

Article

Construction Work-Stage-Based Rule Compliance Monitoring Framework Using Computer Vision (CV) Technology

Numan Khan ¹, Syed Farhan Alam Zaidi ², Jaehun Yang ², Chansik Park ² and Doyeop Lee ^{2,*}

¹ Department of Civil Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi 23460, Pakistan; numanpe@gmail.com

² Construction Technology Innovation Laboratory (ConTI Lab), Department of Architectural Engineering, Chung-Ang University, Seoul 06974, Republic of Korea; syedfarhanalam1993@gmail.com (S.F.A.Z.); yjhoon11@cau.ac.kr (J.Y.); cpark@cau.ac.kr (C.P.)

* Correspondence: doyeop@cau.ac.kr

Abstract: Noncompliance with safety rules is a major cause of unsatisfactory performance in construction safety worldwide. Although some research efforts have focused on using computer vision (CV) methods for safety rule inspection, these methods are still in their early stages and cannot be effectively applied on construction job sites. Therefore, it is necessary to present a feasible prototype and conduct a detailed analysis of safety rules to ensure compliance at the job site. This study aims to extend the validation of safety rule analysis through four case scenarios. The proposed structured classification of safety rules includes categorizing them based on project phases and work stages. The construction phase-related rules are divided into four groups: (1) before work, (2) with intervals, (3) during work, and (4) after work. To validate the proposed framework, this research developed prototypes for each group's scenarios using deep learning algorithms, a storage database to record compliance with safety rules, and an Android application for edge computing, which is required in the "before work" and "after work" groups. The findings of this study could contribute to the development of a compact CV-based safety monitoring system to enhance the current safety management process in the construction industry.

Keywords: construction safety rule; vision intelligence-based monitoring; work-stage-based safety monitoring; deep learning



Citation: Khan, N.; Zaidi, S.F.A.; Yang, J.; Park, C.; Lee, D. Construction Work-Stage-Based Rule Compliance Monitoring Framework Using Computer Vision (CV) Technology. *Buildings* **2023**, *13*, 2093. <https://doi.org/10.3390/buildings13082093>

Academic Editor: Hongping Yuan

Received: 30 June 2023

Revised: 2 August 2023

Accepted: 9 August 2023

Published: 17 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The construction industry takes up approximately 7% of the global workforce; however, it accounts for approximately 30–40% of the total fatalities occurring worldwide [1]. Despite the significant growth in globalization and the development of cutting-edge technologies, construction sites remain hazardous worksites, with many accidents and fatalities still plaguing this industry [2]. On average, approximately one fatality is reported every nine minutes in the construction industry [3]. According to the estimate calculated by the Centre for Construction Research and Training, a 75% probability of an accident leading to disability of a construction worker is expected within their 45-year career [4]. Aside from the humanitarian implications, these incidents result in significant financial losses that could be avoided with special safety precautions [5]. The annual cost of these enormous accidents exceeds USD 48 billion in the United States [6]. Moreover, these injuries and fatalities could result in more excellent experience modification rates, resulting in lower profit margins and increasing operating costs [7]. Correspondingly, such costs have jeopardized the endurance, sustainability, and competitiveness of construction businesses. Based on these statistics, occupational safety and health (OSH) in the construction industry remains a critical concern that must be urgently addressed.

Numerous safety management techniques have been proposed to overcome these adverse outcomes. For instance, safety policy procedures have been globally founded

to improve the construction safety management (CSM) process [8], aiming to foster the adoption of best construction safety practices at job sites. These guidelines could minimize the alarming statistics by fostering the best construction safety practices at job sites. Failure to comply with safety rules or ignoring best practices can substantially increase the probability of workplace accidents. Conversely, the likelihood of accidents reduces when best industrial practices and safety rules are appropriately implemented and managed. Adherence to safety rules and the implementation of industry best practices significantly reduce the likelihood of workplace accidents [9]. However, acquiring a broad range of safety information, identifying relevant rules in a safety database, and manually enforcing them is time-consuming [5].

Moreover, traditional safety management approaches may face challenges in addressing the dynamic and intricate nature of construction sites. Construction projects progress through various stages and work phases, each demanding specific safety considerations. Conventional safety monitoring methods often struggle to keep pace with the constantly changing work environment, leading to potential safety gaps and increased accident risks. As a result, the need for an intelligent and adaptive safety monitoring system becomes evident.

To address these challenges and further enhance construction safety practices, recent advancements in computer vision (CV) technology have captured significant attention. Furthermore, CV has also gotten substantial attention due to its advanced developments in specific aspects, such as enhancements in high-definition cameras, access to high-speed Internet, and advancements in augmented storage. Consequently, CV-based systems are ubiquitous for monitoring project progress [10], productivity analysis [11], and safety monitoring [12]. Nevertheless, while CV technology has demonstrated promising results in project monitoring, its potential for safety rule monitoring, especially in life-threatening circumstances, has not yet been fully realized. Safety protocols at construction job sites involve specific considerations for different work stages and phases, making it crucial to develop a comprehensive framework that allows CV to be harnessed effectively for work-stage-based rule compliance monitoring. This paper aims to bridge this gap and proposes a pioneering framework for work-stage-based safety rule compliance using CV technology. By categorizing and extracting relevant safety rules from Occupational Safety and Health Administration (OSHA) guidelines, tailored to different construction stages and work phases, our framework offers a systematic approach to automated safety monitoring. Through real-life scenarios and prototypes, we validate the efficacy of our framework in improving construction safety practices and minimizing workplace accidents. By amalgamating the power of CV with construction safety management, this research strives to establish a proactive and efficient safety monitoring system that fosters a safer working environment for construction workers and stakeholders alike.

The remainder of this paper is structured as follows. Section 2 gives a comprehensive review of existing safety management techniques in the construction industry. It further explores the significance of safety management and its global impact on improving construction safety management (CSM) processes. Section 3 provides the theoretical foundation and methodology behind the proposed work-stage-based safety rule compliance framework leveraging computer vision (CV) technology. It describes how to identify and organize pertinent safety regulations from OSHA regulations that are appropriate for various work stages and construction phases. Section 4 is about dataset generation. Sections 5–8 discuss the design and implementation of the CV-based safety monitoring system for each construction stage and present real-life scenarios to validate the framework's efficacy. Section 9 presents a discussion of the major findings of this research work. Finally, Section 10 concludes the paper by summarizing the contributions, highlighting practical implications, and outlining potential future research directions to further enhance construction safety using CV technology.

2. Recent Advancements in Construction Safety Monitoring

2.1. Current Safety Monitoring in Construction

Construction safety is becoming increasingly important as more emphasis is placed on developing extra infrastructure, such as complex buildings and large projects [13]. Consequently, its unique nature poses difficulties in terms of safety monitoring and occupational health. Because of its inefficiency in dealings with these risky environments, the construction sector does not have an outstanding global reputation. Driven by this over the last decade, scientists have concentrated on improving safety management by utilizing emerging technologies. Many scholars worldwide have developed different safety planning techniques, such as extracting domain-specific safety guidelines and rules from the safety database and combining these with building information modeling (BIM) [9,14]. Accordingly, numerous researchers have used BIM for risk recognition in the 3D environment during different construction stages. However, this approach still requires manual inputs from the field [15]. The current conventional method for addressing safety concerns is the traditional top-to-bottom control strategy, which is commonly employed for safety monitoring during the construction execution stage at job sites. In this study, the top-to-bottom approach is defined as a safety monitoring process in which safety methods are enforced exclusively from top-level managers downward. Conversely, the bottom-up approach is defined as safety management in which the bottom managers also participate and report the information from the bottom to the top. The traditional top-to-bottom job site observation during construction relies on the physical presence of a safety manager to identify hazards or violations of safety rules [16,17]. Observation of the construction job site is generally performed on a weekly or biweekly basis, depending on the size and complexity, with random visits lasting for 1 to 3 h [18]. The staff visually investigates the equipment, tools, and work area [19] to report the hazards to be eliminated. The success of these observations for safety rule compliance is determined by the expertise and competence of the concerned person [20]. Existing safety observation methods are manual, costly, and time-consuming [21]. Researchers have recently sought to develop advanced approaches such as CV and Internet of Things (IoT)-based automated security management as a solution to these challenges, reducing the challenges associated with manual safety inspection. [8,22–24].

2.2. Technology Advancements in Construction Safety

Technology has been playing an increasingly pivotal role in enhancing job site safety with construction task automation [25,26]. Development and technological advancements have inspired each area of the construction sector, including safety management. Information and communication technologies (ICT) have been widely used in CSM over the past decade [27]. Skibniewski reviewed the use of ICT for CSM over the past decade [28]. One particular area where technology has shown promise is in the application of visualization technology for construction safety training. By harnessing the power of visualization tools, construction professionals can gain a better understanding of potential risks and take proactive measures to mitigate them [7]. These tools facilitate the education of workers regarding safety protocols, assist in safety monitoring [29,30], and aid in job hazard analysis [31].

Moreover, the IoT has attracted significant interest due to its potential deployment and applications in CSM [32]. The IoT allows for the deployment of sensors and connected devices on construction sites, generating real-time data on environmental conditions, equipment performance, and worker behavior. This data can be analyzed to detect safety issues in a proactive manner, enhance real-time hazard identification, and optimize safety protocols [33–35].

Similarly, CV technologies have witnessed rapid advancement and have found increasing applications in CSM. The abundance of digital image data from real construction sites has paved the way for more accurate and efficient CV-based object detection methods. These technologies can analyze images and video footage to identify potential safety hazards, monitor worker behavior, and ensure compliance with safety protocols [36]. Addi-

tionally, CV technology can automate the process of job hazard analysis by automatically detecting and categorizing potential hazards based on visual data [12]. The following section further elaborates on the recent advancements and roles of CV-based technologies in CSM.

2.3. Roles of CV-Based Technologies in Construction

This section summarizes previous studies on the detection and real-world localization of construction entities in digital images extracted from construction sites. The pros and cons of the various approaches are also explored. CV is a field of science that deals with how computers understand high-level information from frames of images [5,16,23]. Advancements in machine learning have enabled computers to better interpret visual data, leading to significant progress in CV [37]. However, traditional machine learning algorithms still face limitations in analyzing raw data, necessitating the development of feature descriptions using expert knowledge and engineering experience [38].

Convolutional neural network (CNN)-based models have proven successful and effective in extracting object information, as demonstrated by their application to Microsoft Common Objects in Context and PASCAL Visual Object Classes (VOC) 2007 datasets [37]. By integrating deep learning methods with collected images and employing CV, it becomes possible to automatically acquire and utilize data for training that may include features not originally designed by human engineers.

In contrast to artificial neural networks, deep learning models consist of multiple processing layers based on neural networks that learn from data at various levels of abstraction [38]. CNN is the most widely used deep learning approach, comprising three types of neural layers: (a) convolutional, (b) pooling, and (c) fully connected. CNN-based deep learning algorithms have found extensive applications in various CV tasks, such as classification, object recognition, object segmentation, and posture analysis. This research study summarizes previous efforts made to adopt deep learning-related technologies as they provide the foundation for CV applications in construction, particularly in recognizing unsafe behavior.

2.4. Image Classification

Image classification is utilized to recognize objects in images by labeling them and determining the probability for a specific visual object class (VOC). In the construction domain, CV commonly employs a combination of hand-crafted features and deep learning algorithms [12]. Various descriptors, such as histogram of optical flow (HOF), scale-invariant feature transform (SIFT), and histogram of oriented gradients (HOG), are popular choices for extracting features from images of construction job sites [39]. However, the dynamic nature of construction sites can impact feature extraction and reduce object detection accuracies due to factors like intra-class variation, complex backgrounds, and changes in viewpoint and scale.

In terms of performance, AlexNet CNN outperformed the scale-invariant feature transform technique with an accuracy of 26.2% and achieved a top-five error rate of 15.3% [37]. As a result, CNN has gained recognition as a prominent classification model in CV, especially with improved detection accuracy obtained through additional training sets from ImageNet. Simonyan and Zisserman [40] introduced the Visual Geometry Group (VGG-16) model, comprising 16 convolutional layers, multiple max-pooling layers, and three fully connected layers. Their model achieved a top-five error rate of 7.3%.

2.4.1. Object Detection

Object detection is a significant part of CV because it identifies the semantic properties of an object and its positions in images. In the field of computer science, Krizhevsky et al. set the foundation for the creation of CNN-based object detection [41]. Object detection is divided into [37] (1) the generation of a set of candidate regions that contain objects, such as with selective search, Deep Mask, RPN, and Edge Boxes, and (2) the use of a CNN

to classify obtained regions into different foregrounds or backgrounds. You only look once (YOLO) version 1–8 [42] and single-shot multi-box detector (SSD) are object detection algorithms that are renowned for their speed of detection at the expense of accuracy [37]. Recent studies on CV-based construction site monitoring have concentrated on limited inspection approaches that use CV to recognize basic objects (if the absence or existence of that object generates hazards). Significant efforts were put to detect PPEs for safety rules compliance at construction sites. For instance, three models based on the YOLO architecture were presented by Nath et al. [43] to detect helmets and safety harnesses. Fang et al. [44] proposed extended Faster R-CNN techniques to recognize non-hardhat users in surveillance videos. Shanti et al. [45] proposed a YOLOv4-based technique to detect safety harnesses, lifelines, and helmets. Similarly, refs. [23,24,46–49] also applied computer vision techniques for PPEs and person detection to comply with OSHA/KOSHA general safety rules.

Various researchers also proposed computer vision-based monitoring systems to detect excavation activities for safety rules compliance at construction sites. For example, Alateeq et al. trained YOLOv5 to detect the PPEs and heavy equipment for monitoring excavation activity at construction sites [50]. Similarly, Arabi et al. [51] proposed a MobileNet-based monitoring system for construction vehicle monitoring. Wang et al. [52] proposed an R-CNN-based technique to predict the hazards among construction workers and equipment. For worker behavior analysis, Anjum et al. proposed an SSD-based technique to detect workers, PPE, and A-type ladders and estimate the worker's height while working at a ladder for compliance with the Korea Occupational Safety and Health Administration (KOSHA) rules, which alert the manager if the workers height is greater than safe height [53]. Instead of falling from a height, researchers also worked on other accidents, such as struck-by and electrocution accidents. Shin et al. [54] trained the YOLOv3 model to detect workers and trucks and calculated the collision proximity to avoid struck-by accidents. Li et al. [55] detected person, PPEs, and electric work equipment, and calculated the safe distance between person/worker and electric equipment.

2.4.2. Object Segmentation

Segmentation is a popular technique used to localize the objects and boundaries of objects. Semantic segmentation is used to detect objects and line curves in images and categorize each pixel into a fixed set of categories without distinguishing between object instances [56]. On the other hand, instance segmentation takes the task a step further by not only detecting objects but also differentiating between individual instances of the same object class [57]. The development of CNN models, which can make pixel-level predictions with a pre-trained network on large-scale datasets, has simplified the task of semantic segmentation. In contrast to image classification and object recognition, semantic segmentation necessitates 2D spatially distributed output masks. Fang et al. used the Mask R-CNN-based approach for detection and mask generation in humans and beams to identify unsafe actions of a person traversing the structural member [23]. Khan et al. developed a correlation-based unsafe behavior detection system pertaining to the mobile scaffold using segmentation through Mask R-CNN [12]. Xiao et al. employed the instance segmentation technique by training Mask R-CNN and tracking the workers at the job site by Kalman filter [58]. Kang et al. proposed a one-stage detector with ResNet-101 as the backbone, a feature pyramid network as the neck, and ProtoNet to predict the five prototype masks to improve the monitoring of outdoor work at construction sites under different weather conditions [59].

