27], [28]. The collected dataset was divided into training, validation, and testing subsets [29]. The training data were used to train the Faster R-CNN model for object detection. The validation data were employed to evaluate the performance of the trained model and to fine-tune its parameters. The testing data were used to assess the model's performance on real-time data, enabling the derivation of conclusive insights regarding the trained model. The dataset was split into 80% for training, 10% for validation, and 10% for testing [30], [31], [32], [33].

Before the data are used to train the artificial intelligence model, a labeling process is required. The assigned labels consist of head and helmet classes. The labeling was performed using LabelImg. Each object was annotated by drawing a bounding box around it and assigning the appropriate classification label for the detected object. The recording results are presented in Figure 5, while the labeling process using LabelImg is shown in Figure 6.

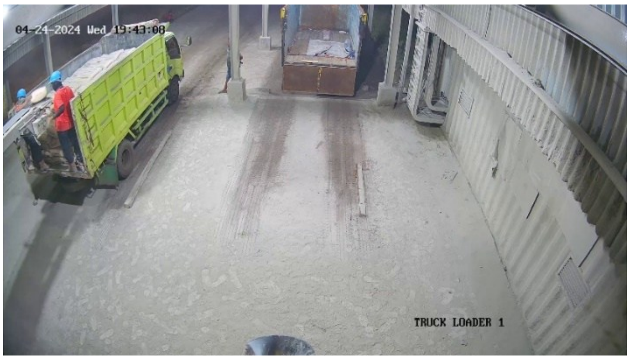

Figure 5. Recording using a camera

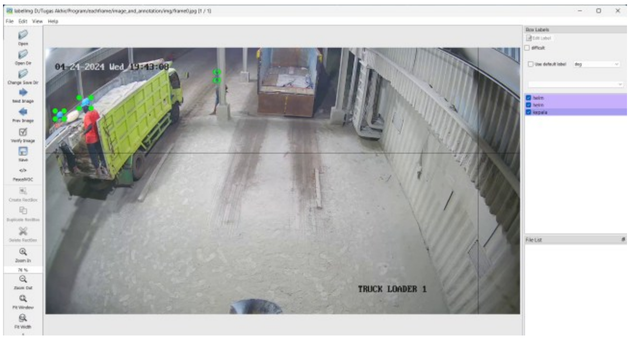

Figure 6. Labeling and classification using LabelImg

Model Tarining

Helmet usage detection in this study is carried out using the Faster R-CNN algorithm. Faster R-CNN generates outputs consisting of object classification results and the bounding box coordinates of each detected object [34], [35]. This algorithm is chosen because, among various Convolutional Neural Network–based detection methods, Faster R-CNN achieves inference times of less than one second [36], [37] which is essential for real-time implementation. Furthermore, Faster R-CNN offers a relatively high mean average precision (mAP) compared to other models [38]. A comparison of mAP and processing speed for several algorithms is presented in Table 1.

Table 1. Comparison of mAP & the processing speed of each model.

Model	mAP (%)	Processing Speed (s)
R-CNN	66	20
Fast R-CNN	70	2
Faster R-CNN	73.2	0.14

Detection was performed using the Faster R-CNN algorithm because it achieves a processing speed of 0.2 seconds on the VOC 2007 dataset, enabling Faster R-CNN to reach equal to or more than 5 frames per second [39], [40]. The Faster R-CNN algorithm learns each label corresponding to the classification of the objects to be detected [41]. The target objects in this study are heads with helmets and heads without helmets. The flowchart for training the Faster R-CNN algorithm is presented in Figure 7.

