

# Image Detection Model for Construction Worker Safety Conditions using Faster R-CNN

Madiah Mohd Saudi<sup>1</sup>, Aiman Hakim Ma'arof<sup>2</sup>, Azuan Ahmad<sup>3</sup>, Ahmad Shakir Mohd Saudi<sup>4</sup>, Mohd Hanafi Ali<sup>5</sup>  
Anvar Narzullaev<sup>6</sup>, Mohd Ifwat Mohd Ghazali<sup>7</sup>

Cyber Security and Systems (CSS) Research Unit, Islamic Science Institute (ISI)  
Universiti Sains Islam Malaysia (USIM), Nilai, Malaysia<sup>1,3</sup>

Faculty of Science & Technology (FST), Universiti Sains Islam Malaysia (USIM), Nilai, Malaysia<sup>2,6,7</sup>  
Institute of Medical Science Technology, Universiti Kuala Lumpur (UNIKL), Kuala Lumpur, Malaysia<sup>4</sup>

Consultation, Research and Development Department<sup>5</sup>

National Institute of Occupational Safety and Health (NIOSH), Bangi, Malaysia<sup>5</sup>

**Abstract**—Many accidents occur on construction sites leading to injury and death. According to the Occupational Safety Health Administration (OSHA), falls, electrocutions, being struck-by-objects and being caught in or between an object were the four main causes of worker deaths on construction sites. Many factors contribute to the increase in accidents, and personal protective equipment (PPE) is one of the defense mechanisms used to mitigate them. Thus, this paper presents an image detection model about workers' safety conditions based on PPE compliance by using the Faster Region-based Convolutional Neural Networks (R-CNN) algorithm. This experiment was conducted using Tensorflow involving 1,129 images from the MIT Places Database (from Scene Recognition) as a training dataset, and 333 anonymous dataset images from real construction sites for evaluation purposes. The experimental results showed 276 of the images being detected as safe, and an average accuracy rate of 70%. The strength of this paper is based on the image detection of the three PPE combinations, involving hardhats, vests and boots in the case of construction workers. In future, the threshold and image sharpness (low resolution) will be two main characteristics of further refinement in order to improve the accuracy rate.

**Keywords**—PPE; OSH; accident; construction site; image detection; faster R-CNN

## I. INTRODUCTION

One of the most dangerous fields to work is the construction industry. The Occupational Safety and Health Administration (OSHA) has outlined safety measures and precautions in the form of a legislative framework for the construction industry. Based on works by [1-4], any worker, especially those in the construction industry, is exposed and vulnerable to accidents that could lead to non-permanent disability (NPD), permanent disability (PD) or death. As stated in work by [5], ignorance of safety procedures and protection such as wearing PPE, failing to understand written safety rules, and many migrant workers, are among the factors that lead to accidents. On top of that, the uniqueness of the industry and of construction site conditions, also play a big role in the cause of accident or death. Moreover, workers on construction sites could help to reduce the risk of accidents by informing their supervisor or employer of any risks that they have spotted so that appropriate control measures could be introduced to

prevent such accidents. Generally, safety performance is measured based on lagging indicators such as Incident Rate (IR), Accident Rate (AR) and Experience Modification Rate (EMR). While works by [6-10], described the examples of safety measures available with different lagging indicators across the world. In addition, Abas and colleagues wrote a comprehensive paper on the factors that affects safety performance on construction projects [11]. These authors identified safety factors which are beneficial when it comes to reducing accident and compensation costs and to increasing productivity, employee awareness attitudes, and project on time completion. As proposed by [12,13], employers should evaluate their employees' knowledge and awareness regarding PPE and other equipment at the construction site, so that proper training and control measures could be implemented to reduce the accident or death risk.

Consequently, this paper presents an image detection model based on safe and dangerous condition facing workers on construction sites in terms of their compliance when it comes to wearing PPE. Existing works for construction sites on safety detection were more focusing on one PPE only such as hardhats, vest or boots. As for our paper, we proposed detection safety conditions based on three combinations of PPEs in the form of hardhats, vests and boots. Furthermore, for us to classify the worker as operating in a safe or dangerous manner, we used the Faster R-CNN algorithm. The PPE considered takes the form of hardhats, boots and vests. These are basic PPE that should be worn all time by construction workers.