Bang et al. [60] proposed an instance segmentation-based proactive proximity-monitoring method using predictions from UAV-acquired video frames to prevent struck-by accidents at construction sites. The method achieves accurate object recognition, predicts future object locations, and generates proactive safety information. Evaluation results demonstrate high precision in object recognition, effective future frame prediction, and reliable proximity estimation. Wang et al. [61] proposed a Mask R-CNN-based technique for rebar image

detection in construction quality inspection. A mask annotation methodology based on BIM and rendering software was proposed to create synthetic datasets, enhancing model performance. Mathur et al. [62] used computer vision and deep learning to detect and segment safety gear and employees in workplace images. The ResNet-101 Mask R-CNN model outperforms ResNet-50, achieving high accuracy in class segmentation. Chen et al. [63] presented a hybrid visual information analysis framework for improving construction site safety management. It combined instance segmentation and pose estimation models to extract on-site entities' information. Using geometric relationship and time series analysis, it successfully identifies hazards, particularly achieving high precision and recall in detecting handrail-related compliance.

2.4.3. Pose Estimation

Human pose estimation aims to infer the location of human joints from photos (e.g., sequences and depth) or skeletal data provided by motion capture technology. Many factors must be considered during pose estimation, such as illumination, viewpoint, and the current contextual backdrop, which may contain noise; thus, the task is arduous [37]. Soltani et al. proposed a framework to enhance productivity and safety using a pose estimation and location data fusion system for excavators using stereo vision in the construction job site [64]. Similarly, Liu et al. [21] examined a computationally effective tracking approach from a 3D human skeleton extracted from stereo videos to observe the situation awareness of individuals to prevent accidents in construction. Assadzadeh et al. used YOLOv5 to detect the excavator and HRNet to perform a 2D pose estimation within the predicted bounding box [65]. Wen et al. utilized a modified Keypoint R-CNN algorithm to extract the 2D pose of an excavator from video frames [66]. They then converted the 2D pose to a 3D pose using kinematics constraints. The 3D pose was reprojected back to a 2D pose using a known camera matrix, and the reprojection error was calculated as the difference between the reprojected 2D pose and the estimated 2D pose. This methodology allowed for accurate excavator pose analysis and has implications for excavator monitoring and control.

Tian et al. [67] proposed a framework based on real-time 3D pose estimation to design dynamic hazardous proximity zones for excavators. The study introduced excavator-specific 3D pose estimation algorithms trained on diverse datasets and utilized post-processing and coordinate transformation to calculate real-time motion statuses and predict trajectories. This research contributed to improving excavator–worker collision prevention and advancing dynamic and specific safety management in construction sites. Zhao et al. [68] introduced computer vision and deep learning technologies to propose the YOLOv5-FastPose model for accurate pose estimation of construction machines. Their model combined YOLOv5m for improved object detection and integrated the FastPose network into the single-machine pose estimation module (SMPE) of AlphaPose. Moreover, their research contributed to enhancing pose estimation performance and adapting human pose estimation models for construction machines.

2.5. CV Techniques and Construction Safety Rules

Following the methodology adopted by Seo et al. [18], and considering the nature of perceptual information, CV-based techniques can be categorized as (a) scene-based, (b) location-based, and (c) action-based risk identification. Scene-based risk identification evaluates risks in a static scene; the required information includes the status of unsafe objects in the scene, the absence of safety equipment, and workers outside the danger zone. Table 1 shows the categorization of risk detection approaches using CV technologies. The three derived approaches from previous efforts are listed with examples, such as a safety vest or helmet for object detection [23], a limited access zone for object tracking [52], and awkward posture recognition [21]. However, the categorical synchronization of the scene-based, location-based, and action-based risk identification with respect to image data-capturing devices has not been presented. As discussed earlier, cognitive and perceptual capabilities play a critical role in risk identification, including unsafe conditions and

acts [69]. For instance, the person who observes risks first needs to understand the scene based on their perceptual capabilities, such as scene recognition or object identification. Subsequently, the observed perceptual information is evaluated with rules and guidelines set by best practices or past experiences to detect unsafe acts and conditions. However, CV techniques that aim to accomplish visual tasks are constrained to extract perceptual information because they lack the evaluation abilities required for the classification of unsafe acts [18]. Hence, CV-based techniques for construction safety monitoring need to consider perceptual information; however, expert knowledge (e.g., safety rules) and perceptual information can enrich the risk recognition process.

Table 1. Categorization of risk detection approaches for CV techniques.

Risk Detection Approaches	Description of Potential CV Techniques	Examples
Scene-based	Identify unsafe acts in static scenes at job sites using object detection	Safety vest, helmet, shoes, etc. [23,43].
Location-based	Detection of unsafe conditions based on movements and locations of entities through object tracking	Congested areas, hazardous materials, and zones of limited access [52,70]
Action-based	Identify violations of safety and health rules related to motions using action recognition.	Improper working at an unsafe speed or shortcuts, improper movements of heavy machines, and awkward postures during task completion [21,71].

2.6. Literature Review Inference

The research on specific safety rule compliance has been limited. CV-based techniques for risk identification need to consider perceptual information, and expert knowledge comparison with perceptual information is required to evaluate and determine the risk. In this domain, there have been enormous improvements in the interpretation of natural language to make it machines friendly, such as binary language [72]. However, owing to inherent complexities in safety rules, determining and implementing appropriate content is still challenging [73]. Therefore, a comprehensive classification of safety rules is mandatory to develop compact CV-based safety management systems [5].

Previous studies provide evidence of the scene-, location-, and action-based risk identification approaches; however, in-depth research is required to understand the relationships between CV-based risk detection with image data capturing devices. Additionally, managing the tremendous number of risks in a huge construction site needs image logistic devices with multiple modes. The cited challenges necessitate an organized classification of safety rules, which could be a base for CV-based safety monitoring systems in construction.

3. Classification of Safety Regulations for CV Systems

3.1. Grounded Theory Methodology

In order to elucidate CV-based safety monitoring in construction, this research implements an inductive approach for OSHA construction regulations analysis. This study follows the procedures in [74] a five-step process designed from ground theory methodology analysis to analyze previous findings. The process involves five steps: research domain definition, selection of source, screening, criteria, and outcomes. Based on the previous facts, ground theory methodology explores people's experiences and establishes a theory of how that process works. The theory is derived solely from the obtained data rather than from other sources (e.g., textbooks or opinions of researchers).

Consequently, construction safety rules were used in a five-step process to produce a theory for an efficient CV-based strategy to improve construction safety monitoring: (1) analysis scope, (2) selecting sources of safety rule investigation, (3) screening, (4) criterion, and (5) coding structure of construction safety rules. The details of this five-step categoric process can be seen in [75]. A definitive sample of 3538 regulations has been rigorously examined to determine the applicability of safety rules pertaining to each project

phase. Group 1, which focuses on ‘safety rules essential to the procurement process,’ includes regulations that specify safety requirements for tools and materials. Group 2, ‘safety rules involved in the preconstruction phase,’ consists of regulations that pertain to design requirements. Group 3, ‘safety rules related to the building phase,’ encompasses regulations that must be implemented during the construction phase. The safety standards involved in the construction phase were chosen for additional examination as the scope of the study determined that the development of CV-based systems pertained to construction safety monitoring.

3.2. Work-Stage-Based Safety Rules Classification

This structured classification prompts tactical considerations for implementing CV technologies to help the transition of conventional construction safety to the digital era. Table 2 illustrates the classification of safety rules based on the project phases. Group 1, named ‘procurement,’ encompasses safety rules related to product specifications. Based on the analysis of 3538 safety standards, 11.10 percent were found to be associated with the procurement phase. The safety rules grouped under Group 2 are associated with job site construction and account for 52.65 percent. Group 3 contains safety regulations that are involved in the design stage and account for 0.31 percent. This lower percentage in the preconstruction stage could be attributed to the long-held perception that the contractor is solely responsible for safety [9,14,76]. Moreover, 15.34 percent of the safety rules were linked to Group 4: general and management. Examples of these rules include the cancellation of permits (1926.1205(e)(3)), providing authorized entrants (1926.1204(e)(4)), medical attendance, examination, and regulations (1926.803(b)). Similarly, 30.12 percent of the rules were related to terminologies/definitions, such as ‘operation and maintenance’ in ‘1926.305(d)(1)’, ‘hand-fed ripsaws’ in ‘1926.304(i)’, ‘communications’ in ‘1926.800(f)’, and ‘landing and placing loads’ in ‘1926.757(e)’.

Table 2. Safety rule classification pertaining to the project phase.

Construction Safety Rules	Group	Types	Quantity of Rules	Shares
OSHA-1926	1	Procurement	302	8.53%
	2	Construction	1863	52.65%
	3	Preconstruction	11	0.31%
	4	General and Management	543	15.34%
	5	Terminologies and definitions	819	23.14%
Total construction safety rules as raw data			3535	100%

The safety standards pertaining to the execution phase of construction were specified for further assessment since the focus of this study is limited to the development of computer vision (CV)-based systems for construction safety monitoring. The classification of safety standards into four groups based on work stages was identified through open codes, axial codes, and selective codes: (1) before work, (2) with intervals, (3) during work, and (4) after work. These groups were characterized as follows: ‘before work’ refers to a set of rules that should be checked before work begins, ‘with interval’ refers to a set that must be checked on a regular basis, ‘during work’ refers to a set of rules that must be followed at all times, and ‘after work’ refers to rules that need to be monitored after the work is done. The additional consideration for dividing the rules into these groups was that temporary facilities (e.g., mobile scaffolds, scaffold systems, and ladders) were categorized under temporary work, while permanent structures such as concrete or steelwork were considered permanent work. The details of the condition criteria and the relationship of the conditions for work-stage-based safety monitoring can be found in the previous article [75].

Table 3 describes the technology implementation and work-stage-based classification to improve compatibility with digital data-capturing devices. Following are the four classes for work-stage-based classification using selective code relationships: (1) a share of 32.95%

was observed in the ‘before work’ class, (2) the ‘during work’ class acquired 41.59 percent of the share, (3) the ‘with interval’ class annexed 8.05%, and (4) the ‘after work’ gained 13.09%. However, management-related classes were observed as 4.2% of the rules.

Table 3. Analysis of Group 3.

Category	Work-Stage-Based Classification	Share
Implementation of CV technology for work-stage-based safety monitoring	Before work	32.95%
	During work	41.59%
	With interval	8.05%
	After work	13.09%
	Management related	4.2%
Total OSHA safety rules		100%

The proposed image data-capturing model for work-stage-based safety rule categorization assists the decision-makers in the safety-planning phase. In this study, image data-capturing devices included fixed devices, tools that could be moved with human effort, and machine-based portable devices. An action camera, a mobile camera, and a drone could be used to identify the safety requirements categorized under the ‘before work’ category. Action cameras, robot dogs, and drones can all identify rules in the ‘with interval’ category. Moreover, the ‘during work’ group could be determined using closed-circuit televisions and 360° cameras. Similarly, the “after work” category necessitates the use of a mobile camera and an action camera.

4. Dataset Description

A total of four datasets, named scaffold checking dataset, construction waste dataset, mobile scaffolding dataset, and rebar dataset, were generated and utilized to train and validate the CV-based systems for safety compliance checking before work, during work, with intervals, and after work, respectively. Image augmentation techniques, including rotation, crop, shear, hue, saturation, brightness, exposure, and blur, were applied to the training set of Dataset A, Dataset B, and Dataset D to prevent overfitting and the dataset was pre-processed and labeled using the Roboflow platform. However, Dataset C was used from the previous research [12], and the augmentation techniques used for that dataset are mentioned in Section 4.3. The details of these datasets are as follows:

4.1. Dataset A

A diverse image dataset of 4567 raw images showcasing various scaffolding scenes was created for training a computer vision-based detection model. The images were obtained from three sources: Google, YouTube, and recorded videos from a scaffolding training institute. A total of 962 images were selected for the experiment, and 4060 annotations were labeled across four classes: missing planks, guardrail missing, person, and wrong installation. The dataset was split into a 70:15:15 ratio for training, validation, and testing, respectively. After the stated augmentation process, the resulting training set had 2016 images. See Figure 1 for detailed information on class instances.

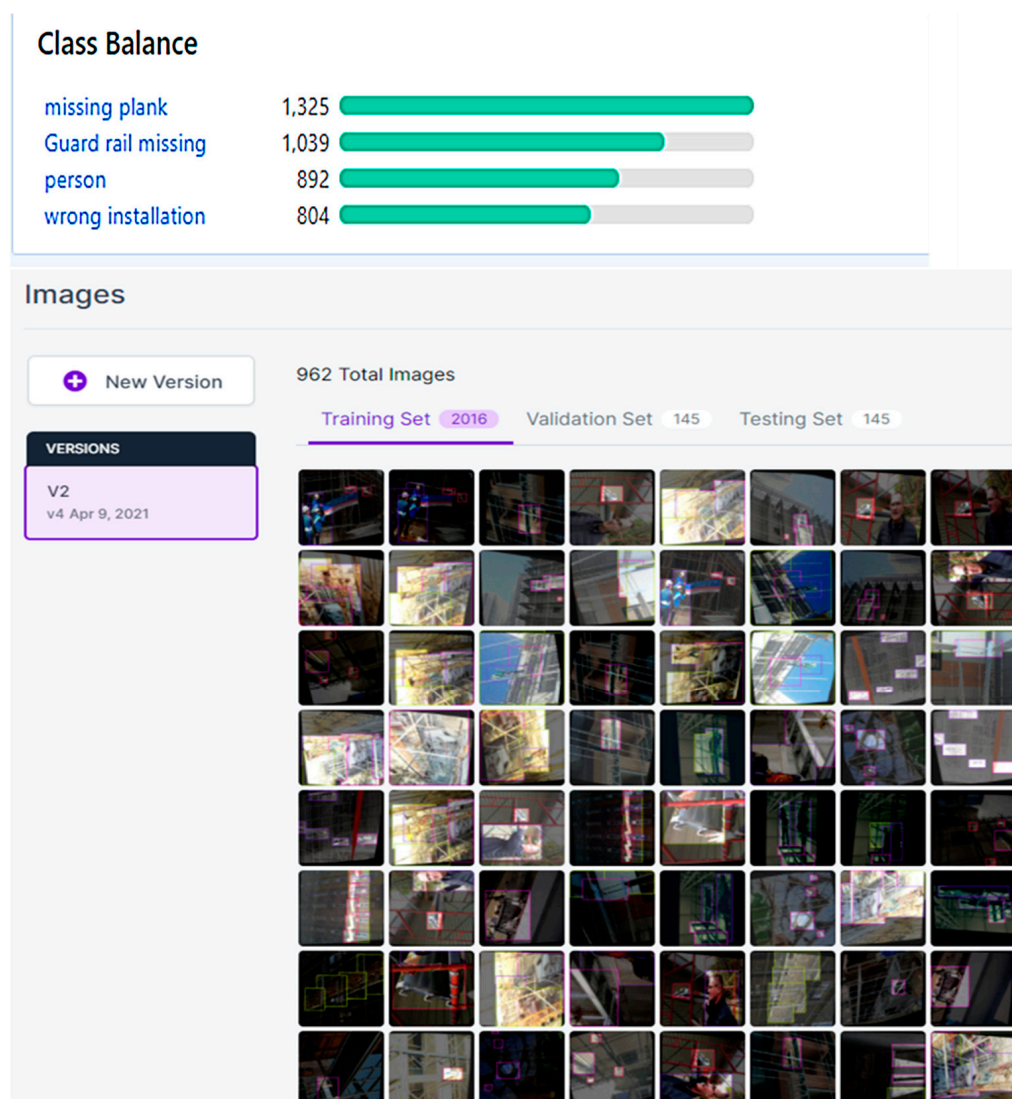


Figure 1. Visualization of dataset pre-processing.