System Training

The performance of the proposed system was evaluated using standard classification metrics, including accuracy, precision, and recall, to assess its capability in distinguishing between helmet and no-helmet usage. Accuracy represents the proportion of correctly classified instances, precision reflects the reliability of positive predictions, and recall measures the model's ability to detect relevant objects. The evaluation was conducted under two distinct conditions: daytime and nighttime. A confusion matrix was employed to systematically compare the predicted labels—helmet, no helmet, and no human—with the corresponding ground truth annotations. Subsequently, the evaluation metrics were computed based on the confusion matrix using the following formulations [42].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

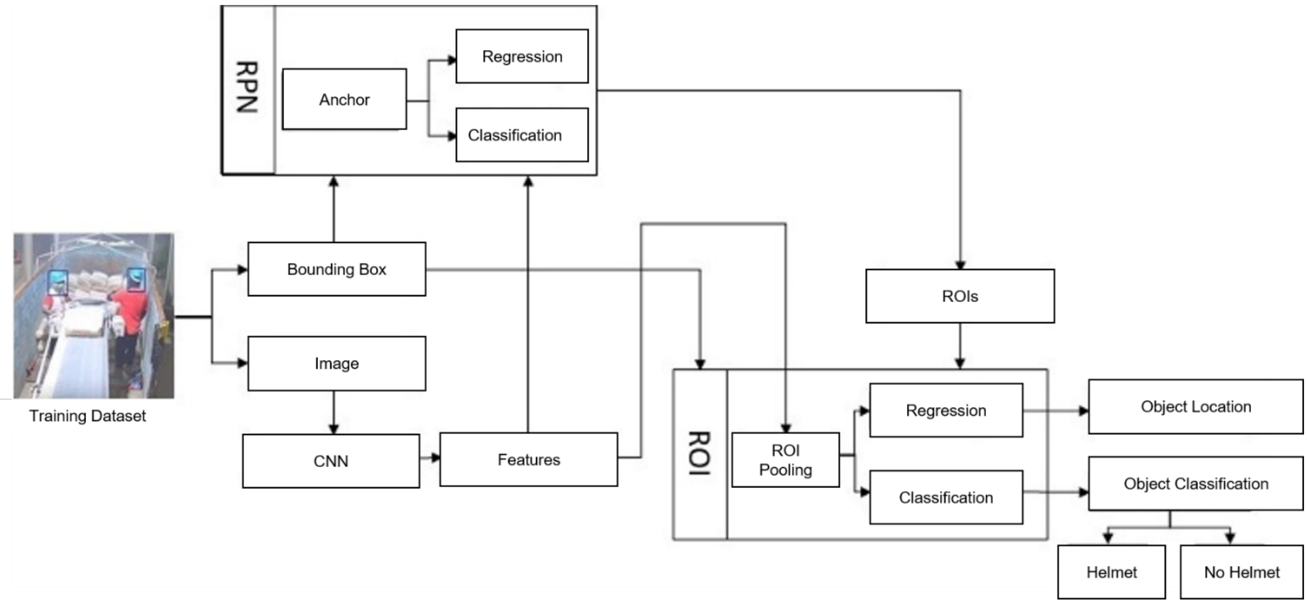

Figure 7. Flowchart of the faster R-CNN algorithm for training the model

The calculation formulas for FN, FP, TN, and TP for the helmet data follow the equations below:

$TP = cell\ (actual\ helmet\ and\ predicted\ helmet)$
$FN = cell\ (actual\ helmet\ and\ predicted\ no\ helmet) + cell\ (actual\ helmet\ and\ predicted\ no\ human)$
$FP = cell\ (actual\ no\ helmet\ and\ predicted\ helmet) + cell\ (actual\ no\ human\ and\ predicted\ helmet)$
$TN = cell\ (actual\ no\ helmet\ and\ predicted\ no\ helmet) + actual\ no\ helmet\ and\ predicted\ no\ human + actual\ no\ human\ and\ predicted\ no\ helmet + actual\ no\ human\ and\ predicted\ no\ human)$

Formula for calculating FN, FP, TN, and TP for no-helmet data:

