Computers in Industry 129 (2021) 103448

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Utilizing safety rule correlation for mobile scaffolds monitoring leveraging deep convolution neural networks

Numan Khan [a], Muhammad Rakeh Saleem [b], Doyeop Lee [a], Man-Woo Park [c], Chansik Park [a,*]

[a] School of Architecture and Building Sciences, Chung-Ang University, Seoul, 06974, Republic of Korea
[b] Department of Architectural Engineering, Pennsylvania State University, 16801, United States
[c] Department of Civil and Environmental Engineering, Myongji University, Gyeonggi-do, 17058, Republic of Korea

ARTICLE INFO

Article history:
Received 25 October 2020
Received in revised form 16 March 2021
Accepted 18 March 2021
Available online 1 April 2021

Keywords:
Safety rule compliance
Deep learning
Safety risk correlation pattern
Risk recognition
Safety monitoring

ABSTRACT

Falls from height (FFH) are still a leading cause of fatalities in the construction industry, which also includes scaffolding-related accidents. Despite regular safety inspections, numerous scaffolding-related accidents occur at the construction site. The current safety monitoring practices are not only impractical but infeasible due to dynamicity of construction environment. Since a separate computer training and detection process is generally required to acquire spatiotemporal reasoning to control a single hazard; thus previous efforts in vision intelligence applications to improve safety monitoring are still limited to specific hazards. Also, in regard to detecting unsafe situations based on extracted correlations from safety rules, to date, previous studies have devoted little attention to this domain. To address these issues, this study proposes a correlation-based approach for mobile scaffold safety monitoring and detecting worker's unsafe behaviors. A deep neural network, Mask R-CNN, was used as classification and segmentation of worker's tasks combined with object correlation detection (OCD) module to identify worker's unsafe behaviors. The approach divides the overall construction worker's safety into two subsets, classification of worker and detection of safe (class-1) and unsafe (class-2) behavior using OCD block. The overall performance was evaluated on set of real scenarios with test results showing 85 % and 97 % precision and recall for class-1 (safe behavior) and 91 % and 65 % precision and recall for class-2 (unsafe behavior). The overall accuracy of 86 % confirms the Mask R-CNN-based OCD module's applicability for detecting worker's unsafe behavior effectively in a construction environment.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Around the globe, the construction industry is suffered from an excessive rate of non-fatal and fatal accidents. Construction site accidents are unplanned events that may lead to injuries, fatalities and occasionally cause property destruction (Haupt, 2001). The United States (US) Department of labor reported 5,250 deaths of workers on the job site in 2018, where the leading cause of fatalities were falls from height (FFH) (OSHA, 2017). The share of fatalities due to falls from scaffolds, ladders, working platforms, and roof edges, is accounted for 60 % (Riggs, 2018). According to the Occupation Safety and Health Administration (OSHA), approximately 65 % of the workers frequently work on scaffold systems (Collins et al., 2014); thus, preventing these accidents can minimize a good share of the total fatalities due to FFH.

Multiple types of scaffoldings are used at construction sites, such as mobile scaffolds, fabricated frame scaffolds, pump jack scaffolds, needle beam scaffolds, horse scaffolds, etc. Despite many concerns about scaffolding-related hazards, safety rules pertaining to scaffoldings are still among the third most frequently violated regulations (Wooten, 2019). Research reveals that falls related to scaffoldings are a vital potential threat in the job site, which causes a significant number of accidents annually (Hoła et al., 2018). A high probability of accidents during construction work using scaffolds has also been reported. Collins et al. investigated the scaffolding project life cycle's safety risk levels using a field survey technique. They developed a plug-in on the 4D-BIM platform to assist decision-makers (Collins et al., 2014). Extensive research studies have been carried out on the structural strength and modes of failure in scaffolding using machine learning (Sakhakarmi et al., 2018), cyber-physical system concepts (Yuan et al., 2016), and strain sensing approach (Cho et al., 2018a), yet, mobile scaffold safety remains unexplored. For instance, Papazoglou et al. provided the data analysis regarding the number of workers and their exposure time to

* Corresponding author.
E-mail address: cpark@cau.ac.kr (C. Park).

https://doi.org/10.1016/j.compind.2021.103448
0166-3615/© 2021 Elsevier B.V. All rights reserved.

construction accidents. Their findings revealed that the individual's risk rate per hour for working on mobile scaffolds is 6.3E-8 for fatal injuries, 3.2E-7 for permanent injuries, and 13.6E-7 for severe recoverable injuries (Papazoglou et al., 2015). In the Republic of Korea, the data published in 2019 by the Korean Occupational Safety and Health Agency (KOSHA), 488 fatalities reported were related to scaffolding work, out of which 45 fatalities were linked to mobile scaffolds (KOSHA, 2019). The above statistics augment the demand for enhanced safety measures to curtail the occurrence of accidents during construction.

The on-site safety monitoring is the last safety management layer; thus, it can play a vital role in preventing accidents (Reason, 2000). The safety manager needs to be highly focused while ensuring a safe work environment as advised by standard practices and regulations during the safety monitoring stage. Avoiding regulations, ignoring best practices, and deviating from a specified way of performing a job increases accident probabilities. Despite extensive efforts to enhance safety monitoring effectiveness by implementing new technologies (Ding et al., 2018; Fang et al., 2018a; Harrison and Dawson, 2015), the current construction safety monitoring still relies on visual inspections and observations, which are tedious, manual, and ineffective (Sakhakarmi et al., 2018; Mneymneh et al., 2018). Among the various methods to manage adequate construction safety, the application of advanced techniques, such as vision-intelligence (Park et al., 2020), smart sensors (Darko et al., 2020), motion tracking technologies (Seo et al., 2015), and augmented reality (Park and Kim, 2013; Tatić and Tešić, 2017), holds immense potential for the proactive prevention of accidents and automatic risk identification.

Recently, vision intelligence has exhibited enormous potential in enhancing the efficiency of various tasks in construction safety monitoring (Ding et al., 2018; Seo et al., 2015; Fang et al., 2018b, 2019), equipment tracking (Soltani et al., 2018; Golparvar-Fard et al., 2013; Memarzadeh et al., 2013), and defect detection (Feng et al., 2017). Moreover, the unprecedented increase in the job site's visual data enables a unique opportunity for vision intelligence systems to be deployed widely, thereby improving safety monitoring (Fang et al., 2018b). Generally, a separate specific vision intelligence algorithm and training procedure is required to detect every single hazard in a scene (Fang et al., 2020), thus limiting the scope of these vision-based systems to a single objective (Tang and Golparvar-Fard, 2017). Extensive research work still needs to be carried out to extend the scope and application of vision intelligence technology.

This study contributes in enhancing the implementation of vision intelligence technology by introducing a correlation-based vision intelligence system for mobile scaffolds' safety monitoring. Following the introduction section, this study reviews the current safety monitoring approaches in construction, existing vision intelligence technologies in construction, state of the art scaffolding monitoring studies in construction, and the need for vision intelligence-based safety monitoring for mobile scaffolds in subsequent sections. Section 3 presents the research methodology and system development of the proposed approach. Section 4 and 5 are intended to explain the experimental validation and critical contributions of the study. Finally, the core conclusion with the future outlook is provided in section 5.

2. Literature review

2.1. Current practices of safety monitoring in construction

Unlike other industries, such as manufacturing, the construction industry comprises dynamic working environments and sometimes complex operating procedures. Consequently, its unique nature creates challenges in safety monitoring and occupational health. Generally, the construction sector and its clients are widely linked with high risks due to the nature of micro, meso, and macro environments specific to construction (Szymański, 2017). Because of the ineffective control of these unique hazardous environments, the construction industry could not build a good reputation globally. Considering this critical issue of job site operations, researchers have focused on enhancing the safety management process by adapting and utilizing new emerging technologies over the last decade.

Traditionally, the safety management process includes safety planning in the pre-construction stage and safety monitoring during the construction stage. Many researchers have proposed an advanced safety planning process by extracting relevant safety rules from the regulation database and integrating those domain-specific rules with building information modeling (BIM) (Khan et al., 2019; Zhang et al., 2015; Heaton et al., 2019). Several researchers also utilized BIM for hazard location detection in the safety monitoring stage (Darko et al., 2020), thereby reducing costs and time. However, this approach still needs manual inspection for generating safety reports from the field (Tang and Golparvar-Fard, 2017). The current standard practice of dealing with safety hazards during physical construction involves monitoring the job site environment. The goal of occupational safety and health monitoring is to effectively ensure a safe working environment by comparing and measuring the best practices in occupational safety and health with a given organization's plans and standards (Seo et al., 2015). Job site safety observations thorough inspection is one of the typical methods used to examine ongoing activities (Reese and Eidson, 2006). The inspections and job site safety observations are generally performed weekly or biweekly depending on the project's size and complexity, and the inspector usually conducts randomly scheduled visits for 1–2 h during the week (Seo et al., 2015). The safety manager must be present physically to identify any potential hazard or safety rule in-compliance during construction. The responsible person (safety inspector or safety manager) visually examines the equipment and the work area (Reese and Eidson, 2006), thus eliminating the potential hazards. The success of these physical visits for safety compliance depends on the safety manager's experience and competence (Building, 2009). In fact, existing inspection methods are painstaking and time-consuming (Liu et al., 2016). To find a solution to these problems, researchers have recently introduced vision-based automated safety management, which can mitigate the existing problems of manual safety inspection (Seo et al., 2015; Fang et al., 2018b, 2019; Liu et al., 2016; Kim et al., 2017).