4.2. Dataset B

A diverse collection of 1003 images and videos from construction sites and various Internet sources was gathered to train deep learning algorithms for object detection in construction waste scenarios. The dataset included categories like brick waste, concrete waste, steel waste, timber waste, and mixed waste, totaling 549 images after filtering for quality. The images were divided into training (70%), validation (15%), and testing (15%) sets. The dataset was labeled with five specified labels. This approach addresses the limitations of existing image datasets in the computer vision community and aims to improve object detection in construction waste scenarios. Figure 2 shows the construction waste dataset.

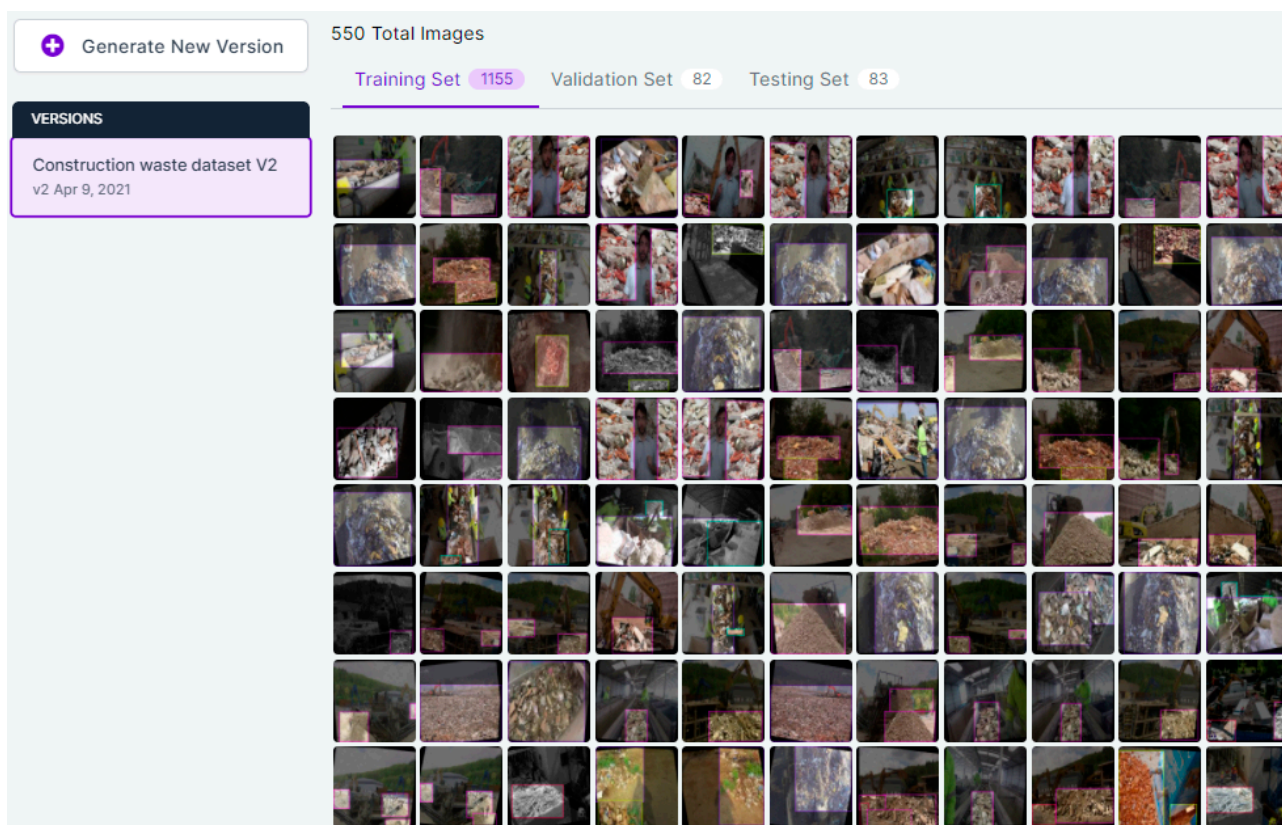


Figure 2. Construction waste dataset.

4.3. Dataset C

A large amount of labeled digital data is required to train a deep learning-based classifier. Therefore, the dataset used in the previous study was employed to validate this group of studies [12]. A total of 938 images were used as the starting point for the experimentation. These 938 images in the database were divided into training and validation sets in a 75:25 ratio, with 75% of the images used to train the model and validate the model during training and 25% used to test the model's precision and accuracy. During each training iteration, data augmentation techniques were applied using five strategies to optimize the performance of the trained network and prevent overfitting. The training images obtained with the VGG annotator were augmented with Gaussian blur, affine transformations, and contrast normalization to reduce the storage space required for the training set. Three classes were defined to recognize unsafe behavior: (1) 'scaffold' for mobile scaffolding without outriggers, (2) 'safescaffold' for mobile scaffolding equipped with outriggers, and (3) 'person' for a worker.

4.4. Dataset D

The authors addressed the lack of a construction steel rebar dataset by creating one based on OSHA safety rules. They used various keywords to collect 849 raw images from Google and YouTube, resulting in a final dataset of 683 images. The dataset was split into training, validation, and testing sets in a 70:15:15 ratio and labeled with two classes: steel with cap and steel without cap using the Roboflow platform. A total of 11,636 annotations were labeled, averaging 17 per image. After augmentation, the training dataset increased to 1440 images, while validation and testing datasets remained the same. Figure 3 shows the health check visualization of the dataset.

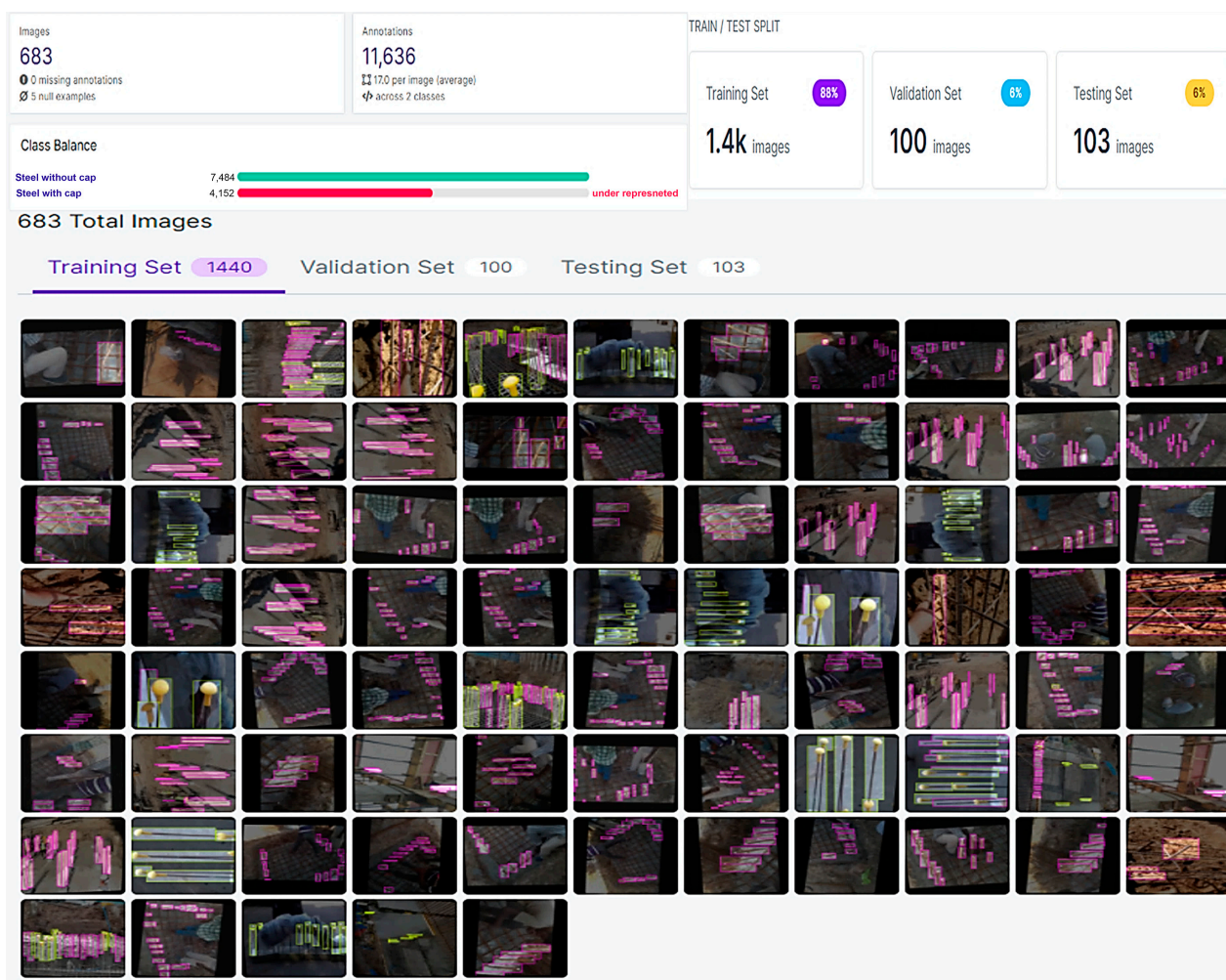


Figure 3. Construction steel rebar dataset labeling process and monitoring.

5. CV-Based System for Safety Compliance Checking before Work

This section presents the motivation for selecting particular examples to validate the proposed concept of a CV-based system for implementing safety rules during each work stage. The object detection and classification model's design, architecture, and deployment for every four cases are provided. The trained model evaluated on test images with prediction, using the CNN and integrated into the Android Studio to make a smartphone-based Android application, was required for work-stage-based safety rules monitoring.

5.1. Motivation for Selecting Scaffolds Checking for Visible Defects as an Example

In 2018, the Department of Labor in the US recorded 5250 worker deaths, with falls from heights being the main cause of death. [77]. According to OSHA, over 65 percent of workers work on scaffold systems on a regular basis [78]. Despite many efforts to implement scaffold safety in job sites, scaffolding safety standards remain the third most violated legislation [79]. According to research, scaffolding-related falls are a substantial potential concern on the job site, accounting for a significant number of accidents each year [80]. Therefore, preventing these accidents could significantly reduce the number of fatalities resulting from falls from heights. To prevent accidents and ensure the quality of the scaffold system, the best practices for checking visible defects in scaffolds are mentioned in the OSHA regulations, as follows:

1926.451(f)(3) Scaffolds and scaffold components shall be inspected for visible defects by competent personnel before each work shift and after any occurrence that could influence the structural integrity of the scaffold.

5.2. Scaffold Monitoring in Construction

Typically, the construction sector has been using safety refresher courses to educate workers on how to safely manage scaffolds [81]. OSHA provides thorough requirements and recommendations for designing, dismantling, and inspecting scaffolds [19]. Many researchers have devoted their interest to developing innovative methodologies for improving scaffold system monitoring on job sites. For instance, Sakhakarmi et al. proposed a machine learning (ML) system that uses strain data from columns of scaffolding to predict safety conditions [82]. To recreate and recognize scaffolds from point cloud data, a three-dimensional feature descriptor was developed to define linear straight-shaped objects [83]. Similarly, Jung et al. used image-processing techniques to create a failure detection method for small-scale scaffolds [84]. While many researchers have explored the structural failure of temporary facilities using IoT technologies, there is still a lack of monitoring for safety rule violations that involve the detection of visible defects in temporary structures before work commences. Thus, this study selected the mentioned case to validate the proposed concept of ensuring safety rules compliance before the start of work.

5.3. System Development

5.3.1. Model Development

YOLO revolutionized real-time object detection by proposing a single-shot approach, offering both speed and accuracy in detecting objects in images and videos [85]. Its subsequent versions, YOLOv2 [86], YOLOv3 [87], and YOLOv4 [88], further improved performance and extended the model's capabilities. Notably, YOLOv5 has achieved real-time object detection at 30 FPS with superior accuracy [89]. This single-shot approach enables faster inference on computationally constrained devices, making it suitable for various real-world applications. With its excellent performance, YOLO v5 continues to push the boundaries of object detection techniques, further advancing the field of computer vision. In this case study, the YOLO v5 (You Only Look Once version 5) architecture is utilized for training the detection model to detect the defects in the scaffold system.

5.3.2. Workflow Design and Model Deployment

Following the dataset preparation process stated in Section 4, this section presents the training of a visible defect detection model using YOLOv5. The model was exported to TensorFlow Lite (TFLite) to detect the mentioned classes. Figure 4 illustrates the design flow for checking safety rules before work commencement. The developed system relied on various open-source frameworks and was built in a cloud-based environment, specifically Google Collaboratory (Colab), and Android Studio. The Colab platform allowed for designing, executing, and validating the system on the same cloud platform. All necessary dependencies and libraries, including the PyTorch and TensorFlow libraries, were cloned and installed in Colab. YOLOv5 required the image data in the yolo (.txt) format, along with a label map consisting of numbers mapped to labels. The training dataset, generated in Roboflow, was exported to Colab. For better training, the training epochs were set to 300, the batch size remained at the default value of 32, and the default dynamic learning was set. After the training of each epoch, the model was validated using the validation set. Multiple models were generated during training, and the best model based on the validation set was chosen for deployment.

The best model is also evaluated on the test dataset, and Figure 5 showcases some examples of the model's inferences. The mAP of 69.1%, precision of 98.8%, and recall of 60.9% were computed on the validation set. Furthermore, on the testing dataset, the mAP was found to be 61.1%, precision was 92.4%, and recall was 54.9%. Subsequently, the downloaded best model is converted into the TFLite format using the TensorFlow Lite converter, which is recommended by Google. This conversion is achieved through the Python API provided by TensorFlow Lite, using the command-line tool [90].

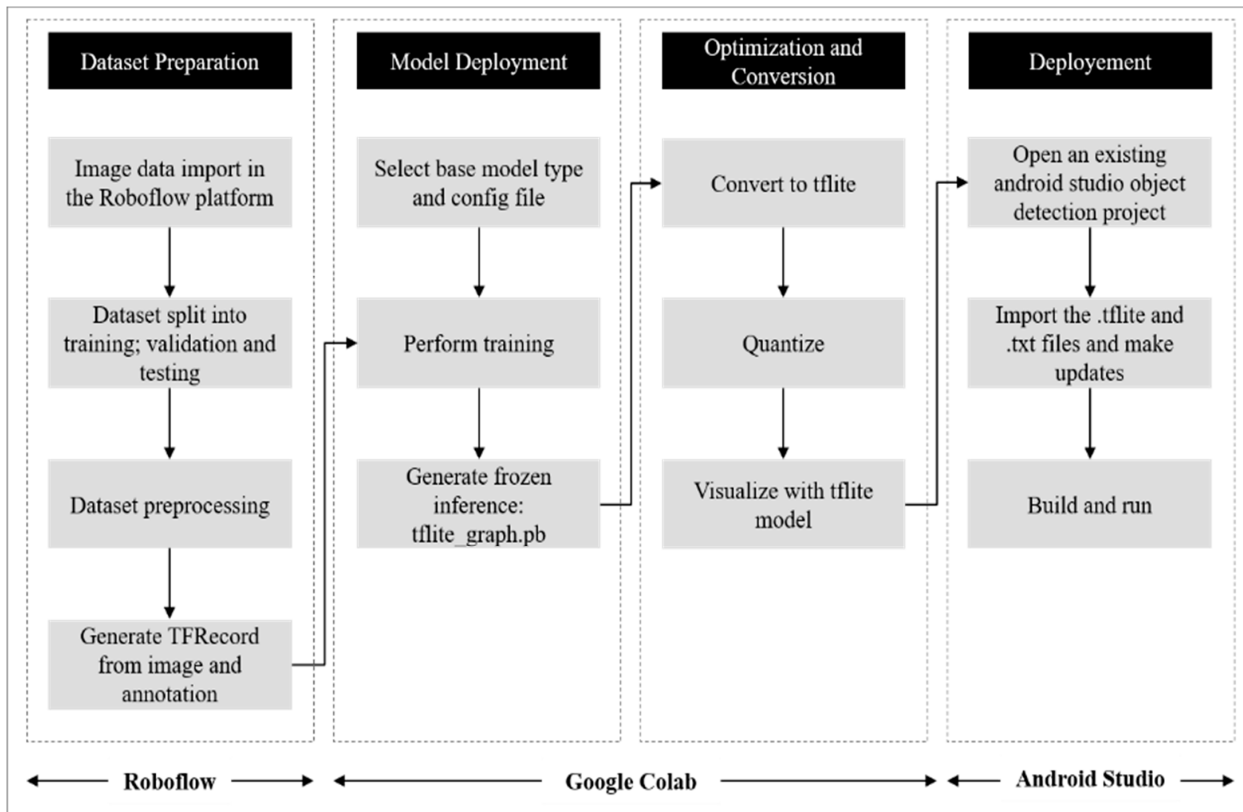


Figure 4. Flowchart of CV-based scaffolding system checking for visible defects before the commencement of work.



