

Received March 21, 2022, accepted March 30, 2022, date of publication April 4, 2022, date of current version April 8, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3164676

# Fall Prevention From Ladders Utilizing a Deep Learning-Based Height Assessment Method

**SHARJEEL ANJUM**<sup>1</sup>, **NUMAN KHAN**<sup>2,3</sup>, **RABIA KHALID**<sup>1</sup>, (Member, IEEE),  
**MUHAMMAD KHAN**<sup>1,4</sup>, **DONGMIN LEE**<sup>1</sup>, AND **CHANSIK PARK**<sup>1</sup>

<sup>1</sup>School of Architecture and Building Sciences, Chung-Ang University, Seoul 06974, South Korea

<sup>2</sup>Department of Construction Engineering, École de Technologie Supérieure, Montreal, QC H3C 1K3, Canada

<sup>3</sup>Healthy Building Research Center, Ajman University, Ajman, United Arab Emirates

<sup>4</sup>Construction Innovation Integration Laboratory (CII-Lab), Civil, Construction and Environmental Engineering Department, The University of Alabama, Tuscaloosa, AL 35487, USA

Corresponding author: Chansik Park (cpark@cau.ac.kr)

This work was supported in part by the Chung-Ang University Young Scientist Scholarship in 2020; in part by the National Research and Development Project for Smart Construction Technology funded by the Korea Agency for Infrastructure Technology Advancement through the Ministry of Land, Infrastructure, and Transport and managed by the Korea Expressway Corporation under Grant 22SMIP-A158708-03; and in part by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT), Korea Government, under Grant NRF-2020R1A4A4078916.

**ABSTRACT** According to the Center for Construction Research and Training (CPWR) and the Korea Occupational Safety & Health Agency (KOSHA), falls from ladders are a leading cause of fatalities. The current safety inspection process to enforce height-related rules is manual and time-consuming. It requires the physical presence of a safety manager, for whom it is sometimes impossible to monitor an entire area in which ladders are being used. Deep learning-based computer vision technology has the potential to capture a large amount of useful information from a digital image. Therefore, this paper presents a deep learning-based height assessment method using a single known value in an image to measure working height, monitor compliance to safety rules, and ensure worker safety. The proposed method comprises (1) extraction of safety rules from the KOSHA database related to the A-type ladder; (2) object detection (Single Shot Multibox Detector SSD) (3) a height-computing module (HCM) to estimate the working height of the worker (how high a worker is from the ground); and (4) classification of worker behavior (using the developed SSD-based HCM) based on the best practices derived from the KOSHA database. The developed algorithm has been tested on four different scenarios based on KOSHA safety rules, with heights ranging from under 1.2 m to over 2 m. Additionally, the proposed method was evaluated on 300 images for binary classification (safe and unsafe) and achieved an overall accuracy of 85.33%, verifying its feasibility for intelligent height estimation and compliance monitoring.

**INDEX TERMS** Falls from ladders, vision intelligence-based monitoring, construction safety rules, vision-based height estimation, deep learning.

## I. INTRODUCTION

Construction work is likely to expose workers to hazardous situations because of its distinctive, dynamic, and complex nature. Compared with the workers in other industries, those in the construction industry are more prone to occupational accidents that could potentially lead to fatalities [1]–[3]. The most common construction accidents include fall from height (FFH), collision with objects, electrocution, and being stuck

in between machinery. Among these, FFH is the most frequent cause of accidents at construction sites. According to the U.S. Bureau of Labor Statistics, the number of fatalities due to falls to a lower level increased by 26% from 2011 to 2016, which includes ladders (836 fatal injuries) and rooftops (763). The most common height of a fall was reported as 30 ft (658 deaths) [4]. In addition, the Center for Construction Research and Training (CPWR) reports that ladder (93)-related accidents were a leading cause of fatal injuries from 2011 to 2019 [5]. Moreover, in South Korea, industrial accidents are increasing, with accidents at construction sites

The associate editor coordinating the review of this manuscript and approving it for publication was Turgay Celik<sup>1</sup>.

accounting for more than one-third of all industrial accidents. According to an industrial accident survey conducted by the Korea Occupational Safety and Health Agency (KOSHA) from 2009 to 2017, fall-related accidents accounted for 47.7–52.1% of the total number of deaths in the construction industry [6]. Likewise, Sim and Kang [7] published a report in 2017, which stated that ladder fatalities accounted for 31% of the total industrial fatalities between 2005 and 2014. The statistics mentioned above illustrate that, in the construction industry, falls from ladders (FFL) are a severe problem in terms of workers' fatalities and injuries. The problems of FFL should be effectively addressed to ensure worker safety.

FFLs are caused by a defective ladder or climbing a ladder with material, climbing a free ladder, or contravening safety rules [8], [9]. The safety rules of the Occupational Health and Safety Administration (OSHA), International Organization of Standardization (ISO), and KOSHA [6], [10], [11] are outlined to reduce FFHs and FFLs. Accidents due to rule contravening result in workers being laid off, low quality of life, and health issues. These factors negatively affect a company's production, finances, and reputation [12]. Unsafe worker behavior has been recognized as a key factor in construction-related accidents [13]. Monitoring the unsafe behavior of workers is important for reducing risks at construction sites. Therefore, safety managers should constantly monitor workers to mitigate the factors mentioned above. However, this monitoring process is manual (safety managers must be physically present at construction workplaces to identify potential hazards or non-compliance to safety rules) and time-consuming, relying on the safety managers' personal experience and competence [14]. The safety training program is another effective and standard approach to reduce unsafe behavior; countries with strong occupational safety rules require workers to have training certification before working at construction sites. Trained workers have adequate knowledge to understand the consequences of unsafe behavior while working at heights; however, they do not take safety rules (such as wearing proper personal protective equipment) seriously, and sometimes unsafe behavior happens out of lack of concern for their safety [15]. To deal with the underestimation and inadequate awareness of risks, researchers have aimed to improve risk management by developing risk management methods for worker safety, which can be categorized into proactive and reactive [16]. Proactive risk analysis methods are preferred for workers' protection, owing to their ability to collect and analyze data in real-time from cameras and sensors. Owing to the complexity of construction sites, the use of sensor technology is limited. Computer vision (CV)-based technology for collecting data to analyze risk from on-site installed cameras is an ideal solution for clogged (material, workers, and equipment) construction sites and preferable for workers who do not want to attach sensors to their bodies [14].

CV-based object detection utilizes two approaches: (1) traditional and (2) deep learning. Before the emergence of deep learning technology, traditional object detection algo-

rithms such as feature descriptors (BRIEF, SIFT, and SURF) were used to extract useful information from digital images [17]. The main problem with traditional approaches is the selection of important features (handcrafted) from images; as the number of classes increases, the feature extraction process becomes more difficult. Thus, researchers introduced convolutional neural network (CNN)-based deep learning technology, which can automatically extract and recognize features from a static image by stacking multiple convolutional and pooling layers.

In recent years, CNNs have played a positive role in CV and pattern recognition [18], [19]. Following the role of CNN in CV, researchers incorporated deep learning-based technology in the construction domain for automated documentation, safety monitoring, hazard detection, and defect detection [14], [20], [21]. For instance, Thakar *et al.* [22] used an object detector (SSD) with non-maximum suppression as a base model for asset monitoring at construction sites. In addition, they used affinity propagation clustering to enhance the performance of SSD, offering an optimum balance between speed and accuracy. Similarly, Ma *et al.* [23] proposed a quality inspection framework by combining an object detector (SSD) used to detect five different types of defects with building information modeling to improve productivity and overcome unnecessary deviations caused by human judgment. Zhong *et al.* [24] proposed a deep learning-based text classification approach by integrating natural language processing and CNN for accident reports and utilized latent dirichlet allocation to understand the factors contributing to construction accidents. Their proposed approach helped safety managers to improve safety at construction sites by investigating accident reports.

The researcher extended the application of CV technology to worker safety and addressed serious FFH problems at construction sites. For instance, Khan *et al.* [25] proposed a mask region-based convolutional neural network (R-CNN-based) detection algorithm to monitor workers' behavior while working on top of mobile scaffolds. Tang *et al.* [26] applied the two-stage object detection algorithm Faster R-CNN to monitor workers' behavior and ensure the safety of their faces, eyes, hands, and feet. Fang *et al.* [27] developed a region-based convolutional neural network (Mask R-CNN) to identify the unsafe behavior of workers crossing the structural support during the construction of deep-pit foundations. Similarly, Wang *et al.* [28] proposed a vision-based system for workers' safety by identifying and analyzing worker-equipment interactions to identify danger zones and generate safety alarms. Likewise, Khan *et al.* [29] developed a tag and Internet of Things (IoT)-based safety hook for worker safety to prevent falls from scaffolding while working on height.

However, few studies have focused on worker safety while working on ladders. For example, Seo *et al.* [30] conducted a study to understand the risks of falls and work-related musculoskeletal disorders while working on ladders by estimating musculoskeletal stress. Ding *et al.* [31] introduced a deep learning-based hybrid model comprising a CNN

and long short-term memory that automatically identified unsafe behavior by detecting workers working on a ladder. Piao *et al.* [32] proposed a dynamic fall risk assessment framework for construction workers that combined CV and the Bayesian network to reduce FFH by automatically detecting risk factors and improving risk assessment efficiency and used working on a ladder as a case study. Chen *et al.* [33] introduced a proactive worker safety risk evaluation framework using position and worker posture as quantitative indicators to classify workers' behavior. The authors used IMU sensors with a vision-based 3D skeleton and ultra-wideband (UWB) to classify workers' behavior as safe or unsafe. Likewise, Han *et al.* [34] collected motion data with a Kinect® depth sensor and investigated motion analysis approaches to automatically recognize unsafe worker behavior while climbing a ladder. These approaches to detect unsafe ladders use motion capturing-based activity recognition models that can intelligently distinguish safety-related behaviors on a ladder. A comparative analysis of existing studies related to worker safety with the proposed study is summarised in Table 1. The comparative analysis is performed based on previous methods, working height estimation, ensuring safety rules at a specified working height, and targeted objects.

Table 1 summarises that, despite the excellent and in-depth research, previous studies exhibit two major limitations. First, checking the safety rules correlated with the specific working height, which is particularly important because workers perform various tasks at different working heights in construction sites, and negligence in compliance with safety rules can lead to hazardous situations. Second, motion capturing-based activity recognition models are more computationally complex than object detection models. Furthermore, motion capturing-based activity recognition models have a higher false detection rate (low accuracy) when the scenes to be monitored are closely correlated [34]. Therefore, an automated and less computationally intensive method to monitor workers' safety while they work on ladders is required. In light of the above findings, we decided to use a vision-based technology for enhancing workers' safety, and this study contributed as follows:

- The manually extracted occupational safety rules correlated with the A-type ladder from the KOSHA expert knowledge database (constituted by the ISO-450001) [35] have been incorporated with CV technology. The integrated technology replaced the manual safety inspection process for real-time monitoring of worker safety.
- A dataset for safety behavior detection is created using 21 videos of working on an A-type ladder. The frames have been extracted and labeled (1825) for Deep learning-based object detection.
- A height-computing module (HCM) leveraging a deep learning-based object detection approach (SSD) has been developed to estimate the working height of a worker with the help of detected object coordinates that

are applied as a source of information. The working height and examination of occupational rules relating to this height could significantly prevent FFLs.