2.2. Vision-based technologies in construction

Vision intelligence technology deals with how computers can obtain a high-level of insight and understanding from videos or images (Fang et al., 2018b). This technology has already been adopted in various construction areas, such as hazard recognition, defect detection, progress monitoring, productivity analysis, and automated documentation (Seo et al., 2015). In construction, artificial vision-based object detection has focused on utilizing the two approaches: (1) hand-crafted features and (2) deep learning methods. Various descriptors are generally used to extract features from images, as Scale Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG), and Histogram of Optical Flow (HOF) (Fang et al., 2020). A summary of significant researches in this domain is enlisted in Table 1. and discussed as follows.

Until recently, resource detection, such as humans, materials, and construction equipment, has come to the surface as a vital development in the vision intelligence domain applied in construction. For example, Memarzadeh et al. (2013) presented a semi-automated method supported by a computer vision-based algorithm for 2D detection of construction workers and equip-

Table 1
Some significant deep learning-based researches in the construction domain.

Authors (Year)	Objective	Target objects	Methodology
Nath et al. (2020)	To detect personal protective equipment (PFE) in real-time.	Hat and Vest,	Three deep learning models built on You Only Look Once (Yolo) architecture.
Fang et al. (2018b)	To determine whether to wear a protective helmet.	None-hard hat use Workers	Detection in far-field surveillance videos using Faster R-CNN.
Fang et al. (2018b)	To detect an individual who does not wear Safety harnesses.	Worker & Safety harness	worker and safety harness detection using deep CNN, Faster-R-CNN.
Luo et al. (2018)	To automatically assess workers and monitor reinforcement installation activity.	Workers doing reinforcement work	Three streams (Red – Green - Blue (RGB), optical flow, and gray stream CNNs) deep CNN model for reinforcement activity monitoring.
Ding et al. (2018)	To recognize workers' unsafe behavior while using the ladder.	Human and Ladder	CNN for extracting the visual features & LSTM for sequencing learning features.
Fang et al. (2018a)	To determine the unsafe act of person traversing and the structural member.	Worker and Structural member	Mask-RCNN for detection and mask generation.
Kim et al. (2019)	Automatically measuring proximities between construction entities.	Worker, Wheel loader, Excavator	Object detection using YOLOV3 and image rectification.

ment from video streams using a Histogram of Oriented Gradients (HOG) and colors. Additionally, Golparvar et al. (Golparvar-Fard et al., 2013) proposed a vision-based system for the earth-moving equipment using the HOG and support vector machine (SVM). Despite the successful implementation of hand-crafted feature-based approaches, the dynamic conditions that prevail in a construction site could affect the features extraction from images. For instance, background clutter, intraclass, variance, and viewpoint scale could lower the accuracy level for object detection (Fang et al., 2020).

Recent research in vision-based monitoring of the construction sites has focused on developing a simple inspection system using deep learning to detect basic hazards. For instance, Nath et al. (2020) presented three deep learning models based on you only look once (YOLO) architecture to detect hat and safety vest in real-time. Mneymneh et al. proposed a novel motion detection algorithm and human classifier to automatically detect non-hard hat people and recognize non-compliance with safety regulations (Mneymneh et al., 2018). Fang et al. (2018a) developed a deep learning-based (Faster R-CNN) approach to detect non-hardhat users from far-field surveillance videos. Additionally, Fang et al. (2018b) developed an automated safety harness recognition method for fall hazards using two CNN architectures.

Similarly, Ding et al. (2018) presented a deep hybrid model, comprised of convolution neural networks (CNN) and long short-term memory (LSTM), that automatically identify unsafe behavior by detecting human climbing on a ladder. Correspondingly, Luo et al. (2018) integrated three streams (Red – Green - Blue (RGB), optical flow, and gray stream CNNs) deep CNN model to assess workers doing reinforcement installation activity automatically. To observe individuals' situation awareness for preventing construction accidents, Liu et al. (2016) investigated a computationally efficient tracking method for three-dimensional (3D) human skeleton extraction from stereo videos.

Nevertheless, the "object" detection ability of the trained models in these research studies is noticeably limited, with a potential reason being the limited datasets used to train the models. Apart from that, the previous efforts revealed the narrow scope of vision-based safety monitoring. Many researchers have focused on exploiting descriptors for machine learning classifiers, and only a few have explored the application of neural networks for object detection in construction. Convolutional neural networks have been used in object detection and classification owing to their advantages in the extraction of image features (Tang and Golparvar-Fard, 2017). Numerous approaches and algorithms have been adopted in construction (e.g., you only look once (YOLO), Faster R-CNN, Mask R-CNN) (Ding et al., 2018; Harrison and Dawson, 2015; Darko et al., 2020; Tang and Golparvar-Fard, 2017; Szymański, 2017; Khan et al., 2019). Although Faster R-CNN proves its significance and exceeds other techniques due to its high accuracy and improved performance, however, a major challenge confronting these CNNs are their large dataset requirements for training. Similarly, an extension of Faster R-CNN, i.e., Mask R-CNN, has been identified to outperform all the developed object detection approaches, as He et al. claimed in 2017 (Zhang et al., 2015). Thus, to attain higher accuracy in terms of detection in a complex construction environment, the Mask-RCNN approach is extended with a correlation technique to automatically recognize the unsafe behavior pertaining to mobile scaffolding.

2.3. State-of-the-art scaffold monitoring studies in construction

Conventionally, the construction industry uses regulations and safety training programs to provide awareness to construction workers on using scaffolds safely (Cho et al., 2018b). OSHA provides thorough and complete requirements and recommendations for designing, dismantling, and inspecting scaffolds (Reese and Eidson, 2006). However, OSHA reported the scaffolding-related violation as one of the top three most frequently violated regulations, which could obviously include rules in-compliance related to mobile scaffolding (Wooten, 2019). Many researchers have developed advanced approaches to enhance the safety monitoring of scaffoldings in construction. Sakhakarmi et al. presented a machine learning approach to predict safety conditions. They used strain data from scaffolding columns to classify scaffolding failure cases reliably predicting scaffold safety (Sakhakarmi et al., 2018; Cho et al., 2018a). Yuan et al. proposed a cyber-physical system to prevent potential failures of temporary structures (Yuan et al., 2016). A data-driven scaffold monitoring system using wireless strain sensors, finite element models, and support vector machine (SVM) was introduced by (Cho et al., 2018b; Cho et al., 2021). However, their developed systems require precise scaffold models as an input to the system and an additional setup of strain sensors to address environmental effects such as humidity and temperature (Cho et al., 2018b). Likewise, F.Song and G.Song developed a percussion-based detection method to minimize the efforts required for sensors installation on scaffolds to detect the cup-lock joint's looseness that can induce collapse.

A 3D local feature descriptor to delineate linear straight-shaped objects was proposed to reconstruct and detect scaffolds from the point cloud (Xu et al., 2018). Likewise, H. Chi et al. developed an image and BIM-enabled approach to automatically monitor scaffolding progress through features extraction from daily site photos (Hung-Lin Chi et al., 2017). Similarly, Jung et al. demonstrated a failure detection approach on small scall scaffolds using image processing techniques (Jung et al., 2018). In another study, Jung focuses on the automated detection of changes caused in temporary structures such as scaffolding by deflection, vibration, accumulated

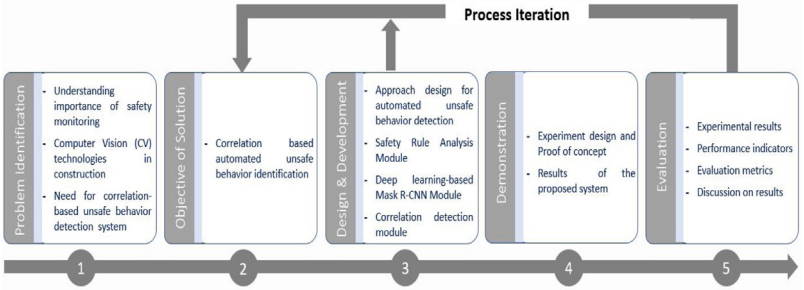

Fig. 1. Design science research process for vision intelligence-based safety monitoring of mobile scaffolds.

stresses, wind, and lateral forces using edge detection techniques in digital image data (Jung, 2014). In order to predict the safety conditions of scaffolds during the operation phase, a multi-level deep learning approach, was proposed based on divide and conquer technique (Sakhakarmi and Park, 2020). In fact, many studies have covered the potential structural failure aspect of temporary structures using sensor technologies; however, monitoring the safety rule in-compliance showing the unsafe use of temporary structures is still weak.

Interestingly, there are very few studies in the literature that focus on safety of mobile scaffolds. For instance, a study endorsed the significance of the individual risk rate per hour for working on mobile scaffolds (Papazoglou et al., 2015). Considering the practical perspective, manual safety monitoring of mobile scaffolds in large projects for outriggers in-compliance is unrealistic, error-prone, resource-intensive, and time-consuming. Contrarily, even though notable developments in construction safety monitoring have been successfully instigated, many aspects are yet to be discovered. Therefore, it is inevitable to increase the computer vision application scope for the automation of the construction's safety process. Recognizing these needs, this study attempts to propose a computer vision-based risk recognition system for mobile scaffoldings.

3. Research methodology and system development

In accordance with previous research studies for construction safety and health monitoring, a deep learning approach was adopted to perform object detection within a workspace at constructin site and automatically detect the unsafe behavior of humans with regard to mobile scaffolding. A design science research process was used, as it is a popular multi-discipline method for research that has been applied in engineering, business, and computer science (Fang et al., 2018b, 2019). The process comprises five steps, as illustrated in Fig. 1: (1) problem identification; (2) objectives of a solution; (3) design and development; (4) experiment and demonstration; and (5) performance evaluation.

The design and development segment of this study includes safety rule analysis (SRA) and two modules that are stated as: 1) the vision intelligence implementation (VII) module; and (2) the object's correlation detection (OCD) module. The SRA is used to extract the safety rules that contain correlation information related to mobile scaffolds; the VII module is intended to develop a Mask R-CNN-based object recognition system; the OCD module is designed to identify the correlation between the entities by meeting certain conditions with respect to predicted bounding boxes of objects (construction entities).