$TP = clel\ (actual\ no\ helmet\ and\ predicted\ no\ helmet)$
$FN = cell\ (actual\ no\ helmet\ and\ predicted\ helmet) + cell\ (actual\ no\ helmet\ and\ predicted\ no\ human)$
$FP = cell\ (actual\ helmet\ and\ predicted\ no\ helmet) + cell\ (actual\ no\ human\ and\ predicted\ no\ helmet)$
$TN = cell\ (actual\ helmet\ and\ predicted\ helmet) + actual\ helmet\ and\ predicted\ no\ human + actual\ no\ human\ and\ predicted\ helmet + actual\ no\ human\ and\ predicted\ no\ human)$

In addition to the previously described metrics, this research also evaluates performance using Average Precision (AP) and mean Average Precision (mAP). Due to the absence of confidence scores required to construct a complete precision–recall curve, AP is approximated using a single-point estimation based on precision and recall, where AP is computed as the product of precision and recall. This approach serves as a simplified representation of the area under the precision–recall curve. Furthermore, mAP is calculated as the mean of the AP values across all evaluated classes. The formulations used to compute AP and mAP are presented as follows.

$$AP \approx Precision\ x\ Recall \quad (4)$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} APi \quad (5)$$

The evaluation was conducted by testing the model using heterogeneous datasets, including daytime and nighttime conditions, as well as scenarios involving one, two, and three employees. Another critical parameter assessed was the detection latency. An effective system is expected to perform detections promptly, thereby minimizing the number of skipped frames in the video sequences.

RESULTS AND DISCUSSIONS

System Prototype

The prototype system was developed through a series of stages, including data collection, data annotation, model training, integration of CCTV with the trained model, and the generation of alerts upon detection of violations.

The equipment utilized in this study comprises CCTV systems, computers, and cloud-based resources, all of which are readily available at PT. Semen Padang. The model development was implemented using the Python programming language, with TensorFlow serving as the primary framework for computational modeling and deep learning tasks.

Results

This study employs the Faster R-CNN model to classify heads as either wearing helmets or not wearing helmets. After undergoing a training process lasting approximately two hours, with a total of 4,100 iterations, the total loss and classification loss curves of the model were obtained, as illustrated in Figure 8 and Figure 9.

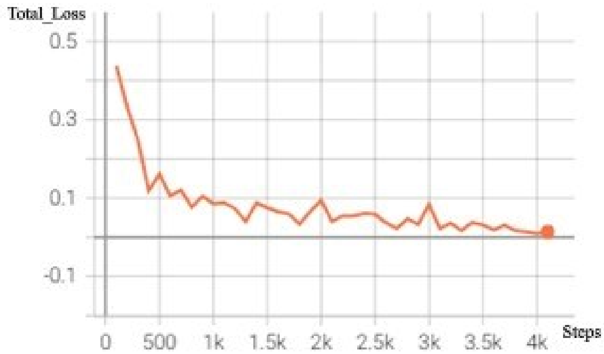

Figure 8. Graph of total loss during training

Figure 9. Graph of classification loss during training

At the final iteration, the training process indicates that the model achieved a total loss of 0.013% and a bounding box classification loss of 0.0005%. These loss values are considerably low compared to those observed in earlier iterations, suggesting that the training process can be terminated. Furthermore, the total loss curve demonstrates a consistent downward trend, indicating that the learning process is stable and efficient.

Daytime and Nighttime Testing

An example of the detection test results under daytime conditions is illustrated in Figure 10. The overall performance of the detection tests is summarized in the confusion matrix for daytime testing, as presented in Table 2.