- The proposed topology validated workers' safe and unsafe behavior using four different cases following the KOSHA rules. The developed algorithm utilized object detection as a base model for the HCM to estimate the working height and compared it with the corresponding occupational safety rule.

## II. RESEARCH METHODOLOGY

Various ladders are used during construction work; however, the scope of this research is focused on A-type ladders. These ladders are primarily used indoors for short-duration work. A fall from these ladders is a particularly severe concern in the industry. Therefore, this study aimed to develop a vision-based HCM that provides an effective solution for automating worker safety monitoring (rule compliance) on an A-type ladder. This section outlines the research process to develop an algorithm. The proposed algorithm can be used to measure the working height from a vision sensor and recognize unsafe worker behavior while they work on an A-type ladder.

A systematic approach deployed in this work comprises of the following four steps (Fig. 1): (1) problem identification and objective, (2) development of the algorithm, (3) experimental setup and results, and (4) evaluation and future work. In the first step, we have identified the problem as FFL. We reiterate that the construction industry must look at an automated solution to compute the working height and monitor the workers' behavior when they are using an A-type ladder. During the development stage (Step 2), dataset preparation, model training, and HCM were performed. Subsequently, in Step 3, four scenarios were evaluated to verify the performance of SSD and HCM modules. Then as the last step, the SSD object detector and HCM have been evaluated and discussed.

### A. PROCESS FLOW

Fig. 2 depicts the workflow of the proposed method, which is described in detail in this section. We extracted frames from the video, and for each frame, the algorithm performed visual recognition to detect a worker and an A-type ladder. Following this, their pixel values are compared, and the intersection between the two bounding boxes is calculated to determine whether the bounding box representing the worker is inside or outside the box, which represents the A-type ladder. If the intersection between the two bounding boxes is true and the worker's bottom-left corner value is less than the ladder bottom-left corner value, then the worker is standing on the ladder. Subsequently, the frame goes through the remaining stages of the algorithm; otherwise, the visual recognition process continues. If a worker is on the ladder, the next step is to determine the height (how high the worker is from the ground). Following this, the KOSHA regulation is examined to classify worker behavior based on the computed height.

TABLE 1. Previous research for workers safety in construction.

Ref	Objective	Height Estimation	Safety Rules Checking at specific working height	Workers Safety	Target Objects	Method
[33]	Automated safety risk evaluation	✗	✗	✓	Worker and Ladder	Fusion of position and posture (IMU Sensor)
[31]	To recognize worker behavior while working on a ladder	✗	✗	✓	Worker and Ladder	CNN and LSTM for features extraction and sequencing learning feature, respectively
[30]	To understand the risk of falls during ladder climbing	✗	✗	✓	Person and Ladder	Vision-based motion captures technology and performed biomechanical analysis to compute musculoskeletal stresses
[32]	To improve the efficiency of risk assessment and reduce the risk of falls through early alarming	✓	✗	✓	Moving Objects, Ladder, and Worker	CV combined with dynamic Bayesian network
[29]	To improve the worker safety	✓	✗	✓	Scaffolding, Worker, and Safety hook	Combined IoT with the Tag base technology to classify workers' behavior
[25]	To recognize unsafe behavior while working on the mobile scaffolding to overcome FFH	✗	✗	✓	Workers and mobile scaffolding	Mask R-CNN is used to detect and generate mask and post-processing detected objects
[36]	To determine whether the worker is wearing a safety harness to overcome FFH	✗	✗	✓	Workers and safety harness	The Faster R-CNN and deep CNN were used to detect worker and safety harnesses
Our	To recognize unsafe behavior at a specific working height on a ladder	✓	✓	✓	Ladder and Worker	CV-based object detection is utilized as a base model for the proposed HCM to examine safety rules and categorize workers' behavior.

The KOSHA defined specific rules for the worker; working at different heights on A-type ladders is explained with an example. A construction worker (working on an A-type ladder up to a maximum height of 1.2 m) exhibits safe behavior when wearing a helmet; otherwise, the behavior is considered unsafe. However, the rules change when the worker works on an A-type ladder at a height greater than 1.2 m but less than or equal to 2 m. In this case, the corresponding KOSHA rule states that the workers must wear a helmet, and those two workers should work together for safe behavior; otherwise, it is unsafe.

**B. PROPOSED METHOD**

To estimate the working height from the ground, we need to select a reference point in all frames of an entire video sequence and then compute the working height with respect to the reference point. Therefore, this study used a fixed

A-type ladder height (1.7 m) as the reference point [37]. The A-type ladder is a modified version of the simple ladder that doesn't inherit the required additional support, as shown in Fig. 3. Fig. 4 depicts the coordinate system of the CV-based approach for height computing using the pixel values of the detected objects.

Fig. 4(a) depicts that the detected objects in a given digital image are a worker with a safety belt and a ladder with outriggers. Fig. 4(b) illustrates the coordinates of interest for the objects detected in Fig. 4(a).

$P_{l,t} = X_t^p, Y_t^p, W_1^p, H_1^p$  and  $P_{l,b} = H_1^p + Y_t^p$  are the top-left and bottom-left coordinates of the person, respectively, and  $L_{l,t} = X_t^l, Y_t^l, W_2^l, H_2^l$  and  $L_{l,b} = H_2^l + Y_t^l$  are the top-left and bottom-left coordinates of the ladder, respectively.

After extracting the coordinates of the detected objects in the digital image, the next step is to determine the working height of the worker on the A-type ladder. The SSD

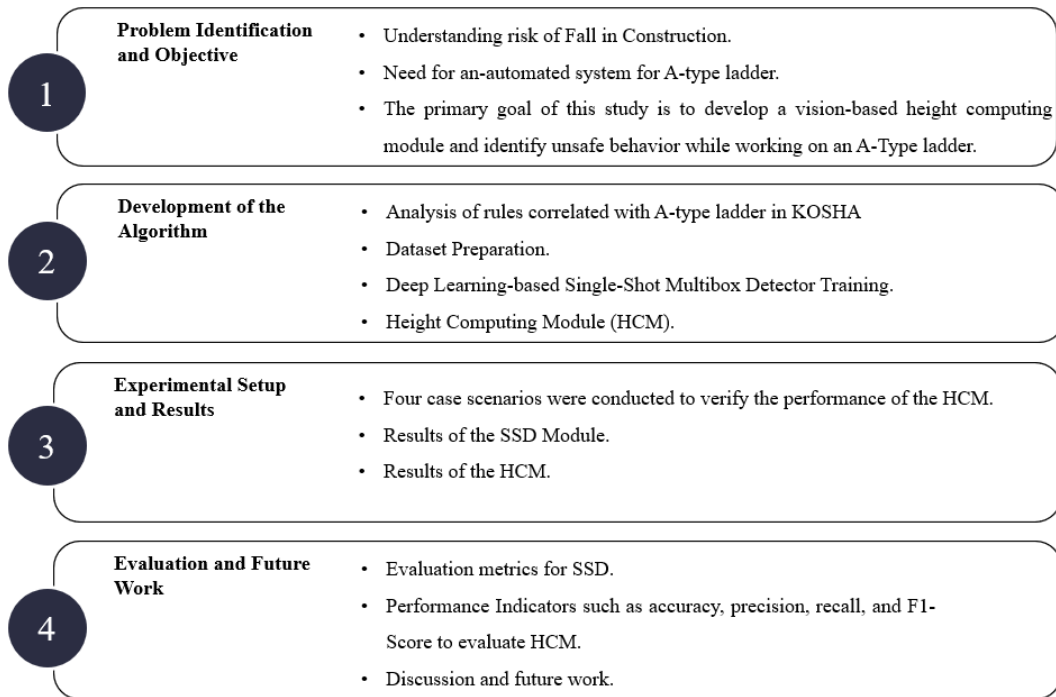


FIGURE 1. Research approach for safety risk identification on ladders utilizing the Deep Learning-based height assessment method.

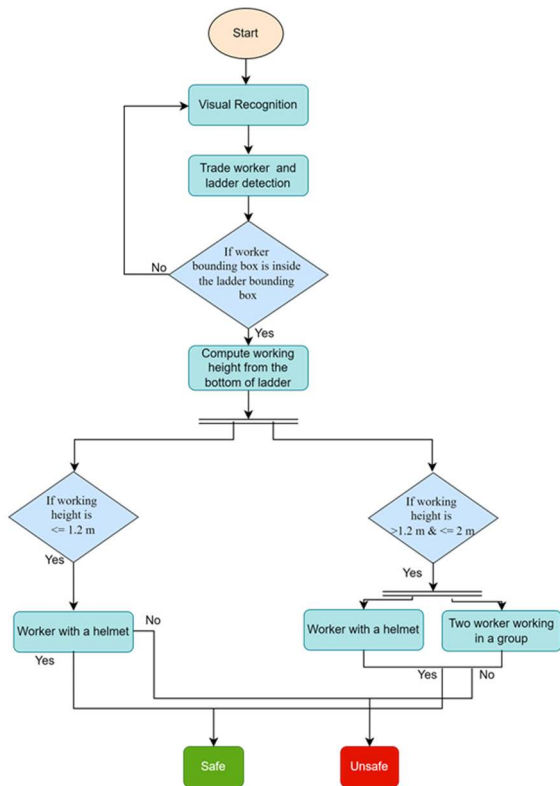


FIGURE 2. Process flow of the proposed algorithm.

(with ResNet50), a deep learning-based detector, is used in this study, which returns the top-left coordinates of the detected objects by default. However, the algorithm requires

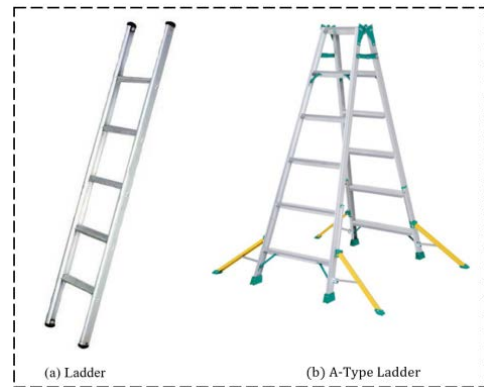


FIGURE 3. Examples of (a) general ladder and (b) A-type ladder.

the bottom-left coordinates of both the worker and A-type ladders to compute the working height. The following equations provide the top-left coordinate of the worker and the ladder:

$$P_{l,t} = (X_t^p, Y_t^p, W_1^p, H_1^p) \quad (1)$$

$$L_{l,t} = (X_t^l, Y_t^l, W_2^l, H_2^l) \quad (2)$$

The bottom-left coordinates of the person and the ladder are extracted using the following equations:

$$P_{l,b} = (H_1^p + Y_t^p) \quad (3)$$

$$L_{l,b} = (H_2^l + Y_t^l) \quad (4)$$

where  $H_1^p$  and  $H_2^l$  are the heights of the bounding boxes corresponding to the worker and the ladder, respectively. The



$Y_t^p$  and  $Y_t^l$  are the top-left y-coordinates of the bounding boxes corresponding to the worker and ladder, respectively.