Fig. 2. explain the coordinate system and process flow of the artificial vision intelligence-based approach for safety monitoring pertaining to mobile scaffolds. A typical scenario of bounding boxes localized for the person on a mobile scaffold is demonstrated in Fig. 2 (a). Fig. 2 (b). illustrates the coordinates of interests such as corner coordinates $\mathbf{S}_{l,t} = (x_l^s, y_t^s)$ and $\mathbf{S}_{\boldsymbol{r,b}} = (x_r^s, y_b^s)$, and centroid coordinate $\mathbf{S}_c = x_c^s, y_c^s)$ for mobile scaffolds. The pursued coordinates for persons are also illustrated in Fig. 2(b), for instance, $\mathbf{P}_{\boldsymbol{l,t}} = (x_l^p, y_t^p)$ and $\mathbf{P}_{\boldsymbol{r,b}} = (x_r^p, y_b^p)$ for corner coordinates and $\mathbf{P}_{\boldsymbol{c}} = (x_c^p, y_c^p)$ for centroid coordinate.

The system flow is discussed in Fig. 2(d). First, the CNN module detects the construction entities such as persons, mobile scaffolds having outriggers, and mobile scaffold with no outriggers in digital image data, as illustrated in the process flow of the vision intelligence-based system for safety monitoring pertaining to mobile scaffolds (see Fig. 2 (c)). After, to classify the behavior as safe or unsafe based on the correlation extracted from the related safety rule, the second step of the system computes the relation between the person and the type of scaffolding to identify if the person is standing on the scaffolding, or not. The system recognizes the person on or near the mobile scaffolding; in the absence of person/s, the system ignores the further processing and declares the scene as safe behavior. If the person class exists, the system search for the mobile scaffold class in the third step; if the detected scaffold is equipped with outriggers, the system comes up with safe behavior, else the y-coordinate of the centroid in the identified bounding boxes are compared. For instance, the y-coordinate of the centroid in person's bounding box $\boldsymbol{y_c^p}$ is less than the y-coordinate of the centroid in scaffolding's bounding box $\boldsymbol{y_c^s}$, then the person is supposed to be on the ground and safe. In order to make sure that the person is on top of the mobile scaffold, the system checks the corner coordinates in fourth step, such as $\boldsymbol{x_r^p}$ is less than $\boldsymbol{x_r^s}$ and $\boldsymbol{x_l^p}$ is greater than $\boldsymbol{x_l^s}$, as stated in Fig. 2 (c). Unsafe behavior will be detected if the condition is satisfied; else, the person will not be assumed on the top of the mobile scaffold. The categorical steps conducted to complete the proposed approach are detailed in the following sections.

3.1. Safety rule analysis

The KOSHA regulations comprise 13 chapters, including various sections, followed by 671 standards, out of which 277 are related to the construction sector (KOSHA, 2021). In 1953, the labor standard law laid the foundation for industrial safety and health policy in Korea. Following this, in 1987, the Korea Occupational Safety and Health Agency (KOSHA) was established motivated by the rapid development of industries in the period of 1970–1980. The corresponding acts were revised to meet the mandatory demand for safety and health standards required for various industries having toxic and complex environments. Since then, KOSHA has collected and analyzed many accident cases, thereby creating an expert knowledge database. Substantial amendments have been made to alleviate the past policies for modern industry compliance and global competition.

However, despite the vigorous enforcement of occupational safety regulation and expert knowledge worldwide, the construction safety management process still relies on traditional practices.

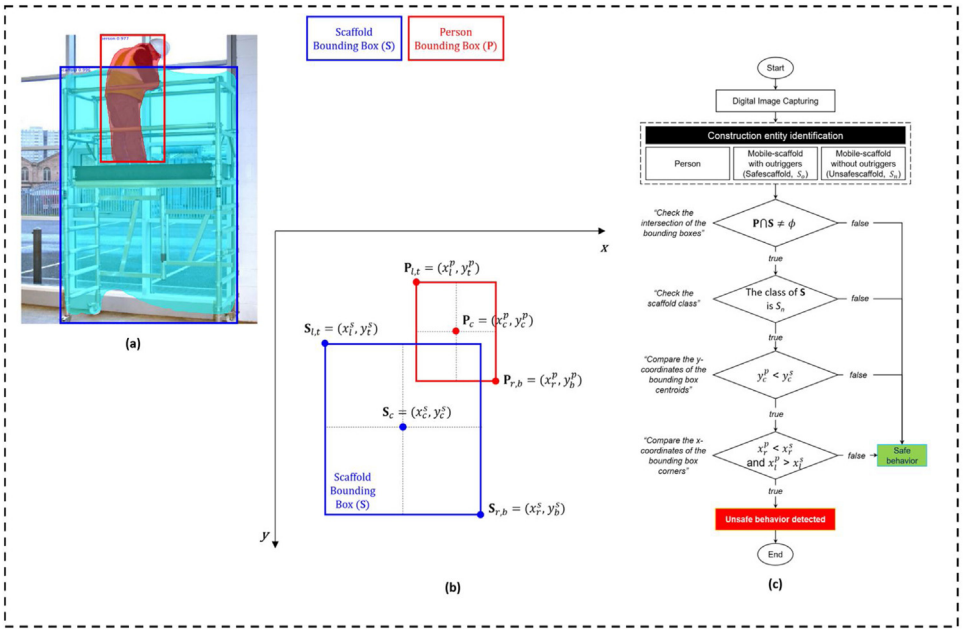

Fig. 2. Vision-intelligence based system for safety monitoring pertaining to mobile scaffolds: (a) illustration of bounding boxes for the proposed system (b) bounding box coordinates of mobile scaffolds and person, (c) system process flow.

Vision intelligence technology can enhance the construction safety enforcement and implementation process by converting the expert knowledge, in this case—the rules, into practicality with deep learning algorithms. Therefore, the scaffolding related KOSHA regulations were analyzed manually by the authors, and relevant rule—having information about the outriggers, to provide structural stability and safety to the mobile scaffolds (used in almost every building construction site), were extracted. Two case scenarios are explained to elaborate on the proposed research concept and identify the proposed system's benefits. The scope of the safety rule analysis is limited to the detection of positive and negative correlations of workers with mobile scaffolds, explained in article 68 (1) of KOSHA.

Article 68 (1): (mobile scaffolding), extracted from KOSHA, states that "when the mobile scaffolding is assembled by the employer, to prevent unexpected sudden movement or turning of the wheels of the mobile scaffold while working, fix the part of the scaffold to a solid facility or install an outrigger." This article of KOSHA demonstrates the two correlations (positive and negative) between the mobile scaffolding and the worker. The extracted case scenarios from this article can be defined as follows:

Case 1 **:** If a person "*P*" is working on the mobile scaffolding with outriggers " S_o " installed, then the relation "*R*" between the two entities is concluded as being positive in this study, and it is safe for work. The mathematical equation can be written as follows:

$$If\ P \cap S_o \in R \tag{1}$$

Case 2 **:** On the contrary, if person "*P*" is working on the mobile scaffolding " S_n " with no outriggers installed, then the relationship "*R*" between *P* and S_n is considered to be negative for this research work, and the equation can be written as under:

$$If\ P \cap S_n \in -R \tag{2}$$

Consequently, the negative relation demonstrates unsafe behavior, and the system needs to automatically record a warning message for the concerned safety manager. Thus, the safety manager responsible for the safety of the mobile scaffolds will no longer need to visit the job site physically.

3.2. Vision intelligence implementation module

This section summarizes the database generation, design flow, and architecture of deep convolutional neural networks (deep CNN). The general architecture of deep CNN comprises a convolutional layer, max-pooling layer, fully connected layer, and softmax layer, which used a SoftMax function (multi-class classification) as an activation function.

3.2.1. Database generation

An enormous number of labeled images are needed to train a CNN classifier. However, obtaining labeled images of construction objects from open-source databases is still challenging. Therefore, raw images of persons and mobile scaffolding (with and without outriggers) with an extensive range of variations, such as lighting illuminations, image orientations, and varying positions, were collected. A total of 1508 images were collected from two sources: (1) Images from Google database were downloaded using crawling techniques by keywords search such as "mobile scaffolding", "mobile scaffolding and worker", "movable scaffolding and worker", "mobile scaffolding with outriggers", "mobile scaffolding without outriggers", "mobile scaffolding in construction", and "construction worker". In addition to the Google search results, (2) a considerable number of images were also collected from one of Tornado group's construction sites for the higher secondary college of technology located in Abu-Dhabi, United Arab Emirates.

A double round review process for the dataset cleaning is conducted to choose candidate images for training and validation. The double round review process intends to collect real images of the cited classes captured in construction sites. Unsuitable images such as incorrect exposure (having insufficient information), irrelevant contents (unrelated images downloaded using the crawler technique), non-realistic contents (sketched not reflecting real objects) are removed in the first round of the data cleaning process. During the second round of the review process, similar images from identical scenes were eliminated from the dataset to avoid overfitting.