Figure 5. Examples of the best model trained on YOLOv5.

Rather than focusing on training models, the TFLite framework aims to facilitate the deployment of models on embedded devices with limited computational power. To successfully deploy the TFLite model on such devices, the metadata of the trained model is crucial. The TensorFlow Lite metadata includes essential information about the model, such as license terms, input details (pre-processing), such as normalization, and output details (post-processing), such as label mapping. Normalization is a common data pre-processing technique used to scale values to a common range without distorting the variations in the original value ranges. In this case, the normalization parameters for the float model were set to a mean and standard deviation of 127.5. Once the model is populated with the necessary information, it can be exported for integration into an Android application.

5.3.3. Android Application Implementation

Following the conversion of the best-fit model to a TFLite model, an integrated development environment is utilized to build the visible defect detection system for scaffolds on embedded edge computers. The application is developed based on an open-source framework cloned from the GitHub repository [91], with a TensorFlow back-end. Essential modifications have been made to both the back-end and front-end of the extracted open-source framework. In the back-end, the system records real-time data on visible unsafe conditions in the scaffolding system, such as missing guardrails, wrong plank installation, and missing planks. This data is stored in a Firebase database. Additionally, a user interface for login has been implemented in the front-end to authenticate users. The downloaded trained model with metadata is embedded within the application framework. To ensure functionality and reliability, the developed application is tested using a virtual smartphone (a built-in function of Android Studio) before deploying it to embedded devices. Once the virtual test is successfully conducted, the application is deployed on a Xiaomi Redmi Note 8. Figure 6 showcases examples of results obtained through edge computing and the user interface.

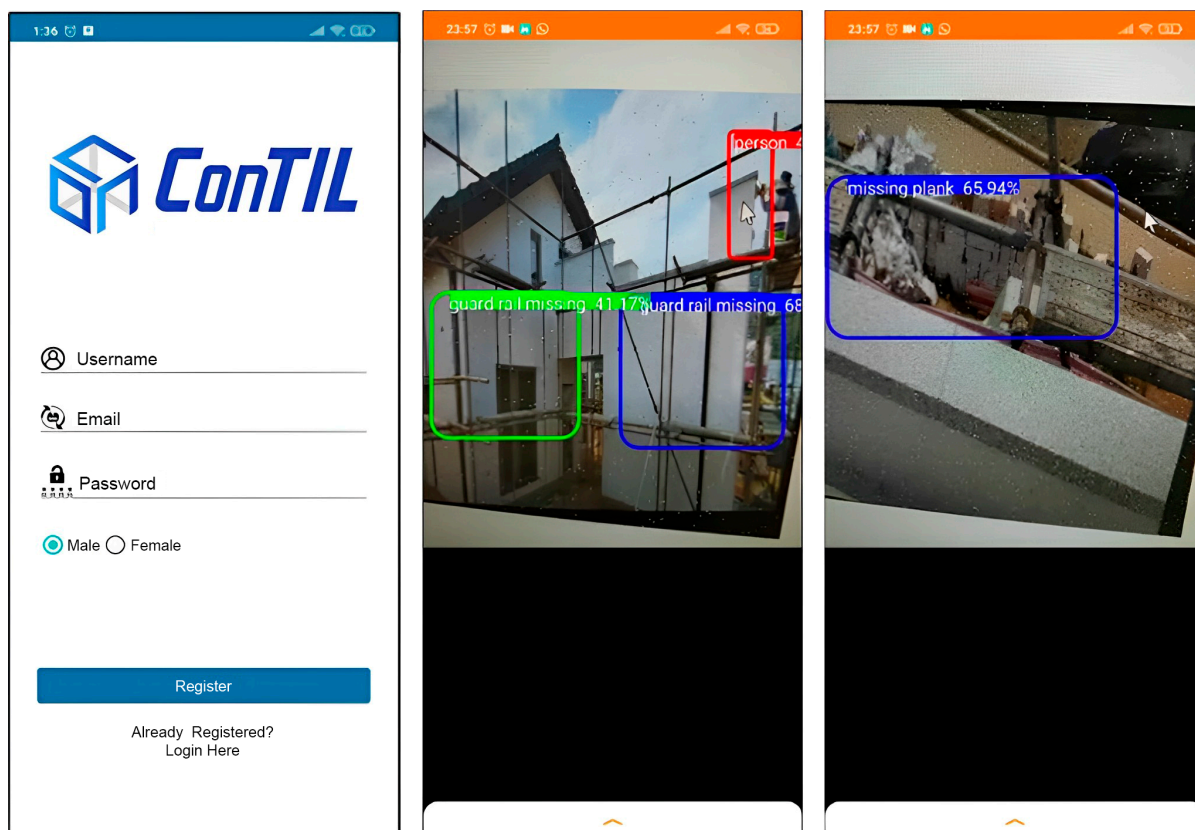


Figure 6. Android-based prototype development for visible defect inspection with a specific interval.

5.4. Firebase Database to Record Visible Defects

Following the detection of unsafe conditions such as missing planks, missing guardrails, and wrong plank installations across the video frames of the smartphone, the next step is to capture the image frame as evidence to assess safety performance. Each detected class in the video stream has an ID that can be traced across the frames. To achieve this, if-else conditions have been added to the back-end programming to record the frame once the model detects the mentioned classes, excluding the person class. The identified results are stored in a cloud-based real-time Firebase database. The Firebase real-time database is a cloud-hosted NoSQL database that enables real-time data synchronization. The Firebase SDKs for cloud storage integrate Google security to facilitate data uploading and retrieval in the application. To establish a connection between the Android application and the Firebase database, a project is created on the Firebase database. Users can sign up through the application after completing the necessary registration and authentication procedures provided by Firebase. In the Android Studio, a trial application is developed and executed on a video containing visible defects such as missing guardrails and missing planks. The Android application seamlessly communicates with Firebase and records the detected defects of the corresponding frame as a snapshot. Subsequently, the system uploads the snapshot to the specified folder in the Firebase cloud storage.

6. CV-Based Safety Rule Monitoring at Intervals

6.1. Motivation for Selecting Construction Waste Monitoring as an Example

Concurrent involvement of various resources is required in construction job sites, which makes the workplace complex [92]. Debris and trash, such as bricks, wooden chunks, and cement blocks, are generated as by-products during construction operations. These materials can restrict movement and create hazardous obstacles, increasing the risk of slip, trip, and fall incidents [93]. Slips and trips often result in falling accidents, and falls are considered a common cause of severe injuries in construction sites [93,94]. Slip, trip, and fall incidents contribute significantly to work-related accidents and injuries, including falls from heights and musculoskeletal disorders. These incidents occur when individuals lose balance while walking on workplace surfaces [94,95]. The lower extremities of the body, such as the foot or ankle, are more susceptible to injuries in slip and trip events. Construction workers are more likely to experience fractures after trips or slips, including falls from heights, as well as contusions from falls on the same level and strains or sprains resulting from trips or slips without falls from heights [94]. Instances where construction workers quickly regain balance and prevent injuries after a trip or slip are referred to as near-misses [96].

Slips and trips serve as precursors to collisions or falls, leading to severe injuries, permanent disabilities, and even fatalities among construction workers in extreme cases [96]. The occurrence of slips and trips is closely related to the level of housekeeping in the workplace. Proper housekeeping practices, including controlling, cleaning, and appropriately disposing of waste materials, maintaining clear aisles, and keeping work areas and exits clean, are essential safety measures that significantly improve construction site safety [73–77]. The construction industry employs various safety management techniques, including job safety planning, hazard control at the source and along worker paths, technical and engineering controls at the worksite, managerial controls, and the use of safety protective equipment [96]. The identification and recognition of near-miss incidents are crucial for promoting safety management and reducing potential risks.

To control housekeeping, OSHA construction safety regulations recommend the following: **1926.502(j)(6)(ii)** ‘Excess mortar, broken or scattered masonry units, and all other materials and debris shall be kept clear from the work area by removal at regular intervals.’

Safety rule 1926.502(j)(6)(ii) explains the housekeeping of job sites, including the removal of debris, excess mortar, and scattered masonry units at regular intervals to prevent tripping hazards. However, it is not practical for safety and health representatives to examine and control the housekeeping status of every work zone. The proposed system offers

an engineering technology-based approach to control construction waste and minimize or eliminate hazards by monitoring the workplace environment. This case study focuses on a CV-based method that monitors the generation of construction waste at specific intervals. A diverse dataset of construction waste, including masonry work, steelwork, and wooden work, was collected from videos. The CV-based technique is used to train a waste detection model, which is then evaluated with a test dataset to assess its ability to detect waste and identify unsafe conditions at a construction site. The system can raise an alarm by predicting potential near-misses based on the detection of unsafe conditions caused by generated waste, allowing safety practitioners to take corrective actions and control the identified hazards in the workplace.

6.2. System Development

6.2.1. Model Development

Detection using deep learning techniques has gained significant attention among researchers worldwide due to its wide range of practical applications [37]. To achieve these applications, GPUs, CPUs, IoT clusters, or embedded computers are required. Model scaling techniques have been developed to efficiently detect objects with high accuracy and real-time inference on various devices, aiming to minimize computational requirements. Recently, a novel deep learning model called scaled YOLOv4 was introduced, which is suitable for both small and large networks without compromising accuracy and speed [97]. When compared to other prominent object detection algorithms, the scaled YOLOv4 model has demonstrated superior performance in terms of optimal speed and accuracy, making it the preferred choice for object detection [98]. To achieve high object detection accuracy, a YOLOv4-large model designed for cloud GPUs was developed. In the scaled YOLOv4 approach, the authors applied a ‘CSP-ised’ technique to a portion of the network and ‘CSP-ised’ a cross-stage partial network as defined in [97]. They emphasized the importance of increasing the number and depth of stages in the CNN backbone and neck for detecting large objects in large images while noting that increasing the width has only a marginal effect. Therefore, they scaled up the number of stages and input size, adjusting the depth and width to maintain real-time inference speed. The experiments conducted with YOLOv4-P6 and YOLOv4-P7 achieved real-time performance at 30 FPS (with a width scaling factor of 1) and 16 FPS (with a width scaling factor of 1.25), respectively [97]. This study focuses on using scaled YOLOv4 for detecting construction waste in a video stream obtained from unmanned vehicles such as drones or robot dogs.

6.2.2. Model Training

Experiments were performed using an open-source framework for scaled YOLOv4, available on GitHub and Roboflow [98]. The following steps were performed to train the construction waste detection model based on scaled YOLOv4. The YOLOv4 repository was cloned to the Colab environment, and all dependencies were loaded, including the Torch Mish activation function for CUDA. The annotated dataset in the Roboflow platform was exported in the YOLOv5 PyTorch format. To import the data into the Colab notebook, the generated curl link was pasted to import the images and annotations in the .txt format. After importing the data to the Colab notebook, the scaled YOLOv4 architecture was executed. The basic scaled YOLOv4-CSP was used for training because the Colab notebook offers a single GPU to train the dataset. The batch size and epochs were set to 16 and 500, respectively. The model was trained for 200 epochs with a learning rate of 0.001. After fine-tuning all layers for another 300 epochs, the learning rate dropped by 10. Additionally, the learning momentum and weight decay were set at 0.0001 and 0.9, respectively.

An example of the image dataset used is shown in Figure 7a, and the corresponding labels are shown in Figure 7b. In the training process, the algorithm passes the training data into a data loader for each training batch, which augments the data using a mosaic data loader. Mosaic data augmentation is a novel technique that combines four images into four random-ratio tiles. The mosaic data augmentation is visualized with the labels

in Figure 7c. Mosaic augmentation helps the model learn and address the small object problem, where tiny objects are not as accurately detected as large objects. It also allows the model to detect various types of images in different parts of the frame. This enables the identification of objects that are not in their normal perspective.

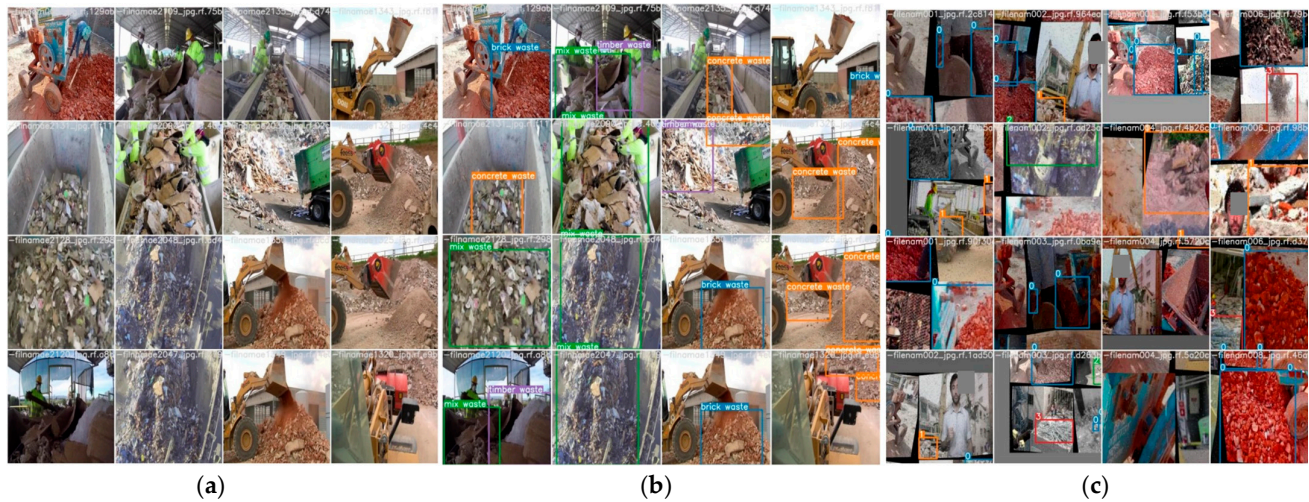


Figure 7. Examples of images in the training dataset: (a) training images, (b) labeled images, and (c) augmentations using mosaic data.

6.3. Model Evaluation and Results

In this section, three performance indicators were used as evaluation metrics to quantitatively measure the performance of the proposed system. In object localization and classification, three commonly cited metrics, namely precision, recall, and mean average precision (mAP), were used to assess the performance of the trained model. Precision and recall are commonly used metrics for evaluating classification models. Precision determines the percentage of relevant cases among those found, while recall calculates the total sum with the total of related instances retrieved [23]. The performance indicators considered for the trained model's efficiency and reliability are calculated using Equations (1)–(3) and are indicated as follows:

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}} \quad (1)$$

$$\text{Recall} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}} \quad (2)$$

$$\text{AP} = \frac{1}{11} \sum_{\text{Recall}} \text{Precision}(\text{Recall}_i) \quad (3)$$

The general definition of average precision (AP) is the area under the precision–recall curve, and mAP is then calculated as the average of AP values across different classes or instances, providing an overall measure of performance for a detection model [99]. The mAP measures the ability of a model to accurately predict bounding boxes at specific confidence levels, such as mAP@0.5. The confidence values represent the intersection over union (IoU); the percentage of overlap between a predicted box and the ground truth bounding box is typically calculated at a threshold of 0.5–1. IoU. In this case, the IoU confidence value is set to 0.5, and the average for 11 points interpolated AP is calculated. The performance of the trained scaled YOLOv4 model is measured using TensorBoard, which visualizes the TensorFlow graph and plots the quantitative metrics along with additional data such as images. Figure 8 displays the precision, recall, and mAP@0.5 of the scaled YOLOv4 model trained on the validation set of the construction waste dataset.