A fixed height of the A-type ladder used as the reference point. The height of the A-type ladder and its corresponding pixel value is given as follows:

$$\text{Ladder\_Height} = 5.5 \text{ ft or } 1.7 \text{ m} \quad (5)$$

$$\text{Ladder\_Pixel\_value} = 100\% \quad (6)$$

The height of the ladder bounding box  $H_2^l$  can be written as

$$H_2^l = l_{bbh1} + l_{bbh2} \quad (7)$$

where  $l_{bbh1}$  and  $l_{bbh2}$  are the first and second halves of the ladder bounding box pixel values, respectively; and

$$l_{bbh1} = L_{l,b} - P_{l,b} \quad (8)$$

Using Equation (8), the value of  $l_{bbh2}$  can be determined as

$$l_{bbh2} = H_2^l - l_{bbh1} \quad (9)$$

The working height in pixels (i.e., the height at which the worker works on the ladder) is given as follows:

$$\text{Working\_height\_pixels} = \frac{l_{bbh1}}{H_2^l} \times \text{Ladder\_Pixel\_value} \quad (10)$$

The actual working height in feet can be obtained using Equations (5), (6), and (10) as follows:

$$\text{Working\_height\_feet} = \frac{\text{Working\_height\_pixels}}{\text{Ladder\_Pixel\_value}} \times \text{Ladder\_Height} \quad (11)$$

The working height in meters is given as follows

$$\text{Working\_Height} = \text{Working\_height\_feet} \times 0.3048 \quad (12)$$

After obtaining the working height, the algorithm tests it against the corresponding KOSHA regulation to predict worker behavior.

### C. ANALYSIS OF KOSHA RULES CORRELATED WITH A-TYPE LADDER

The KOSHA regulations consist of 13 chapters with multiple sections, including 657 standards, of which 277 are associated with the construction industry [38]. The Labor Standards Act of 1953 created a framework for Korea's industrial safety and health standards, which possesses an ISO-45001 accreditation, to establish safety and health management systems (KOSHA 18001) at work [35]. Inspired by the rapid rise of industries between 1970 and 1980, KOSHA was founded in 1987. KOSHA amended the corresponding rules and regulations to meet the mandatory safety and health requirements in numerous industries having toxic and complex working environments. Since then, KOSHA has compiled and examined several cases, resulting in creating an expert knowledge database. Significant changes have been made to improve policies in compliance with modern industry practices, facing

global competition. However, despite the strict implementation of occupational safety regulations, certain practices still follow traditional approaches, such as working with A-type ladders. The ban on the use of mobile/portable ladders imposed by the Ministry of Employment and Labor in Korea and KOSHA was lifted as of March 2019. Mobile/portable ladders can now be used following defined safety rules. Furthermore, it is recommended that a fall prevention device must be installed when an A-type/mobile ladder is used. Several safety measures have been proposed for the use of ladders. Typically, all ladders must be used on flat, solid, and non-slip floors [39]. The rules employed in the developed SSD-based HCM are listed below. We have designed four case scenarios based on these.

- If the working height is less than or equal to 1.2 m, the worker must wear a helmet.
- If the working height is greater than 1.2 m but less than 2 m, the worker must wear a helmet and work in a group of two workers. The use of the topmost rung is prohibited.

The CV algorithm is considered a better approach for implementing expert knowledge (safety regulations) at construction sites. Therefore, we manually analyzed and extracted KOSHA rules correlated with A-type ladders. In this study, we developed a vision-intelligence-based HCM to prevent FFL. Four case scenarios, listed in Table 2, are considered to classify a worker's behavior as safe or unsafe.

*Case 1:* If a worker (W) work at a height less than or equal to 1.2 m on a ladder without outriggers ( $L_w$ ) and is wearing a helmet ( $h$ ), then this should be classified as safe behavior ( $B_s$ ) and represented as

$$\text{if } W \cap L_w \ \& \ h \in B_s \quad (13)$$

*Case 2:* If a worker (W) works at a height less than or equal to 1.2 m on a ladder without outriggers ( $L_w$ ) and is not wearing a helmet ( $h$ ), then this should be classified as unsafe behavior ( $B_u$ ) and represented as

$$\text{if } W \cap L_w \ \& \ \neg h \in B_u \quad (14)$$

*Case 3:* If a worker (W1) works at a height greater than 1.2 m but less than 2 m on a ladder without outriggers ( $L_w$ ) while wearing a helmet ( $h$ ), and if two workers are working in a group (W1, W2), then this should be classified as safe behavior and represented as

$$\text{if } W_1 \cap L_w \ \& \ h \ \& \ W_2 \in B_s \quad (15)$$

*Case 4:* If a worker (W1) works at a height greater than 1.2 m but less than or equal to 2 m on a ladder without outriggers ( $L_w$ ) and works without the support of a co-worker while not wearing a helmet ( $h$ ), then this should be classified as unsafe behavior and represented as

$$\text{if } W_1 \cap L_w \ \& \ \neg h | \neg W_2 \in B_u \quad (16)$$

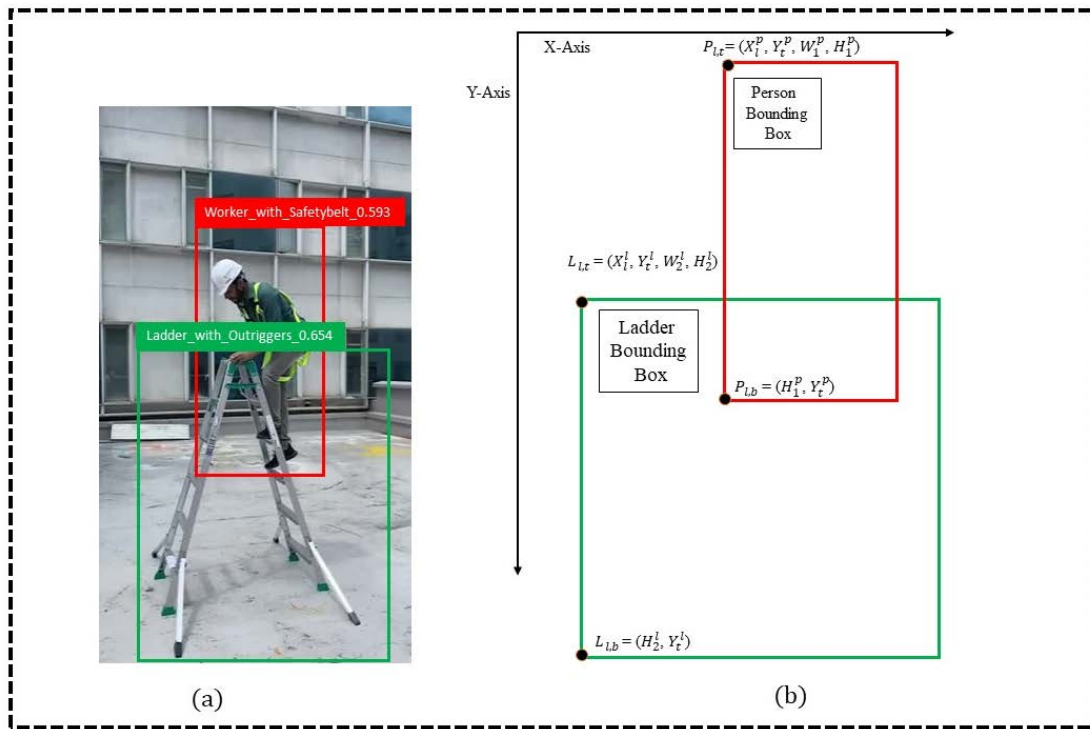


FIGURE 4. CV-based approach (a) Examples of the detected objects. (b) Bounding box coordinates of a person and A-type ladder.

TABLE 2. Selected case scenarios for predicting the behavior of workers on A-type ladders.

Case No.	Working Height	KOSHA Rule	Behavior
1	$H \leq 1.2$ m	Worker with helmet	Safe
2	$H \leq 1.2$ m	Worker without helmet	Unsafe
3	$1.2$ m $< H \leq 2.0$ m	1. Worker with helmet 2. Two workers works in a group	Safe
4	$1.2$ m $< H \leq 2.0$ m	1. Worker without a helmet 2. Worker works on an A-type ladder without the support of co-worker.	Unsafe

#### D. DEVELOPMENT OF THE ALGORITHM

This section summarizes the necessary steps in developing the algorithm, such as dataset preparation, deep-learning-based model selection, HCM, and model training.

##### 1) DATASET PREPARATION

Many digital images covering a wide range of patterns are required to train vision-based object detection models. However, obtaining labeled datasets from open-source websites is difficult because vision technology is relatively new in the construction industry. The dataset for this study consists of 21 videos collected from the Construction Technology Innovation Laboratory (ConTil) of Chung-Ang University, Seoul, South Korea. To prepare an image recognition dataset from these videos, the authors used the command line tool Fast-forward MPEG (ffmpeg) to extract frames from each video. After extracting the frames, performed data cleaning

to obtain an appropriate and useful dataset for training a deep learning-based SSD model. Each frame has been manually reviewed during the data cleaning process to determine whether it is suitable for training the model. Improper or unsuitable frames, such as those with incorrect exposure or repeated images, have been removed. A total of 1,825 images were obtained after the cleaning process. These images were imported into an image labeling application in MATLAB, which is used to label the ground truth in the images. The input dataset is divided into the following five class labels.

1. Ladder with outriggers
2. Ladder without outriggers
3. Worker with helmet
4. Worker without helmet
5. Worker with safety belt

The labeled dataset was randomly shuffled and split into training and evaluation datasets in an 80:20 ratio, such that



FIGURE 5. Dataset labeling for the SSD-based object detector.

the training and testing set contained 1,460 and 365 images, respectively. The resolution of the dataset is  $404 \times 720$  pixels. Fig. 5 shows the labeled dataset.

## 2) DEEP LEARNING MODEL SELECTION

This study used a deep learning algorithm as the backbone model for object detection. There are two types of object detectors: one- and two-stage detectors. The primary difference between one- and two-stage object detectors is that one-stage detectors use only a single CNN to predict classes and offsets from anchor boxes without requiring proposal generators. In contrast, two-stage object detectors achieve prediction in two stages: the first generates region proposals with a high score, and the second provides the final prediction. One-stage object detectors are preferred for real-time object detection due to a fixed inference time, unlike two-stage detectors with a variable inference time [40]. Existing state-of-the-art deep learning object detectors, such as Faster R-CNN [41], YOLO[42], and SSD [43], can be used to detect objects effectively in the presence of illumination as well as occlusions. This study used an SSD with ResNet50 to detect workers' and A-type ladders in construction sites, as the SSD outperformed the state-of-the-art object detector Faster R-CNN.