A total of 938 images were finalized for the experiment. Among them, 253 were related to the person's class, 548 were corresponding to scaffolds with outriggers, and 481 were of mobile scaffolds

Fig. 3. Dataset labeling Process: (a) VGG Image Annotator (VIA) annotation tool, (b) class-aware annotation, (c) instance aware annotation.

without outriggers. It is important to note that some of the images include two or more classes; thus, that image is counted multiple times but not more than three times (as the number of classes is 3), which made the total sum (1,282) exceeding than the total images (938). The database of 938 images is split into training and validation folders with a ratio of 75:25, where 75 % (703 images) were used to train the model and 25 % (235 images) for validation to test the model for its accuracy and precision. Additionally, the validation set was utilized for testing the object correlation detection block. To increase the performance of the trained network and prevent overfitting, data augmentation with five techniques were applied at each iteration of training. The training image built with VGG annotator was augmented with affine, gaussian blur, contrast normalization, and flipped-right-left with a 50 % chance to reduce the storage used for training images. Three classes were defined to detect the correlation among the resources in selected scenarios as a case example: "scaffold" for mobile scaffolding without outriggers, "safescaffold" for mobile scaffolding having outriggers, and "person" for a person. All the images were labeled manually using an online annotation tool—VGG Image Annotator (VIA; see Fig. 3(a).) (Dutta and Zisserman, 2021). The dataset was labeled using class-aware and instance-aware annotation techniques. In class-aware, images were labeled using the predefined classes such as scaffold, safescaffold, and person (see Fig. 3(b).). The class-aware annotation is intended to label the data for the detection of interested objects (classes), as cited above. In contrast, for instance-aware annotation, every object is labeled as an instance in each image (see Fig. 3(c).). The instance-aware annotation helps in the semantic segmentation of each object in the image (Dai et al., 2021). However, the scope of this study is limited to class-aware annotation.

3.2.2. Design flow

The proposed system framework for the correlation-based unsafe behavior detection leveraging a CNN is shown in Fig. 4. It is evident from the figure that the image frame of the construction site is passed through the CNN block, performing predictions to detect the person and scaffolding. Following the object detection process, an object's correlation detection module is presented for the post-processing of CNN prediction between either of these classes to check for subjectivity. The correlation block checks the matches between the person and scaffolding to detect the person's unsafe behavior. In this research work, unsafe behavior is defined as a situation in which a person is working on scaffolding with no outriggers.

Conversely, safe behavior is defined as a situation in which a person is working on the scaffolding equipped with outriggers. The proposed algorithm can classify safe and unsafe behavior based on the correlation block, by computing matches certain conditions and conducting overlapping between the localized position of person and scaffolding. The output results can warn if unsafe behavior is detected on behalf of any construction site worker.

3.2.3. Mask and region-based CNN

In recent years, many techniques have been introduced related to region-based CNN for instance segmentation, as discussed in (Ren et al., 2021). Recently, a new approach Mask R-CNN (He et al., 2017), has been developed based on its predecessor with adding instance segmentation on top of Faster R-CNN. Mask R-CNN, as one of hybrid technique, incorporate Faster R-CNN and Fully Convolutional Network (FCN) for object detection and localization with pixel-based masking for instance segmentation, as shown in Fig. 4. It is a two-stage network for enhancing better subjectivity of objects under consideration using feature pyramid network (FPN). The first stage is the region proposal network (RPN), responsible for generating random proposals by scanning over an entire image, predicting an output class, and regression bounding box. The second layer takes features from the RPN and FPN layers and performs convolution to generate masks for each region of interest (RoI) that are added to realize pixel-based segmentation. The output prediction shows masks overlaid on top of bounding boxes and class results.

The training of the Mask R-CNN does not require an RPN to retrain unless specified, because it uses the pre-trained weights from Faster R-CNN. A Mask R-CNN has three output layers corresponding to the SoftMax classifier, bounding box regressor, and a mask branch. We define the multi-task loss on each sampled RoI during training as in Eq. (3), Where the classification loss, $\boldsymbol{L_{cls}}$, and bounding-box loss, $\boldsymbol{L_{box}}$, are identical to those defined in (Ren et al., 2021).

$$L = L_{cls} + L_{box} + L_{mask} \tag{3}$$

The mask branch has a $\boldsymbol{Km^2}$ -dimensional output for each RoI, which encodes $\boldsymbol{K}$ binary masks of resolution $\boldsymbol{m \times m}$, one for each of $\boldsymbol{K}$ number of classes. The mask loss, $\boldsymbol{L_{mask}}$, is defined as a per-pixel sigmoid and the average binary cross-entropy loss as defined in (He et al., 2017). Here, $\boldsymbol{L_{mask}}$, is the mask loss defined in Eq. (4) as a per-pixel sigmoid and the average binary cross-entropy loss on the k^{th} mask given by:

$$L_{mask} = -\frac{1}{m^2} \sum_{1 \le i,j \le m} [y_{ij} \log \hat{y}^{k}_{ij} + (1 - y_{ij}) \log(1 - \hat{y}^{k}_{ij})] \tag{4}$$

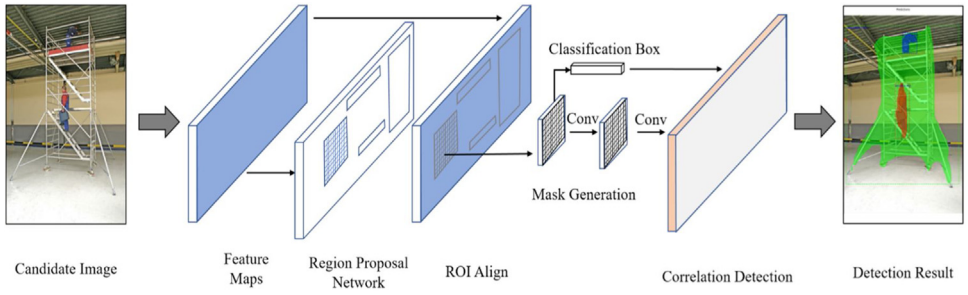

Fig. 4. Vision intelligence-based correlation detection framework.

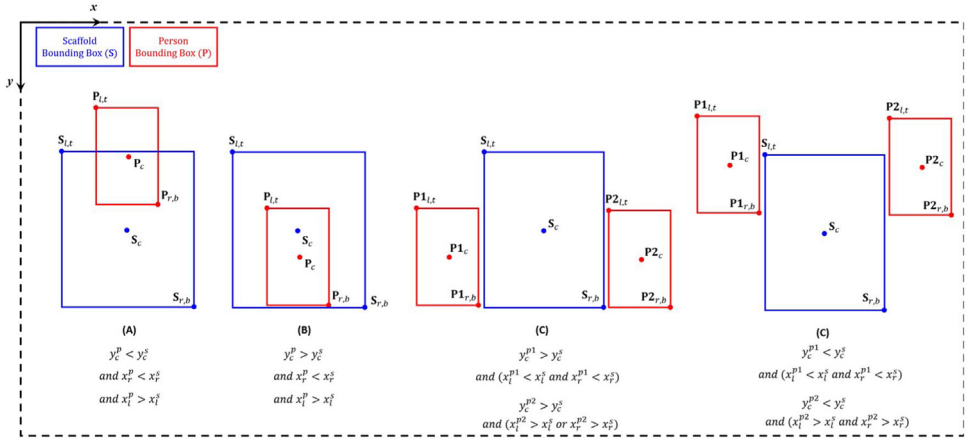

Fig. 5. Example of bounding box coordinates: (A) person is on mobile scaffolding, (B) person is on the ground or near to the ground, (C) one or more persons are at ground level at the right or left of the mobile scaffolding, (D) one or more persons are in the background or not relevant to the mobile scaffolding.

where y_{ij} is the label of a cell (i, j) in the true mask for the region of size $m \times m$; $\hat{y}_{ij}^{k}$ is the predicted value of the same cell in the mask learned for the ground-truth class k.

3.3. Object correlation detection module

The CNN is used to perform object detection; however, automated recognition of unsafe behavior in a construction site is difficult to achieve using only CNN predictions. Therefore, an additional OCD block is added to determine the correlation between the construction entities. In this study, it is necessary to obtain the particular position of a person corresponding to the location of mobile scaffold. Initially, CNN in this study is responsible for detecting persons and two sub-classes of scaffolds as previously defined: (1) safescaffold; and (2) scaffold. After the detection process, the OCD module substantiates if the person is standing on the scaffolding or not. It then computes the relation between the person and the type of scaffolding to classify the behavior as safe or unsafe.

To understand the behavior of a person/s working with mobile scaffolds, checking for the presence of an outrigger is required. If a person is working on scaffolding with no outriggers, then this behavior is recognized as unsafe, and a warning is required to the site management. Conversely, if the person is working on scaffolding with outriggers, then this behavior is recognized as safe. The procedure for detecting the person and scaffolding classes are described earlier. The relation between them is computed through matching certain conditions and positions of the bounding boxes (person and scaffolding), as depicted in Fig. 5. The bounding box overlapping for correlation detection between the construction entities is one of the significant contributions proposed by this study.

The system detects and masks all the persons and scaffolding in a given image, as illustrated in the vision intelligence-based correlation detection framework (see Fig. 4.). The extraction and comparison of the corner coordinates and y-coordinate of the center in specified bounding boxes are graphically illustrated in Fig. 5. After detection and segmentation, the system identifies and recognizes the person class on or near the mobile scaffolding; in the absence of a person/s, the system ignores further processing and declares the scene as safe behavior. If the person class exists, the y-coordinate of the center in the identified bounding boxes are compared.

The OCD module is developed for multiple scaffolding and multiple persons; however, only four scenarios are exemplified in Fig. 5. for an understanding of the readers:

For instance: in case (A), if the y-coordinate of the center in person's bounding box " y_c^p " is lesser than that of the centroid of mobile scaffolding bounding box " y_c^s ", with further conditions along x-axis such as " x_r^p " is lesser than " x_r^s ", and " x_l^s " is greater than " x_l^p ", then the person is considered to be working on the mobile scaffolding.

In case (B), if the " y_c^p " is greater than that of the mobile scaffolding " y_c^s ", with supplementary conditions along their relative coordinate at x-axis such as " x_l^p " is greater than " x_l^s ", and " x_r^p " is less than " x_r^s ", then the person is supposed to be on the ground or near to the ground, as portrayed in Fig. 5(B).

