MALAYSIAN JOURNAL OF
CYBERSECURITY AND APPLICATIONS

ENHANCING SAFETY AT CONSTRUCTION SITES THROUGH FASTER R-CNN MODELING

Madihah Mohd Saudi[1], Ratih Ardiati Ningrum[2], and Malikhah[2]*

[1]Cyber Security and Systems (CSS) Research Unit, Faculty of Science and Technology (FST),
Universiti Sains Islam Malaysia (USIM)
Bandar Baru Nilai, Malaysia.

[2]Faculty of Advanced Technology and Multidiscipline, Airlangga University
Kampus C UNAIR, Surabaya, Indonesia

Emails: madihah@usim.edu.my*, ratih.an@ftmm.unair.ac.id, meli2511@gmail.com,

ABSTRACT
Accidents and deaths at construction sites are on the rise, including incidents where buildings under construction have unexpectedly collapsed, impacting nearby vehicles and workers. There is an increasing demand for more manpower and authorized personnel for safety inspections at these sites. To mitigate accidents and fatalities, the use of sensors, mobile technologies, and machine learning algorithms is essential. Furthermore, Personal Protective Equipment (PPE) monitoring systems rely on networked cameras, AI inference engines, and data communication channels, making them cyber-physical assets that are vulnerable to cyber threats such as data tampering, model manipulation, and system compromise. Hence, this paper presents a model designed to identify safety hazards based on construction workers' compliance with wearing PPE during working hours. To ensure the images used are not compromised via adversarial attacks, the images used for training and testing are consolidated through secure model deployment and continuous integrity monitoring to preserve detection accuracy and prevent false safety compliance. The model has been integrated into a website as an image surveillance tool to monitor safety conditions at construction sites. The model was developed using 7,000 images from the MIT Places Database (for scene recognition) as a training dataset and 400 images sourced from Google for testing. It was built using TensorFlow 1.15, Python scripting, and the Faster R-CNN algorithm. The evaluation results of the model indicate a 75% accuracy rate. The novelty of this model lies in its integration of four(4) components of PPE, its practical application for mobile usage, and a secure model against adversarial attacks. Supervisors at construction sites can utilize their mobile phones and drones as surveillance tools to enhance the safety of their workers. In future developments, the model will also incorporate external factors contributing to accidents and fatalities, such as environmental conditions, electrical hazards, and falls.

Keywords: PPE; construction site; image detection; faster R-CNN, adversarial attacks.

1. INTRODUCTION

The unique conditions of the construction industry significantly contribute to the occurrence of accidents and fatalities. According to research conducted by sources [1], [2], [3], [4], all workers are at risk of incidents that may result in non-permanent disability (NPD), permanent disability (PD), or even death. A lack of knowledge regarding safety procedures and protective measures, such as the use of personal protective equipment (PPE) and an inability to comprehend written safety regulations, are major factors contributing to these accidents [5], [6], [7], [8], [9]. While adversarial attacks are deliberate manipulations of input data or model behaviour that exploit vulnerabilities in machine learning systems, causing object detection models to generate incorrect or misleading predictions and thereby undermining the reliability of AI-driven decision-making [10]. In construction-site PPE detection, recent studies report that adversarial patches, visually modified helmets or safety vests, and poisoned training datasets can successfully evade deep learning-based detectors, resulting in false safety compliance and posing significant risks to occupational safety [11]. This paper addresses adversarial threats comprehensively by designing a secure PPE detection model with robust training mechanisms and implementing continuous monitoring to detect anomalies and mitigate adversarial manipulation throughout system operation.

Safety performance is typically evaluated using lagging indicators, such as Incident Rate (IR), Accident Rate (AR), and Experience Modification Rate (EMR). However, various safety measures and lagging indicators are employed globally [12], [13], [14], [15], [16]. For instance, Abas and his colleagues have identified key factors

that influence safety performance at construction sites, which can help reduce accidents while enhancing productivity and awareness among construction workers [17].

Supervisors should assess their employees' knowledge and understanding of personal protective equipment (PPE) and other safety gear at construction sites. This evaluation will help implement proper training and control measures to reduce the risk of accidents or fatalities, as suggested by previous studies [18], [19]. This paper introduces an image detection model that evaluates safe and dangerous conditions at construction sites based on compliance with PPE usage. Unlike existing research that focuses on a single type of PPE, this study incorporates a combination of four essential PPE components: hard hats, vests, boots, and gloves, which are critical for the safety of construction workers. The modelling is built on the Faster R-CNN algorithm, which classifies workers as either being in a safe or dangerous situation based on their PPE usage. Once the model is developed, it will be integrated into a website and drone systems for real-time monitoring and evaluation.

This paper is organized as follows. Section 2 explains related work, Section 3 presents the method used in this research, Section 4 consists of the findings and their evaluation, and Section 5 concludes the paper and makes suggestions for future work.