Moreover, SSD exhibits much better accuracy than other single-stage models, as claimed by Wei *et al.* [41]. An input image passes through a single CNN operation, following which the relevant features are extracted from this image and the target objects are detected, as shown in Fig. 6. The HCM post-process the CNN predictions to compute the working height of the worker on the A-type ladder. The algorithm then determines whether the workers' behavior is safe or unsafe. This study classifies worker behavior based on the scenarios discussed in Section II-C.

## 3) HEIGHT COMPUTING MODULE (HCM)

Although this study utilized a ResNet-50 CNN in the SSD for object detection, the working height on the ladder can

be determined by post-processing the ResNet-50 classified output. Therefore, an additional module (HCM) is introduced for post-processing to measure the working height. SSD is utilized as a base model for HCM; the algorithm checks the correlated safety KOSHA rule and determines whether the worker behavior is safe or unsafe. The computational steps involved are as follows. First, the SSD detects the target objects in the construction site images, as stated in the dataset preparation Section II-D (1). Next, the bounding box values and labels of the detected objects are input to the HCM, which checks whether the worker is working on a ladder. Then, the HCM computes the working height (based on the equations derived in Section II-B). Finally, the result of the SSD-based HCM is utilized to cross-check the corresponding safety rule in order to classify worker behavior.

In HCM, class labels are assigned after computing the working height based on the bounding box names of the person and ladder. To determine whether a person stands on an A-type ladder, the HCM check the intersection between the bounding boxes of the detected objects, which requires two position vectors (defined as "Person" and "Ladder") as input; the detector then returns the area (i.e., a scalar value) of the intersection. For better accuracy, the HCM compares the pixel values of the bottom-left coordinates of the person and the ladder bounding boxes. When the intersection area between the two bounding boxes is greater than "1" and the pixel value of the bottom-left coordinate of the person bounding box is less than that of the bottom-left coordinate of the ladder bounding box, then it is evaluated that the worker is working on the ladder. The bounding box coordinates of a person standing at different heights on the ladder are shown in Fig. 7. Note that the proposed HCM can simultaneously identify multiple workers and ladders; moreover, it assigns a unique I.D. to the worker when the person bounding box intersects the ladder bounding box; otherwise, it skips further processing.

For the cases represented in Fig. 7(a) and (b), if the person bounding box is inside the ladder bounding box, the HCM



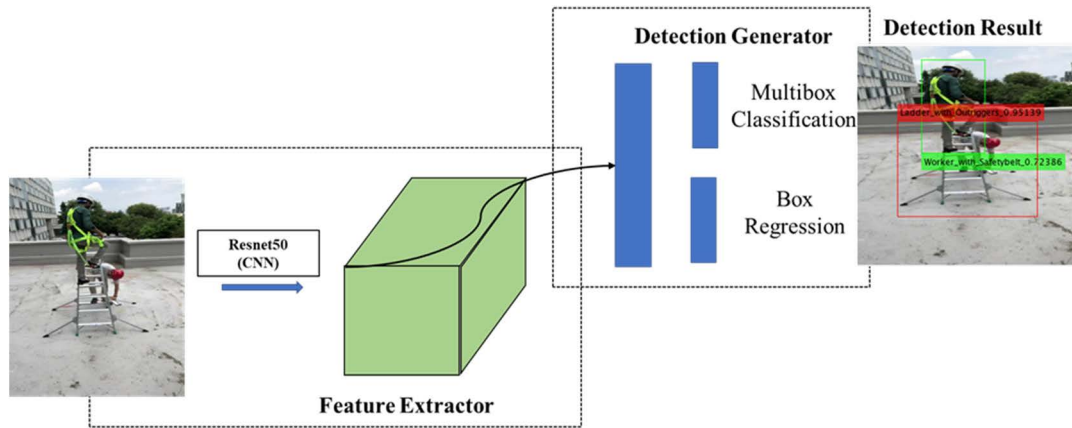


FIGURE 6. Architecture of SSD (modified from [43]).

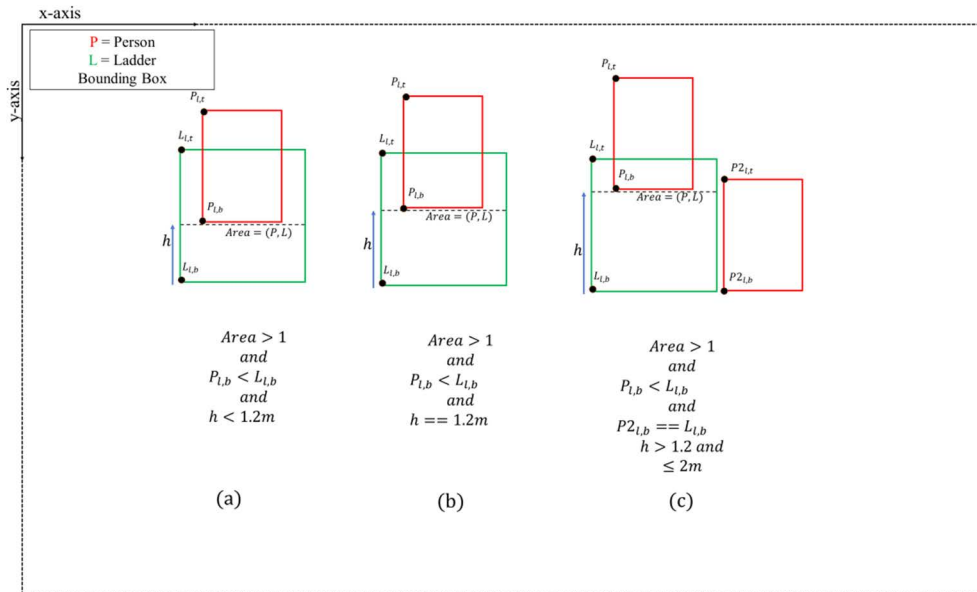


FIGURE 7. Representation of bounding box coordinates at different working heights. (a) Person on an A-type ladder of height less than 1.2 m. (b) Person on an A-type ladder of height equal to 1.2 m. (c) Person on an A-type ladder of height greater than 1.2 m and less than 2 m.

determines the intersection between the two bounding boxes, and the detector returns the corresponding area. If the area is greater than “1”, then the HCM compared the bottom-left y-coordinate of the person bounding box  $P_{l,b}$  with the bottom-left y-coordinate of the ladder bounding box  $L_{l,b}$ . If  $P_{l,b} < L_{l,b}$ , it implies that the person is standing on the ladder. If the computed working height (H) is less than or equal to 1.2 m, the HCM checks the class label of the person. When the class label is “worker with helmet,” then the behavior is classified as safe, otherwise, it is classified as unsafe.

The HCM follows a similar procedure for the case shown in Fig. 7(c). First, it determines whether the person  $P_1$  is on a ladder. Next, it checks if person  $P_2$  is present (this is performed by comparing the bottom-left y-coordinate of the bounding box that represents a second person  $P_{2l,b}$  with that

of the ladder bounding box  $L_{l,b}$ . If  $P_{2l,b} = L_{l,b}$ , this implies that the second person works in a group to hold an A-type ladder). If  $1.2\text{ m} < H < 2\text{ m}$ , the HCM checks the class label of the person. If the class label is “worker with helmet” and two workers are present, then the behavior is classified as safe, otherwise it is classified as unsafe.

Algorithm I presents the pseudocode of the proposed method, which identifies and classifies objects, intended to recognize unsafe behavior when using an A-type ladder. The pseudocode clearly outlined the important steps of the algorithm. It takes videos as input from the IP camera and predicts the output as safe or unsafe. Line 1 to 4 extract frames from an input video and pass them to the trained model. Lines 5 and 6 extract the coordinates of the person and the ladder bounding boxes. The intersection between the person and ladder

**TABLE 3.** Model training parameters.

Parameters	Values
Momentum	0.9000
Network Size	$300 \times 300, 3$
Number of classes	5
Mini-batch Size	16
Learning rate drop period	0
Maximum number of epochs	300

bounding boxes is obtained in lines 7 and 8. The working height of the ladder is computed from lines 9 (i) to 9 (ix). Finally, lines 11 and 12 performed a rule-based comparison between the computed height and converted KOSHA safety rules (Section II-C) to determine unsafe behavior.

#### 4) MODEL TRAINING

Training an SSD requires the following input arguments: pre-processed data, layer graphs, and training options. Pre-processed data (i.e., modified input data) corresponds to the prerequisites of the selected model. In this study, the size of the input image and bounding boxes are modified. In addition, SSD layers are utilized as layer graphs, which need input parameters of image size, number of classes, and network architecture. The input image size is set to  $300 \times 300 \times 3$ , and the number of classes is set to 5. ResNet50 is used as the base network (a pre-trained CNN). The default parameters of the training options, such as the momentum, initial learning rate, mini-batch size, learning rate schedule, learning rate drop factor, and maximum number of epochs, were modified. The input network size set to  $300 \times 300, 3$ . The initial learning rate, mini-batch size, and stochastic gradient descent with momentum were set to 0.001, 16, and 0.9000 respectively. The execution environment was set to a GPU for fast training.

In addition, a piecewise learning rate schedule was used; the learning rate drop period and maximum number of epochs were set to 30 and 300, respectively. The training was performed on a Windows 10 Pro Intel®Core i9 10th generation, 3.30 GHz processor with 256 GB RAM. Furthermore, we trained, tested, and evaluated the proposed algorithm using MATLAB R2020b. The model training parameters are listed in Table 3.

### III. RESULTS

The Android mobile application “IP Webcam” was used as an IP camera to send real-time video data from a smartphone camera to MATLAB using the Hypertext Transfer Protocol (HTTP) as a wireless communication protocol. The developed model was deployed on a local system (Core i9 10th generation) and received a real-time video from an IP camera to feed the developed algorithm.

All objects have been identified using an SSD-based deep learning model. Fig. 8(a) shows the accurate detection of a worker wearing a safety belt and helmet working on a ladder, a worker wearing a helmet holding a ladder, and a ladder with outriggers. In contrast, in Fig. 8(b), the identified object is a ladder without outriggers. The model accurately detected a worker without any safety equipment working on an A-type

ladder, as shown in Fig. 9(a). In Fig. 9(b), a worker wearing a safety belt and helmet correctly identified as working on an A-type ladder with outriggers. These results demonstrate that our trained model successfully identified various objects irrespective of the viewing angle. The detected objects were post-processed by the HCM to determine the height. The HCM determines the working height on the ladder and cross-checks the converted corresponding KOSHA safety rule (Section II-C), and categorizes the worker behavior as safe (if no rule violation) and unsafe (in case of violation of safety rules) on the local system. The safety manager is notified about workers’ visual safety status and behavior on an A-type ladder. We detail the results obtained from the SSD-based HCM for the four scenarios in real-time.