The x-axis of Figure 8a,b represents the training iterations, while the y-axis of Figure 8a represents precision and the y-axis of Figure 8b represents recall. During the experiment, the observed precision was 73% at approximately 500 iterations, the recall was 86%, and mAP was noted as 83%. However, the performance of the model decreased by up to five percent for the testing set, for example, the precision was observed at 69%, recall at 83%, and mAP at 79%. The program automatically tests these indicators against the validation dataset to select the best model based on performance.

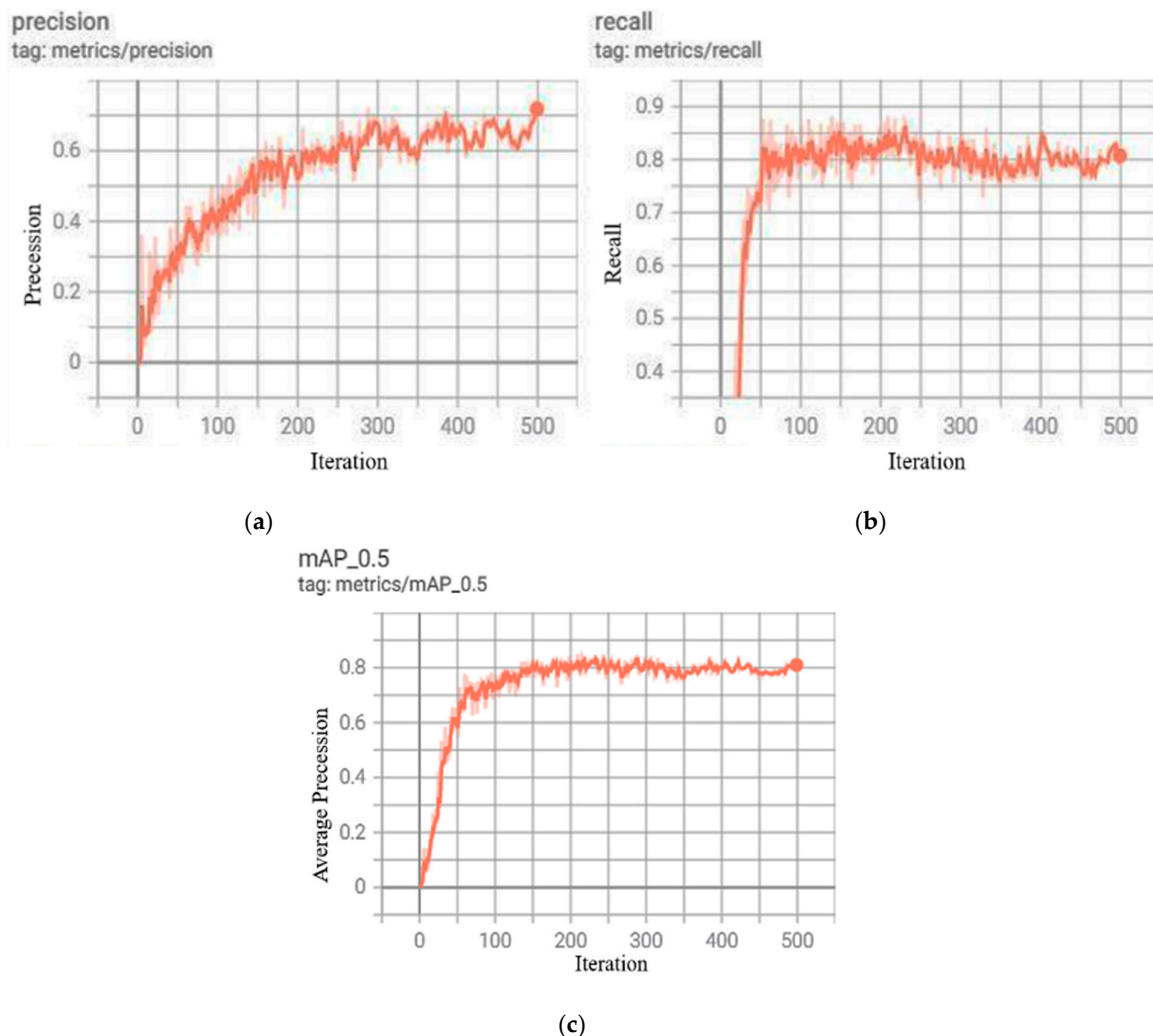


Figure 8. Performance graphs of scaled YOLOv4 for 500 iterations; (a) precision, (b) recall, and (c) mAP.

The developed model was evaluated on the test dataset to generate predictions using scaled YOLOv4 for five classes: brick, concrete, mixed, steel, and timber wastes. The best model was selected from the trained models based on its performance against the validation dataset. The chosen model was then applied to the test dataset, and the results are depicted in Figure 9. Accordingly, the false predictions and missed predictions made by the proposed system are depicted in Figure 10. In Figure 10a, an excavator can be seen demolishing a house, resulting in the generation of mixed waste. However, the system fails to correctly identify this category. This could be due to the relatively higher amount of concrete waste present in the image compared to other types of waste. Similarly, Figure 10b shows waste that should be categorized as timber waste; however, the system does not detect any results for this image. This could be attributed to the background wall having the same color as the waste, making it appear like a painted sketch on the wall. To address

these limitations, future improvements could involve expanding the image dataset with more high-quality images and synthetic data.



Figure 9. Examples of true positive prediction brick, concrete, steel, and timber wastes.

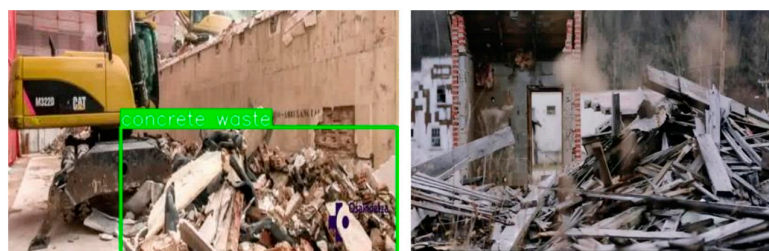


Figure 10. Examples of (left) false prediction and (right) miss detection.

6.4. Deep Learning Model Deployment for Periodic Checking of Safety Rules

With the recent advancements in robots and drone technologies, CV could be a game-changer for the construction industry. This paper aims to adopt these technologies to ensure compliance with construction safety rules by monitoring and verifying adherence to these rules periodically. To demonstrate the feasibility of this approach, a simple case study based on OSHA rules for monitoring construction waste is selected. An object detection model has been trained to recognize five classes of waste: brick, concrete, steel, timber, and mixed. The development procedure is divided into three parts: (1) streaming the video to a local server, (2) running an object detection model on the streaming video, and (3) displaying and reporting unsafe conditions related to construction waste. Figure 11 illustrates the complete procedure, from data collection to reporting the results.

Owing to the nonavailability of a drone or a robot dog, the proposed concept has been validated using the IP camera application in smartphones. A smartphone (Samsung S8+) was used, leveraging the ability to send live-stream videos. This video stream was then accessed frame-by-frame programmatically, using OpenCV libraries in Python. The process was initiated by executing a Python code on the server (local machine) to receive the live stream from the IP camera. The smartphone needed a connection with the local machine over a Wi-Fi network. Creating a Wi-Fi hotspot is an excellent option to connect the IP camera and local machine, allowing minimum lagged stream. However, this option was ignored because the smartphone was connected to the same Wi-Fi network. The local machine IP was configured after establishing a connection between the IP camera and the local machine. The IP camera then successfully forwarded the live stream to the server. Once the live feed of the IP camera is programmatically accessed, any deep learning framework can be applied to the live feed to display the inferences. The weight file of the trained model was imported into the server, and a Python code designed to apply the deep learning framework on the live feed was executed. The detection model trained on the scaled YOLOv4 framework was executed frame-by-frame and was displayed on the main window.

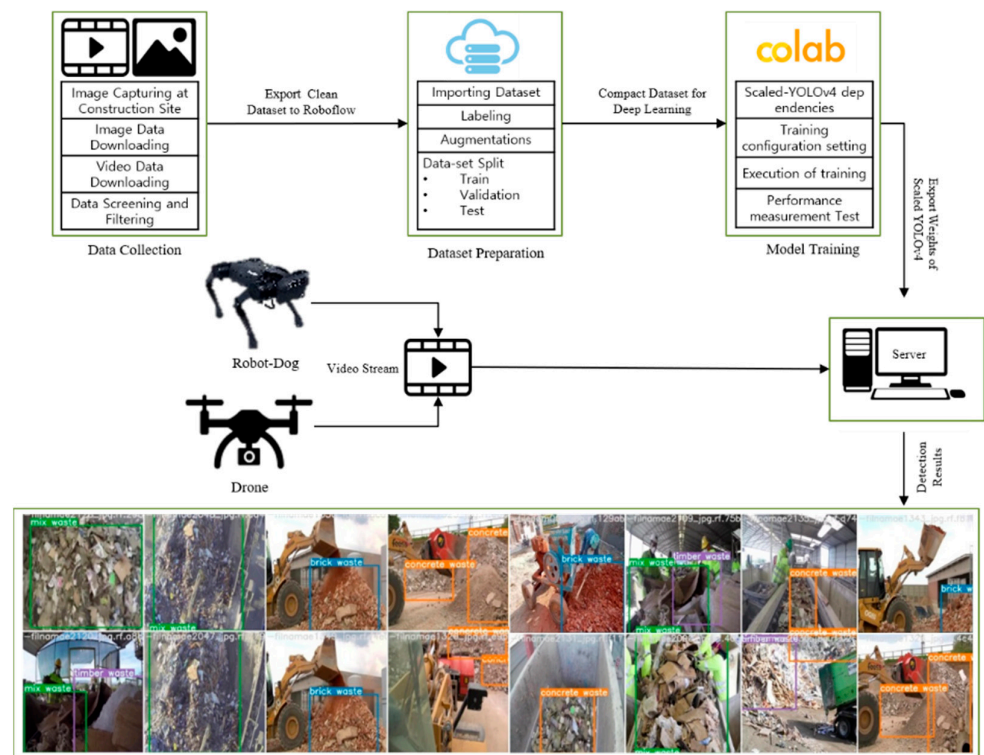


Figure 11. Framework of a CV-based construction waste detection system for safety rule compliance at specific time intervals.

7. CV-Based Safety Rule Compliance Monitoring during Work Execution

7.1. Motivation for Selecting Mobile Scaffolding as an Example

Falls from elevated working platforms, scaffolds, ladders, and roof edges account for 60% of the total fatalities [100]. Various forms of scaffolds, such as mobile, manufactured frame, pump jack, needle beam, and horse scaffolds, are used on construction sites. Extensive research has been conducted to develop failure mode detection systems that consider the structural strength of scaffolding. These approaches include the use of machine learning [82], cyber-physical system concepts [101], and strain sensing methods [102]. However, the safety of mobile scaffolds has not been fully explored, and accidents related to mobile scaffolds remain a significant concern. In 2019, according to statistics released by the Korean Occupational Safety and Health Agency, there were 488 deaths associated with scaffolding work, with 45 of them specifically related to activities involving mobile scaffolds [103].

Avoiding safety best practices and deviating from predetermined methods of executing tasks may increase the risk of accidents. The utilization of modern techniques, such as vision intelligence [5], smart sensors [104], and augmented reality [105], holds great promise for proactively preventing accidents and automatically identifying risks. Computer vision (CV) has recently demonstrated significant potential in enhancing the efficiency of various tasks in construction safety monitoring [10,18,25,106], such as equipment tracking and defect detection. However, few studies have focused specifically on mobile scaffolds' safety and health. For instance, one study examined the hourly risk rate of working on mobile scaffolds [107]. Manual safety monitoring of mobile scaffolds, particularly for checking outriggers in large projects, is impractical, resource-intensive, error-prone, and time-consuming in practice. Despite significant advancements in construction safety monitoring, there are still many unexplored areas. Therefore, this study aims to address the detection of unsafe behavior when workers are working with mobile scaffolds without outriggers.

7.2. System Development

7.2.1. Design Process

The system presented for the detection of unsafe behavior during work utilizes Mask R-CNN. In this system, frames from a video of the construction site are passed through a CNN to perform predictions and detect two classes: mobile scaffold and person. Following the object detection of these classes, an object correlation detection module is provided to reveal unsafe behavior. In this study, unsafe behavior is defined as when a person is working on the top of a mobile scaffold without outriggers installed. Conversely, when a person is working on scaffolding with outriggers, it is considered safe behavior. The developed system predicts whether the observed behavior in the frame is safe or unsafe by comparing specific criteria and ensuring overlap between the localized positions of a person and scaffolding. If unsafe behavior is detected, the system can provide a warning to the relevant authorities.

In recent years, several deep learning techniques related to region-based CNN have been introduced. Mask R-CNN combines Faster R-CNN and the fully convolutional neural network to detect objects in a scene and localize them with pixel-based masking, known as segmentation. The region proposal network serves as the initial step, generating random proposals by scanning the entire frame, producing class outputs, and predicting regression bounding boxes. In the second layer, convolutional features from the region proposal network and feature pyramid network layers are used to predict masks for each region of interest, which are then combined to achieve segmentation. Masks are applied on top of bounding boxes and class results to provide output predictions. Mask R-CNN has three output layers: the SoftMax classifier, bounding box regressor, and mask branch. During training, a multi-task loss is defined on each sampled region of interest (RoI), as shown in Equation (3), where L_{cls} represents the classification loss and L_{box} represents the bounding box loss; both are detailed in [108].

$$L = L_{cls} + L_{box} + L_{mask} \quad (4)$$

The mask branch of Mask R-CNN generates a Km^2 -dimension for each region of interest (RoI), encoding K binary masks of size $m \times m$, representing each of the K classes. The mask loss, denoted as L_{mask} , is calculated using a per-pixel sigmoid and the mean binary cross-entropy loss, as described in [56]. Here, L_{mask} is the mask loss defined in Equation (4) as a per-pixel sigmoid and the mean binary cross-entropy loss on the k^{th} mask presented by the following:

$$L_{mask} = -\frac{1}{m^2} \sum_{1 \leq i, j \leq m} \left[y_{ij} \log \hat{y}_{ij}^k + (1 - y_{ij}) \log (1 - \hat{y}_{ij}^k) \right] \quad (5)$$

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

7.2.2. Model Development

The proposed system and the process of extracting coordinates for mobile scaffolding safety monitoring during the construction phase are illustrated in Figure 12. Figure 12a demonstrates the corner coordinates $\mathbf{S}_{l,t} = (x_l^s, y_t^s)$ and $\mathbf{S}_{r,b} = (x_r^s, y_b^s)$ and centroid coordinate $\mathbf{S}_c = (x_c^s, y_c^s)$ of mobile scaffolds having no outriggers. Similarly, the coordinates for individuals are also explained; for instance, $\mathbf{P}_{l,t} = (x_l^p, y_t^p)$ and $\mathbf{P}_{r,b} = (x_r^p, y_b^p)$ for the corner coordinates and $\mathbf{P}_c = (x_c^p, y_c^p)$ for the centroid coordinate. The flow of the developed system is discussed in Figure 12b. In the image scene, the CNN module first identifies the entities such as persons, mobile scaffolds with outriggers, and mobile scaffolds without outriggers, as demonstrated in Figure 12a. The relationship between the detected person and mobile scaffold is then computed in the second phase to determine if a worker's position defines the behavior as safe/unsafe. The system identifies the worker near or on

the mobile scaffold; the system avoids the additional processing and states the frame as safe in the worker's absence. If the person class exists in the scene, the system searches for the mobile scaffold class equipped with outriggers, and thus the system indicates the behavior as safe; otherwise, the y-coordinates of the centroid are evaluated. For instance, if the y-coordinate of the centroid in the person's bounding box y_c^p is less than the y-coordinate of the centroid in the scaffolding bounding box y_c^s , then the person is assumed to be safe. To ensure the person is on top of the mobile scaffold, the system checks the corner coordinates, such as if x_r^p is less than x_r^s and x_l^p is greater than x_l^s , as stated in Figure 12b. If the criteria are met by the system, unsafe behavior will be identified; otherwise, the person would not be supposed to be on the mobile scaffold.