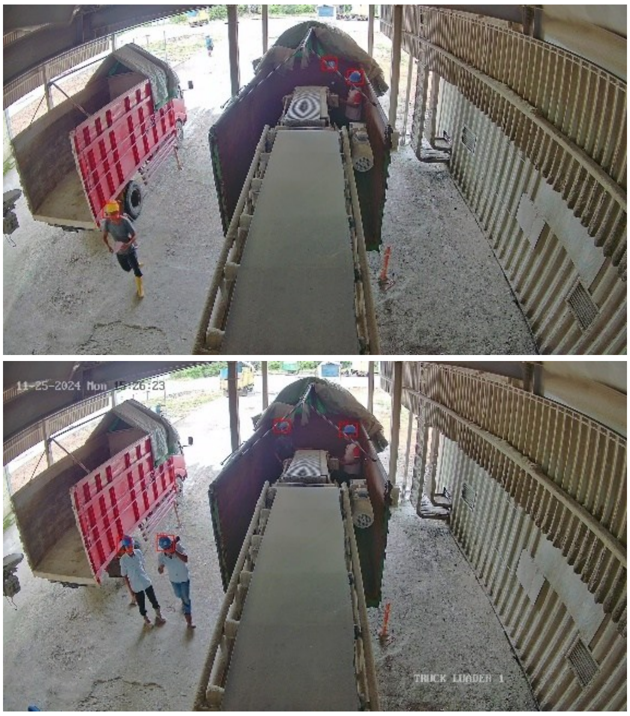

Figure 10. Results of the daytime testing

Table 2. Confusion matrix for daytime testing

		Predicted		
		Helmet	No Helmet	No Human
Actual	Helmet	95	0	19
	No Helmet	0	54	25
	No Human	0	0	76

Subsequently, the results of the tests conducted under nighttime conditions are presented in Figure 11. The corresponding detection performance is summarized in the confusion matrix for nighttime testing, as shown in Table 3.

Figure 11. Results of the nighttime testing

Table 3. Confusion matrix for nighttime testing

		Predicted		
		Helmet	No Helmet	No Human
Actual	Helmet	75	0	3
	No Helmet	0	44	12
	No Human	0	0	41

The performance of the system was evaluated under both daytime and nighttime conditions using the confusion matrix method. During daytime testing, the helmet class achieved an accuracy of 92.9%, whereas the non-helmet class reached an accuracy of 90.7%. In contrast, nighttime testing yielded higher accuracies, with the helmet class achieving 98.3% and the non-helmet class 93.1%. The detailed accuracy, precision and recall results for each class are summarized in Table 4.

Table 4. Accuracy, Precision and Recall per Class

Class	Accuracy	Precision	Recall
Daytime			
- Helmet	92.9%	100%	83.3%
- No Helmet	90.7%	100%	68.4%
Nighttime			
- Helmet	98.3%	100%	96.2%
- No Helmet	93.1%	100%	78.6%

The accuracy of the helmet class is consistently higher than that of the no-helmet class under both daytime and nighttime conditions. This can be attributed to the more distinctive color and shape characteristics of helmets compared to bare heads. The relatively stable geometric structure and stronger visual features of helmets enable the model to learn discriminative features more effectively than those of non-helmet instances.

Furthermore, the evaluation results indicate that the accuracy obtained under daytime conditions is lower than that achieved during nighttime testing. A similar trend is observed in the mAP values. Based on the estimation approach, the mAP is calculated to be 75.8% for daytime testing and 87.4% for nighttime testing, as presented in Table 5.

Table 5. AP and mAP (approximation)

Class	AP	mAP
Daytime		
- Helmet	83.3%	75.8%
- No Helmet	68.4%	
Nighttime		
- Helmet	96.2%	87.4%
- No Helmet	78.6%	

The higher accuracy and mAP values observed during nighttime testing indicate that the model performs better under nighttime conditions compared to daytime scenarios. This difference can be attributed to environmental factors affecting image quality. During daytime, the observed objects are occasionally affected by excessive sunlight exposure, resulting in glare within the CCTV camera's field of view. In contrast, nighttime conditions benefit from more controlled illumination provided by artificial lighting, which is sufficiently stable to support clearer object observation by the CCTV system.