#### A. CASE 1: WORKER ON A-TYPE LADDER FOR SAFE BEHAVIOR ( $H \leq 1.2M$ )

This experimental scenario demonstrates the safe behavior of a worker on an A-type ladder working at a height less than or equal to 1.2 m. Fig. 10(a) shows a worker wearing a helmet working on a ladder with outriggers at the height of 1.01 m (safe behavior). Fig. 10(b) depicts a worker wearing a helmet working on a ladder with outriggers at the height of 1.2 m (safe behavior). This scenario is deemed safe because it fulfils the worker safety requirement as per the KOSHA rule in equation 13 (Section II-C).

#### B. CASE 2: WORKER ON A-TYPE LADDER FOR UNSAFE BEHAVIOR ( $H \leq 1.2M$ )

This experimental scenario demonstrates the unsafe behavior of a worker on an A-type ladder working at a height less than or equal to 1.2 m. Fig. 11(a) depicts a worker without a helmet working on a ladder without outriggers at the height of 0.52 m (unsafe behavior). Fig. 11(b) shows a similar example as in Fig. 11(a), except that the working height is 1.2 m (unsafe behavior). This scenario is classified as unsafe because it contravenes the safety rule shown in equation 14 (Section II-C).

#### C. CASE 3: WORKER ON A-TYPE LADDER FOR SAFE BEHAVIOR ( $1.2 m < H < 2 m$ )

This experimental scenario demonstrates the safe behavior of a worker on an A-type ladder working at a height greater than 1.2 m and less than 2 m. Fig. 12(a) depicts the two workers performing work together, with worker 1 wearing a safety belt and helmet and standing on a ladder, and the second is holding the ladder. Fig. 12(b) shows a similar example as in Fig. 12(a) except the working height is 1.31 m (safe behavior). This scenario is classified as safe because it fulfils the worker safety requirement according to the KOSHA rule in equation 15 (Section II-C).

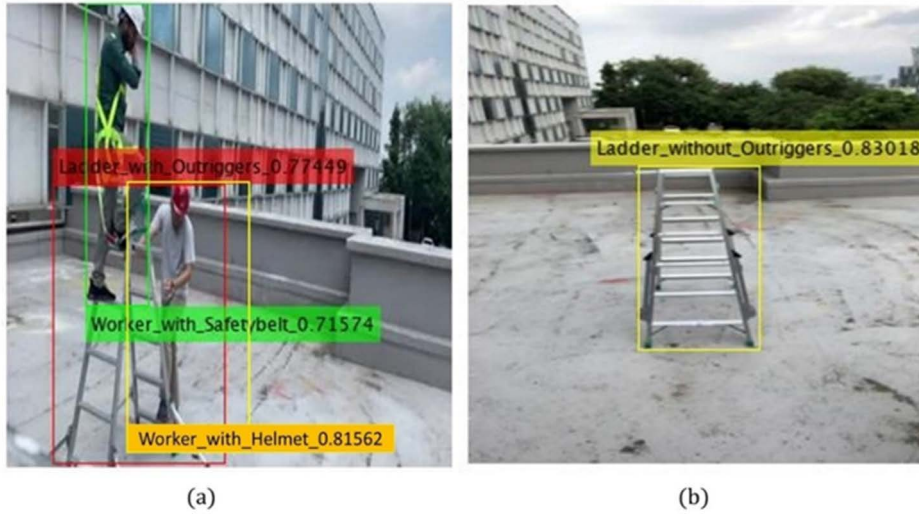
#### D. CASE 4: WORKER ON A-TYPE LADDER FOR UNSAFE BEHAVIOR ( $1.2 m < H < 2 m$ )

This experimental scenario demonstrates the unsafe behavior of a worker on an A-type ladder working at a height greater

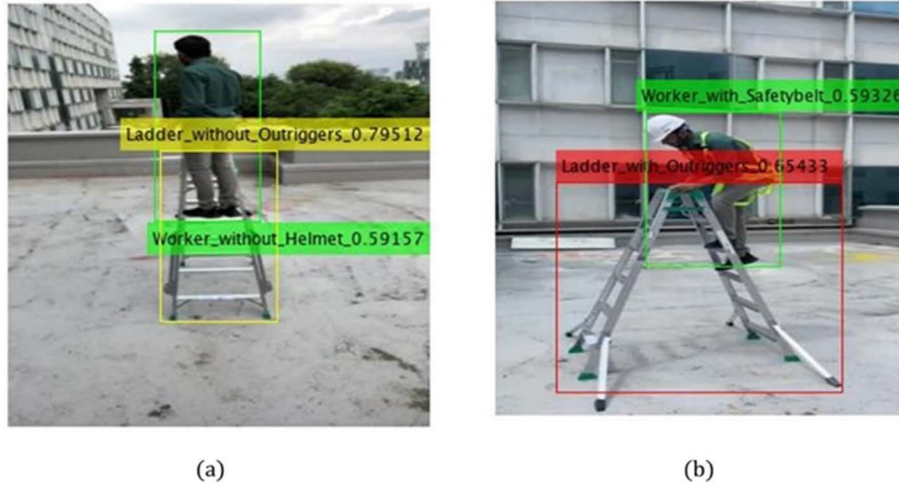
**Algorithm 1** Pseudocode for The Proposed Method**Input:** Video frames**Output:** Prediction of safe or unsafe working behavior

1. Load trained model.
2. Extract frames from real-time video.
3. Pass each frame to the trained model.
4. The model performs object detection and returns the corresponding bounding box values, labels, and scores.
5. Extract the top-left bounding box coordinates as x, y, w, h.
  - i.  $P_{l,t} = (X_t^p, Y_t^p, W_1^p, H_1^p)$  // top-left person coordinates
  - ii.  $L_{l,t} = (X_t^l, Y_t^l, W_2^l, H_2^l)$  // top-left ladder coordinates
6. Extract the bottom-left y-coordinates of the detected objects.
  - i.  $P_{l,b} = (H_1^p, Y_t^p)$  // bottom-left person coordinates
  - ii.  $L_{l,b} = (H_2^l, Y_t^l)$  // bottom-left ladder coordinates
7. Store the class labels “ladder with outriggers” and “ladder without outriggers” in a 2D list as [row, colum]. // to find the intersection
8. For i=1 to size of detected labels:
  - i. Store the bounding box value of row → in the person variable
  - ii. Store the bounding box value of i → in the ladder variable
  - iii. area = rectint(person, ladder) // gives the intersection between person and ladder
9. If (area > 1 and  $P_{l,b} < L_{l,b}$ ):
  - i. Ladder\_Height = 5.5ftor1.7m
  - ii. Ladder\_Pixel\_value = 100%
  - iii.  $H_2^l = lbbh1 + lbbh2$  //divide height of ladder bounding box into two parts
  - iv.  $lbbh1 = L_{l,b} - P_{l,b}$  //value of ladder bbox half1
  - v.  $lbbh2 = H_2^l - lbbh1$  //pixel value of ladder bbox half2
  - vi.  $Working\_height\_pixels = \frac{lbbh1}{H_2^l} \times Ladder\_Pixel\_value$
  - vii.  $Working\_height\_feet = \frac{Working\_height\_pixels}{Ladder\_Pixel\_value} \times Ladder\_Height$
  - viii.  $Working\_Height = Working\_height\_feet \times 0.3048$
  - ix.  $Working\_Height = abs(Working\_Height)$  // absolute value
10. If ( $Working\_Height \leq 1.2$ ) // in meters
  - i. If string compare labels(i) == “worker with helmet”
    - “Safe Behavior”
  - ii. Else:
    - “Unsafe Behavior”
11. Else If ( $Working\_Height > 1.2$  and  $\leq 2.0$ ):
  - i. Count=0 // to check how many workers are there
  - ii. Flag = false // to check the rules
  - iii. If string compare labels(i) == “worker with helmet”:
    - $Flag = true$
    - $Count = Count + 1$ //
  - iv. Else If string compare labels(i) == “worker without helmet”:
    - $Count = Count + 1$ //
  - v. Else If string compare labels(i) == “worker with safety belt”:
    - $Count = Count + 1$ //
  - vi. Else: // do nothing
  - vii. If Count ==2 and flag: // means all KOSHA rules have been checked
    - “Safe Behavior”
  - viii. Else:
    - “Unsafe Behavior”
12. Else:

This process continues until the person is found to be on the ladder.



**FIGURE 8.** Object detection in real-time. (a) Two workers and a ladder with outriggers. (b) Ladder without outriggers.



**FIGURE 9.** Object detection in real-time. (a) Worker without a helmet on an A-type ladder without outriggers. (b) Worker with helmet and safety belt on an A-type ladder with outriggers.

than 1.2 m and less than 2 m. Fig. 13 (a) depicts a worker with a helmet but working as an individual at the height of 1.68 m (unsafe behavior as per the corresponding safety rule). Fig. 13(b) shows a worker on a ladder without outriggers working at the height of 1.7 m. Although the worker is wearing a helmet and a safety belt, the behavior is classified as unsafe as the safety rule states that two workers should be working in a group (in equation 16, Section II-C).

#### IV. EVALUATION METRICS

We evaluated the efficiency of the trained model using the following metrics: precision, recall, F1-score, true positive rate (TPR), false positive rate (FPR), and average precision (Equations (17)–(22)). Precision quantifies the number of predicted true positives, whereas recall or TPR signifies the correct identification of true positives. The FPR indicates

when the model classifies the positive class inaccurately. The average precision is an important evaluation metric that demonstrates the overall usefulness of the algorithm through a single numerical value.

$$Precision(P) = \frac{TP}{(TP + FP)} \tag{17}$$

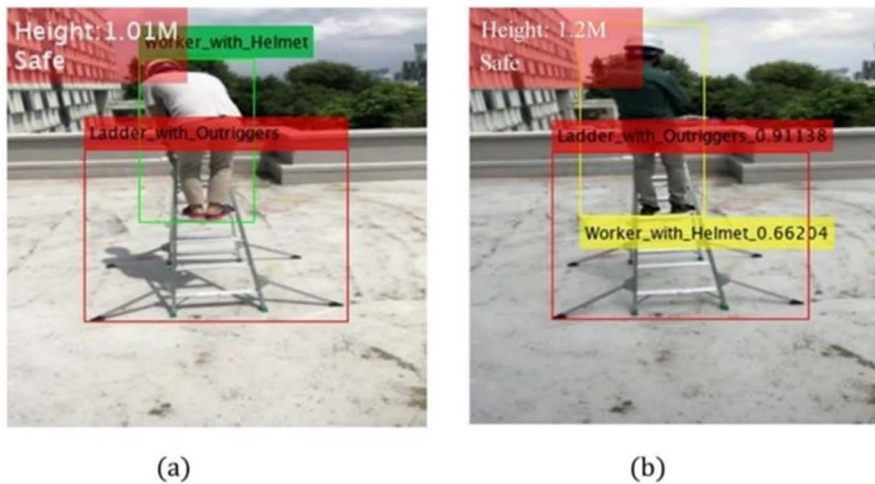
$$Recall(R) = \frac{TP}{(TP + FN)} \tag{18}$$

$$F1 - Score = 2 \frac{(P * R)}{(P + R)} \tag{19}$$

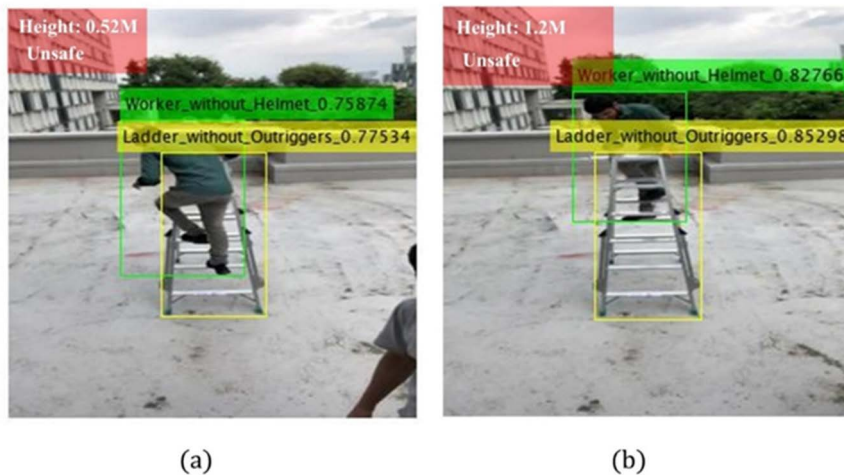
$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{20}$$

$$TruePositiveRate (TPR) = \frac{TP}{TP + FN} \tag{21}$$





**FIGURE 10.** Examples of predicted safe behavior at a working height  $H \leq 1.2$  m on an A-type ladder with outriggers. (a) Worker with a helmet at  $H = 1.01$  m. (b) Worker with safety belt and helmet at  $H = 1.2$  m.