Case (C) represents a scenario in which multiple persons (P1 and P2) are working at the right and left of the mobile scaffolding. Intended for the person (P1) in this scenario, " y_c^{p1} " is greater than that of the mobile scaffolding " y_c^s ", while " x_l^{p1} " and " x_r^{p1} " is lesser than " x_l^s " and " x_r^s ", respectively. Whereas, for the person (P2), " y_c^{p2} " is greater than that of the mobile scaffolding " y_c^s ", while " x_l^{p2} " and " x_r^{p2} " is greater than " x_l^s " and " x_r^s ", respectively, as illustrated in Fig. 5 (C).

Table 2
Algorithm for image overlapping using matches.

```
Required:
   safe_site_boxes, unsafe_boxes, person_boxes, safe_overlap,
   unsafe_overlap, match.
Where
   doOverlap = checking overlap of person with scaffold
   match     = classify type of scaffold for safe or unsafe category
   class_ids = 1. scaffold
               2. Safe scaffold
               3. Person
get safe_site_boxes(boxes, class_ids): → get bboxes of safe scaffold object
get unafe_boxes(boxes, class_ids): → get bboxes of unsafe scaffold object
get person_boxes(boxes, class_ids): → get bboxes of person object
while person:
            get safe_overlap(P_{l,t}, P_{r,b}, S_{l,t}, S_{r,b})x: → compute center point w.r.t safe
      object
     if (y_c^p < y_c^s): → check if person is standing on safe scaffold or ground
        if(x_r^p< x_r^s & x_l^p > x_l^s): → Person is standing on safe scaffold with fully
     overlap
            return safe
        else:
            return unsafe → Person is standing aside in the foreground of the
       image
     else:
            return safe → Person is standing on the ground
   get unsafe_overlap (P_{l,t}, P_{r,b}, S_{l,t}, S_{r,b}): → compute center point w.r.t unsafe
object
     if(y_c^p < y_c^s): → check if person is standing on unsafe scaffold or ground
        if(x_r^p< x_r^s & x_l^p > x_l^s ): → Person is standing on unsafe scaffold with
     fully overlap
            return unsafe
        else:
            return safe → Person is standing aside in the foreground of the
       image
     else:
            return safe → Person is standing on the ground
```

Case (D) illustrates if one or more persons are in the background and not relevant to the mobile scaffolding, in this situation, " y_c^{p1} " is greater than the " y_c^s " of mobile scaffolding contemplating y-coordinate, while, " x_l^{p1} " and " x_r^{p1} " is lesser than their corresponding coordinates such as " x_l^s " and " x_r^s", respectively. The conditions for a person (P2) reflected in the mentioned scenario portrays that " y_c^{p2} " is greater than the " y_c^s ", however; " x_l^{p2} " and " x_r^{p2} " have greater value than the respective coordinates, " x_l^s " and " x_r^s " contemplating x-axis. If any of these conditions are satisfied, then the system detects correlation and recognizes the corresponding behavior.

The algorithm for the proposed correlation detection module is presented in Table 2. To identify the negative correlations, the vision intelligence system needs to check the scaffolding class and decide accordingly. The proposed algorithm includes three layers: 1) definition and detection of the specified classes, 2) extraction and computing coordinates of the detected classes, 3) visualize the safe and unsafe behavior. In the first layer, the developed algorithm gets the bounding boxes of the construction entities such as scaffold, safescaffold, and person, as transcribed by the pseudo-code in Table 2. In the second layer of the developed algorithm, the y-coordinate of the center and corner coordinates in corresponding localized bounding boxes are extracted, and matches are computed. To identify the correlation, the overlapping and matches are computed between the person to safe scaffold and unsafe scaffold. The third layer of the algorithm filter and visualize the result of safe and unsafe behavior based on the correlation.

4. Experimental validation

4.1. Training setup

The experiments were performed using an open-source Mask R-CNN library with ResNet-50 weights pre-trained using the COCO dataset. A pre-trained model contains the weights representing the features such as edges and lines of the concerned larger dataset on which the model is previously trained, learned features are transferable to the model of another dataset. The coco dataset is

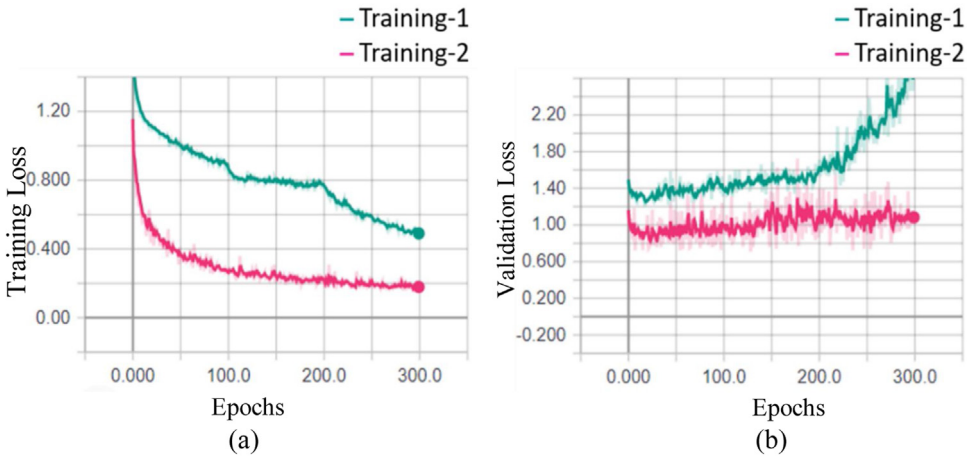

Fig. 6. Visualization of comparative plots:(a) training loss plots of training-1 and training-2, (b) validation loss plots of training-1 and training-2.

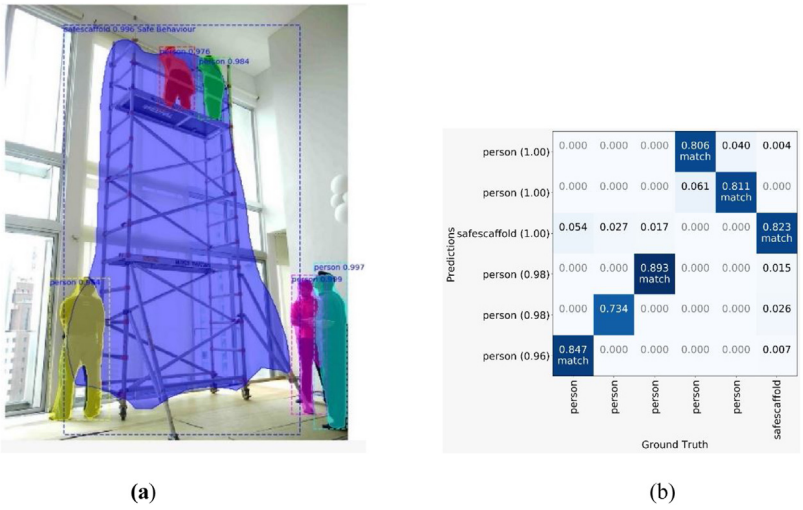

Fig. 7. Validation of the trained model for case 1: (a) detection results of mobile- scaffolding with outriggers and persons, and (b) confusion matrix of the image.

a large-scale collection of 328,000 images that contains 91 types of common objects (Lin et al., 2021). In this study, a model trained on a coco dataset is adopted as an initial model for faster convergence, time-saving, and better accuracy. The training was conducted on a server equipped with i7−9700k CPU, an NVIDIA RTX 2080Ti GPU, and 16 GB RAM. Each mini-batch had two images, and the training images were resized to 800 pixels when the image size had different dimensions.

The model was trained with 0.001 learning rate for 100 epochs and was decreased by factor of 10 when all layers were fine-tuned for additional 200 epochs with 300 steps per epoch. The iteration represents a particular update of a model's weights throughout the training process (Google Developers, 2020). Likewise, epoch determines one iteration over the entire dataset. The optimizer used was stochastic gradient descent (SGD) to minimize the cost function. Additionally, weight decay and learning momentum were set as 0.0001 and 0.9, respectively. Image augmentation was applied during training at each step to reduce the storage used for training images, increase the trained network's performance, and prevent overfitting. The training curves shown in Fig. 6 depicts the training performance for two different training setups. Training-1 was performed with a standard procedure of the same learning rate and learning momentum, while Training-2 indicates the gradually decreasing learning rate to prevent the model from exploding. The learning rate was reduced gradually to minimize overfitting, and L-2 regularization was applied to reduce cost function. This approach helps in updating weights efficiently at the beginning of the training phase, prevents them from exploding, and change drastically.

4.2. Validation of CNN classifier

The model is tested on the validation dataset to perform predictions and validate the performance of the CNN classifier. As depicted in Figs. 7 and 8, the detection results of the trained model for three classes, namely, −safescaffold, scaffold, and person. The examples of quantitative results are also shown as confusion matrices for two images with multiple classes from the validation set. Each class may have multiple instances, and each instance of a class is segmented individually (recall the labeling section of the study: the labeling was done with class-aware and instance-aware), as portrayed in Figs. 7(a) and 8 (a). Explicitly, in the problem of statistical classification and field of vision intelligence, a confusion matrix (error matrix) offers the visualization of a model's performance. The trained model results are shown as confusion matrices for the given images in Figs. 7 (b) and 8 (b). These figures explain the match between ground-truths and predictions for each instance of the detected classes if they crossed a certain threshold score. The x-axis represents the ground-truth label, whereas the y-axis represents the predicted label. The numbers in the confusion matrix correspond to the correlation of the actual object to the predic-

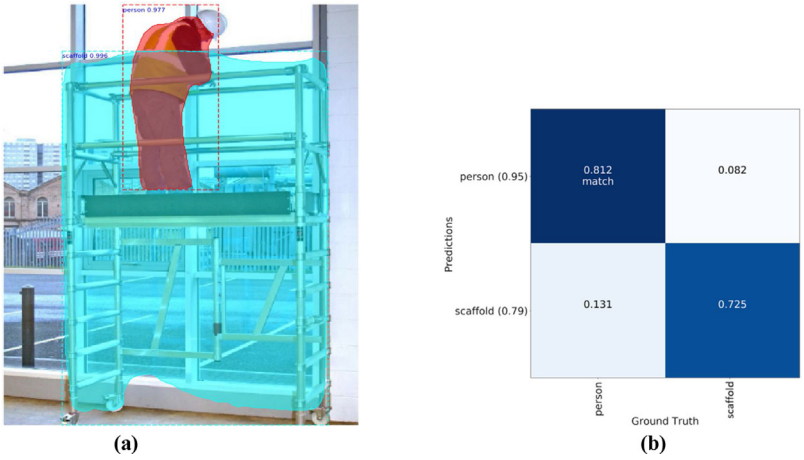

Fig. 8. Validation of the trained model for case 2: (a) detection results of mobile- scaffolding with outriggers and persons, and (b) confusion matrix of the image.

