The XXXIVth Annual International Occupational Ergonomics and Safety Conference
Virtual Conference
September 15-16, 2022

Vision-Based Personal Protective Equipment (PPE) Recognition for Construction Worker Safety Compliance Monitoring

Shashank Muley[1*], Chao Wang[1], Srikanth Sagar Bangaru[2], Fereydoun Aghazadeh[3]

[1]Bert S. Turner Department of Construction Management, Louisiana State University, Baton Rouge, LA, USA

[2]ISC Constructors LLC, Baton Rouge, Louisiana, USA

[3]Department of Mechanical and Industrial Engineering, Louisiana State University, Baton Rouge, LA, USA

* Corresponding Author's Email: smuley1@lsu.edu

Author Note: Shashank Muley is currently a Ph.D. student in the Bert S. Turner Department of Construction Management, Louisiana State University. **Chao Wang** is an Associate Professor in the Bert S. Turner Department of Construction Management, Louisiana State University. **Srikanth Sagar Bangaru** is a Corporate Project Controls Data Scientist at ISC Constructors LLC. **Fereydoun Aghazadeh** is a Professor in the Department of Mechanical and Industrial Engineering, Louisiana State University.

Abstract: The construction industry environment is volatile and dynamic in nature. Industries' reliance on technology and self-sustained tools that can facilitate participants' ability to perform activities with high productivity and resolute safety is increasing. One of the areas where this advancement of automation in construction is strengthening its stronghold is Construction Safety. Construction workers are vulnerable to injuries due to the nature of their work. Object recognition and machine learning (MI) are frequently used techniques in automation. PPE (Personal Protective Equipment) recognition is a well-explored area in MI for safe practice and procedural detection. Helmet or Safety Hard Hat which is one of the most vital and enforced PPE to protect participants from head injuries and fatal traumatic incident caused by a falling object, trip, and fall or collision like stuck-by an object have caught the attention of MI. Still, the ability to recognize the craft level of a participant on a site-specific project has not been explored yet. This study is intended to reduce that gap by presenting a deep learning model built on Detectron 2 a computer vision model zoo using faster R-CNN architecture. This ability to recognize PPE and participants' craft will facilitate safety professionals to identify PPE compliance with craft specifications to include an additional layer of details in safety assessment surveillance. Developed Vision-Based Personal Protective Equipment (PPE) Recognition model using 261 high-resolution images to train, validate, and test in 80:10:10 ratios for efficiency of object recognition. Supervision and Pipefitters were the specific craft selected in correspondence to white and other color hard hats. Data to train this model was collected from multiple sources such as crowd-sourced, web-scrapping, and personal captures. Data preprocessing was conducted to bring in robustness, resulting in high recognition accuracy under certain conditions. Among the tested data set, the average precision (AP50) was 85%, making this model satisfactory compared to its counterpart.

Keywords: Object Detection, Machine Learning, Construction Safety

1. INTRODUCTION

Workers who may be in a residential, commercial, or industrial setting are continuously exposed to different levels of risk and hazards at their workplace. This reflects in the U.S. Bureau of Labor Statistics for fatal cases. In 2019, 5333 workers died on their job site, which counts for 3.5 per 100,000 full-time equivalent workers, and every 99 minutes, there was one casualty of a construction worker (BLS, 2020). Due to the high frequency of workplace accidents, safety has become the focal point of many organizations, specifically general contractors, as they must keep the productivity and health of the workforce in compliance with clients' regulations (Singh & Mishra, 2021). A major cause of workplace injuries and accidents in the construction industry is workers' behavior leading to unsafe acts (Han & Lee, 2013). Most of these unsafe acts are the result of unsafe behavior and negligence from construction workers Behavior-based safety (BBS) is an effective way to observe and record participants' unsafe actions (Wirth & Sigurdsson, 2008) but proper use of Personal Protective Equipment (PPE) can make a massive mark in the reduction of accidents and injuries. In the United States, 24% of all construction fatalities were traumatic brain injuries (Colantonio et al., 2009; Konda et al.,2016). Therefore, improving monitoring of unsafe use of PPE and specifically hard hat compliance according to standards will make a drastic difference in traumatic head injuries and accidents for construction workers. On-site safety supervision and monitoring enhance the extent of construction site safety.