**FIGURE 11.** Examples of predicted unsafe behavior at a working height  $H \leq 1.2$  m on an A-type ladder without outriggers. (a) Worker without a helmet at  $H = 0.52$  m. (b) Worker without a helmet at  $H = 1.2$  m.

$$\text{FalsePositiveRate (FPR)} = \frac{FP}{FP + TN} \quad (22)$$

where TP, FN, FP, and TN represent the true positive, false negative, false positive, and true negative, respectively.

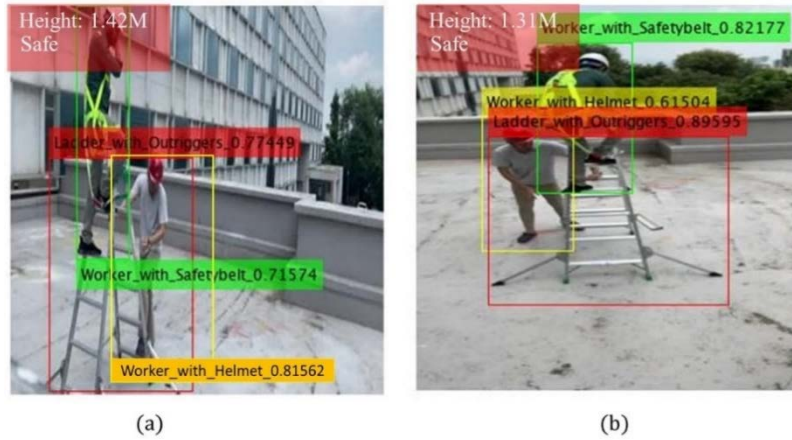
#### A. EVALUATION OF SSD

The trained model has been evaluated using an average precision indicator on a test dataset comprising 365 images. The five classes considered in this study are ladder without outriggers, ladder with outriggers, worker with helmet, worker with safety belt, and worker without a helmet. Figures 14(a)–14(e) illustrate the precision-recall curves. The recall was plotted on the X-axis and the precision on the Y-axis, which was evaluated at a threshold of 0.3. Fig. 14 (a) depicts the average precision of the class ladder without outriggers as 98% (confirming the ability to detect an object).

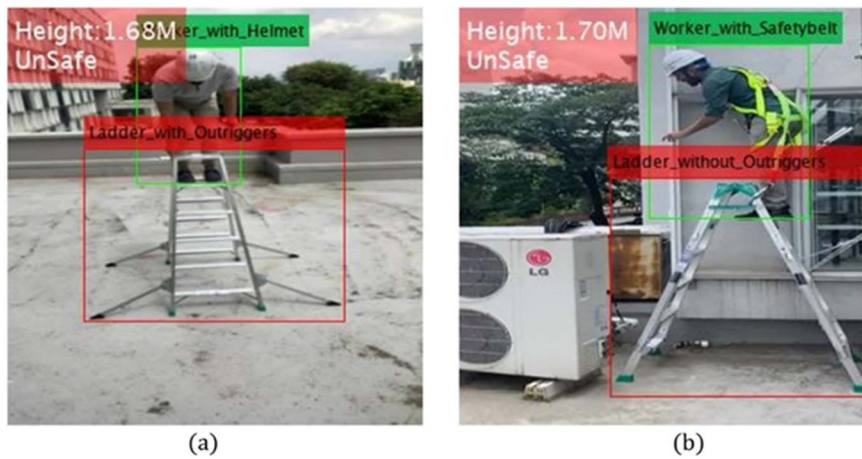
Similarly, Figs. 14 (b) and (c) show the average precision of the class ladder with outriggers as 99% and worker with safety belt as 90% (confirming the ability to recognize objects). The average precision of class workers without helmets and workers with helmets was 84% and 70%, respectively. These values appear to be low; however, the lower average precision of these classes compared with the other classes is due to the imbalanced distribution in the dataset.

#### B. EVALUATION OF HCM

We evaluated the proposed algorithm using four performance indicators on a set of 300 images. These images are divided into class-1 (160 images) for safe behaviors and class-2 for unsafe behaviors (140-images). Both classes are assigned binary numbers, i.e., 0 to safe and 1 to unsafe, to compare the ground truth and prediction. Fig. 15 shows an  $(n \times n)$



**FIGURE 12.** Examples of predicted safe behavior at a working height of  $1.2\text{ m} < H < 2\text{ m}$  on an A-type ladder with outriggers. (a) Two workers with helmets worker with a safety belt on ladder) work in a group at  $H = 1.42\text{ m}$ . (b) Two workers with helmets (worker with a safety belt on ladder) work in a group at  $H = 1.31\text{ m}$ .



**FIGURE 13.** Examples of predicted unsafe behavior at a working height  $1.2\text{ m} < H < 2\text{ m}$  on an A-type ladder. (a) Single worker with helmet working on a ladder with outriggers at  $H = 1.68\text{ m}$ . (b) Single worker with safety belt and helmet working on a ladder without outriggers at  $H = 1.70\text{ m}$ .

confusion matrix, where  $n$  is defined as the number of classes. In this study,  $n = 2$  (safe and unsafe behavior). The columns represent the ground truth, and the rows represent the target predictions. The SSD-based HCM correctly identified (TP) safe behavior as 137, while the unsafe behavior in actual but classified as safe (FP) was 21. Similarly, the algorithm correctly predicted the scene as an unsafe behavior (TN) of 119; however, the scene predicted as unsafe with a safe behavior in actual (FN) is 23. Table 4 summarises the performance indicators of the proposed algorithm. This algorithm achieved precision, recall, F1-score, and overall accuracy of 86.7%, 85.6%, 86.4%, and 85.33%, respectively.

Additionally, to validate the effectiveness of HCM in classifying behavior, the receiver operating characteristic (ROC) and area under the curve are shown in Fig. 16. The 300 images were divided into a set of 5-(k-fold) and performed the prediction on each k-fold to determine the TPR and FPR. The ROC curve was plotted for each fold, with values ranging from 0 to 1, with the calculated TPR and FPR. The green ROC curve shows the average of all the fold AUC values as

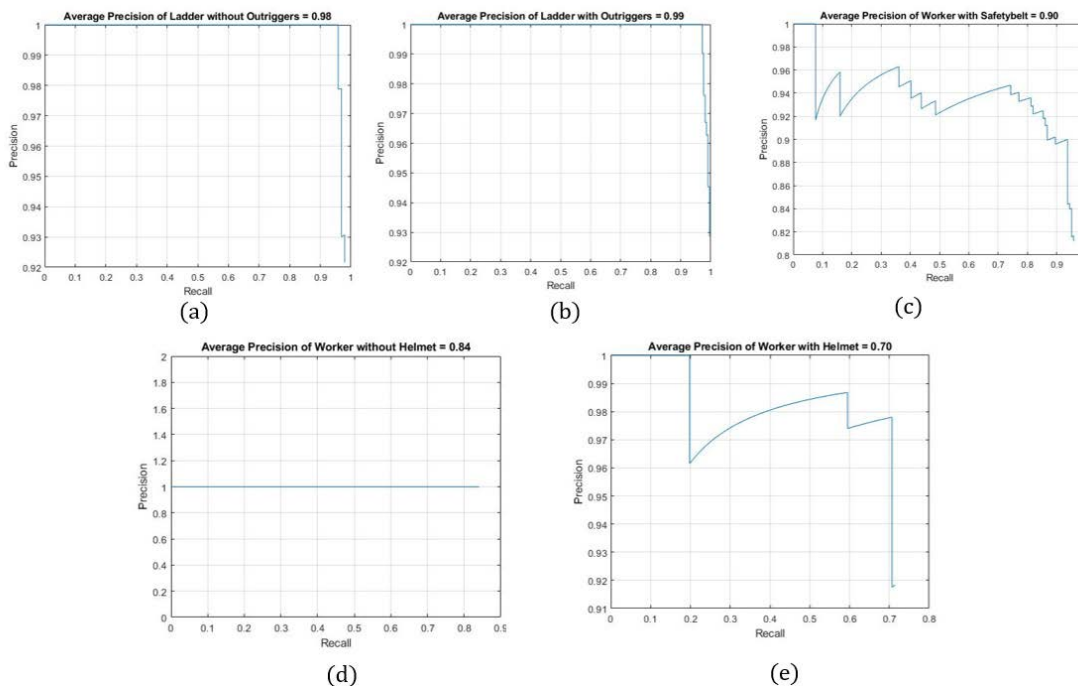
**TABLE 4.** Performance results of SSD-based HCM.

Performance Matrix	Percentage
Precision	86.7%
Recall	85.6%
F1-score	86.4%
Accuracy	85.33%

0.84, demonstrating that HCM can effectively identify unsafe behavior.

## V. DISCUSSION AND FUTURE WORK

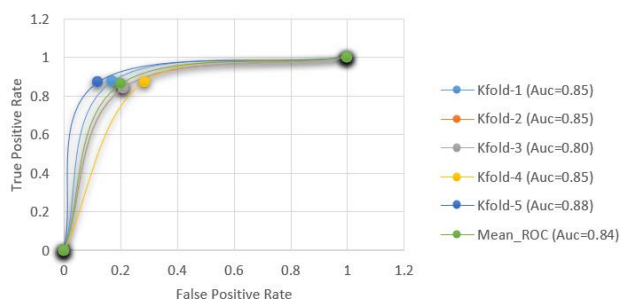
The proposed method can compute the working height using vision sensors (cameras) and proactively identify unsafe behavior in the case of negligence in compliance with rules. This method can also be used as a safety intervention as it is developed for safety monitoring and as a source to highlight unsafe behavior at different working heights on the ladder. When workers are aware that they are being monitored



**FIGURE 14.** Precision-Recall curves for the classes: (a) ladder without outriggers, (b) ladder with outriggers, (c) worker with helmet, (d) worker with safety belt, and (e) worker without helmet.