This paper is organized as follows. Section II explains related work, Section III presents the method used in this research, Section IV consists of the findings and their evaluation, and Section V concludes the paper and makes suggestions for future work.

## II. RELATED WORKS

According to work by [14], falling objects among construction workers is ranked as the one of the highest incident that occurred at construction sites. Hence, we need a solution to lower the death risk among such workers. There is a small number of existing works related to the construction system and to quality of project management such as [15,16,

23]. Work by [15] tended to focus on project progress monitoring, and communication between employees, while work by [16] proposed a quality management project for construction projects. Work by [23] proposed optimisation modelling for repetitive works on construction site by using an Unmanned Aerial Vehicle (UAV), and work by [26] detected workers on construction sites by using UAV and RFID concepts. Different scopes of work using UAVs were summarised by Ham and colleagues [27] such as for progress monitoring, building inspection, building measurement, surveying, safety inspection, structural damage assessment and geo-hazard investigation. As far as safety inspection concerned, existing works tend to focus on one piece of PPE only in the form of the hardhat. There are few existing studies related to image detection about construction sites, as summarised in Table I.

TABLE I. SUMMARISATION OF EXISTING WORKS RELATED TO CONSTRUCTION SITES

Author	Detection Method	Dataset	Challenges	Strength
[17]	Used: ResNet-152 and Faster R-CNN.	Training: ImageNet & MS COCO 2014 dataset. Testing: 3,241 images	Limited to features of construction worker's body only	Considering varying poses of the images
[18]	Used: Deep CNN and Faster R-CNN.	Training: 81,000 images. Testing: 19,000 images. Both were self-collected from 25 different construction projects.	Dedicated to existing construction workers.	Produce highly accurate results based on image size with average of more than 95%
[19]	Used: Google Inception v3	Training: 1208 images Testing: 27 images	It needs bigger image dataset.	Produced accuracy rate of 90%.
[20]	Used: Enhanced Faster R-CNN	Training: 9,500 images Testing: 1500 images	Full coverage of image sources.	Produced accuracy rate of more than 95%.
[21]	Used: Histogram of oriented gradients (HOG)	Training: 100 images (hardhat) & 1,800 dataset (people condition) Testing: Hundreds of self-collected images.	Limited to hardhat and worker images in standing position.	Achieved overall 94.3% precision.
[22]	Used: HOG and Circle Hough Transform (CHT).	Training: 954 images. Testing: 200 images.	Improvement for image detection.	Detection based on colours.
This research	Used: Faster R-CNN.	Training: MIT Places Database 1129 images Testing: 333 images (263 MIT, 65 self-collected, 5 google)	Improvement of accuracy rate based on picture size refinement.	Detection safety conditions based on 3 combinations of PPEs in the form of hardhats, vests and boots.

Based on Table I and on our summarised works, we have identified that the most common challenges for researchers in the future would be to have more PPE detection on construction sites or at the workplace, more dataset resources, the detection of object blockage endangering the target image, and different image positions and backgrounds. From the existing works, the best recommendations for image detection algorithm performance is the use of the Faster R-CNN algorithm. Hence, our research has tackled these challenges and used the Faster R-CNN algorithm as our detection algorithm.

### III. METHOD

For this research, the setup, software and hardware used for the experiment as displayed in Table II, and the research processes are as depicted in Fig. 1.

TABLE II. EXPERIMENT SETUP

Hardware/ software	Description
Lenovo Legion i7-7700HQ @ 2.80 GHz	Computer specification.
Microsoft Windows 10 Home	Operating system that is used by the computer to run the project
NVIDIA® GeForce® GTX 1050 Ti	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.

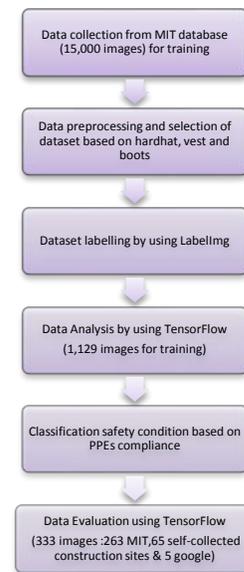


Fig. 1. Overall Research Processes.