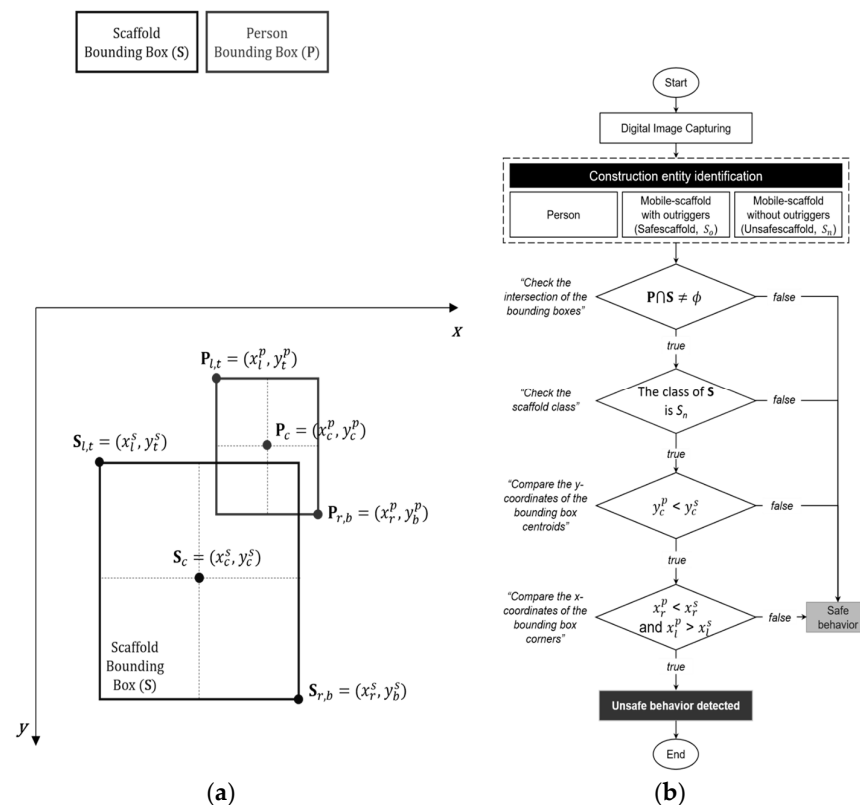


Figure 12. CV-based system for 'during work' based safety monitoring: (a) bounding box coordinates; (b) system flow.

7.3. Experimental Validation and Results

Experiments were conducted with an Intel i7-9700k processor, an NVIDIA RTX 2080Ti GPU, and 16 GB of RAM. The model was trained with a 0.001 learning rate for 100 epochs and was decreased by a factor of 10 until all layers were fine-tuned for additional 200 epochs with 300 steps per epoch. Also, weight decay and learning momentum were set as 0.0001 and 0.9, respectively.

The postprocessing module was used to determine the correlation between a person and the type of scaffolding to detect safe/unsafe behaviors. Figure 13a shows the detection of persons and scaffoldings equipped with outriggers. Conversely, Figure 13b reveals that the person working on the mobile scaffolds has no outriggers, and the system predicts the behavior as unsafe per the OSHA rule.

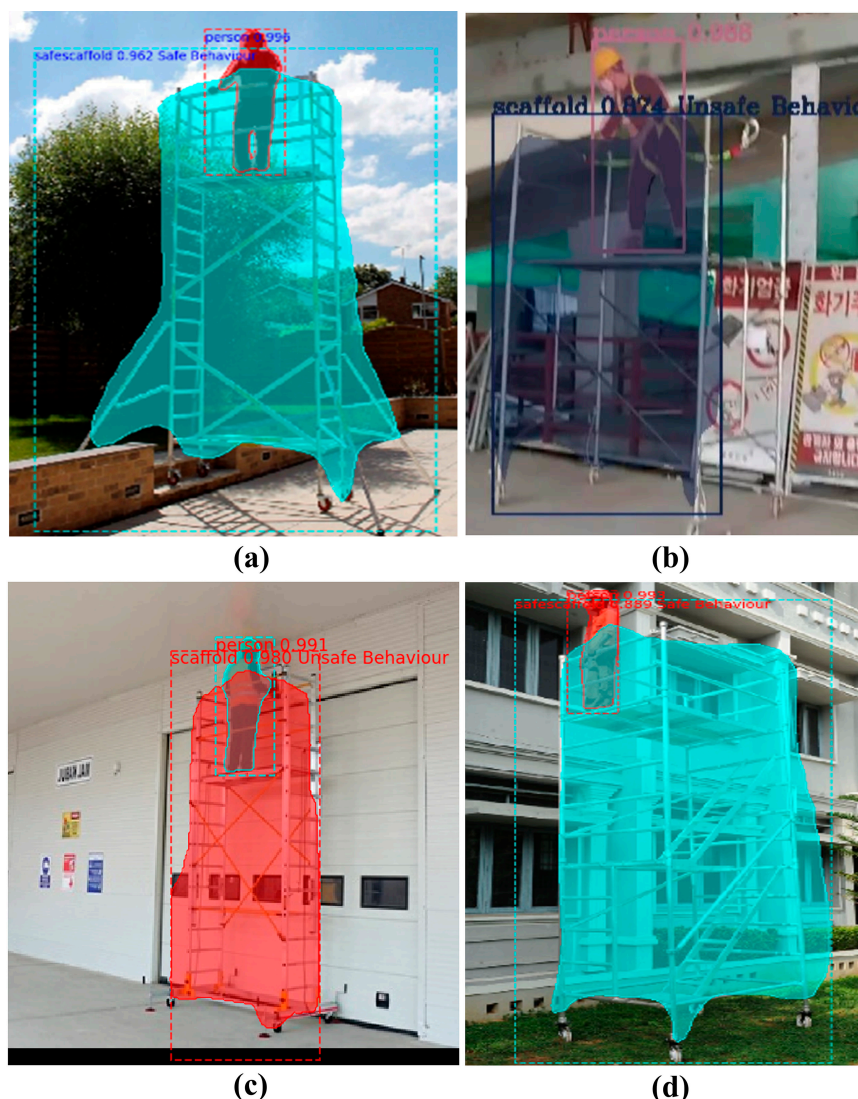


Figure 13. Results: (a) True positive prediction example: a person working on mobile scaffolds having outriggers. (b) True positive prediction example: a person working on mobile scaffolds without outriggers. (c) Examples of false prediction: persons and mobile scaffolds without outriggers. (d) Examples of false prediction: mobile scaffolds with outriggers and persons.

Similarly, false predictions can be seen in Figure 13c,d. The outriggers in Figure 13c are not the same as those used in training; therefore, this style of outriggers must be added to the dataset to enhance detection results. Accordingly, the person in Figure 13d is standing on a mobile scaffolding with no outriggers, which is unsafe behavior; however, the system classifies the behavior as safe. This could be due to the building's complex background, which includes windows and cantilever constructions that resemble pipes and lines. These weaknesses will be addressed in future research by expanding the dataset.

Evaluation of the Developed System

The developed system is assessed with four performance metrics on the validation set as this evaluation was carried out for the OCD block. The validation set (235 images) was grouped into two groups: (1) safe behaviors and (2) unsafe behaviors. The images containing person/s working on the top of the mobile scaffoldings with no outriggers were classified as unsafe behavior (Class-2), whereas the remaining images were kept as safe behavior (Class-1). Table 4 shows the findings of the performance indicators; the overall accuracy noted was 86%. The precision and recall calculated during the experiment

for Class-1 were 85% and 97%, respectively. For Class-2, the precision and recall were calculated to be 65% and 76%, respectively. Overall, the accuracy of the classification on the testing set was 86%.

Table 4. Performance indicators of the developed system.

Classification	Binary Classification	Precision (%)	Recall (%)
Safe (158)	0	85	97
Unsafe (77)	1	91	65

8. CV-Based System for Post-Work Construction Safety Compliance

8.1. Motivation for the Selection of Protruding Rebar Safety as a Case Study

Falls are a significant contributor to workers' injuries and death in construction sites [109]. Previous studies have shown that falls from heights account for 30% of fatalities and 48% of fatal and nonfatal injuries [23]. Nonfatal injuries owing to falls from heights are often debilitating and lead to permanent disability and long-term pain [12]. OSHA has identified four common causes of falls from height in construction sites: (1) unprotected edges, floor holes, and wall openings; (2) improper scaffold construction; (3) unguarded protruding steel rebar; and (4) misuse of portable ladders [110]. Extensive research has been conducted to explore the significance of safety and health, considering unprotected edges, floor, and wall openings, scaffold safety, and the safe use of ladders [12,39,95,111]; however, safety concerns associated with protruding steel rebar remain relatively under-researched. This is despite the fact that protruding steel bars create a severe threat to workers' safety when unguarded [110].

Reinforcement bars are an essential component of concrete work used in building construction and civil engineering projects. During the execution of concrete work in the construction of a permanent structure, reinforcement bars are among the initial components that need to be placed. Thus, they restrict movement and generate hazardous obstacles for those who prefer using shortcuts when going about their daily work tasks, such as traversing steel bars in the columns and beams that may contribute to trips and falls. Trips lead to falling accidents, and falls are assumed to be the primary cause of severe injuries on construction sites [93,94,112]. Thus, the safety of protruding rebar is a crucial issue on construction sites and needs to be addressed by leveraging recent emerging technology-based solutions. Safety practitioners have attempted to use advanced technologies, such as IoTs, BIM, and CV [5]. These attempts are notable but still suffer limitations. Therefore, the safety rule pertaining to protruding rebar safety in construction is extracted from the OSHA regulations owing to its cited significance.

1926.701(b) Reinforcing steel. All protruding reinforcing steel, onto and into which employees could fall, shall be guarded to eliminate the hazard of impalement.

This class is a set of safety rules that require strict compliance once a segment or permanent work is completed. According to the 'after work' rule set, 1926.701(b) shows that the reinforcing steel onto or into which persons can fall should be guarded. This set of safety rules can be ensured through the digital image data collected using a mobile camera or an action camera on the concerned persons once the task is completed.

8.2. Model Development

The training and algorithm optimization of the protruding rebar detection model after dataset preparation is presented in this section. The development of a deep learning model with a limited dataset is challenging; in this study, a popular deep learning technique is used, such as Faster R-CNN with Inception V2 meta-architecture. This technique was chosen because of easy implementation, less training time, and higher accuracy [113]. Moreover, Dai studied the object detection problem of resource-constrained devices employing the coco dataset. SSD, compared with different architectures such as deep lab-VGG, Inception V2, and MobileNet, have been tested. Similarly, Faster R-CNN (300 and 600) is evaluated

with these cited architectures. SSD is substantially faster but has lower accuracy; however, the Faster R-CNN with Inception V2 is slower in terms of real-time detection but has higher accuracy [114]. In addition, Galvez et al. compared MobileNet SSD-V1 with Faster R-CNN Inception V2; the results revealed that the former model is the best fit for real-time application, whereas the latter is appropriate for accurate detection [115]. Consequently, in detecting the rebar cap, it is better to have higher accuracy than real-time speed; thus, Faster R-CNN with Inception V2 is selected for the training process.

8.2.1. Faster R-CNN with Inception V2 Meta-Architecture

The Faster R-CNN consists of three key components: (1) convolutional layers for feature extraction, (2) region proposal network (RPN) for object detection, and (3) two fully connected layers to classify objects and generate their coordinates [116]. The distinction between Faster R-CNN and Fast R-CNN is the region proposal; the RPN is combined with the architecture in Faster R-CNN, unlike the external selective search in Fast R-CNN [115]. Faster R-CNN is derived from Fast R-CNN using a region proposer, which offers faster computation than the predecessor and SSD [117].

The Faster R-CNN uses object proposals, convolution, and max-pooling layers to produce a convolution feature map. An RoI extracts the feature vector for each object proposal, which is then processed by a concatenation of fully connected layers resulting in two output layers. The first layer uses SoftMax probability to rate N object categories, while the second layer generates four real-valued numbers against each N object category, including the 'background' category. The RoI pooling layer uses max-pooling to alter each RoI feature in a slight feature map. An RoI is a rectangular window in the convolution feature map determined by four variables (r , c , h , w). The RoI max-pooling operates by dividing the RoI window into a grid ($H \times W$) of subwindows with an estimated size ($h/H \times w/W$). The max-pooling values are located inside the related output grid cell in the entire subwindows. The RoI pooling layer's output is fed to two fully connected layers, which determine the object's class using the object's coordinates with the linear and SoftMax activation functions. Faster R-CNN [118] has two core processes when performing object detection. A feature extractor model processes images, identifies RPN, and extracts category bounding box proposals using medium-level layers in parallel. These proposals are used to crop features from the medium feature map and predict category labels. Faster R-CNN avoids cropping proposals directly from the image and instead reruns the crops over the feature extractor, resulting in more replicated computations.

Inception V2 is a CNN with complex architecture. It uses convolution as input, with three filter sizes (1×1 , 3×3 , and 5×5) and max-pooling. The resulting outputs are concatenated and passed to the next inception module. The GoogleNet neural network uses nine similar inception modules to increase efficiency and reduce the expensive computations of deep learning networks, as explained in [119]. Additionally, the Inception V2 architecture improves computational speed by factorizing a 5×5 convolution into two 3×3 convolutions, as the latter is 2.78 times less expensive.

8.2.2. Training Implementation and Model Deployment

The model was trained on a server with an i7-9700k CPU, RTX 2080Ti GPU, and 16 GB RAM using a virtual environment created on Anaconda for Python. The dataset was formulated in Roboflow, and a pre-trained model was imported from GitHub. TensorFlow and its object detection API were used to train the Faster R-CNN inception model with a learning rate of 0.00002 for 300,000 epochs, and each trained result was recorded every 50 steps. Image augmentation techniques, including flip augmentation, were applied during training. The resulting model was exported as an inference graph (.pb) and a set of checkpoints, which were then converted to TFLite for deployment on the mobile application.

8.2.3. Model Evaluation and Results

In this section, to quantitatively measure the performance of the proposed system, two performance indicators, for instance, precision and recall, are used as evaluation metrics. The performance metrics for precision and recall are calculated using Equations (1) and (2). In this study, the IoU confidence value used is 0.5. Performance measurement of the trained Faster R-CNN model has been visualized using TensorBoard. The trained model is evaluated on the validation dataset. Consequently, the TFRecord file of detections on the validation dataset is exported and then compared with the ground truth to compute a recall and precision score at IoU = 0.5. Table 5 shows the precision and recall at 0.5 IoU. On the validation dataset, the results obtained were 48% and 81% for precision and recall, respectively. Moreover, the observed precision and recall for the test set were 44% and 76% for ‘steel without cap.’ However, the noted precision and recall during the experiment for ‘steel with cap’ were 50% and 60%, respectively. The precision and recall calculation are listed in Table 5.