Ground Truth \ Prediction	Safe Behavior (0)	Unsafe Behavior (1)
Safe Behavior (0)	TP(137)	FP(21)
Unsafe Behavior (1)	FN (23)	TN(119)

**FIGURE 15.** Confusion Matrix for SSD-based HCM using binary classification.



**FIGURE 16.** ROC and AUC for SSD-based HCM using binary classification.

continuously, they are more likely to follow the safety rules. Safety managers’ are responsible for protecting workers from hazardous situations. Therefore, they enforce safe practices by physically visiting the construction site. However, it is not possible for a safety manager to be ubiquitous; therefore,

an automated height assessment method should be developed to assist workers and construction companies.

Researchers have previously utilized motion-capturing-based activity recognition models to classify the behavior of workers on ladders. Similar research to ours [33] developed an integrated system using the position of a worker (collected using an ultra-wide-band system) and posture (using IMU and 3D skeleton) to classify workers’ behavior based on the activities being performed disregarding the safety rules check. In addition, they performed a comparative analysis for safety risk evaluation separately for posture, position, and fusion to determine the best approach. The fusion-based approach achieved an accuracy of 83%. Nevertheless, their approach achieved good accuracy, but as the worker movement is too complex and unpredictable, the system might show low accuracy for the unseen behavior of the worker. In comparison with their approach, our developed algorithm achieved an average accuracy of 85.33% and an F1 score of 86.4% (Section IV-B) in classifying worker behavior while working on the ladder. The overall accuracy demonstrated that the developed algorithm could intelligently compute the working height using pixel values, check the safety rules (outlined by the KOSHA) at specific working heights, and identify workers’ behavior.

During the experiment (Section III), the developed algorithm was tested in four experimental scenarios to verify its effectiveness. The first and second (A, B) scenarios demonstrated the safe and unsafe behavior of a worker on an A-type ladder working at a height less than or equal to 1.2 m. The third and fourth (C, D) scenarios demonstrated the safe and



unsafe behavior of a worker on an A-type ladder working at a height greater than 1.2 m but less than 2 m. The proposed method can be deployed at a construction site to recognize unsafe behavior (in real-time) as safety management and intervention system. This research not only provides an easy and automatic way to recognize unsafe behavior using CV and safety regulations but also provides insights for determining height using imaging data. The HCM can be easily adopted in other engineering domains.

Despite the effectiveness of this algorithm, it has several limitations. As we used 2D images obtained from 2D CCTV cameras, workers standing behind an A-type ladder were misidentified as working on the ladder. This is because 2D cameras cannot identify the actual position of an object. This limitation can be overcome by using stereo vision cameras. These cameras can collect depth and distance information of workers and A-type ladders at construction sites; moreover, such cameras enable accurate computation of the distance between a worker and an A-type ladder. Now the proposed algorithm can only predict the behavior of workers working at a height in the 1.2–2.0 m range on A-type ladders. However, in future work, we plan to extend this algorithm to predict the behavior of workers working at the height of up to 3.5 m to cover all KOSHA regulations associated with the A-type ladder. We plan to develop an early risk assessment framework with the safety risks index by considering risks and severity to classify risks as low, medium, and high for more advanced practical usability in managing risks while working at a height on the ladder [32]. We plan to create a larger dataset by collecting images from various construction sites to detect relevant objects for a more practical application of CV-based safety monitoring. Moreover, this method requires a reference point (ladder height) to estimate the working height. We are trying to overcome this limitation by utilizing the objects' distance from the camera using depth information to estimate the reference point automatically.

## VI. CONCLUSION

This research focuses on a deep learning-based height estimation method for worker safety monitoring in real-time to predict worker safety status at A-type ladders. This paper presents an automated solution to facilitate safety management and overcome manual safety monitoring to reduce FFLs in construction sites. The main aspects of this study are as follows:

- 1) KOSHA Rules Analysis: We manually extracted the safety rules correlated with the A-type ladder from the KOSHA expert database, converted them into a computer program, and integrated them with the CV technology to provide workers with an effective and automated safety method while working on A-type ladders.
- 2) Dataset: A dataset for safety behavior detection has been prepared using 21 videos of working on an A-type ladder. The frames were extracted (1825 images) and

labeled for deep learning-based object detection (SSD with ResNet-50).

- 3) HCM: Workers may put their lives in danger by breaching safety rules while working at different heights on ladders. We tried to create a safer working environment by integrating the safety rules correlated with the A-type ladder from the KOSHA expert database with a deep learning-based object detection system. The coordinates of the detected objects have been used to estimate the working height.
- 4) Experiments: The proposed topology validated workers' safe and unsafe behaviors, following the KOSHA rules, using four different cases. The developed algorithm utilizes object detection as a base model for the HCM to estimate the working height and compare it with the corresponding occupational safety rules. The case scenarios demonstrated that the proposed height assessment method performed better in all cases to classify worker behavior as safe or unsafe.
- 5) Evaluation of Model: The detection model was tested on a dataset (365 images) categorized into five classes: ladder without outriggers, ladder with outriggers, worker with helmet, worker with safety belt, and worker without helmet, which achieved average precision values of 98%, 99%, 70%, 90%, and 84%, respectively.
- 6) Evaluation of HCM: We evaluated the performance of the HCM using a set of images (300 for all case scenarios) for binary classification (safe and unsafe). The proposed algorithm achieved precision, recall, F1-scores, and overall accuracies of 86.7%, 85.6%, 86.4%, and 85.33%, respectively.

The findings of this study show that the proposed approach can accurately classify worker behavior as safe or unsafe at a specified working height on an A-type ladder. The proposed SSD-based HCM has produced convincing evidence that this algorithm could help to estimate the working height and automate the current safety monitoring process. The proposed method protects workers from injuries and fatalities and improves productivity, quality, and worker determination. Furthermore, it has the potential to improve the return on investment by overcoming the FFL, which leads to the high amount of insurance and fines from the occupational agencies. Moreover, the HCM module can be used in other engineering domains for height estimation using a vision camera, and it can be generalized with minor changes in assessing the safety conditions according to the different occupational safety measures. However, future research should focus on overcoming the limitations of the current method, as discussed in Section V.

## ACKNOWLEDGMENT

Chansik Park would like to express his gratitude to Junsung Park and Dr. Doyeop Lee, who assisted in the extraction of safety rules.