2. RELATED WORKS

According to research [20], falls are one of the leading causes of incidents at construction sites, particularly among construction workers. Therefore, it is essential to find solutions that can reduce the risk of fatalities in this workforce. Limited studies are addressing the construction system and the quality of project management, as noted in works [21], [22], [23]. Work by [21] focused on project progress monitoring and improving communication among employees, while [20] proposed a quality management framework specifically for construction projects. Additionally, [23] introduced optimization modelling for repetitive tasks at construction sites using Unmanned Aerial Vehicles (UAVs). Another study [24] utilized UAVs and RFID technology for worker detection at construction sites. Ham and colleagues [25] summarized various applications of UAVs, including progress monitoring, building inspections, measurement, surveying, safety inspections, structural damage assessments, and geo-hazard investigations. However, most existing studies on safety inspections tend to concentrate solely on a single piece of personal protective equipment (PPE), specifically hard hats. There is a scarcity of research focused on image detection in construction environments, as indicated in Table I.

Table 1: Summary of Existing Works Related to Construction Sites

No	Author	Method	Dataset
[26]	W. Fang et al. (2018)	To identify real-time construction sites of employee and excavators, a new method called Enhanced Faster R-CNN is developed.	10,000 images collected and labelled for training and 1500 labelled data for testing
[27]	Q. Fang et al. (2018)	A deep CNN uses the features to detect region and fast R-CNN is used to classify the images.	They collected 100,000 image frames from 25 different construction sites.81,000 images were used for training. All images were annotated.
[28]	Gil, Lee, & Jeon (2018)	Categorizing the image into 27 job types by using Google inception V3.	1208 images for training and 27 images from validation
[23]	(S. Bng & H. Kim, 2020)	Dense underline technology to describe the location, status and movement of building resources photos, the sentences were created. First, the UAV flight data must import the picture's latitude , longitude, date and time. Finally, the database was organized. For construction sites, the data are place, time and definition.	UAV used the data from six separated construction sites to collect 1431 images. It used 8601 captions about the image regions. 1000 and 365 photos for training and testing were used
[29]	(Luo et al., 2020)	Stacked Hourglass Network (HG), Cascaded Pyramid Network (CPN), and an ensembled model (HG-CPN) used.	The researcher used 70% of images for training (4483 images), and 10% for validation (641 images), and 20% for testing (1281 images).

[30]	(Saudi et al., 2020)	Faster R-CNN Inception v2 COCO is used to train the images in the TensorFlow	1,129 images were used for training and 333 images used for testing, of which 263 images taken from MIT, 65 images were taken from the construction site and 5 images from Google

Based on Table I and our summarized findings, we have identified several common challenges that researchers may face in the future. These challenges include the need for improved personal protective equipment (PPE) detection on construction sites or in the workplace, the availability of more dataset resources, the detection of object blockages that may obstruct the target image, and variations in image positions and backgrounds. While many recent works [27], [28] prioritise lightweight single-stage detectors to achieve real-time performance, such approaches may compromise detection precision and increase false positives in cluttered construction scenes. In this paper, Faster R-CNN is adopted to address these limitations, as its two-stage architecture enables more accurate region proposal generation and refined bounding box localisation, which are critical for reliably identifying PPE in densely populated and visually complex construction environments. From our review of existing literature, we recommend the Faster R-CNN algorithm as the most effective approach for image detection performance in the context of controlled settings and in detecting larger PPE items. Therefore, our paper addresses these challenges and utilizes the Faster R-CNN algorithm as our chosen detection method.

3. METHOD

Table 2 : Experiment Setup

Hardware/Software	Description
Ryzen 5 3600 4.2 Ghz	Processor specification
Microsoft Windows 10	Operating system that is used by the computer to run the project
NVIDIA® GeForce® RTX 2060	Graphic card specification to run Tensorflow software.
16 GB DDR4-2666 (1333 MHz)	RAM and central processing unit specification.
Tensorflow 1.15	Open source software used for training and testing the images. It consists of the Faster R-CNN algorithm. Tensorflow ran inside Anaconda for the usage of Tensorflow-GPU, which is faster than Tensorflow-CPU.
Anaconda	Virtual environment for Python code
LabelImg	Tool written in Python code for graphical image labelling, and for image training and testing.