ISBN: 9781938496608
DOI: https://doi.org/10.47461/isoes.2022_muley

Developing and using advanced technologies such as computer-aided vision and sensors are predominantly the two methods most adopted techniques in safety supervision that can effectively recognize unsafe acts and behavior (Wei et al., 2019). Computer-aided vision-based technology occupies a dominant position due to feasibility and in comparison, too expensive sensor-based techniques (Wu, J. et al., 2019). The construction safety inspection is largely dependent on manual inspection, which could be time-consuming, error-prone, and costly for large job sites where multiple activities and personnel are simultaneously conducting operations (Akhavian & Behzadan, 2016). The use of vision-based monitoring in collaboration with human expertise can be effective and increase the efficiency of PPE compliance on the job site. Several studies have been published for Hard Hat and PPE compliance detection (Wu, J. et al., 2019) and human identity using face recognition. Although there has been considerable research and techniques developed in PPE recognition specifically for Hard Hat recognition, there is a need to improve the object detection model to improve its classification for safety compliance. This study uses a vision-based detection model to identify site-specific craft considered color to be used as craft identifiers for hard hat safety compliance with satisfactory results.

2. OBJECTIVE

This study aims to develop a machine-learning model for automated PPE compliance checking and craft identification. Even though previous studies have proposed various A.I. models for PPE detection, none of the studies have considered color to be used as a craft identifier for hard hats into consideration. Also, this study improves the performance of the PPE detection model by creating a robust dataset and hyperparameter tuning. This study will facilitate an improved safety monitoring system and compliance techniques.

3. LITERATURE REVIEW

This study aimed to develop a machine-learning model for automated PPE compliance checking and craft identification. Even though the objective of safety monitoring using PPE detection is to mitigate injuries and incidents rates that primarily arise from unsafe acts or conditions. PPE detection techniques can be categorized into sensor-oriented and vision-based techniques. Vision-based techniques use images or videos from the construction site using cameras which are processed in PPE detection models for evaluation. Many researchers have attempted to improve the accuracy, speed, and robustness of these detection models using machine learning and object detection algorithms

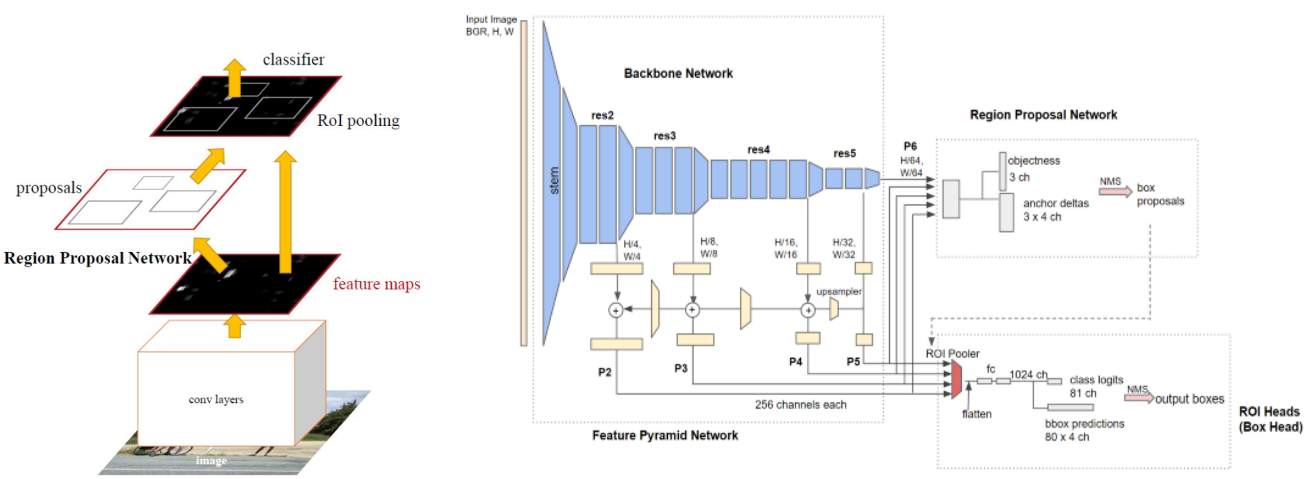

Figure 1: Faster R-CNN Framework (Left), Detectron2 with base R-CNN Architecture (Right)