The precision values for both the helmet and no-helmet classes reach 100% under both daytime and nighttime testing conditions, indicating that the model consistently performs correct inter-class classification without producing false positive predictions. However, the relatively lower recall values suggest that a number of objects remain undetected. This indicates that the model still requires improvement, particularly in terms of object detection capability.

Detection Speed Testing

The detection speed evaluation was carried out by measuring the time required by the system to detect an image. Using computational resources equipped with an Nvidia T4 provided by Google Colaboratory—featuring 12 GB of RAM, 15 GB of GPU RAM, and 112 GB of disk storage—the system achieved a detection speed of 0.10 seconds. This detection time is sufficiently fast and meets the requirements for real-time detection. A comparison of the detection performance from this study with previous related studies is presented in Table 6.

Table 6. Comparison of Detection Results

Model	Accuracy		Detection
	Helmet	No Helmet	Time (s)
Xu Li et al. [43]	81%	74%	-
Laily et al. [44]	95%	-	0.92
Our model	95%	92%	0.10

DISCUSSIONS

The primary objective of this study is to develop a real-time system for monitoring safety helmet compliance in industrial environments using the Faster R-CNN algorithm, while achieving a balance between detection accuracy and computational efficiency. The experimental results demonstrate that the proposed system is capable of achieving high accuracy under both daytime and nighttime conditions, with superior performance observed during nighttime testing.

When compared with previous studies, the results of this research indicate several notable improvements. For instance, the study conducted by Xu Li et al. [43], which also utilized the Faster R-CNN architecture, reported accuracy levels of 81% for helmet detection and 74% for non-helmet detection across highly heterogeneous datasets involving multiple environments, weather conditions, and lighting variations. In contrast, the proposed system achieves higher accuracy, particularly under controlled industrial conditions. This suggests that dataset consistency and environmental control play a significant role in improving detection performance.

Similarly, the study by Laily et al. [44], which employed the Mask R-CNN method, achieved a high detection accuracy of 95%, comparable to the results obtained in this study. However, the detection time reported in their work was approximately 0.92 seconds per frame, which is significantly slower than the 0.10 seconds achieved in this research. This indicates that the proposed Faster R-CNN–based system offers a more favorable trade-off between accuracy and detection speed, making it more suitable for real-time industrial applications.

In addition, recent studies on PPE detection using deep learning, such as those based on YOLO architectures [45] have demonstrated faster inference times due to their one-stage detection framework. However, these methods often exhibit lower localization accuracy compared to two-stage detectors like Faster R-CNN. The findings of this study are consistent with prior work [15] which highlights that Faster R-CNN provides higher precision in object localization, particularly for structured objects such as safety helmets. This explains the consistently high precision values (100%) observed in this study, indicating the model's robustness in minimizing false positive detections.

Another important finding is the performance difference between daytime and nighttime conditions. The results show that nighttime detection yields higher accuracy and mAP values compared to daytime conditions. This phenomenon can be explained by illumination stability. Previous studies have reported that excessive lighting, glare, and shadow effects can degrade object detection performance in outdoor or high-exposure environments. In this study, the controlled artificial lighting at night reduces visual noise and enhances feature extraction, enabling the model to perform more effectively. This finding reinforces the importance of environmental factors in real-world computer vision applications.

Despite the promising results, several limitations remain. The relatively lower recall values indicate that some objects are not successfully detected, suggesting that the model may still struggle with certain object variations or occlusions. This limitation is consistent with findings in previous studies [46] where insufficient dataset diversity negatively impacts detection robustness. Therefore, increasing the dataset size and incorporating greater variability in terms of object position, scale, and lighting conditions are essential steps for improving model generalization.

Furthermore, the current system is limited to detecting only two classes: helmet and no-helmet. In practical industrial scenarios, comprehensive PPE compliance monitoring requires the detection of additional equipment such as safety vests, gloves, glasses and safety shoes. Recent research [47] has demonstrated the feasibility of multi-class PPE detection using advanced deep learning architectures. Therefore, extending the proposed model to support multi-class detection represents an important direction for future work.