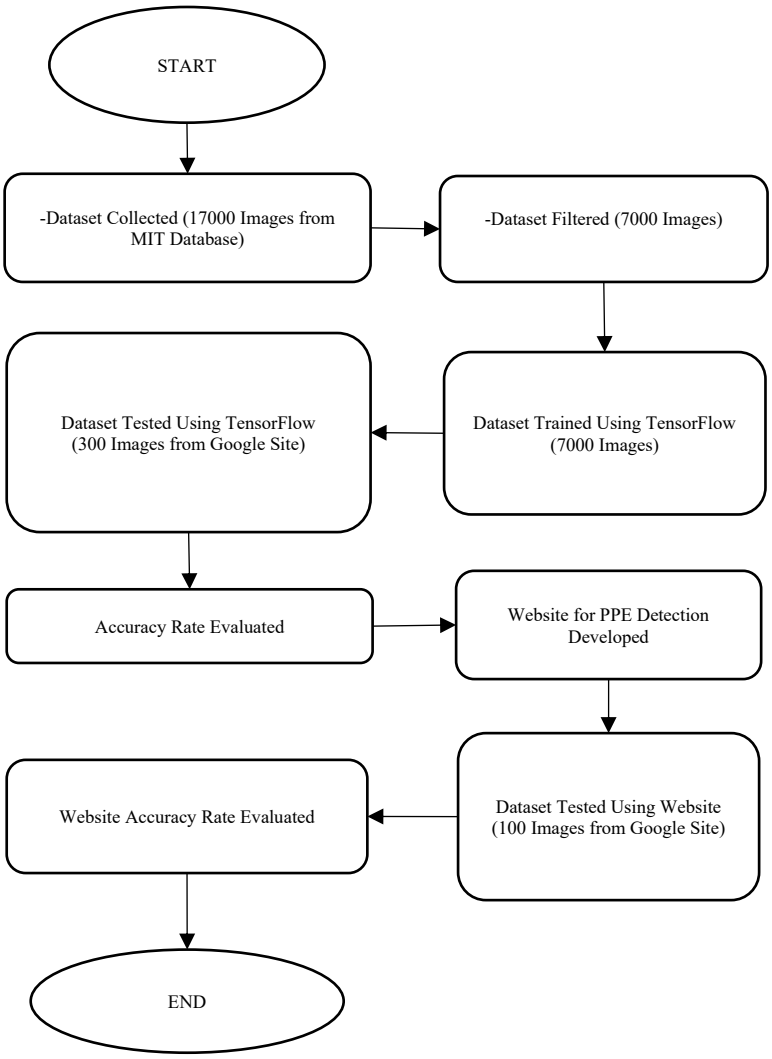

Fig. 1: Overall Research Processes.

The images utilized in this paper were sourced from the MIT Database [30], [31]. After a comprehensive filtering and selection process, which involved removing duplicates, inaccurate images, and low-resolution photos of construction sites, 7,000 images from the MIT Places Database (used for scene recognition) were chosen as training datasets. These images focused on personal protective equipment (PPE) components, including hard hats, vests, boots, and gloves. Additionally, 300 images from Google Sites were set aside for testing purposes, and another 100 images from Google Sites were specifically for testing on the developed website. The labeling of these images was performed using LabelImg, a Python-based tool that features a Qt graphical interface (see Fig. 2). The annotations were saved in XML format following the PASCAL VOC standards. This research categorized images into two classes: safe and unsafe. An image is deemed safe when a worker wears appropriate PPE, such as hard hats, vests, and boots (gloves are optional). Conversely, an image is classified as unsafe when these elements are missing.

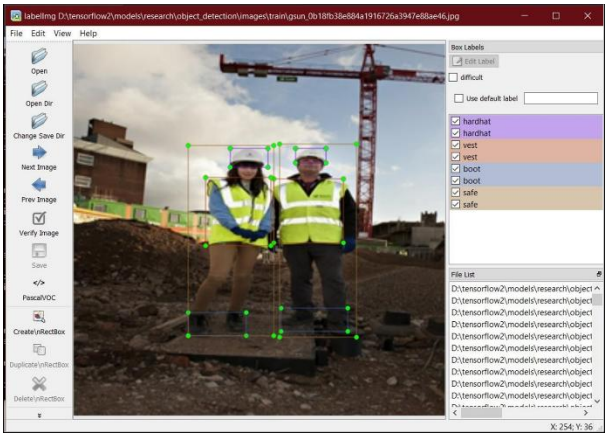

Fig. 2 : Examples of Dataset Labeling of the Safety Condition using LabelImg.

The image was analyzed, trained, and classified using TensorFlow. For this analysis, we employed the Faster R-CNN Inception v2 COCO model to train and classify the images. After completing the training, we tested the model on a test dataset that included 300 images sourced from Google Sites, and we evaluated its accuracy. Additionally, we developed a website for detecting personal protective equipment (PPE). Compared with single-stage detectors, Faster R-CNN demonstrates superior region proposal quality and localisation accuracy, making it particularly effective for precise detection of personal protective equipment in complex construction scenes where high spatial accuracy and reduced false positives are prioritised over real-time performance. To mitigate the risk of adversarial attacks that manipulate visual inputs, this paper incorporates secure data handling and continuous integrity monitoring throughout the training and testing phases to ensure that construction-site PPE images are not maliciously altered prior to model inference. Adversarial robustness is further strengthened through systematic validation of training and testing datasets to detect anomalous patterns that may indicate data poisoning or adversarial perturbations. During model deployment, continuous performance monitoring is implemented to identify abnormal detection behaviours that could arise from adversarial image manipulation or evasion attempts. By integrating secure dataset management, robust model training, and ongoing monitoring mechanisms, the proposed Faster R-CNN–based PPE detection system enhances resilience against adversarial image tampering in safety-critical construction environments.