Table 5. Precision and recall calculation at 0.5 IoU on the validation set.

Classes	Category	Precision	Recall
0	Steel without cap	0.44	0.77
1	Steel with cap	0.51	0.61

Overall, owing to the complex nature of the construction site, the initial results of the developed model are quite satisfactory and achieved good precision and recall across all classes. Figure 14 shows the loss against each epoch until 240 k epochs. Because the loss is continuously dropping, the training was increased from 200 k to 300 k. The designed Python code automatically tests the training and validation loss indicators to choose the best model, and an optimum best model is extracted based on performance. Here, the model weight generated at 232 k epoch with a total loss of 0.1334 is extracted.

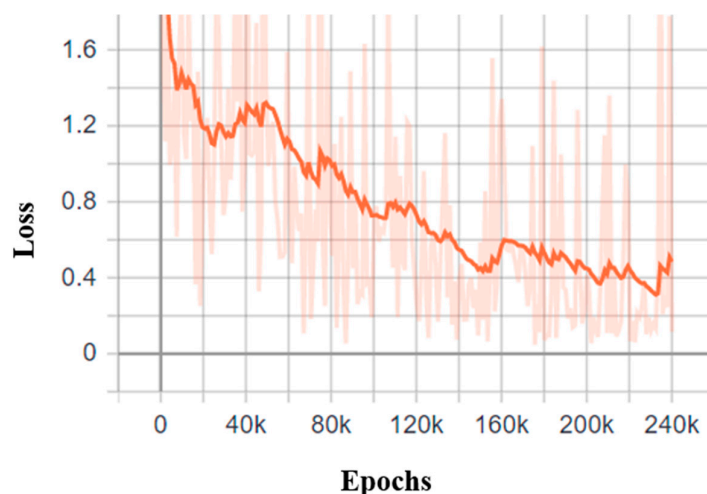


Figure 14. Training/validation total loss during the training process.

The developed model is assessed using a test set to substantiate the performance of the optimized Faster R-CNN algorithm. This section presents the detection results for two classes: ‘steel with cap’ and ‘steel without cap’. Figure 15b,d indicates the true-positive predictions for the two classes. Figure 15a,c shows the actual image from the testing dataset. Figure 16b,c shows the actual prediction using the trained model. Accordingly, the false predictions and missed predictions using the proposed system are shown in Figure 16, which shows a complex picture of the site. The system fails to recognize all the classes in the image. This may be attributed to the smaller size of the classes. The weaknesses

identified in this case study will be addressed in a future study by expanding the dataset by adding higher-quality images and synthetic data.

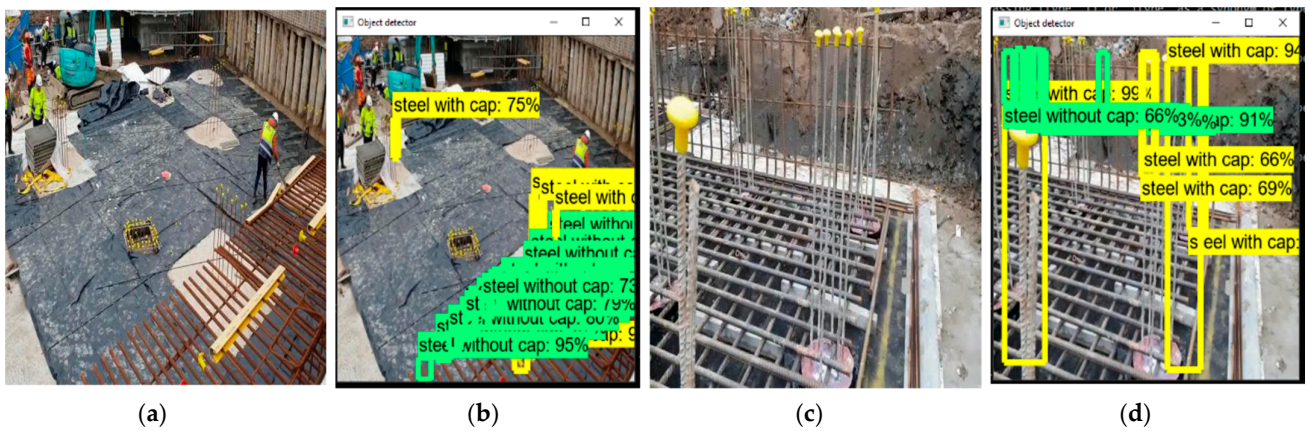


Figure 15. Examples of true positive prediction for steel with cap and steel without cap.



Figure 16. Example results of (a) false prediction and (b) miss detection.

8.2.4. Android Implementations

To deploy the model on the embedded device, the developed Faster R-CNN model is converted to TFLite format using a two-step procedure: (1) conversion of the .h5 file into the frozen graph.pb file format and (2) the full-scale TensorFlow model.pb into the TFLite model. TensorFlow Lite is the framework designed to visualize inference on small devices, thus avoiding a round-trip data computation to and from the server. This allows real-time detection and function without an Internet connection [120]. This framework is not used to train the models; rather, it is used with the model during production. An Android-supported smartphone was selected to execute the converted model. The snapshots of the Android-based prototype for visible defect inspection are shown in Figure 17.

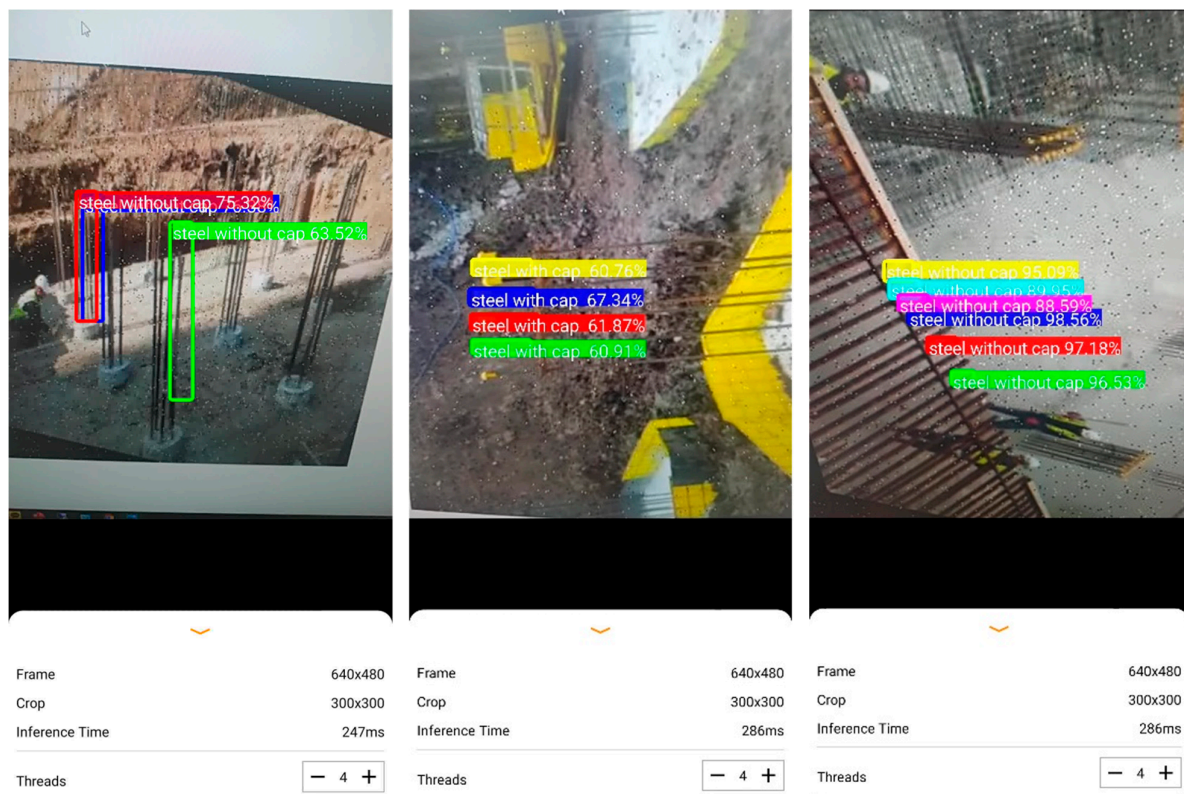


Figure 17. Android-based prototype development for visible defect inspection with a specific interval.

8.3. Firebase Database to Record the Visible Defects

Following the detection of unsafe conditions such as missing planks and guardrails, as well as the incorrect installation of planks across the video frames of the smartphone, the next step involves recording the image frame as evidence to assess the safety performance. Every detected class in a video stream has an ID that can be traced across the frames. If-else conditions have been added in the back-end programming to record the frame once the model detects cited classes, excluding the person class.

A cloud-based real-time Firebase database is used to record the detected results. The real-time Firebase database is a cloud-hosted NoSQL database that supports users to sync their data in real-time. Users can sign up/sign in using the authentication module in the Firebase database. Cloud storage for Firebase is a powerful, simple, and cost-effective object storage service built for Google scale. The Firebase SDKs for cloud storage add Google security to upload and download data in the developed application. First, a project on the Firebase database to enable connection with an Android application is established. Users can sign up through an application after the registered authentication procedure by Firebase. A trial application developed in the Android Studio is executed on a video with visible defects, such as missing guard rails and planks. The Android application automatically communicates with the Firebase and records the cited defects of the concerned frame as a snapshot. Afterward, the system uploads it into the specified folder on the Firebase cloud storage.

9. Discussion

The aim of this study was to address the challenges faced by traditional safety management approaches at construction sites and explore the potential of computer vision (CV) technology to enhance safety monitoring practices. The findings of this research make a significant contribution to the field of safety management, offering practical and theoretical implications.

This study explored the potential of CV technology in enhancing construction safety monitoring. CV has shown promising advancements in project progress and productivity analysis, but its application to life-threatening safety rule monitoring remained limited. This study fills this gap by proposing a comprehensive framework for work-stage-based rule compliance monitoring using CV technology. The prototype development section demonstrated the feasibility of the proposed classification method. The creation of novel datasets and the deployment of object detection and classification models on edge devices showcased the potential practical applications of CV in safety monitoring. For instance, the integration of IP cameras with deep learning techniques enables the remote detection of unsafe conditions with specific intervals, thereby enhancing safety monitoring capabilities at construction sites.

Conversely, this study presented positive results, but there are certain limitations that need to be acknowledged. The research focused on specific safety aspects related to scaffolding, construction waste, mobile scaffolding, and protruding steel rebar to validate the framework. Future studies should explore additional critical safety aspects and expand the dataset to cover a broader range of construction scenarios to enhance the feasibility and practicality of the proposed approach. Additionally, the effectiveness of the proposed CV-based safety monitoring system should be tested and validated on a larger scale in real-world construction projects. Conventional monitoring systems utilized CCTV cameras, which are installed at construction sites. Existing and proposed CV-based methods for safety management access these CCTV cameras to apply algorithms for safety rule compliance. Furthermore, it is crucial to consider the technical aspects and limitations of CV technology, such as the use of stereo cameras, real sense cameras, and Lidar to provide depth information and improve the detection of certain safety hazards. Although, these devices are very expensive and cannot be used for ordinary monitoring. Nevertheless, future research should explore the integration of such advanced devices into the safety monitoring framework to enhance its accuracy and reliability.

One of the main challenges posed by a limited dataset is the potential for overfitting. With a smaller number of samples, the risk of the model memorizing the data rather than learning meaningful patterns increases. This can lead to reduced generalization performance when the model encounters new, unseen data. Furthermore, the limited dataset used in this study might not cover all possible variations and edge cases present in the real-world environment. As a result, the model's performance might be compromised in situations not well-represented in the dataset. Collecting more diverse data or collaborating with other research groups to merge datasets could address this limitation. Also, careful model selection, regularization methods, and transfer learning can improve the model's performance.

This study has provided a foundational framework for CV-based safety monitoring. However, it is essential to keep in mind that construction sites are dynamic environments with unique challenges. As construction practices evolve, new safety rules and considerations may arise. Therefore, ongoing research and development efforts are necessary to ensure that the CV-based safety monitoring system remains adaptive and up to date with the changing construction industry.

10. Conclusions and Future Work

This study has successfully demonstrated the potential of CV-based systems to complement traditional safety monitoring approaches. Through the rigorous analysis and validation of OSHA safety rules, a vision intelligence-powered work-stage-based system was developed, providing a structured and adaptive framework for safety compliance monitoring during different construction phases. The findings of this study contribute to the body of knowledge on safety management and have important practical and theoretical implications. The OSHA safety rules are analyzed and validated for a vision intelligence-powered work-stage-based system.

- The rules are classified into two layers. In the first layer, five groups are revealed in the analysis: (1) procurement phase (8.54%), (2) preconstruction phase (0.31%), (3) construction phase (52.67%), (4) general rules (15.35%), and (5) terminologies and definitions (23.15%).
- In the second layer, the construction phase-related rules are further classified into four classes: before work, with intervals, during work, and after work. The percentages of each class are 32.95, 8.05, 41.59, and 13.09, respectively. In addition to these four classes, 4.3% of the safety rule set is categorized into general management-related rules.
- The prototype development phase validated the proposed classification, exhibiting the practical applicability of CV models for detecting various safety hazards. The prototypes are developed on the construction phase-related rules that include the development:
 - A novel dataset of scaffold systems with visible defects was developed, and a lightweight object detection model was trained using YOLOv5. A Firebase database was integrated with an Android application to record unsafe conditions for safety practitioners and decision-makers automatically.
 - An object classification model was deployed on remote devices to detect construction waste. An IP camera was integrated with deep learning techniques to automatically and remotely detect unsafe conditions.
 - A dataset for three classes was created to detect unsafe behaviors related to mobile scaffolding using CCTV for real-time monitoring of safety rule compliance.
 - A dataset of protruding steel rebar with and without caps was created. The collected dataset was processed using a double-shot detector, Faster R-CNN, and the best-fit model was converted into a TFLite model for edge computing.

In the future, the author intends to extend this research in two directions: theoretical analysis and technical prototype optimization. Work-stage-based safety rules will be compared against devices suitable for capturing specific scenes, considering technical details such as stereo cameras for detecting depth information. The established dataset will be enriched with novel CNN-based methods to enhance safety monitoring at job sites, producing an extensive dataset to improve the practical application of a CV-based construction safety monitoring system. Another framework based on accident case databases will be designed to develop a practical system capable of dealing with real construction problems. The author also presents Faster R-CNN and Mask R-CNN as having good detection performance and plans to use double-stage detectors in future research.