Overall, the findings of this study confirm that the Faster R-CNN–based approach is effective for real-time helmet compliance monitoring in industrial environments. The combination of high accuracy, fast detection speed, and integration with automated alert systems highlights its practical applicability. Moreover, the study contributes to bridging the gap identified in previous research by not only focusing on algorithm performance but also implementing a complete real-time monitoring system tailored to industrial needs.

 https://doi.org/10.25077/ajeeet.v6i1.218

CONCLUSIONS

This study successfully developed a real-time safety helmet compliance detection system using the Faster R-CNN algorithm. The proposed system achieved high detection accuracy, reaching 92.9% for the helmet class and 90.7% for the no-helmet class under daytime conditions, and 98.3% and 93.1% respectively under nighttime conditions, while maintaining a fast detection time of 0.10 seconds per frame using the NVIDIA T4. These findings indicate that the proposed approach is well-suited for real-time monitoring of workers' compliance with helmet usage, enabling automatic alert notifications to supervisors via Android devices when violations—such as the absence of a helmet—are detected. However, the proposed model still requires further improvement, particularly in terms of object detection performance. This can be addressed by increasing the size and diversity of the dataset, including variations in object position, distance, and lighting conditions. Furthermore, the model has the potential to be extended to detect a more comprehensive range of PPE, such as safety shoes, safety vests, and other protective equipment.

REFERENCES

[1] Kementerian Ketenagakerjaan RI, *Profil Keselamatan dan Kesehatan Kerja Nasional Indonesia Tahun 2022*. 2022.

[2] S. Pengendalian, I. Perseroan, T. K. Perusahaan, and T. Sosial, "Laporan tahunan 2022," 2022.

[3] H. Chen, Y. Li, H. Wen, and X. Hu, "YOLOv5s-gnConv: detecting personal protective equipment for workers at height," *Front. Public Heal.*, vol. 11, no. September, 2023, doi: 10.3389/fpubh.2023.1225478.

[4] C. B. Souto Maior *et al.*, "Personal protective equipment detection in industrial facilities using camera video streaming," *Saf. Reliab. - Safe Soc. a Chang. World - Proc. 28th Int. Eur. Saf. Reliab. Conf. ESREL 2018*, no. 2006, pp. 2863–2868, 2018, doi: 10.1201/9781351174664-359.

[5] Ö. Hatipo and A. Köksal, "Ki ş isel Koruyucu Donan ı m Tespiti Detection of Personal Protective Equipment".

[6] R. Amilton *et al.*, "OPEN ACCESS THE IMPORTANCE OF PPE USE IN CIVIL CONSTRUCTION : CASE STUDY," 2019.

[7] L. Hallonqvist, "Detection of safety equipment in the manufacturing industry using image recognition," 2021.

[8] F. Zhafran, E. S. Ningrum, M. N. Tamara, E. Kusumawati, A. C. Neural, and N. Cnn, "Computer Vision System Based for Personal Protective Equipment Detection , by Using Convolutional Neural Network," *2019 Int. Electron. Symp.*, pp. 516–521, 2019.

[9] W. Zhang, C. Yang, F. Jiang, X. Gao, and X. Zhang, "Safety Helmet Wearing Detection Based on Image Processing and Deep Learning," no. 2, pp. 343–347, 2020, doi: 10.1109/CISCE50729.2020.00076.

[10] S. M. Ahsan, "PPE detector : a YOLO-based architecture to detect personal protective equipment (PPE) for construction sites," 2022, doi: 10.7717/peerj-cs.999.

[11] C. S. de Oliveira, C. Sanin, and E. Szczerbicki, "Video Classification Technology in a Knowledge-Vision-Integration Platform for Personal Protective Equipment Detection: An Evaluation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10751 LNAI, pp. 443–453, 2018, doi: 10.1007/978-3-319-75417-8_42.