Using this website, we conducted further testing with 100 images from Google Sites, utilizing real-time webcam detection. Finally, we carried out an evaluation of the accuracy of the built website. We trained the images using the Faster R-CNN Inception V3 model based on the COCO dataset, as illustrated in Fig. 3. The strength of Faster R-CNN lies in its capability to reuse results from the Convolutional Neural Network (CNN) during the regional proposal process. This means that only one CNN needs to be trained, allowing for regional proposals to be generated with minimal additional computational cost.

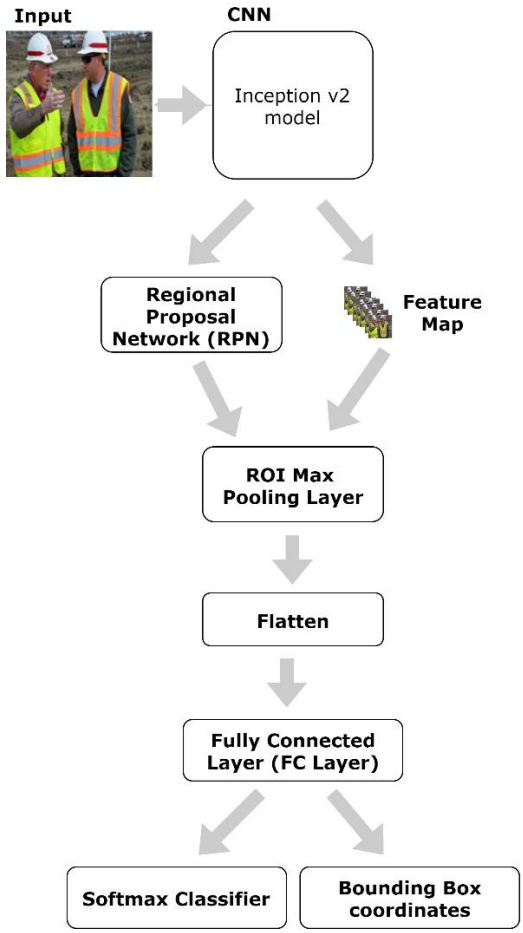

Fig. 3 : Faster R-CNN Design for classification and localization safety at construction.

After an image is input, the Faster R-CNN provides classifications and bounding box coordinates for the specified classes within the images. The architecture of the Faster R-CNN can be divided into four main components: the deep feature extractor (CNN), the Regional Proposal Network (RPN), ROI Pooling, and the classification layer. Labeled images serve as input for the Faster R-CNN. In this case, we utilized Inception V2 as the deep feature extractor, which is one of the architectures of CNN. In the Inception V2 architecture, the convolution layer has a special type of linear operation. In the convolution layer, an array containing numbers (kernel) is implemented on the input as an image containing an array of numbers called tensors. Calculations performed on each element of the kernel matrix and the input tensor are summed to obtain a feature map output. Different kernel sizes produce different extraction results. The calculation for convolution propagation was calculated using equation 1, where x is the input, a is the output after convolution, k is the kernel index, w is the kernel or weight filter, and b is the bias.

$$a_{ij}^{(k)} = \sum_{s=0}^{m-1} \sum_{t=0}^{n-1} W_{st}^{(k)} x_{(i+s)(j+t)} + b^{(k)} \quad (1)$$

By using Inception v2, the dimensions of the feature map can be reduced. Next, the RPN generates region proposals, determining whether the anchor box is a hard hat, a vest, boots, or gloves. Accurate proposal regions are obtained using bounding box regression. Faster R-CNN uses RPN to generate a detection box so that object detection can be done faster. After the proposal region and feature map are generated, the proposal region is used to cut the feature map and generate a set of ROI. These ROIs then pass through the max pooling layer, where the highest value in each filter will be selected, and the other values will be discarded. The output is then flattened to become a 1-dimensional vector array. It is connected to a fully connected layer, which produces the final output, namely the probability values of each class and bounding box coordinates. The probability value of each class is calculated using the softmax activation function because this study tries to solve multi-class problems. The softmax activation function can be calculated using equation 2, where $\vec{z}$ is the input vector, in this case, the output of the fully connected layer, e^{z_i} , which is the standard exponential function for the input vector, e^{z_j} the standard exponential function for the vector output, and k is the number of classes.

$$softmax\,(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \quad (2)$$

Bounding box coordinates specifying the location of each object that needed to be detected. Finally, the probability values of each class and the bounding box coordinates were combined with the original image. This implementation was done using TensorFlow. We used a loss function algorithm for an image to evaluate how much faster the R-CNN modeling of the dataset is. The loss function algorithm is defined in equation 3.