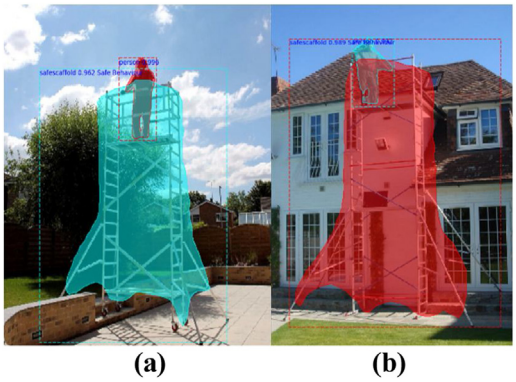

Fig. 9. Examples of true positive prediction: (a), and (b) mobile- scaffolding with outriggers and persons working (case 1).

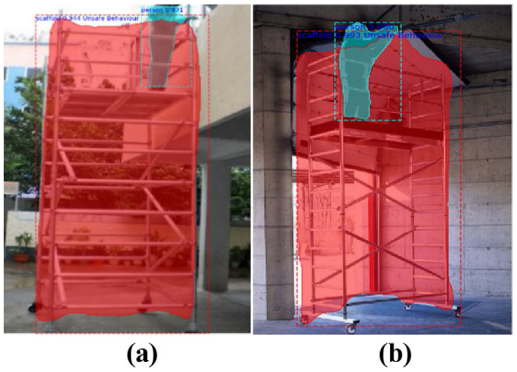

Fig. 10. Examples of true positive prediction: (a), and (b) mobile- scaffolding without outriggers and person working (case 2).

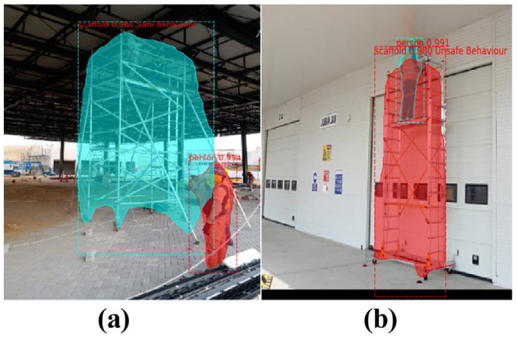

Fig. 11. Examples of false prediction: (a), and (b) mobile- scaffolding without outriggers and persons.

tions. The numbers in the brackets outside the confusion matrix show the aggregated correlation score for any particular instance of the detected classes, and the score in the boxes inside the confusion matrix corresponds to the individual correlation score. The number in each box represents the match between the ground truth and predicted result and shows the match for the results of the person/s and scaffold greater than 80 % on average. For our testing, the minimum score for matching ground-truths and predictions was selected to be 80 %.

4.3. Detection of safety risk correlation patterns

The network architecture, described in section 3.2.3, was implemented on the dataset generated from various sources, considering the KOSHA regulation-based case scenarios for mobile scaffolding. The detections were performed using the deep convolutional model on construction job site images with a focus on recognizing unsafe behavior. The post-processing module was used to establish a correlation between a person and the type of scaffolding to distinguish between safe and unsafe behavior. Fig. 9(a) and (b) shows case 1, which includes the detection of persons and scaffoldings with outriggers. The system first detects the scaffolding and persons; following this, it determines the relationship using the OCD block discussed in section 3.3. The safe behavior detection result pertaining to case 1 shows the positive correlation between the person and mobile scaffolding equipped with outriggers. On the contrary, Fig. 10(a) and (b) depicts the negative correlation (case 2) of the persons working on top of the mobile scaffoldings with no outriggers, and the system predicts the behavior being unsafe as per KOSHA rules; both Figs. 9 and 10 indicates true predictions.

Similarly, the false predictions of the proposed system can be observed in Figs. 11 (a, b) and 12 (a, b). In Fig. 11(a), one worker is standing in front of the mobile scaffold equipped with outriggers, and two persons, equipped with safety harnesses, are working on the top of the mobile scaffold. However, the system classifies the practice as unsafe behavior (false result), contrary to the real situation. This may be due to the occlusion of outriggers caused by the person standing in front of the scaffold. Likewise, in Fig. 11(b), these outriggers differ from those used in training; therefore, images that belong to these types of outriggers must be added to the dataset to improve detection results. Correspondingly, Fig. 12 (a) depicts

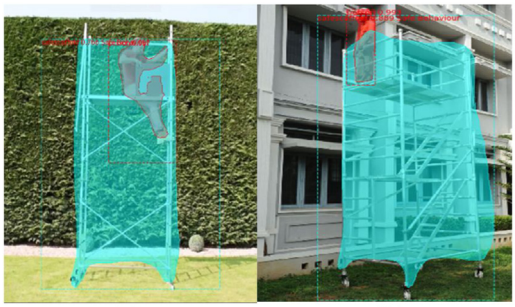

Fig. 12. Examples of false prediction: (a), and (b) mobile- scaffolding with outriggers and persons.

the person climbing on the mobile scaffold with no outriggers and almost reaching the top; however, the system presents the behavior as safe, although it is absolutely unsafe. This might be the reason of the classification model, as the background is all green and noisy, making the classification model to detects it as a safe mobile scaffold. In Fig. 12(b), the person is standing on a mobile scaffold having no outriggers, which are unsafe for work, but the system cited the behavior as a safe behavior. The reasons behind this might be the complex background, such as windows and cantilever structure of the buildings, which looks like pipes and lines. As this research portion belongs to the ongoing vision intelligence project, the weaknesses identified would be overcome in future work by increasing the dataset.

Fig. 13 shows the images with miss detections, and these are the result of occluded and noisy images with low resolution and extreme near field. The causes of missed detection of these scenes could be the noisy environment, digital data quality, and zoom visualization.

Considering the zoom visualization in Fig. 13(a), the system was trained for the mid-near field images. Thus the model is unable to detect the instant near field data. Fig. 13(b) shows the missed detection of a person working on the mobile scaffold; the reason for this missed detection could be the tilt position of the worker. The system also shows the unsafe mobile scaffold as a safe behavior due to the extra pipes near the mobile scaffold.

4.4. Proposed system evaluation

The performance evaluation aimed to check the ability of the proposed system for detecting construction entities and recognizing the behavior (safe/unsafe) based on performance indicators such as precision (equation-4), recall (equation-5), F1 score(equation-6), and accuracy (equation-7). Precision and recall are the metrics for classification models; the former identifies the fraction of relevant instances among those retrieved, whereas the latter calculates the number of relevant instances retrieved over the total amount (Fang et al., 2018b). Similarly, the F1-score is a measure for the accuracy of the test used in the statistical analysis of binary classification, and accuracy in binary classification is the sum of true positives and true negatives divided by the total number of instances. The performance metrics taken into consideration for the overall system reliability and robustness are given below:

Table 3
Performance Results of the Trained Model.

Performance Metrics	Percentage
Mean average precision	70.30 %
Precision	92.35 %
Recall	91.89 %
F1-score	92.12 %
Accuracy	87.28 %

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive\ (FP)} \tag{4}$$

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negetive\ (FN)} \tag{5}$$

$$F1 - Score = 2.\frac{(Precision)\ .\ (Recall)}{(Precision)\ + (Recall)} \tag{6}$$

$$Accuracy = \frac{True\ Positive\ (TP) + True\ Negetive\ (TN)}{True\ Positive\ (TP) + False\ Negetive\ (FN) + True\ Negetive\ (TN) + False\ Positive\ (FN)} \tag{7}$$

4.4.1. Evaluation of detection model

Table 3 presents the results of the performance indicators taken into account for the evaluation of the model. The trained model exhibits a precision and recall of 92.35 % and 91.89 %, respectively, with an overall mean average precision (mAP) of 70.3 %, having a 50 % intersection over union (IoU). The results revealed promising outputs for the test data, and the model was selected to perform detection on the real construction sites data.