Use of deep learning techniques to train a model with images of known PPE and then evaluate an unknown image with PPE for an automated monitoring system (Nath and Behzadan, 2020). Capturing moving objects from the video, which are then input into a CNN model for classification of a hard hat, head, and person using K-Nearest Neighbors (KNN) have been

ISBN: 9781938496608
DOI: https://doi.org/10.47461/isoes.2022_muley

used (Wu, H. & Zhao, 2018). Proper use of the helmet was also detected by detecting gestures to determine the location of the helmet on the worker's head (Chen & Demachi, K. 2020).

Object detection techniques have been confined to two prominent families in recent years: Region-Based Convolutional Neural Networks (R-CNNs) and You Only Look Once (YOLO). YOLO with SSD and R-CNN were compared for their detection performance for the same data set (Xie et al., 2018). R-CNN takes a long time to train because training is done in multiple stages. Besides training, the prediction stage is also slow. Therefore, a more sophisticated Faster R-CNN model (Girshick, 2015) was incorporated to tackle these issues. Fast R-CNN is trained as a single model instead of three separate modules, providing faster execution Figure 1 (Left). This architecture takes the images and proposes candidate regions, then passes them through a popular pre-trained image classification model e.g., ResNet (He et al., 2016), VGG-16 (Simonyan & Zisserman, 2014) to extract features from the candidates. The extracted features undergo a Region of Interest (RoI) pooling layer, followed by two fully connected layers. In this study, we used the Detectron2-based R-CNN model which is Facebook A.I. Research's next-generation software system that implements state-of-the-art object detection algorithms Figure 1 (Right).

4. METHODOLOGY

To achieve the objective mentioned in section 2 the methodology adopted in this study is as shown in figure 2. The methodology includes data collection, preparation, model training, and performance evaluation.