$$L\,(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}\,(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^*\, L_{reg}(t_i, t_i^*) \quad (3)$$

where,
i = Mini-batch anchor index.
p_i = Predicted probability of anchor i being an object.
p_i^* = Ground-truth label. When the anchor is positive, the value is 1.0; otherwise, 0 when the anchor is negative.
t_i = Vector representing the 4 parametrized coordinates of the predicting bounding box.
t_i^* = Ground-truth box associated with a positive anchor.
L_{cls} = Log loss over two classes (object, non-object).
$L_{reg}(t_i, t_i^*)$ = Equal to $R(t_i - t_i^*)$ where R is the robust loss function (smooth L_1).
$p_i^* L_{reg}$ = The regression is activated only when the anchor is positive ($p_i^* = 1$), deactivated when the anchor is negative ($p_i^* = 0$).
cls = Consist of $\{p_i\}$
reg = Consist of $\{t_i\}$.
The term cls is normalized by mini-batch size, which N_{cls} = 256, and the term reg is normalized by the number of anchor locations, such that N_{reg} ~2,400, and both are weighted by the balancing parameter λ. Both terms are equally weighted with default λ = 10. The parameterizations of 4 coordinates for bounding box regression are the following:

$$t_x = \frac{x - x_a}{w_a}\,, \quad t_y = \frac{y - y_a}{h_a} \quad (4)$$
$$t_w = \log(w/w_a)\,, t_h = \log(h/h_a) \quad (5)$$
$$t_i^* = \frac{x^* - x_a}{w_a}\,, \quad t_y^* = \frac{y^* - y_a}{h_a} \quad (6)$$
$$t_w^* = \log(w^*/w_a)\,, t_h^* = \log(h^*/h_a) \quad (7)$$

Where:
x, y = Coordinates of box's center
w, h = Width and height of the box's center
x, y, w, h = Variables for predicted box
x_a, y_a, w_a, h_a = Variables for anchor box
x^*, y^*, w^*, h^* = Variables for ground truth box

In our paper, we used Faster R-CNN to identify and assess safety conditions based on personal protective equipment (PPE) compliance. We classified safety conditions as either "safe" or "unsafe" (dangerous), depending on whether workers were adhering to the requirements for wearing hard hats, vests, boots, and gloves at the construction site, as shown in Fig. 4. After completing this safety classification, we evaluated the results using a dataset of 300 images. The formulation for assessing safety conditions was as follows.

Fig. 4 : Safety Classification Condition.

It is a safe image of a worker's PPE when they wear helmets, vests, and boots (optional gloves), and when they are unsafe when either is inaccessible. The following equation can be used to represent this condition:

$$Z = A \cap B \cap C \cup D \quad (8)$$
$$X = A \cup B \cup C \cup D \quad (9)$$

where,
Z = Safe
X = Unsafe
A = Hardhat/Helmet
B = Vest
C = Boot
D = Gloves

The evaluation of this research is based on accuracy as follows.

$$Accuracy = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} * 100 \quad (10)$$

where:
Tp= True positive (number of worker correctly classified as safe).
Fp= False positive (number of worker incorrectly classified as unsafe).
Tn= True negative (number of worker correctly classified as unsafe).
Fn= False negative (number of workers incorrectly detected as safe.

The findings based on these formulations are explained in next section.

4. RESULTS AND DISCUSSION

Based on the experiment conducted, a total of 7,000 images related to personal protective equipment (PPE) were used to train a model to classify conditions as safe or unsafe. This dataset included hard hats, boots, vests, and gloves. For evaluation, 300 images from Google Sites were utilized, along with 100 images for testing on the developed website.

During the training phase, the model detected 8,953 hard hats, 2,355 pairs of boots, 5,102 vests, and 425 gloves from the 7,000 images. Table III summarizes the experimental results from the data training and the 100 images used for testing, while Fig. 5 illustrates examples of the evaluation image results. The image training process lasted a total of 15 hours, with a total accuracy loss of 0.5 or lower, as shown in Fig. 6. In the evaluation of the 300 images, the model successfully detected 248 hard hats, 89 boots, 109 vests, 20 gloves, and identified 223 safe conditions, resulting in an overall accuracy of 75%.

Table 3: Experiment Results

PPE	Total Detected Images (data training)	Total Detected Images (testing of 100 images)
Hardhat	8953	243
Vest	5102	188
Boots	2355	76
Gloves	425	24
Safe	1026	75