For this research, the images were collected from the MIT Database [24]. From fifteen thousand construction images, 1,129 images were selected as the training dataset based on PPE components, which inclusive of hardhats, vests, and boots. These images were then labelled using LabelImg python scripting (see Fig. 2) and further analysed, trained and classified using Tensorflow. LabelImg is written in Python in order to label the images, together with Qt graphical interface. The annotations were saved as an XML file in PASCAL VOC. During the image analysis, we trained and classified the images by using the Faster R-CNN Inception v2 COCO model. This model uses fast R-CNN with shared convolutional feature layers, and a unified model composed of RPN (region proposal network) as depicted in Fig. 3.

The strength of Faster R-CNN is based on its ability to reuse the CNN results for the regional proposal process. Hence, only one CNN needs to be trained, and regional proposals can be made almost cost-free computationally [25]. Once the image has been inserted, the Faster R-CNN produces the classifications and bounding box co-ordinates of the specified classes in the images. In our research, this algorithm helps us to identify and to assign the safety condition based on PPE compliance. The safety condition was decided upon, as being either safe or unsafe (dangerous) based on worker compliance in terms of wearing a hardhat, vest and boots at the construction site as depicted in Fig. 4. Once this safety classification was completed, the evaluation was carried out with the aid of 333 images. For safety conditions, the formulation was as follows.

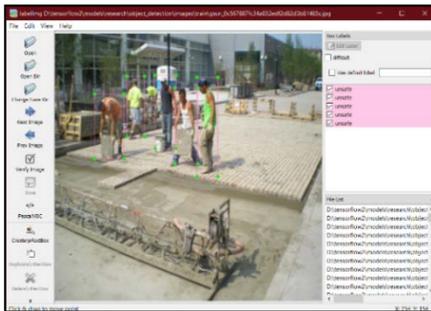


Fig. 2. Examples of Dataset Labelling of the Safety Conditions.

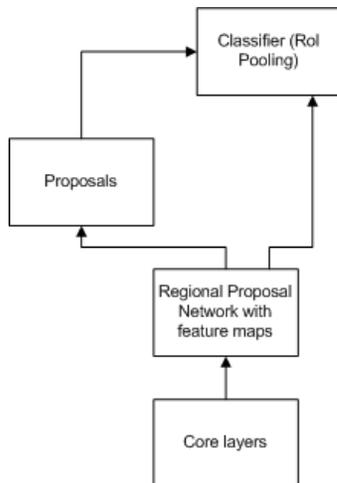


Fig. 3. Faster R-CNN Design.



Fig. 4. Safety Classification Condition.

For PPE, there are three main variables, which are hardhat, vest and boots. The formula is as follows:

$$\text{Let } \alpha_i \text{ be a hardhat I, and } \alpha = \bigcap_{i=1}^n \alpha_i, \beta_j \text{ be a vest j,}$$

$$\text{and } \beta = \bigcup_{i=1}^m \beta_i, \gamma_k \text{ be as boots } \gamma = \bigcap_{i=1}^p \gamma_i$$

Let PPE be the PPE classification and T be the target image. S is the safety model and it can be defined as the following function:

$$f(PPE, T) = S \tag{1}$$

where,  $PPE(\alpha, \beta, \gamma) = \alpha \cap \beta \cap \gamma$

$$f(PPE_i, T_j) = S_{ij} \tag{2}$$

where, PPE represents the PPE classification, T represents the target image and S is the safety model.

$$PPE(\alpha, \beta, \gamma) = \alpha \cap \beta \cap \gamma \tag{3}$$

$$\alpha = \alpha_1 \cup \alpha_2$$

$$\beta = \beta_1 \cup \beta_2$$

$$\gamma = \gamma_1 \cup \gamma_2$$

$$\begin{bmatrix} PPE_i & \alpha & \beta & \gamma \\ \vdots & \ddots & \ddots & \vdots \\ PPE_n & \dots & \delta_n & \end{bmatrix}$$

where,  $\alpha_1, \alpha_2$  : with hardhat, without hardhat

$\beta_1, \beta_2$  : with vest, without vest

$\gamma_1 \cup \gamma_2$  : with boots, without boots

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

$$Accuracy = \frac{Tp+Tn}{Tp+Tn+FP+FN} \times 100 \tag{4}$$

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 the next section.