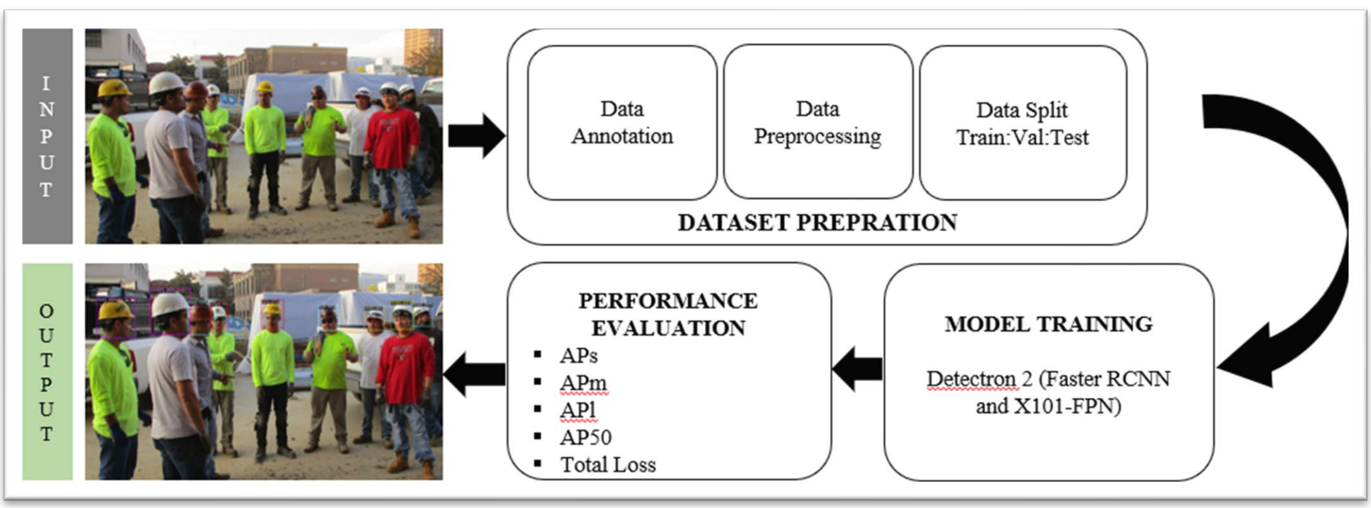

Figure 2 Object Detection Methodology

4.1. Dataset Preparation

A balanced dataset for Safety Hard Hat & Craft detection was developed. Images were collected using image scrapping from google, the public dataset, and individual collection. 261 high-resolution 1280 x 720 pixels and JPEG color format images were selected out of 1200 image sets, including personals wearing a multi-color hard hat. This pre-processing elimination of images was a vital step as it provided us with a diverse image data set with white, blue, and other colors and no hard hat specifications required to train site-specific craft and PPE detection models. Craft-specific hard hats color was manually checked in each image and labeled providing us with four classes (HHPBC- Hard Hat Pipefitter Blue Craft, HHWS- Hard Hat White Supervisor, HHO- Hard Hat Other Craft, NHH- No Hard Hat) (Figure 3) using image labeling tool. The dataset was divided into three sets by the ratio of 80:10:10 data split for training, validation, and testing with 210, 25, and 26 images (Figure 4).

ISBN: 9781938496608
DOI: https://doi.org/10.47461/isoes.2022_muley

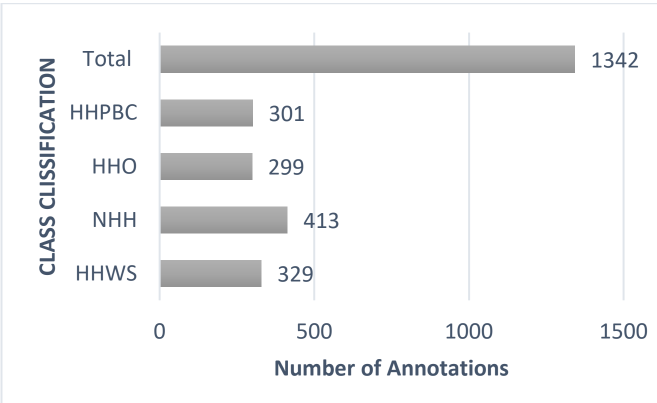

Figure 3: Data set labeling annotation distribution

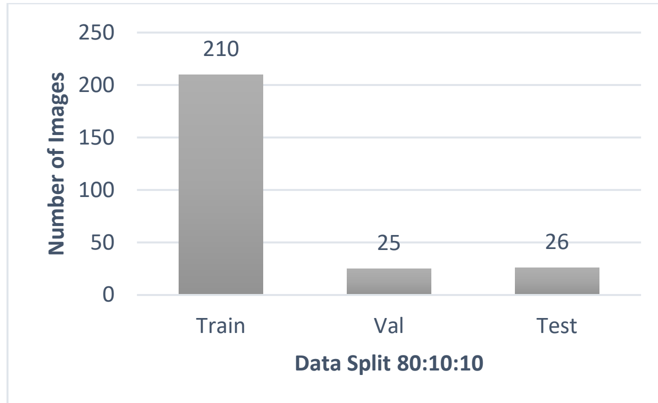

Figure 4: Data split (Train: Val: Test) 80:10:10

The data set was processed to ensure learning and prediction were made on a homogenous set of images. The preprocessing stage included an auto-orient that enables image visualization as they are stored in the storage location. The second step was to resize the images to 416x416 pixels, which helped improve the accuracy and speed of the model compared to the 1280x720 pixel set. Once the processing was completed, generated annotations were exported in COCO JSON format which was further used for training. Detectron2 needs COCO JSON format as a COCO dataset and JSON structure used by Microsoft, Google, and Facebook (Rostianingsih et al., 2020).