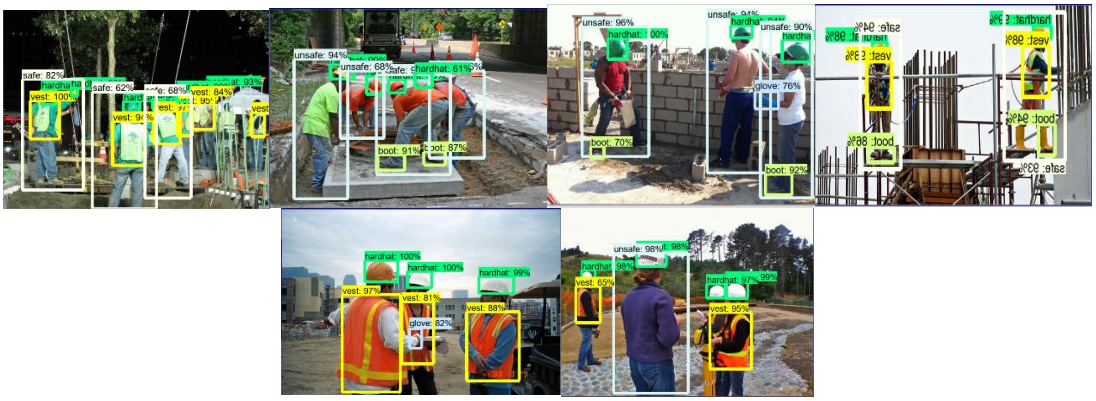

Fig. 5: Examples of the Tested Images.

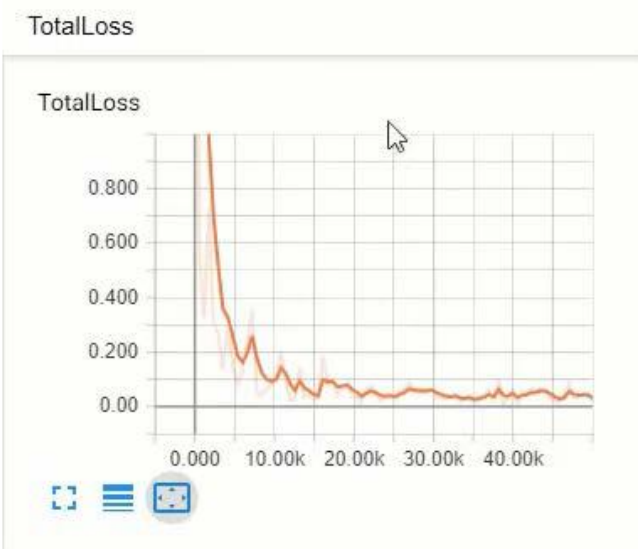

Fig. 6 : Total Loss Accuracy Rate.

Direct comparison with existing studies is not undertaken in this work because the proposed model uniquely incorporates a combined classification of four PPE components: hard hats, safety boots, vests, and gloves

into explicit safe and unsafe conditions, a feature that is not simultaneously addressed in prior research. Unlike most existing PPE detection studies that focus solely on the presence or absence of individual PPE items, this approach evaluates overall worker safety status by aggregating multiple PPE components into a unified safety assessment. During training, the model learned from 7,000 images containing diverse PPE instances, enabling it to distinguish safe and unsafe conditions based on multi-PPE detection rather than isolated object recognition. As a result, the evaluation outcomes, including an overall accuracy of 75% derived from safety condition classification, are not directly comparable with previous works that do not employ a similar safe–unsafe PPE combination framework.

For the additional 100 images tested on the developed website, 75 cases were classified as safe, while 25 were classified as unsafe, yielding an overall accuracy of 67.3%. These results indicate that the developed website and training model can facilitate real-time detection of PPE, making the process faster and more efficient. Several factors influenced the accuracy rate of our model. In addition to the model itself, elements such as the training dataset, input image resolution, and training configurations, including batch size, image resizing, learning rate, and learning rate decay, also played significant roles. In our study, achieving a 70% accuracy rate for detecting safe conditions is considered a strong result for real-time detection. Comparing our results with other existing studies can be challenging and subjective due to the different experimental settings used by each work. Nevertheless, selecting the optimal detection algorithm and configuration is essential for effective image detection, as these two factors strike a balance between speed and accuracy. We opted for the Faster R-CNN algorithm because it demonstrated superior accuracy compared to other algorithms, as shown in Table III. Furthermore, our experimental findings indicate that the developed formulations enable easy identification and measurement of construction workers' compliance with wearing personal protective equipment (PPE).

Furthermore, Table 3 illustrates the detection capability and accuracy trend of the proposed model across different PPE components and safety conditions during both training and testing phases. During training, the model demonstrates strong learning capacity for frequently occurring PPE items, particularly hard hats (8,953 detections) and vests (5,102 detections), indicating robust feature representation for these dominant safety elements. In the testing phase using 100 images, this learning translates into higher detection counts for hard hats (243) and vests (188), reflecting better detection accuracy for PPE items with richer training samples. In contrast, boots (76 detections) and especially gloves (24 detections) show lower detection rates during testing, which can be attributed to their relatively fewer training samples and inherent challenges such as small object size, occlusion, and visual similarity to the background. The “Safe” category, which represents a combined assessment of PPE compliance, records 75 detections in testing, demonstrating the model’s ability to aggregate multiple PPE detections into an overall safety classification. Overall, the table indicates that detection accuracy is positively correlated with the volume and visual clarity of training data, while smaller and less frequent PPE items remain more challenging, highlighting areas for future dataset expansion and model refinement

This paper mitigates adversarial attacks by enforcing secure data handling and integrity checks on all training and testing images to prevent unauthorised manipulation prior to model processing. Robust training strategies, including controlled dataset validation and consistency checks, are applied to reduce the impact of adversarial perturbations and data poisoning. During system operation, continuous monitoring of model performance is implemented to detect abnormal inference patterns that may indicate adversarial evasion attempts. Together, these measures enhance the resilience of the proposed PPE detection system against adversarial attacks in safety-critical construction environments.