[12] M. Riaz *et al.*, "Enhancing Workplace Safety: PPE_Swin—A Robust Swin Transformer Approach for Automated Personal Protective Equipment Detection," *Electron.*, vol. 12, no. 22, 2023, doi: 10.3390/electronics12224675.

[13] R. G. B. Features, "Moving Object Detection Based on Fusion of Depth Information and RGB Features," 2022.

[14] C. Avilés-cruz, A. Ferreyra-ramírez, and A. Zúñiga-lópez, "for Human Activity Recognition," 2019, doi: 10.3390/s19071556.

[15] S. Chen and W. Wang, "Detection of Safety Helmet Wearing Based on Improved Faster R-CNN," 2020.

[16] M. M. Saudi, A. Hakim, A. Ahmad, A. Shakir, M. Saudi, and M. H. Ali, "Image Detection Model for Construction Worker Safety Conditions using Faster R-CNN," vol. 11, no. 6, pp. 246–250, 2020.

[17] L. Jiang, "Research on application of a Faster R-CNN based on upper and lower layers in face detection by," vol. 2022, pp. 1–95, 2022.

[18] V. C. Sensors *et al.*, "Face Detection in Nighttime Images Using Region-Based Convolutional Neural Network," 2018, doi: 10.3390/s18092995.

[19] T. Zhang *et al.*, "Recent Advances in Video Analytics for Rail Network Surveillance for Security , Trespass and Suicide Prevention — A Survey," 2022.

[20] T. V. Detection, "Towards an End-to-End Framework of CCTV-Based Urban Traffic Volume Detection and Prediction," vol. D, pp. 1–23, 2021.

[21] M. Shakhnoza, U. Sabina, M. Sevara, and Y. Cho, "Novel Video Surveillance-Based Fire and Smoke Classification Using Attentional Feature Map in Capsule Networks," pp. 1–17, 2022.

[22] I. Hipiny, H. Ujir, R. Hassan, A. Arabi, and A. Aimran, "An image and video dataset of nesting green sea turtles with annotated data," *Sci. Data*, pp. 1–7, 2024, doi: 10.1038/s41597-024-04336-3.

[23] "24 Computational Intelligence and Neuroscience - 2021 - Maraghi - Scaling Human-Object Interaction Recognition in the Video.pdf."

[24] W. Wu, Y. Yin, X. Wang, D. Xu, and S. Member, "Face Detection With Different Scales Based on Faster R-CNN," pp. 1–12, 2018, doi: 10.1109/TCYB.2018.2859482.

[25] A. Heredia and G. Barros-gavilanes, "Video processing inside embedded devices using SSD-Mobilenet to count mobility actors," *2019 IEEE Colomb. Conf. Appl. Comput. Intell.*, pp. 1–6.

[26] B. Karbouj and G. A. Topalian-rivas, "ScienceDirect Comparative Performance Evaluation of One-Stage and Two-Stage Object Detectors for Screw Head Detection and Classification in Disassembly Processes," vol. 122, pp. 527–532, 2024, doi: 10.1016/j.procir.2024.01.077.

[27] S. Patel, "Marigold Flower Blooming Stage Detection in Complex Scene Environment using Faster RCNN with Data Augmentation," no. May, 2023, doi: 10.14569/IJACSA.2023.0140379.

[28] "No Title," 2020.

[29] Z. Sheng, K. Tian, Q. Tian, and H. Qu, "A Faster R-CNN based High-Normalization Sample Calibration Method for Dense Subway Passenger Flow Detection," *2018 11th Int. Congr. Image Signal Process. Biomed. Eng. Informatics*, pp. 1–5, 2018.

[30] R. Soekarta, M. Yusuf, J. Visman, and M. F. Hasa, "IMPLEMENTATION OF DEEP LEARNING FOR PERSONAL PROTECTIVE EQUIPMENT (PPE) DETECTION IN WORKERS IN THE OIL INDUSTRY USING THE YOLOv5 ALGORITHM," vol. 7, no. 2, pp. 96–106, 2025.