#### IV. FINDINGS

Based on the experiment conducted, 1,129 images related to PPE that included hardhat, vest and boots were trained in order to classify conditions as being safe or unsafe. Then, 333 anonymous self-collected images from construction sites were used for evaluation.

During the training, from 1,129 images, a total of 2,373 hardhats, 1,023 pairs of boots and 1,478 vests were detected. In terms of the evaluation of 263 images, a total of 156 hardhats, 49 boots, 73 vests and 123 safe conditions were detected with an overall accuracy of 70%. In the case of a further 70 images self-collected from construction sites, 53 cases were detected as being safe. Table III summarises the experimental results, while Fig. 5 shows examples of the evaluation image results. For this experiment, a total of 6 hours was used for image training, and the total accuracy loss was 0.5 or less, as displayed in Fig. 6.

Many factors contributed to the accuracy rate. Apart from the model itself, other factors such as the training dataset, input image resolution, and training configurations including batch size, input image resize, learning rate, and learning rate decay, also affected the accuracy rate [28]. In our case, the ability to detect safe conditions with an accuracy rate of 70% is considered as a good result in terms of real-time detection. It is hard and very subjective to make comparisons with any other existing works due to the different settings of the experiment for each existing work. However, the selection of the best detector algorithm and the best configuration are crucial in terms of image detection. These two factors contribute to the best balance of speed and accuracy. We chose the Faster R-CNN algorithm due to its better accuracy rate compared to that of other existing algorithms, as summarised in Table I. In addition, based on the experiment conducted, the formulations developed mean that the construction workers' compliance with wearing PPE can easily be identified and measured.

TABLE III. EXPERIMENT RESULTS

PPE	Total Detected Images (Training)	Total Detected Images (Testing of 70 Images)
Hardhat	2372	156
Vest	1478	73
Boot	1023	49
Safe	572	53



Fig. 5. Examples of the Tested Images.



Fig. 6. Total Loss Accuracy Rate.

#### V. CONCLUSION

Based on the experiment conducted, there are a few considerations. These are the threshold value assigned during the data configuration settings, the momentum optimizer value, and the belief that image resize and image sharpness (low resolution) could be further adjusted or improved for better accuracy. All these elements are among the components of the Faster R-CNN algorithm. Nonetheless, this paper has successfully developed formulations and an image detection model relating to construction workers' compliance when it comes to wearing PPE in the workplace. This will help create a safer and healthier environment on construction sites. For future work, the threshold, momentum optimizer, image resize and image sharpness will be further refined and improved to obtain an improved accuracy rate.

#### ACKNOWLEDGMENT

The authors would like to express their gratitude to National Institute Of Occupational Safety And Health (NIOSH) Malaysia, UNIKL and Universiti Sains Islam Malaysia (USIM) for the support and facilities provided. This research paper project is under NIOSH grant: [03-16/03/DRONE(E)/2019/01] / [USIM/NIOSH-UNIKL/FST/052003/42119].