5. CONCLUSION

This paper has demonstrated the feasibility of detecting personal protective equipment (PPE) worn by workers through a proposed classification system that differentiates between safe and unsafe PPE images. When tested using 300 construction site images sourced from Google, the proposed model achieved a 75% accuracy rate. However, there were instances where the model failed to identify individuals as workers, and some workers not wearing PPE went undetected. This classification system aims to enhance safety inspections on construction sites and contribute to creating a secure and stable working environment. Furthermore, this paper demonstrates that integrating multi-component PPE detection into explicit safe and unsafe classifications provides a more comprehensive and practical assessment of construction-site safety than conventional presence-based approaches. By incorporating a secure model design with continuous monitoring and data integrity controls, the proposed system effectively mitigates adversarial image manipulation, ensuring reliable PPE compliance detection in safety-critical environments.

REFERENCES

[1] S. Kumar, H. Gupta, D. Yadav, I. A. Ansari, and O. P. Verma, "YOLOv4 algorithm for the real-time detection of fire and personal protective equipments at construction sites," *Multimed. Tools Appl.*, vol. 81, no. 16, pp. 22163–22183, 2022, doi: 10.1007/s11042-021-11280-6.

[2] A. Darda'u Rafindadi, M. Napiah, I. Othman, H. Alarifi, U. Musa, and M. Muhammad, "Significant factors that influence the use and non-use of personal protective equipment (PPE) on construction sites—Supervisors' perspective," *Ain Shams Eng. J.*, vol. 13, no. 3, p. 101619, 2022, doi: https://doi.org/10.1016/j.asej.2021.10.014.

[3] J. Li, X. Zhao, G. Zhou, and M. Zhang, "Standardized use inspection of workers' personal protective equipment based on deep learning," *Saf. Sci.*, vol. 150, p. 105689, 2022, doi: https://doi.org/10.1016/j.ssci.2022.105689.

[4] S. Chen and K. Demachi, "Towards on-site hazards identification of improper use of personal protective equipment using deep learning-based geometric relationships and hierarchical scene graph," *Autom. Constr.*, vol. 125, p. 103619, 2021, doi: https://doi.org/10.1016/j.autcon.2021.103619.

[5] N. D. Nath, A. H. Behzadan, and S. G. Paal, "Deep learning for site safety: Real-time detection of personal protective equipment," *Autom. Constr.*, vol. 112, p. 103085, 2020, doi: https://doi.org/10.1016/j.autcon.2020.103085.

[6] T. Wright, A. Adhikari, J. Yin, R. Vogel, S. Smallwood, and G. Shah, "Issue of Compliance with Use of Personal Protective Equipment among Wastewater Workers across the Southeast Region of the United States," 2019. doi: 10.3390/ijerph16112009.

[7] T. K. M. Wong, S. S. Man, and A. H. S. Chan, "Critical factors for the use or non-use of personal protective equipment amongst construction workers," *Saf. Sci.*, vol. 126, p. 104663, 2020, doi: https://doi.org/10.1016/j.ssci.2020.104663.

[8] G. Masoud and E. Behzad, "Unmanned Aerial Systems (UAS) for Construction Safety Applications," in *Construction Research Congress 2016*, in Proceedings. , 2016, pp. 2642–2650. doi: doi:10.1061/9780784479827.263.

[9] H. Vitharana, S. de silva, and S. De Silva, "Health hazards, risk and safety practices in construction sites – a review study," *Eng. J. Inst. Eng. Sri Lanka*, vol. 48, p. 35, Jul. 2015, doi: 10.4038/engineer.v48i3.6840.

[10] D. Wen, K. Peng, K. Yang, Y. Chen, R. Liu, J. Zheng, A. Roitberg, D. P. Paudel, L. V. Gool, and R. Stiefelhagen, "RoHOI: Robustness Benchmark for Human-Object Interaction Detection," *arXiv*, Jul. 2025. doi: 10.48550/arxiv.2507.09111.

[11] S. Kumar, M. Poyyamozhi, B. Murugesan, N. Rajamanickam, R. Alroobaea, and W. Nureldeen, "Investigation of Unsafe Construction Site Conditions Using Deep Learning Algorithms Using Unmanned Aerial Vehicles," *Sensors*, vol. 24, no. 20, p. 6737, 2024. https://doi.org/10.3390/s24206737

[12] D. Dyck and T. Roithmayr, "Great Safety Performance: An Improvement Process Using Leading Indicators," *AAOHN J.*, vol. 52, pp. 511–520, Jan. 2005, doi: 10.1177/216507990405201205.