4.4.2. Evaluation of OCD block

The proposed approach is additionally evaluated for the OCD block with four performance indicators on the validation set. The validation set (235-images) were grouped into two classes: (1) *class-1* was for safe behaviors (158-images), and (2) *class-2* was comprised of unsafe behaviors (78-images). The model was trained on the images of persons (P), mobile scaffolds with outriggers (S_o), and mobile scaffolds with no outriggers (S_n); recall Eqs. 1 and 2. Thus, the images that contain person/s working on the top of the mobile scaffolds having no outriggers (A+ S_n) was classified in class-2, while the rest of the images (A+S_o, A, S_n, S_o), and images that have empty (S_n, S_o) were kept in class -1. Unique identification numbers are assigned to each image, and both the classes (class-1& class-2) are converted to binary numbers for comparison between the ground truth and predications. Fig. 14. shows confusion matrix ($n \times n$) for classification model, where n is the number of target

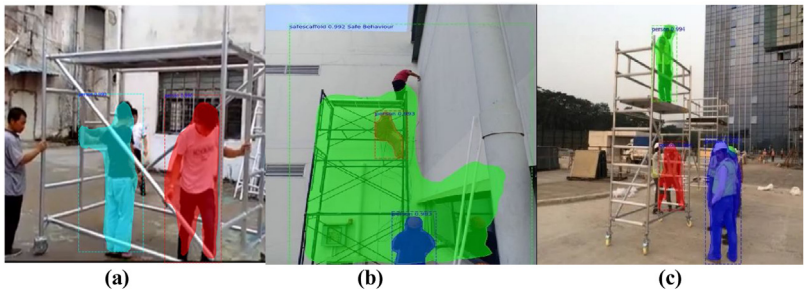

Fig. 13. Examples of missed detection: (a) missed detection of mobile scaffolding, (b) missed detection of a person, (c) missed detection due to occlusion.

Table 4
Performance Results of the Proposed System.

Behavior (Images)	Classification	Binary Numbers	Precision	Recall	F1-score	Overall Accuracy
Safe (158)	Class-1	0	0.85	0.97	0.91	0.86
Unsafe (78)	Class-2	1	0.91	0.65	0.76	

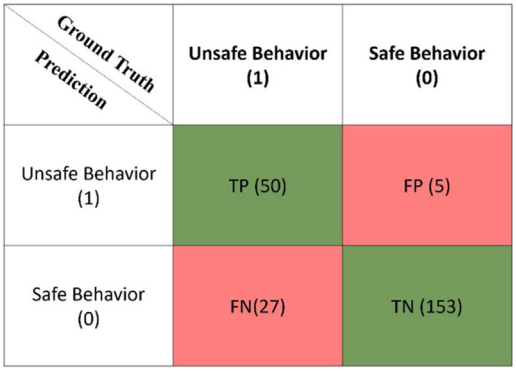

Fig. 14. Confusion matrix for the model using binary classification.

classes (safe behavior or unsafe behavior) which is (2 × 2) in this study. The columns represent the ground truths, with the predictions of the targets in rows. The unsafe behavior correctly classified (true positive -TP) by the model were fifty in numbers, while the safe behavior in actual but classified as unsafe by the model (false positive – FP) were 5. The model predicts the scene as a safe situation and indeed having no unsafe situation (true negative -TN) were 153; however, the scene predicted as a safe and indeed has an unsafe situation in actual (false negative -FN) were noticed 27 in numbers.

Table 4 presents the results of the performance indicators for the evaluation of the proposed system. The overall accuracy recorded during the evaluation process was 86 %. Similarly, the other indicators were individually calculated for classified classes; for instance, the precision, recall, and F1-score observed for class-1 were 85 %, 97 %, and 91 %, respectively. However, the precision, recall, and F1-score for class-2 were 91 %, 65 %, and 76 %, respectively.

5. Conclusion and future outlook

Safety in construction is directly proportional to the safety of workers' lives. Generally, construction safety comprises two phases: safety planning and safety monitoring, of which the latter is more critical because of its nature as a final layer of safety management, yet it depends on the manual practices. Thus far, few research studies have considered vision intelligence applications as a vital technology to tackle the safety management issue in construction. In order to put some effort, this study contributes to extend the scope of vision intelligence application by introducing a correlation identification between the construction entities. The results revealed that the proposed approach offered excellent accuracy in detecting an individual's unsafe behavior pertaining to mobile scaffolds not equipped with outriggers; however, several limitations in this study still exist, which need to be addressed in the future.

5.1. Summary of findings and contributions

This work extracts construction safety rules related to mobile scaffolds that correlate with other construction entities (classes). Thus a new dataset for three classes, such as (1) construction worker (person), (2) mobile scaffolds with outriggers, and (3) mobile scaffolds with no outriggers, is developed. The effectiveness of the collected dataset is enumerated through deep convolution neural networks with the additional OCD block. The key findings are listed as follows:

1 Images were collected from two different sources for the development of the dataset and model training.
2 The developed correlation-dependent vision intelligence system was successfully implemented on two scenarios from KOSHA safety rules.
3 The preliminary results of the detection using mask R-CNN algorithms were presented. The detection model's quantitative findings revealed 70.30 % of mean average precision (mAP), 92.35 % for precision, 91.89 % for recall, 92.12 % for F1-score, and 87.28 % for overall accuracy.
4 The trained model is extended with additional OCD block to determine the behavior (safe/unsafe) based on workers' correlation with mobile scaffolds equipped having outriggers or with no outriggers.
5 The proposed approach is additionally evaluated with four performance indicators. The overall accuracy noted during the evaluation process was 86 %, whereas the rest of the three indicators were individually calculated: for instance, the precision, recall, and F1-score observed for class-1 were 85 %, 97 %, and 91 %, respectively. Correspondingly, the precision, recall, and F1-score for class-2 were noted 91 %, 65 %, and 76 %, respectively.

Contemplating the detection model outturn through several performance indicators, the authors are convinced to recommend the convolution neural network (CNN)-based image detection algorithms as a promising approach to recognized hazards in the construction job site. Motivated by the facts and figures gained through the performance indicators, the authors are persuaded by the outcomes of the OCD block and expect that the proposed system could facilitate and enhance the current safety management process on the construction site.

5.2. Limitations and future outlook

Even though this research presented a novel vision intelligence system leveraging deep learning algorithms to put some efforts in the same domain; yet, it is not deprived of limitations. Firstly, this study has implemented a deep learning technique on a limited dataset. Secondly, the system was not able to detect all the unsafe behaviors. The reason might be the limited dataset and captured angle of the 2D images; thus, significant improvement is needed to detect accurate coordinates of the construction entities (for instance: images from CCTV and using a stereo camera can overcome this issue). A larger dataset is required from real construction sites to improve the results and prevent overfitting. Third, detecting unsafe behavior of persons below the midpoint of the mobile scaffolds bounding box was not considered in this study. Because the height of the mobile scaffold used in the construction site varies, and the site itself is dynamic in nature, it could lead to occlusion problems. Solving these issues by incorporating threshold for different types of mobile scaffold and techniques for occlusion such as random erasing, cutout, hide and seek on the training dataset will enable vision intelligence-based real-time monitoring systems to be much more efficient prolong in their application. Finally, this study did not consider the structural failure of the mobile scaf-

folds, which is also crucial for safety in construction and could be addressed by using strain sensors or stereo vision with AI.

In the future, we intend to create a larger dataset from real construction sites for improved detection of construction objects for the practical implementation of a computer vision-based construction safety system in the job site. To obtain a real reflection of the construction job site and more practical results, the same framework will be extended to accident cases from the database. Furthermore, we expect that the developed system can automatically assess safety performance when integrated with other technologies, such as blockchain and the Internet of Things.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2020R1A4A4078916), and Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (National Research for Smart Construction Technology: Grant 20SMIP-A158708-01).

References

Building, B., 2009. International Occupational Safety and Health Information Centre. ILO Cat. Publ. Data, pp. 9–21 http://www.ilo.org/public/english/protection/safework/cis/products/bulletin.htm.

Cho, C., Sakhakarmi, S., Kim, K., Park, J.W., 2018a. Scaffolding modelling for real-time monitoring using a strain sensing approach. ISARC 2018 - 35th Int. Symp. Autom. Robot. Constr. Int. AEC/FM Hackathon Futur. Build. Things, International Association for Automation and Robotics in Construction I.A.A.R.C), http://dx.doi.org/10.22260/isarc2018/0007.

Cho, C., Kim, K., Asce, A.M., Park, J., Asce, A.M., Cho, Y.K., 2018b. Data-Driven Monitoring System for Preventing the Collapse of Scaffolding Structures., http://dx.doi.org/10.1061/(ASCE)CO.1943-7862.0001535.

C. Cho, J. Park, K. Kim, S. Sakhakarmi, Machine Learning for Assessing Real-Time Safety Conditions of Scaffolds, 2021 n.d.

Collins, R., Zhang, S., Kim, K., Teizer, J., 2014. Integration of safety risk factors in BIM for scaffolding construction. Proc. 2014 Int. Conf. Comput. Civ. Build. Eng., 307–314.

J. Dai, K. He, J. Sun, Instance-aware Semantic Segmentation via Multi-task Network Cascades, 2021 n.d.

Darko, A., Chan, A.P.C., Yang, Y., Tetteh, M.O., 2020. Building information modeling (BIM)-based modular integrated construction risk management – critical survey and future needs. Comput. Ind. 123, 103327, http://dx.doi.org/10.1016/j.compind.2020.103327.

Ding, L., Fang, W., Luo, H., Love, P.E.D., Zhong, B., Ouyang, X., 2018. A deep hybrid learning model to detect unsafe behavior: integrating convolution neural networks and long short-term memory. Autom. Constr. 86, 118–124, http://dx.doi.org/10.1016/j.autcon.2017.11.002.

A. Dutta, A. Zisserman, The VIA Annotation Software for Images, Audio and Video, 2021 n.d. http://www.robots.ox.ac.uk/ (accessed March 8, 2021).

Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T.M., An, W., 2018a. Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. Autom. Constr. 85, 1–9, http://dx.doi.org/10.1016/j.autcon.2017.09.018.

Fang, W., Ding, L., Luo, H., Love, P.E.D., 2018b. Falls from heights: a computer vision-based approach for safety harness detection. Autom. Constr. 91, 53–61, http://dx.doi.org/10.1016/j.autcon.2018.02.018.

Fang, W., Zhong, B., Zhao, N., Love, P.E.D., Luo, H., Xue, J., Xu, S., 2019. A deep learning-based approach for mitigating falls from height with computer vision: convolutional neural network. Adv. Eng. Informatics 39, 170–177, http://dx.doi.org/10.1016/j.aei.2018.12.005.