4.2. Model Training

Instead of developing a Faster R-CNN model from scratch, Detectron2 was used as an MI platform to speed up the development cycle. Detectron2 is Facebook A.I. Research's next-generation software system that implements state-of-the-art object detection algorithms. It is also a common practice to use a base model pre-trained on a large image set such as ImageNet (Deng et al., 2009) as the feature extractor part of the network. Detectron2 provides many such base models (Detectron2, 2016). However, for Faster R-CNN, two commonly used models are R101-FPN and X101-FPN. These two pre-trained models were selected because of their efficacy in predicting Average Precision (A.P.) compared to others. They are 42% and 43%, respectively. As, X101- FPN had better box A.P. on the ImageNet benchmark it was used for the Faster R-CNN model. Therefore, this approach was implemented on the PPE detection dataset.

The detectron-2-based R-CNN model was trained using Google Colab as a virtual environment which provided access to GPUs with a maximum lifetime of 12 hours for the virtual portal. The dataset for training, validation, and testing included 261 images 12 hours of the runtime was sufficient for model development. Timeframe to train this model was minimal with 1-1.5 hours, well within the maximum lifetime of free google colab access. Model training with multiple hyperparameters is the most time-consuming aspect of object detection. After trying multiple hyperparameters the best results for this model were achieved by following hyperparameters as described in Table 1.

Table 1: Hyperparameters used for model training

Iterations	2000
Images per Batch	44
Gamma	0.05
Number of Classes	5
Base Learning Rate	0.001

ISBN: 9781938496608
DOI: https://doi.org/10.47461/isoes.2022_muley

4.3. Model Evaluation

Model evaluation is conducted based on the model's performance on the test dataset. Metrics used for evaluation are average precision medium (APm), and Intersection-Over-Union (IoU). The IoU and APm are measured as Equations 1 and 2, respectively. Where T.P., F.P., and F.N. are true positive, false positive, and false negative. IoU or Jaccard Index is the intersection ratio over the union of predicted and ground truth labels/annotations. Whereas APm is the average precision percentage for medium objects.

$$IoU(Jaccard\ Index) = \frac{2TP}{2TP + FP + FN} \quad (1)$$

$$IAPm = \frac{1}{N}\sum_{i=1}^{N} API_i \quad (2)$$

5. RESULTS

Results for this hard hat and craft compliance training model can be evaluated by using average precision, training loss, and val. accuracy, val. loss, and average precision (A.P.) as prime parameters defined by the Common Object in Context (COCO) evaluation matrix (COCO, 2021).

Table 2: Average Precisions achieved for training model

AP	AP50	AP75	APs	APm	APl
52.443	85.093	62.227	40.259	58.756	75.305

The Average Precision is the average of multiple intersections over union (IoU) values. AP50 and AP75 represent the average over multiple IoU thresholds with 50% and 75% respectively, COCO uses 10 IoU thresholds with 50%: 5%: 95% for evaluation (COCO, 2021). Using the hyperparameters combination the model trained provided average precision of small objects (A.P.s) of 40.26, average precision of medium objects at 58.77, and average precision of large objects of 75.305 which are satisfactory for the dataset used in this study. AP50 and AP70 which represent 50% and 75% overlap of segmentation mask values were 85.093 and 62.227 (Table 2). 85.093 AP50 value can be interpreted as 85% of the majority of hard hat & craft classes detected in the validation subset. The average precision of the individual class was in the range of 46 to 56, with the lowest being HHWS (Hard Hat White Supervision) and the highest being HHO (Hard Hat Other Craft) (Table 3).