[13] J. Lin and A. Mills, "Measuring the occupational health and safety performance of construction companies in Australia," *Facilities*, vol. 19, pp. 131–139, Mar. 2001, doi: 10.1108/02632770110381676.

[14] H. Jafri, mohamad wijayanuddin Ali, A. Ahmad, and M. Kamsah, "Effective occupational health and safety performance measurements," Jan. 2005.

[15] S. Salminen, J. Saari, K. Saarela, and T. Räsänen, "Organizational factors influencing serious occupational accidents.," *Scand. J. Work. Environ. Health*, vol. 19, pp. 352–357, Nov. 1993, doi: 10.5271/sjweh.1463.

[16] Construction Industry Development Board (CIDB), "Safety and Health Assessment System in Construction(SHASSIC) CIS 10:2008."

[17] N. H. Abas, N. Yusuf, N. A. Suhaini, N. Kariya, H. Mohammad, and M. F. Hasmori, "Factors Affecting Safety Performance of Construction Projects: A Literature Review," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 713, no. 1, p. 12036, 2020, doi: 10.1088/1757-899X/713/1/012036.

[18] T. D. Smith and D. M. DeJoy, "Safety climate, safety behaviors and line-of-duty injuries in the fire service," *Int. J. Emerg. Serv.*, vol. 3, no. 1, pp. 49–64, Mar. 2014, doi: 10.1108/IJES-04-2013-0010.

[19] Y. Li, Y. Ning, and W. T. Chen, "Critical Success Factors for Safety Management of High-Rise Building Construction Projects in China," *Adv. Civ. Eng.*, vol. 2018, no. 1, p. 1516354, Jan. 2018, doi: https://doi.org/10.1155/2018/1516354.

[20] M. O. Sanni-Anibire, A. S. Mahmoud, M. A. Hassanain, and B. A. Salami, "A risk assessment approach for enhancing construction safety performance," *Saf. Sci.*, vol. 121, pp. 15–29, 2020, doi: https://doi.org/10.1016/j.ssci.2019.08.044.

[21] K. H. H. Adel Khelifi, "A Mobile Device Software to Improve Construction Sites Communications "MoSIC," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 11, 2016, doi: 10.14569/IJACSA.2016.071108.

[22] P. Nguyen *et al.*, "Construction Project Quality Management using Building Information Modeling 360 Field," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 11, Jan. 2018, doi: 10.14569/IJACSA.2018.091028.

[23] S. Bang and H. Kim, "Context-based information generation for managing UAV-acquired data using image captioning," *Autom. Constr.*, vol. 112, p. 103116, 2020, doi: https://doi.org/10.1016/j.autcon.2020.103116.

[24] M. Abbas, B. E. Mneymneh, and H. Khoury, "USE OF UNMANNED AERIAL VEHICLES AND COMPUTER VISION IN CONSTRUCTION SAFETY INSPECTIONS," *Proc. Int. Struct. Eng. Constr.*, vol. 3, Oct. 2016, doi: 10.14455/ISEC.res.2016.183.

[25] Y. Ham, K. K. Han, J. J. Lin, and M. Golparvar-Fard, "Visual monitoring of civil infrastructure systems via camera-equipped Unmanned Aerial Vehicles (UAVs): a review of related works," *Vis. Eng.*, vol. 4, no. 1, p. 1, 2016, doi: 10.1186/s40327-015-0029-z.

[26] W. Fang, L. Ding, B. Zhong, P. E. D. Love, and H. Luo, "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach," *Adv. Eng. Informatics*, vol. 37, pp. 139–149, 2018, doi: https://doi.org/10.1016/j.aei.2018.05.003.

[27] Q. Fang *et al.*, "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos," *Autom. Constr.*, vol. 85, pp. 1–9, 2018, doi: https://doi.org/10.1016/j.autcon.2017.09.018.

[28] D. Gil, G. Lee, and K. Jeon, "Classification of Images from Construction Sites Using a Deep-Learning Algorithm," in *Proceedings of the 35th International Symposium on Automation and Robotics in Construction (ISARC)*, J. Teizer, Ed., Taipei, Taiwan: International Association for Automation and Robotics in Construction (IAARC), Jul. 2018, pp. 176–181. doi: 10.22260/ISARC2018/0024.

[29] H. Luo, M. Wang, P. K.-Y. Wong, and J. C. P. Cheng, "Full body pose estimation of construction equipment using computer vision and deep learning techniques," *Autom. Constr.*, vol. 110, p. 103016, 2020, doi: https://doi.org/10.1016/j.autcon.2019.103016.

[30] M. M. Saudi *et al.*, "Image Detection Model for Construction Worker Safety Conditions using Faster R-CNN," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 6, pp. 246–250, Jun. 2020, doi: 10.14569/IJACSA.2020.0110632.

[31] "MIT Places Database for Scene Recognition." [Online]. Available: ttp://places.csail.mit.edu/index.html