[31] L. Wei, P. Liu, H. Ren, and D. Xiao, "Research on

helmet wearing detection method based on deep learning," *Sci. Rep.*, pp. 1–15, 2024, doi: 10.1038/s41598-024-57433-z.

[32] Y. Nanda, K. Umat, and W. Gunawan, "YOLOv10 for Real-Time Detection of Personal Protective Equipment on Construction Workers," vol. 17, no. 2, pp. 131–139, 2025.

[33] Z. Wang, Y. Wu, L. Yang, A. Thirunavukarasu, C. Evison, and Y. Zhao, "Construction Sites Using Deep Learning Approaches," pp. 1–22, 2021.

[34] Y. Wang, Z. Liu, and W. Deng, "Anchor Generation Optimization and Region of," 2019, doi: 10.3390/s19051089.

[35] M. R-cnn, "Crack Detection and Comparison Study Based on Faster R-CNN," 2022.

[36] M. Hou, X. Dong, J. Li, G. Yu, R. Deng, and X. Pan, "PDC : Pearl Detection with a Counter Based on Deep Learning," 2022.

[37] H. Possible and R. P. Network, "Improved Faster R-CNN Tra ffi c Sign Detection Based Regions Proposal Network," 2019.

[38] L. G. Divyanth, P. Soni, C. M. Pareek, and R. Machavaram, "Detection of Coconut Clusters Based on Occlusion Condition Using Attention-Guided Faster R-CNN for Robotic Harvesting," 2022.

[39] S. Srivastava, A. V. Divekar, C. Anilkumar, I. Naik, V. Kulkarni, and V. Pattabiraman, "Comparative analysis of deep learning image detection algorithms," *J. Big Data*, 2021, doi: 10.1186/s40537-021-00434-w.

[40] S. Ren, K. He, and R. Girshick, "Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks," pp. 1–9.

[41] X. Cao, P. Wang, C. Meng, and X. Bai, "Region Based CNN for Foreign Object Debris," pp. 1–14, doi: 10.3390/s18030737.

[42] W. Amelia, S. 'Asyarina Ramadhani, B. Sunaryo, R. Desnita, and N. F. Nilakesuma, "Harnessing CNN for Early Breast Cancer Detection: Enhancing Precision in Image-Based Diagnosis," *Ingénierie des systèmes d Inf.*, vol. 30, no. 2, Feb. 2025, doi: 10.18280/isi.300219.

[43] X. Li, T. Hao, F. Li, L. Zhao, and Z. Wang, "applied sciences Faster R-CNN-LSTM Construction Site Unsafe Behavior Recognition Model," 2023.

[44] T. Online, M. E. Laily, F. Nur, G. Qorik, and O. Pratamasunu, "Jurnal Politeknik Caltex Riau Deteksi Penggunaan Alat Pelindung Diri (APD) Untuk Keselamatan dan Kesehatan Kerja Menggunakan Metode Mask Region Convolutional Neural Network (Mask R-CNN)," vol. 8, no. 2, pp. 279–288, 2022.

[45] D. G. Lema, R. Usamentiaga, and D. F. García, "Low-cost system for real-time verification of personal protective equipment in industrial facilities using edge computing devices," *J. Real-Time Image Process.*, vol. 20, no. 6, pp. 1–18, 2023, doi: 10.1007/s11554-023-01368-7.

[46] H. Gu and N. Konz, "A SYSTEMATIC STUDY OF THE FOREGROUND - BACKGROUND IMBALANCE PROBLEM IN DEEP LEARNING FOR OBJECT".

[47] J. Airton, L. Cabrejos, and A. Roman-gonzalez, "Artificial Intelligence System for Detecting the Use of Personal Protective Equipment," vol. 14, no. 5, pp. 580–585, 2023.