Table 3: Average Precisions of Individual Class

Category	AP
HHO	55.292
HHPBC	54.099
HHWS	46.941
NHH	53.439

ISBN: 9781938496608

DOI: https://doi.org/10.47461/isoes.2022_muley

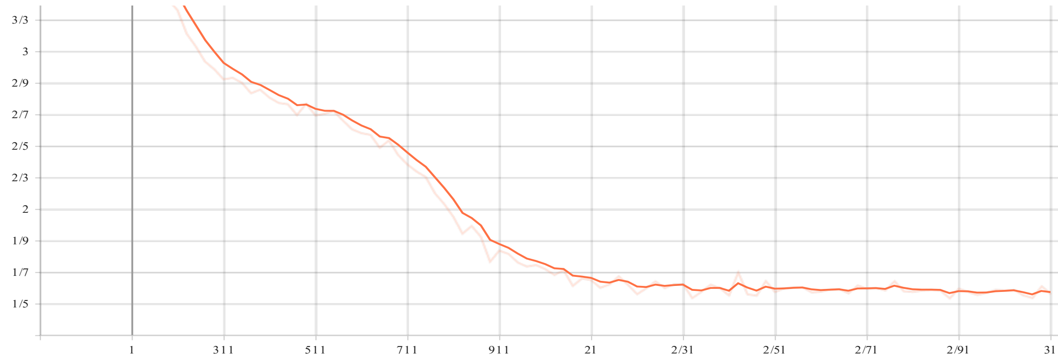

Figure 5: (a)

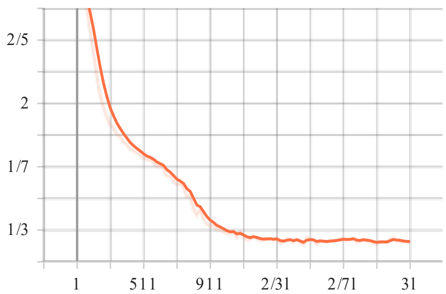

Figure 5(b)

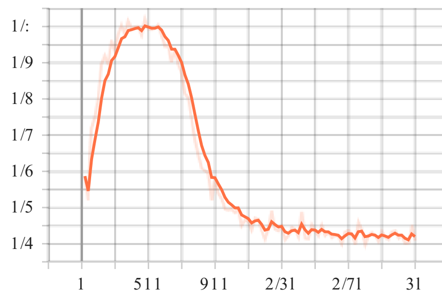

Figure 5(c)

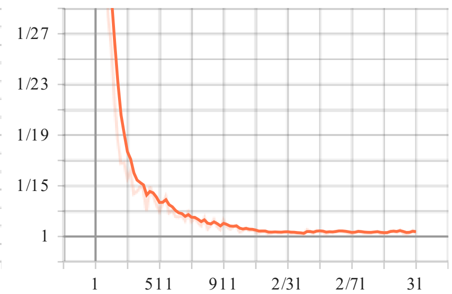

Figure 5(d)

Table 4: Evaluation Parameters of the trained model

Figure	**(5a)**	**(5b)**	**(5c)**	**(5d)**
Iterations	**Total Loss**	**Loss_cls**	**Loss_box_reg**	**Loss_rpn_cls**
2000	0.373	0.085	0.254	0.003

Total Loss after 2000 iterations reached 0.373 which does not show signs of overfitting and its value is reduced over time (Table 4). A similar trend was observed in loss per class which decreased over time to 0.085 but did not reduce to zero (Table 4). Localization loss of model loss_box_reg which is the predicted location of AOI vs true location was at 0.254 plateauing at the 2000th iteration (Figure 5c). Loss in the region proposed network reached 0.003 which is the permissible limit (Figure 5d). In nutshell, the evaluation of this trained and tested model depicted satisfactory results.

6. CONCLUSION

This study explored the object detection model for PPE compliance and craft identification of construction workers using hard hats and color as craft classification for detection. Detectron 2's faster R-CNN is implemented with an in-house developed and annotated safety hard hat and craft classification dataset. After preprocessing stage, model training with hyperparameters described in (Table 1) provides a detection accuracy of 85.093 for AP50. Although the model's detection capabilities in this research are satisfactory, robust detection models for dynamic construction environments still need to be made. Figure 6 showcases model accuracy on random site images collected.

Figure 6: Results of model accuracy with some false positive on-site images

7. DISCUSSION

To mitigate construction injuries and fatalities, it is vital to bolstering policies regarding the proper use of PPE as they are the final line of defense against site incidence, body injury, or fatalities. This study addresses the importance of vision-based object detection techniques which will supplement improved PPE monitoring and compliance for a hard hat with craft detection using hard hat color as a classification. This would also facilitate safety managers to deploy an alert notification system that will notify safety managers if non-essential craft personnel is entering a high-risk craft-specific work area on the construction site. As occupational safety and health regulatory bodies such as OSHA, NIOSH does not have a specific policy regarding the appearance or color of hard hats worn by construction workers as their general standards. To better implement, this model for multiple sites and construction environments future research should improve dataset adaptability for different disciplines and construction sites. To improve PPE detection effectiveness and robustness, assembling a dataset that can be trained to adapt to changing construction environments, sites, and PPE specifications is something future research should focus on.