Fang, W., Ma, L., Love, P.E.D., Luo, H., Ding, L., Zhou, A., 2020. Knowledge graph for identifying hazards on construction sites: integrating computer vision with ontology. Autom. Constr. 119, 103310, http://dx.doi.org/10.1016/j.autcon.2020.103310.

Feng, C., Liu, M.-Y., Kao, C.-C., Lee, T.-Y., 2017. Deep Active Learning for Civil Infrastructure Defect Detection and Classification., pp. 298–306, http://dx.doi.org/10.1061/9780784480823.036.

Golparvar-Fard, M., Heydarian, A., Niebles, J.C., 2013. Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers. Adv. Eng. Informatics 27, 652–663, http://dx.doi.org/10.1016/j.aei.2013.09.001.

Google Developers, 2020. Machine Learning Glossary (accessed September 12, 2020) https://developers.Google.com/machine-learning/glossary.

Harrison, J., Dawson, L., 2015. Occupational Health: Meeting the challenges of the next twenty years. Saf. Health Work 7, 1–7, http://dx.doi.org/10.1016/j.shaw.2015.12.004.

Haupt, T., 2001. The Performance Approach to Construction Worker Safety and Health., pp. 372 http://purl.fcla.edu/fcla/etd/ufe0000322.

He, K., Gkioxari, G., Dollar, P., Girshick, R., 2017. Mask R-CNN. Proc. IEEE Int. Conf. Comput. Vis., 2980–2988, http://dx.doi.org/10.1109/ICCV.2017.322, 2017-Octob.

Heaton, J., Parlikad, A.K., Schooling, J., 2019. Design and development of BIM models to support operations and maintenance. Comput. Ind. 111, 172–186, http://dx.doi.org/10.1016/j.compind.2019.08.001.

Hoła, A., Sawicki, M., Szóstak, M., 2018. Methodology of classifying the causes of occupational accidents involving construction scaffolding using Pareto-Lorenz analysis. Appl. Sci. 8, http://dx.doi.org/10.3390/app8010048.

Hung-Lin Chi, J., Chai, Jian, Wu, Changzhi, Zhu, C.L., Wang, Xiangyu, 2017. Scaffolding Progress Monitoring of LNG Plant Maintenance Project Using BIM and Image Processing Technologies. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8002505.

Jung, Y., 2014. An Approach to Automated Detection of Failure in Temporary Structures Using Image Processing.

Jung, Y., Oh, H., Jeong, M.M., 2018. An approach to automated detection of structural failure using chronological image analysis in temporary structures an approach to automated detection of structural failure using chronological image analysis in temporary structures. Int. J. Constr. Manag. 19, 178–185, http://dx.doi.org/10.1080/15623599.2017.1411457.

Khan, N., Ali, A.K., Skibniewski, M.J., Lee, D.Y., Park, C., 2019. Excavation Safety Modeling Approach Using B.I.M. and VPL., pp. 2019.

Kim, H., Kim, H., Hong, Y.W., Byun, H., 2017. Detecting construction equipment using a region-based fully convolutional network and transfer learning. J. Comput. Civ. Eng. 32, 04017082, http://dx.doi.org/10.1061/(asce)cp.1943-5487.0000731.

Kim, D., Liu, M., Lee, S.H., Kamat, V.R., 2019. Remote proximity monitoring between mobile construction resources using camera-mounted UAVs. Autom. Constr. 99, 168–182, http://dx.doi.org/10.1016/j.autcon.2018.12.014.

KOSHA, 2019. Practicale Guide for Scaffolfing in Construction Sites. Korean Occup. Saf. Heal. Agency Rep., pp. 1–230 http://oktoeng.com/bbs/board.php?bo_table=b6_30&wr_id=300.

KOSHA, KOSHA | Legislation | Occupational Safety and Health, Korean Occup. Saf. Heal. Agency Rep. 2021 (n.d.). http://www.kosha.or.kr/english/legislation/occupationalSafetyAndHealth.do?mode=list (accessed August 25, 2020).

T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C.L. Zitnick, P. Dolí, Microsoft COCO: Common Objects in Context, 2021 n.d.

Liu, M., Han, S., Lee, S., 2016. Tracking-based 3D human skeleton extraction from stereo video camera toward an on-site safety and ergonomic analysis. Constr. Innov. Inf. Process. Manag. 16, 348–367, http://dx.doi.org/10.1108/CI-10-2015-0054.

Luo, H., Xiong, C., Fang, W., Love, P.E.D., Zhang, B., Ouyang, X., 2018. Convolutional neural networks: computer vision-based workforce activity assessment in construction. Autom. Constr. 94, 282–289, http://dx.doi.org/10.1016/j.autcon.2018.06.007.

Memarzadeh, M., Golparvar-Fard, M., Niebles, J.C., 2013. Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors. Autom. Constr. 32, 24–37, http://dx.doi.org/10.1016/j.autcon.2012.12.002.

Mneymneh, B.E., Abbas, M., Khoury, H., 2018. Vision-Based Framework for Intelligent Monitoring of Hardhat Wearing on Construction Sites., http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000813.

Nath, N.D., Behzadan, A.H., Paal, S.G., 2020. Deep learning for site safety: real-time detection of personal protective equipment. Autom. Constr. 112, http://dx.doi.org/10.1016/j.autcon.2020.103085.

OSHA, 2017. Commonly Used Statistics. Occupational Safety and Health Administration (accessed January 31, 2020) https://www.osha.gov/data/commonstats.

Papazoglou, I.A., Aneziris, O., Bellamy, L., Ale, B.J.M., Oh, J.I.H., 2015. Uncertainty assessment in the quantification of risk rates of occupational accidents. Risk Anal. 35, 1536–1561, http://dx.doi.org/10.1111/risa.12354.

Park, C.S., Kim, H.J., 2013. A framework for construction safety management and visualization system. Autom. Constr. 33, 95–103, http://dx.doi.org/10.1016/j.autcon.2012.09.012.

Park, C., Lee, D., Khan, N., 2020. An Analysis on Safety Risk Judgment Patterns Towards Computer Vision Based Construction Safety Management., pp. 52, http://dx.doi.org/10.3311/CCC2020-052.

Reason, J., 2000. Human errors models and management. BMJ Br. Med. Journal, 4–5 http://130.88.20.21/trasnusafe/pdfs/HumanErrorsModelsandManagement.pdf.

Reese, Charles D., Eidson, J.V., 2006. Handbook of OSHA Construction Safety and Health. CRC Press Taylor & Francis Group, http://dx.doi.org/10.1201/9781420006230.

S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2021 n.d. http://image-net.org/challenges/LSVRC/2015/results (accessed September 14, 2020).

Riggs, J., 2018. HSE Statistics Should Remind Us About the Dangers of Working at Height. Planning, BIM & Construction Today, pp. 1 (accessed May 13, 2020) https://www.pbctoday.co.uk/news/health-safety-news/hse-statistics-should-remind-us-about-the-dangers-of-working-at-height/39615/.

Sakhakarmi, S., Park, J.W., 2020. Multi-level-Phase deep learning using divide-and-Conquer for scaffolding safety. Int. J. Environ. Res. Public Health 17, 2391, http://dx.doi.org/10.3390/ijerph17072391.

Sakhakarmi, S., Park, J., Asce, A.M., Cho, C., 2018. Enhanced Machine Learning Classification Accuracy for Scaffolding Safety Using Increased Features., http://dx.doi.org/10.1061/(ASCE).

Seo, J., Han, S., Lee, S., Kim, H., 2015. Computer vision techniques for construction safety and health monitoring. Adv. Eng. Informatics. 29, 239–251, http://dx.doi.org/10.1016/j.aei.2015.02.001.

Soltani, M.M., Zhu, Z., Hammad, A., 2018. Framework for location data fusion and pose estimation of excavators using stereo vision. J. Comput. Civ. Eng. 32, http://dx.doi.org/10.1061/(ASCE) CP.1943-5487.0000783.

Szymański, P., 2017. Risk management in construction projects. Procedia Eng. 208, 174–182, http://dx.doi.org/10.1016/j.proeng.2017.11.036.

Tang, Shuai, Golparvar-Fard, Mani, 2017. Joint reasoning of visual and text data for safety hazard recognition. Comput. Civ. Eng. 3 (2017), 326–334, http://dx.doi.org/10.1061/9780784407943.

Tatić, D., Tešić, B., 2017. The application of augmented reality technologies for the improvement of occupational safety in an industrial environment. Comput. Ind. 85, 1–10, http://dx.doi.org/10.1016/j.compind.2016.11.004.

Wooten, Jeremiah, 2019. OSHA Reveals Top 10 Health & Safety Violations for 2019 (accessed May 13, 2020) https://inspectioneering.com/blog/2019-12-18/8922/oshas-top-ten-safety-and-health-violations-for-2019.

Xu, Y., Tuttas, S., Hoegner, L., Stilla, U., 2018. Reconstruction of scaffolds from a photogrammetric point cloud of construction sites using a novel 3D local feature descriptor. Autom. Constr. 85, 76–95, http://dx.doi.org/10.1016/j.autcon.2017.09.014.

Yuan, X., Anumba, C.J., Parfitt, M.K., 2016. Cyber-physical systems for temporary structure monitoring. Autom. Constr. 66, 1–14, http://dx.doi.org/10.1016/j.autcon.2016.02.005.

Zhang, S., Sulankivi, K., Kiviniemi, M., Romo, I., Eastman, C.M., Teizer, J., 2015. BIM-based fall hazard identification and prevention in construction safety planning. Saf. Sci. 72, 31–45, http://dx.doi.org/10.1016/j.ssci.2014.08.001.