#### REFERENCES

- [1] Nath, Nipun D., Amir H. Behzadan, and Stephanie G. Paal. "Deep learning for site safety: Real-time detection of personal protective equipment." *Automation in Construction*, 112, 103085, 2020.
- [2] Wright, Tamara, Atin Adhikari, Jingjing Yin, Robert Vogel, Stacy Smallwood, and Gulzar Shah. "Issue of compliance with use of personal protective equipment among wastewater workers across the southeast region of the United States." *International Journal of Environmental Research and Public Health*, 16, no. 11, 2019.
- [3] Wong, Tom Ka Man, Siu Shing Man, and Alan Hoi Shou Chan. "Critical factors for the use or non-use of personal protective equipment amongst construction workers." *Safety science*, 126: 104663, 2020.
- [4] M. Gheisari and B. Esmaeili, "Unmanned Aerial Systems (UAS) for Construction Safety Applications," in *Construction Research Congress 2016: Old and New Construction Technologies Converge in Historic San Juan - Proceedings of the 2016 Construction Research Congress*, CRC 2016, 2016, pp. 2642–2650.
- [5] V. H. P. Vitharana, G. H. M. J. 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, no. 3, p. 35, Jul. 2015.
- [6] Dyck D and Roithmayr T., "Great Safety Performance: an improvement process using leading indicators. - PubMed - NCBI," *AAOHN J.*, vol. 52, no. 12, pp. 511–520, Dec-2004.
- [7] 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.
- [8] Hasnora Jafri, Mohd Wijayanuddin Ali, Arshad Ahmad, and Mohd Zaki Kamsah, "Effective occupational health and safety performance measurements," in *Int. Conf. of Chemical and Biochemical Eng. (Sabah)*, 2005, pp. 702–708.
- [9] S. Salminen, J. Saari, K. L. Saarela, and T. Rasanen, "Organizational factors influencing serious occupational accidents," *Scand. J. Work. Environ. Heal.*, vol. 19, no. 5, pp. 352–357, 1993.
- [10] Construction Industry Development Board (CIDB), "Safety and Health Assessment System in Construction (SHASSIC) CIS 10:2008," 2008.
- [11] 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," in *IOP Conf. Series: Materials Science and Engineering* 713 (2020) 012036, 2020.
- [12] 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, Jan. 2014.
- [13] Y. Li, Y. Ning, W. T. Chen, and D. W. M. Chan, "Critical Success Factors for Safety Management of High-Rise Building Construction Projects in China," 2018.
- [14] Sanni-Anibire, Muizz O., Abubakar S. Mahmoud, Mohammad A. Hassanain, and Babatunde A. Salami. "A risk assessment approach for enhancing construction safety performance." *Safety science*, vol. 121, pp. 15–29, 2020.
- [15] A. Khelifi and K. Hesham Hyari, "A Mobile Device Software to Improve Construction Sites Communications 'MoSIC,'" *IJACSA Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 11, 2016.
- [16] P. Thanh Nguyen et al., "Construction Project Quality Management using Building Information Modeling 360 Field," *IJACSA Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 10, 2018.
- [17] H. Son, H. Choi, H. Seong, and C. Kim, "Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks," *Autom. Constr.*, vol. 99, pp. 27–38, Mar. 2019.
- [18] 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, Aug. 2018.
- [19] D. Gil, G. Lee, and K. Jeon, "Classification of Images from Construction Sites Using a Deep-Learning Algorithm," in *35th International Symposium on Automation and Robotics in Construction (ISARC 2018)*, 2018, pp. 176–181.
- [20] Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T.M. and An, W., "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos," *Autom. Constr.*, vol. 85, pp. 1–9, Jan. 2018.
- [21] M.-W. Park, N. Elsafty, and Z. Zhu, "Hardhat-Wearing Detection for Enhancing On-Site Safety of Construction Workers," *J. Constr. Eng. Manag.*, vol. 141, no. 9, p. 04015024, Sep. 2015.
- [22] A. H. M. Rubaiyat et al., "Automatic Detection of Helmet Uses for Construction Safety," in *2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW)*, 2017, pp. 135–142.
- [23] S. Bang and H. Kim, "Context-based information generation for managing UAV-acquired data using image captioning," *Autom. Constr.*, vol. 112, p. 103116, Apr. 2020.
- [24] "MIT Places Database for Scene Recognition." [Online]. Available: <http://places.csail.mit.edu/index.html>.
- [25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
- [26] M. Abbas, B. Eddine Mneymneh, and H. Khoury, "USE OF Unmanned Aerial Vehicles and Computer Vision in Construction Safety Inspections," in *Proceedings of the Third Australasia and South-East Asia Structural Engineering and Construction Conference (ASEA-SEC-3)*, 2016.
- [27] Ham, Youngjib, Kevin K. Han, Jacob J. Lin, and Mani Golparvar-Fard. "Visual monitoring of civil infrastructure systems via camera-equipped Unmanned Aerial Vehicles (UAVs): a review of related works." *Visualization in Engineering* 4, no. 1 (2016): 1.
- [28] Jonathan Hui, "Object detection: speed and accuracy comparison (Faster R-CNN, R-FCN, SSD, FPN, RetinaNet and YOLOv3)," 2018.[Online]. Available: [https://medium.com/@jonathan\\_hui/object-detection-speed-and-accuracy-comparison-faster-r-cnn-r-fcn-ssd-and-yolo-5425656ae359](https://medium.com/@jonathan_hui/object-detection-speed-and-accuracy-comparison-faster-r-cnn-r-fcn-ssd-and-yolo-5425656ae359).