8. REFERENCES

Akhavian, R. and Behzadan, A.H. (2016). Smartphone-Based Construction Workers' Activity Recognition and Classification. Automation in Construction, 71, 198-209.

BLS, (2020, December). One worker died every 99 minutes from a work-related injury in 2019. Retrieved from: https://stats.bls.gov/opub/ted/2020/one-worker-died-every-99-minutes-from-a-work-related-injury-in-2019.htm

Census of fatal occupational injuries summary, 2020. (2021, December 16). Retrieved October 10, 2022, from https://www.bls.gov/news.release/cfoi.nr0.htm

Chen, S. and Demachi, K. (2020). A Vision-Based Approach for Ensuring Proper Use of Personal Protective Equipment (PPE) in Decommissioning of Fukushima Daiichi Nuclear Power Station. Appl. Sci. 2020, 10, 5129.

COCO (2021). "COCO - common objects in context", Cocodataset.org. Accessed on: Dec 21, 2021. [Online]. Available: https://cocodataset.org/

Colantonio, A., McVittie, D., Lewko, J., & Yin, J. (2009). "Traumatic brain injuries in the construction industry," Brain Injury, vol. 23, no. 11, pp. 873-878, 2009.

ISBN: 9781938496608

DOI: https://doi.org/10.47461/isoes.2022_muley

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.

Detectron2, "Detectron2 model zoo," https://github.com/facebookresearch/detectron2/blob/main/MODEL_ZOO.md, accessed: 2022-10-10.

Girshick, R. (2015). Fast r-cnn In: Proceedings of the IEEE international conference on computer vision pp 1440-1448 https://doi. org/10.1109. ICCV.

Han, S.U. and Lee, S.H. (2013) A Vision-Based Motion Capture and Recognition Framework for Behavior-Based Safety Management. Automation in Construction, 35, 131-141. https://doi.org/10.1016/j.autcon.2013.05.001

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Konda, S., Tiesman, H. M., & Reichard, A. A. (2016). Fatal traumatic brain injuries in the construction industry, 2003− 2010. American journal of industrial medicine, 59(3), 212-220.

Nath, N.D. and Behzadan, A.H. (2020). Deep Learning Detection of Personal Protective Equipment to Maintain Safety Compliance on Construction Sites. In Construction Research Congress 2020: Computer Applications; American Society of Civil Engineers: Reston, VA, USA, 2020; pp. 181–190.

Rostianingsih, S., Setiawan, A., & Halim, C. I. (2020). COCO (creating common object in context) dataset for chemistry apparatus. Procedia Computer Science, 171, 2445-2452.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Singh, A. and Mishra, S.C. (2021) Safety Performance & Evaluation Framework in Indian Construction Industry. Safety Science, 134, 105023

Wei, R., Love, P. E., Fang, W., Luo, H., & Xu, S. (2019). Recognizing people's identity in construction sites with computer vision: A spatial and temporal attention pooling network. Advanced Engineering Informatics, 42, 100981.

Wirth, O., & Sigurdsson, S. O. (2008). When workplace safety depends on behavior change: Topics for behavioral safety research. Journal of safety Research, 39(6), 589-598.

Wu, H.; Zhao, J. (2018). Automated Visual Helmet Identification Based on Deep Convolutional Neural Networks; Elsevier Masson SAS:Amsterdam, The Netherlands, 2018; Volume 44, pp. 2299–2304.

Wu, J., Cai, N., Chen, W., Wang, H., & Wang, G. (2019). Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset. Automation in Construction, 106, 102894.

Xie, Z.; Liu, H.; Li, Z.; He, Y. (2018). A convolutional neural network-based approach towards real-time hard hat detection. In Proceedings of the 2018 IEEE International Conference on Progress in Informatics and Computing, Suzhou, China, 14–16 December 2018; pp. 430–434.

ISBN: 9781938496608

DOI: https://doi.org/10.47461/isoes.2022_muley