## REFERENCES

- [1] R. Y. Sunindijo and P. X. W. Zou, "Political skill for developing construction safety climate," *J. Construct. Eng. Manage.*, vol. 138, pp. 605–612, May 2012, doi: [10.1061/\(ASCE\)CO.1943-7862.0000482](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000482).
- [2] N. M. Hanapi, M. M. M. Kamal, M. I. Ismail, and I. A. P. Abdullah, "Identifying root causes and mitigation measures of construction fall accidents," *Gading J. Social Sci.*, vol. 17, no. 1, pp. 65–79, 2013.
- [3] P. Kines and K. L. Mikkelsen, "Effects of firm size on risks and reporting of elevation fall injury in construction trades," *J. Occupational Environ. Med.*, vol. 45, no. 10, pp. 1074–1078, Oct. 2003, doi: [10.1097/01.jom.0000085887.16564.3a](https://doi.org/10.1097/01.jom.0000085887.16564.3a).
- [4] U.S. Bureau of Labor Statistics. (2018). *Fatal Work-Related Falls*. Accessed: Sep. 21, 2021. [Online]. Available: <https://www.bls.gov/opub/ted/2018/fatal-work-related-falls-to-a-lower-level-increased-26-percent-from-2011-to-2016.htm>
- [5] CPWR. (2021). *Fatal Injury Trends in the Construction Industry*. [Online]. Available: <https://www.cpwr.com/wp-content/uploads/DataBulletin-February-2021.pdf>
- [6] Ministry of Employment and Labor. (2019). *Korea Occupational Safety and Health Agency*. Accessed: Sep. 21, 2021. [Online]. Available: <https://www.kosha.or.kr/kosha/index.do>
- [7] H.-H. Sim and K.-S. Kang, "A study on the death accident analysis of ladder and prevention measures for fall accidents," *J. Korea Saf. Manage. Sci.*, vol. 19, no. 4, pp. 95–104, 2017, doi: [10.12812/kmsm.2017.19.4.95](https://doi.org/10.12812/kmsm.2017.19.4.95).
- [8] R. A. Partridge, A. S. Virk, and R. E. Antosia, "Causes and patterns of injury from ladder falls," *Acad. Emergency Med.*, vol. 5, no. 1, pp. 31–34, Jan. 1998, doi: [10.1111/J.1553-2712.1998.TB02571.X](https://doi.org/10.1111/J.1553-2712.1998.TB02571.X).
- [9] N. Gromicko and K. Shepard. (2021). *Ladder Safety*. Accessed: Sep. 21, 2021. [Online]. Available: <https://www.nachi.org/ladder-safety.htm>
- [10] International Organization for Standardization. *ISO—ISO 45000 Family—Occupational Health and Safety*. Accessed: Feb. 11, 2022. [Online]. Available: <https://www.iso.org/iso-45001-occupational-health-and-safety.html>
- [11] Occupational Safety & Health Administration. *Portable Ladder Safety*. Accessed: Nov. 10, 2021. [Online]. Available: <https://www.osha.gov/laws-regs/regulations/standardnumber/1926/1926.1053>
- [12] Safety Company. (Feb. 10, 2015). *Direct vs Indirect Costs of Work-Related Injuries—SafetyCompany*. Accessed: Feb. 14, 2022. [Online]. Available: <https://www.safetycompany.com/safetyblog/direct-vs-indirect-costs-of-workrelated-injuries/>
- [13] H. Li, M. Lu, S.-C. Hsu, M. Gray, and T. Huang, "Proactive behavior-based safety management for construction safety improvement," *Saf. Sci.*, vol. 75, pp. 107–117, Jun. 2015, doi: [10.1016/j.ssci.2015.01.013](https://doi.org/10.1016/j.ssci.2015.01.013).
- [14] J. Seo, S. Han, S. Lee, and H. Kim, "Computer vision techniques for construction safety and health monitoring," *Adv. Eng. Informat.*, vol. 29, no. 2, pp. 239–251, Apr. 2015, doi: [10.1016/J.AEI.2015.02.001](https://doi.org/10.1016/J.AEI.2015.02.001).
- [15] L. Hou, S. Wu, G. K. Zhang, Y. Tan, and X. Wang, "Literature review of digital twins applications in construction workforce safety," *Appl. Sci.*, vol. 11, no. 1, pp. 1–21, 2021, doi: [10.3390/app11010339](https://doi.org/10.3390/app11010339).
- [16] W. Xu and T.-K. Wang, "Dynamic safety prewarning mechanism of human-machine-environment using computer vision," *Eng., Construct. Archit. Manage.*, vol. 27, no. 8, pp. 1813–1833, Jul. 2020, doi: [10.1108/ECAM-12-2019-0732](https://doi.org/10.1108/ECAM-12-2019-0732).
- [17] N. O'Mahony, "Deep learning vs. traditional computer vision," in *Proc. Sci. Inf. Conf.*, vol. 943, 2020, pp. 128–144.
- [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2012.
- [19] J. Schmidhuber, "Deep Learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Oct. 2015, doi: [10.1016/j.neunet.2014.09.003](https://doi.org/10.1016/j.neunet.2014.09.003).
- [20] W. Fang, L. Ding, B. Zhong, P. E. D. Love, and H. Luo, "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach," *Adv. Eng. Informat.*, vol. 37, pp. 139–149, Aug. 2018, doi: [10.1016/J.AEI.2018.05.003](https://doi.org/10.1016/J.AEI.2018.05.003).
- [21] T. Kong, W. Fang, P. E. D. Love, H. Luo, S. Xu, and H. Li, "Computer vision and long short-term memory: Learning to predict unsafe behaviour in construction," *Adv. Eng. Informat.*, vol. 50, Oct. 2021, Art. no. 101400, doi: [10.1016/j.aei.2021.101400](https://doi.org/10.1016/j.aei.2021.101400).
- [22] V. Thakar, H. Saini, W. Ahmed, M. M. Soltani, A. Aly, and J. Y. Yu, "Efficient single-shot multibox detector for construction site monitoring," in *Proc. IEEE Int. Smart Cities Conf. (ISC2)*, Sep. 2018, pp. 1–6, doi: [10.1109/ISC2.2018.8656929](https://doi.org/10.1109/ISC2.2018.8656929).
- [23] G. Ma, M. Wu, Z. Wu, and W. Yang, "Single-shot multibox detector- and building information modeling-based quality inspection model for construction projects," *J. Building Eng.*, vol. 38, Jun. 2021, Art. no. 102216, doi: [10.1016/J.JOBE.2021.102216](https://doi.org/10.1016/J.JOBE.2021.102216).
- [24] B. Zhong, X. Pan, P. E. D. Love, L. Ding, and W. Fang, "Deep learning and network analysis: Classifying and visualizing accident narratives in construction," *Autom. Construct.*, vol. 113, May 2020, Art. no. 103089, doi: [10.1016/j.autcon.2020.103089](https://doi.org/10.1016/j.autcon.2020.103089).
- [25] N. Khan, M. R. Saleem, D. Lee, M.-W. Park, and C. Park, "Utilizing safety rule correlation for mobile scaffolds monitoring leveraging deep convolution neural networks," *Comput. Ind.*, vol. 129, Aug. 2021, Art. no. 103448, doi: [10.1016/J.COMPIND.2021.103448](https://doi.org/10.1016/J.COMPIND.2021.103448).
- [26] S. Tang, D. Roberts, and M. Golparvar-Fard, "Human-object interaction recognition for automatic construction site safety inspection," *Autom. Construct.*, vol. 120, Dec. 2020, Art. no. 103356, doi: [10.1016/J.AUTCON.2020.103356](https://doi.org/10.1016/J.AUTCON.2020.103356).
- [27] W. Fang, B. Zhong, N. Zhao, P. E. D. Love, H. Luo, J. Xue, and S. Xu, "A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network," *Adv. Eng. Informat.*, vol. 39, pp. 170–177, Jan. 2019, doi: [10.1016/J.AEI.2018.12.005](https://doi.org/10.1016/J.AEI.2018.12.005).
- [28] M. Wang, P. K.-Y. Wong, H. Luo, S. Kumar, V.-S.-K. Delhi, and J. C.-P. Cheng, "Predicting safety hazards among construction workers and equipment using computer vision and deep learning techniques," in *Proc. Int. Symp. Automat. Robot. Construct. (IAARC)*, May 2019, pp. 399–406, doi: [10.22260/aisarc2019/0054](https://doi.org/10.22260/aisarc2019/0054).
- [29] M. Khan, R. Khalid, S. Anjum, N. Khan, S. Cho, and C. Park, "Automation in construction tag and IoT based safety hook monitoring for prevention of falls from height," *Autom. Construct.*, vol. 136, Jan. 2022, Art. no. 104153, doi: [10.1016/j.autcon.2022.104153](https://doi.org/10.1016/j.autcon.2022.104153).
- [30] J. Seo, S. Han, S. Lee, and T. J. Armstrong, "Feasibility of on-site biomechanical analysis during ladder climbing," in *Proc. Construct. Res. Congr.*, 2014, pp. 739–748.
- [31] L. Ding, W. Fang, H. Luo, P. E. D. Love, B. Zhong, and X. Ouyang, "A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory," *Autom. Construct.*, vol. 86, pp. 118–124, Feb. 2018, doi: [10.1016/j.autcon.2017.11.002](https://doi.org/10.1016/j.autcon.2017.11.002).
- [32] Y. Piao, W. Xu, T.-K. Wang, and J.-H. Chen, "Dynamic fall risk assessment framework for construction workers based on dynamic Bayesian network and computer vision," *J. Construct. Eng. Manage.*, vol. 147, no. 12, pp. 1–16, 2021, doi: [10.1061/\(asce\)co.1943-7862.0002200](https://doi.org/10.1061/(asce)co.1943-7862.0002200).
- [33] H. Chen, X. Luo, Z. Zheng, and J. Ke, "A proactive workers' safety risk evaluation framework based on position and posture data fusion," *Autom. Construct.*, vol. 98, pp. 275–288, Feb. 2019, doi: [10.1016/j.autcon.2018.11.026](https://doi.org/10.1016/j.autcon.2018.11.026).
- [34] S. Han, S. Lee, and F. Peña-Mora, "Vision-based detection of unsafe actions of a construction worker: Case study of ladder climbing," *J. Construct. Eng. Manage.*, vol. 27, no. 6, pp. 635–644, 2013, doi: [10.1061/\(asce\)cp.1943-5487.0000279](https://doi.org/10.1061/(asce)cp.1943-5487.0000279).
- [35] *Korea Occupational Safety and Health Agency | Business | Certification | 03. KOSHA 18001*. Accessed: Feb. 10, 2022. [Online]. Available: <https://www.kosha.or.kr/english/business/KOSHA18001.do>
- [36] W. Fang, L. Ding, H. Luo, and P. E. D. Love, "Falls from heights: A computer vision-based approach for safety harness detection," *Autom. Construct.*, vol. 91, pp. 53–61, Jul. 2018, doi: [10.1016/J.AUTCON.2018.02.018](https://doi.org/10.1016/J.AUTCON.2018.02.018).
- [37] J. Shao, S. K. Zhou, and R. Chellappa, "Robust height estimation of moving objects from uncalibrated videos," *IEEE Trans. Image Process.*, vol. 19, no. 8, pp. 2221–2232, Aug. 2010, doi: [10.1109/TIP.2010.2046368](https://doi.org/10.1109/TIP.2010.2046368).
- [38] *Kosha Occupational Safety and Health Agency. (2015). Legislation*. Accessed: Sep. 24, 2021. [Online]. Available: <https://www.kosha.or.kr/english/legislation/occupationalSafetyAndHealth.do?mode=list>
- [39] *Post Safety Work Instructions for Mobile Ladders*. Accessed: Sep. 24, 2021. [Online]. Available: [http://www.moel.go.kr/info/etc/dataroom/view.do?bbs\\_seq=20190300518](http://www.moel.go.kr/info/etc/dataroom/view.do?bbs_seq=20190300518)
- [40] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-YOLOv4: Scaling cross stage partial network," 2021, *arXiv:2011.08036*.
- [41] S. Ren, K. He, R. Girshick, and J. Sun. *Faster R-CNN: Towards Real-Time Object Detection With Region Proposal Networks*. Accessed: Sep. 21, 2021. [Online]. Available: <https://github.com/>

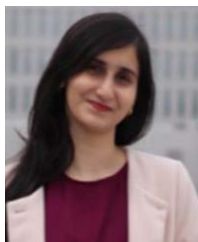
- [42] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- [43] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9905, 2016, pp. 21–37, doi: [10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2).



**SHARJEEL ANJUM** received the bachelor's degree in computer science (BSCS) from the University Institute of Information Technology (UIIT), Arid Agriculture University, Pakistan. He is currently pursuing the master's degree from the School of Architecture and Building Science, Chung-Ang University, Seoul, South Korea. His research interests include machine/deep learning, the IoT, smart applications for construction management, and worker safety.



**NUMAN KHAN** received the B.S. degree in civil engineering and the M.S. degree in engineering management from the Ghulam Ishaq Khan Institute of Science and Technology (GIKI), Topi, Pakistan, in 2015, and the Ph.D. degree in architectural engineering from Chung-Ang University, Seoul, South Korea. He currently holds a postdoctoral position with the Department of Construction Engineering, École de Technologies Supérieure, Montreal, Canada. His research interests include BIM, computer vision, construction quality, rule-based modeling, construction fire safety, human safety in construction, and construction 4.0.



**RABIA KHALID** (Member, IEEE) received the bachelor's degree from the Department of Electrical Engineering, Pakistan Institute of Engineering & Applied Sciences (PIEAS). She is currently pursuing the master's degree with the School of Architecture and Building Science, Chung-Ang University, Seoul, South Korea. Her research interests include vision intelligence, IoT-based construction management systems, and construction safety.



**MUHAMMAD KHAN** received the master's degree in civil engineering from Dong-A University. He is currently pursuing the Ph.D. degree with the Civil, Construction and Environmental Engineering Department, The University of Alabama, USA. His research mainly focuses on workers' safety at the construction site utilizing sensors and computer vision technologies. In this context, his current research will identify the different accident risk factors that trigger fall accidents at the construction site and propose a safety framework by integrating different technologies to mitigate risks. In addition, he will be working to develop a digital twin model for PM dust emission, control, and monitoring in real-time.



**DONGMIN LEE** received the B.E. and Ph.D. degrees in civil and architectural engineering from Korea University, Seoul, South Korea. He has been an Assistant Professor with the School of Architecture and Building Science, Chung-Ang University, since 2021. His research interests include the integration of construction equipment, method, planning, scheduling, and control to support a better human-robot collaborative working environment. In this context, his current research focuses on improving project performance (e.g., cost, schedule, quality, safety, and sustainability) in the built environment by developing and testing of a digital twin of physical assets (e.g., robots, workers, and materials), which can be used to simulate "what-if" scenarios using AI-based techniques (e.g., deep reinforcement learning).



**CHANSIK PARK** received the B.E. and M.E. degrees in architecture from Chung-Ang University, Seoul, South Korea, the M.S. degree from the University of Colorado at Boulder, and the Ph.D. degree from the University of Florida with a major in construction management. He has been a Professor with the School of Architecture and Building Science and the former Dean of the Graduate School of Construction Engineering, Chung-Ang University, since 1995. He is one of the founders and the former President of KICEM, the Founder of ICCEPM, and the Vice President of Building Smart Korea.

...