Author Contributions: Conceptualization: N.K., D.L. and C.P.; methodology: N.K., S.F.A.Z. and J.Y.; data curation: N.K. and J.Y.; formal analysis: N.K., S.F.A.Z. and J.Y.; visualization, N.K., J.Y. and S.F.A.Z.; writing—original draft preparation: N.K.; writing—review and editing: S.F.A.Z., J.Y., D.L. and C.P.; supervision: D.L. and C.P.; project administration: C.P. and D.L., funding acquisition: C.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was conducted with the support of the “National R&D Project for Smart Construction Technology (No.23SMIP-A158708-04)” funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure, and Transport, and managed by the Korea Expressway Corporation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sunindijo, R.Y.; Zou, P.X.W. Political Skill for Developing Construction Safety Climate. *J. Constr. Eng. Manag.* **2012**, *138*, 605–612. [[CrossRef](#)]
2. Le, Q.T.; Pedro, A.; Park, C.S. A Social Virtual Reality Based Construction Safety Education System for Experiential Learning. *J. Intell. Robot. Syst. Theory Appl.* **2015**, *79*, 487–506. [[CrossRef](#)]
3. Lingard, H. Occupational Health and Safety in the Construction Industry. *Constr. Manag. Econ.* **2013**, *31*, 505–514. [[CrossRef](#)]
4. Schwatka, N.V.; Rosecrance, J.C. Safety Climate and Safety Behaviors in the Construction Industry: The Importance of Co-Workers Commitment to Safety. *Work* **2016**, *54*, 401–413. [[CrossRef](#)]
5. Park, C.; Lee, D.; Khan, N. An Analysis on Safety Risk Judgment Patterns Towards Computer Vision Based Construction Safety Management. In Proceedings of the Creative Construction e-Conference 2020, Online, 28 June–1 July 2020; Volume 52. [[CrossRef](#)]
6. Corso, P.; Finkelstein, E.; Miller, T.; Fiebelkorn, I.; Zaloshnja, E. Incidence and Lifetime Costs of Injuries in the United States. *Inj. Prev.* **2006**, *12*, 212–218. [[CrossRef](#)]
7. Jeelani, I.; Albert, A.; Han, K. Improving Safety Performance in Construction Using Eye-Tracking, Visual Data Analytics, and Virtual Reality. In Proceedings of the Construction Research Congress 2020: Safety, Workforce, and Education—Selected Papers from the Construction Research Congress 2020, Tempe, AZ, USA, 8–10 March 2020; American Society of Civil Engineers (ASCE): Atlanta, GA, USA, 2020; pp. 395–404.
8. Fang, W.; Zhong, B.; Zhao, N.; Love, P.E.D.; Luo, H.; Xue, J.; Xu, S. A Deep Learning-Based Approach for Mitigating Falls from Height with Computer Vision: Convolutional Neural Network. *Adv. Eng. Inform.* **2019**, *39*, 170–177. [[CrossRef](#)]
9. Khan, N.; Ali, A.K.; Skibniewski, M.J.; Lee, D.Y.; Park, C. Excavation Safety Modeling Approach Using BIM and VPL. *Adv. Civ. Eng.* **2019**, *2019*, 1515808. [[CrossRef](#)]
10. Golparvar-Fard, M.; Peña-Mora, F.; Savarese, S. Automated Progress Monitoring Using Unordered Daily Construction Photographs and IFC-Based Building Information Models. *J. Comput. Civ. Eng.* **2015**, *29*, 04014025. [[CrossRef](#)]
11. Roberts, D.; Golparvar-Fard, M. End-to-End Vision-Based Detection, Tracking and Activity Analysis of Earthmoving Equipment Filmed at Ground Level. *Autom. Constr.* **2019**, *105*, 102811. [[CrossRef](#)]
12. Khan, N.; Saleem, M.R.; Lee, D.; Park, M.W.; Park, C. Utilizing Safety Rule Correlation for Mobile Scaffolds Monitoring Leveraging Deep Convolution Neural Networks. *Comput. Ind.* **2021**, *129*, 103448. [[CrossRef](#)]
13. Liu, P.; Chi, H.-L.; Li, X.; Guo, J. Effects of Dataset Characteristics on the Performance of Fatigue Detection for Crane Operators Using Hybrid Deep Neural Networks. *Autom. Constr.* **2021**, *132*, 103901. [[CrossRef](#)]
14. Zhang, S.; Sulankivi, K.; Kiviniemi, M.; Romo, I.; Eastman, C.M.; Teizer, J. BIM-Based Fall Hazard Identification and Prevention in Construction Safety Planning. *Saf. Sci.* **2015**, *72*, 31–45. [[CrossRef](#)]
15. Tang, S.; Golparvar-Fard, M. Joint Reasoning of Visual and Text Data for Safety Hazard Recognition. *Comput. Civ. Eng.* **2017**, *3*, 326–334. [[CrossRef](#)]
16. Khan, N.; Lee, D.; Baek, C.; Park, C.S. Converging Technologies for Safety Planning and Inspection Information System of Portable Firefighting Equipment. *IEEE Access* **2020**, *8*, 211173–211188. [[CrossRef](#)]
17. Chansik, P.; Doyeop, L.; Numan, K. *An Analysis on Safety Risk Judgment Patterns towards Computer Vision Based Construction Safety Management*; Periodica Polytechnica; Budapest University of Technology and Economics: Budapest, Hungary, 2020; pp. 31–38.
18. Seo, J.; Han, S.; Lee, S.; Kim, H. Computer Vision Techniques for Construction Safety and Health Monitoring. *Adv. Eng. Inform.* **2015**, *29*, 239–251. [[CrossRef](#)]
19. Reese, C.D.; Eidson, J.V. *Handbook of OSHA Construction Safety and Health*; CRC Press Taylor & Francis Group: Boca Raton, FL, USA, 2006; ISBN 9780849365461.
20. Cayet, T.; Rosental, P.A.; Thébaud-Sorger, M. How international organisations compete: Occupational safety and health at the ILO, a diplomacy of expertise. *J. Mod. Eur. Hist.* **2009**, *7*, 174–196. [[CrossRef](#)]
21. Liu, M.; Han, S.; Lee, S. Tracking-Based 3D Human Skeleton Extraction from Stereo Video Camera toward an on-Site Safety and Ergonomic Analysis. *Constr. Innov.* **2016**, *16*, 348–367. [[CrossRef](#)]
22. Kim, H.; Kim, H.; Hong, Y.W.; Byun, H. Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning. *J. Comput. Civ. Eng.* **2017**, *32*, 04017082. [[CrossRef](#)]
23. Fang, W.; Ding, L.; Luo, H.; Love, P.E.D. Falls from Heights: A Computer Vision-Based Approach for Safety Harness Detection. *Autom. Constr.* **2018**, *91*, 53–61. [[CrossRef](#)]
24. Fang, Q.; Li, H.; Luo, X.; Ding, L.; Luo, H.; Li, C. Computer Vision Aided Inspection on Falling Prevention Measures for Steeplejacks in an Aerial Environment. *Autom. Constr.* **2018**, *93*, 148–164. [[CrossRef](#)]
25. Suddin, K.; Ani, M.; Ismail, A.; Ibrahim, M. Investigation the Safety, Health and Environment (SHE) Protection in Construction Area. *Int. Res. J. Eng. Technol.* **2015**, *2*, 624–636.
26. Ju, C.; Rowlinson, S. Institutional Determinants of Construction Safety Management Strategies of Contractors in Hong Kong. *Constr. Manag. Econ.* **2014**, *32*, 725–736. [[CrossRef](#)]
27. Ozumba, A.O.U.; Shakantu, W. *Adaptation: A Lens for Viewing Technology Transfer in Construction Site Management*. In *Product Design*; IntechOpen: London, UK, 2020.
28. Skibniewski, M.J. Information Technology Applications in Construction Safety Assurance. *J. Civ. Eng. Manag.* **2014**, *20*, 778–794. [[CrossRef](#)]

29. Guo, H.; Yu, Y.; Skitmore, M. Visualization Technology-Based Construction Safety Management: A Review. *Autom. Constr.* **2017**, *73*, 135–144. [[CrossRef](#)]
30. Guo, B.H.W.; Yiu, T.W.; González, V.A. Predicting Safety Behavior in the Construction Industry: Development and Test of an Integrative Model. *Saf. Sci.* **2016**, *84*, 1–11. [[CrossRef](#)]
31. Zhang, S.; Frank, B.; Jochen, T. Ontology-Based Semantic Modeling of Construction Safety Knowledge: Towards Automated Safety Planning for Job Hazard Analysis (JHA). *Autom. Constr.* **2015**, *52*, 29–41. [[CrossRef](#)]
32. Rossi, A.; Vila, Y.; Lusiani, F.; Barsotti, L.; Sani, L.; Ceccarelli, P.; Lanzetta, M. Embedded Smart Sensor Device in Construction Site Machinery. *Comput. Ind.* **2019**, *108*, 12–20. [[CrossRef](#)]
33. Xu, M.; Nie, X.; Li, H.; Cheng, J.C.P.; Mei, Z. Smart Construction Sites: A Promising Approach to Improving on-Site HSE Management Performance. *J. Build. Eng.* **2022**, *49*, 104007. [[CrossRef](#)]
34. Rao, A.S.; Radanovic, M.; Liu, Y.; Hu, S.; Fang, Y.; Khoshelham, K.; Palaniswami, M.; Ngo, T. Real-Time Monitoring of Construction Sites: Sensors, Methods, and Applications. *Autom. Constr.* **2022**, *136*, 104099. [[CrossRef](#)]
35. Katika, T.; Konstantinidis, F.K.; Papaioannou, T.; Dadoukis, A.; Bolierakis, S.N.; Tsimiklis, G.; Amditis, A. Exploiting Mixed Reality in a Next-Generation IoT Ecosystem of a Construction Site. In Proceedings of the 2022 IEEE International Conference on Imaging Systems and Techniques (IST), Taiwan, China, 21–23 June 2022; pp. 1–6.
36. Arshad, S.; Akinade, O.; Bilal, M.; Bello, S. Computer Vision and IoT Research Landscape for Health and Safety Management on Construction Sites. *J. Build. Eng.* **2023**, *76*, 107049. [[CrossRef](#)]
37. Fang, W.; Love, P.E.D.; Luo, H.; Ding, L. Computer Vision for Behaviour-Based Safety in Construction: A Review and Future Directions. *Adv. Eng. Inform.* **2020**, *43*, 100980. [[CrossRef](#)]
38. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
39. Fang, W.; Ma, L.; Love, P.E.D.; Luo, H.; Ding, L.; Zhou, A. Knowledge Graph for Identifying Hazards on Construction Sites: Integrating Computer Vision with Ontology. *Autom. Constr.* **2020**, *119*, 103310. [[CrossRef](#)]
40. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015.
41. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
42. Terven, J.; Cordova-Esparza, D. A Comprehensive Review of YOLO: From YOLOv1 to YOLOv8 and Beyond. *arXiv* **2023**, arXiv:2304.00501.
43. Nath, N.D.; Behzadan, A.H.; Paal, S.G. Deep Learning for Site Safety: Real-Time Detection of Personal Protective Equipment. *Autom. Constr.* **2020**, *112*, 103085. [[CrossRef](#)]
44. Fang, Q.; Li, H.; Luo, X.; Ding, L.; Luo, H.; Rose, T.M.; An, W. Detecting Non-Hardhat-Use by a Deep Learning Method from Far-Field Surveillance Videos. *Autom. Constr.* **2018**, *85*, 1–9. [[CrossRef](#)]
45. Shanti, M.Z.; Cho, C.S.; de Soto, B.G.; Byon, Y.J.; Yeun, C.Y.; Kim, T.Y. Real-Time Monitoring of Work-at-Height Safety Hazards in Construction Sites Using Drones and Deep Learning. *J. Saf. Res.* **2022**, *83*, 364–370. [[CrossRef](#)]
46. Hung, H.M.; Lan, L.T.; Hong, H.S. A Deep Learning-Based Method for Real-Time Personal Protective Detection. *Quy Don Tech. Univ.-Sect. Inf. Commun. Technol.* **2019**, *13*, 23–34.
47. Huang, L.; Fu, Q.; He, M.; Jiang, D.; Hao, Z. Detection Algorithm of Safety Helmet Wearing Based on Deep Learning. *Concurr. Comput.* **2021**, *33*, e6234. [[CrossRef](#)]
48. Wang, Z.; Wu, Y.; Yang, L.; Thirunavukarasu, A.; Evison, C.; Zhao, Y. Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. *Sensors* **2021**, *21*, 3478. [[CrossRef](#)] [[PubMed](#)]
49. Han, K.; Zeng, X. Deep Learning-Based Workers Safety Helmet Wearing Detection on Construction Sites Using Multi-Scale Features. *IEEE Access* **2021**, *10*, 718–729. [[CrossRef](#)]
50. Alateeq, M.M.; Fathimathul, F.R.; Ali, M.A.S. Construction Site Hazards Identification Using Deep Learning and Computer Vision. *Sustainability* **2023**, *15*, 2358. [[CrossRef](#)]
51. Arabi, S.; Haghghat, A.; Sharma, A. A Deep-Learning-Based Computer Vision Solution for Construction Vehicle Detection. *Comput.-Aided Civ. Infrastruct. Eng.* **2020**, *35*, 753–767. [[CrossRef](#)]
52. Wang, M.; Wong, P.; Luo, H.; Kumar, S.; Delhi, V.; Cheng, J. Predicting Safety Hazards among Construction Workers and Equipment Using Computer Vision and Deep Learning Techniques. In Proceedings of the 36th International Symposium on Automation and Robotics in Construction, ISARC 2019, Banff, AB, Canada, 21–24 May 2019; International Association for Automation and Robotics in Construction (I.A.A.R.C): Edinburgh, UK, 2019; pp. 399–406.
53. Anjum, S.; Khan, N.; Khalid, R.; Khan, M.; Lee, D.; Park, C. Fall Prevention from Ladders Utilizing a Deep Learning-Based Height Assessment Method. *IEEE Access* **2022**, *10*, 36725–36742. [[CrossRef](#)]
54. Shin, Y.-S.; Kim, J. A Vision-Based Collision Monitoring System for Proximity of Construction Workers to Trucks Enhanced by Posture-Dependent Perception and Truck Bodies' Occupied Space. *Sustainability* **2022**, *14*, 7934. [[CrossRef](#)]
55. Li, H.; Qiu, J.; Yu, K.; Yan, K.; Li, Q.; Yang, Y.; Chang, R. Fast Safety Distance Warning Framework for Proximity Detection Based on Oriented Object Detection and Pinhole Model. *Measurement* **2023**, *209*, 112509. [[CrossRef](#)]
56. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision 2017, Venice, Italy, 22–29 October 2017; pp. 2980–2988. [[CrossRef](#)]

57. Hafiz, A.M.; Bhat, G.M. A Survey on Instance Segmentation: State of the Art. *Int. J. Multimed. Inf. Retr.* **2020**, *9*, 171–189. [CrossRef]
58. Xiao, B.; Xiao, H.; Wang, J.; Chen, Y. Vision-Based Method for Tracking Workers by Integrating Deep Learning Instance Segmentation in off-Site Construction. *Autom. Constr.* **2022**, *136*, 104148. [CrossRef]
59. Kang, K.S.; Cho, Y.W.; Jin, K.H.; Kim, Y.B.; Ryu, H.G. Application of One-Stage Instance Segmentation with Weather Conditions in Surveillance Cameras at Construction Sites. *Autom. Constr.* **2022**, *133*, 104034. [CrossRef]
60. Bang, S.; Hong, Y.; Kim, H. Proactive Proximity Monitoring with Instance Segmentation and Unmanned Aerial Vehicle-Acquired Video-Frame Prediction. *Comput.-Aided Civ. Infrastruct. Eng.* **2021**, *36*, 800–816. [CrossRef]
61. Wang, H.; Ye, Z.; Wang, D.; Jiang, H.; Liu, P. Synthetic Datasets for Rebar Instance Segmentation Using Mask R-CNN. *Buildings* **2023**, *13*, 585. [CrossRef]
62. Mathur, S.; Jain, T. Segmenting Personal Protective Equipment Using Mask R-CNN. In Proceedings of the 2023 11th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), Jaipur, India, 10–11 February 2023; pp. 1–6.
63. Chen, S.; Dong, F.; Demachi, K. Hybrid Visual Information Analysis for On-Site Occupational Hazards Identification: A Case Study on Stairway Safety. *Saf. Sci.* **2023**, *159*, 106043. [CrossRef]
64. Soltani, M.M.; Zhu, Z.; Hammad, A. Framework for Location Data Fusion and Pose Estimation of Excavators Using Stereo Vision. *J. Comput. Civ. Eng.* **2018**, *32*, 04018045. [CrossRef]
65. Assadzadeh, A.; Arashpour, M.; Li, H.; Hosseini, R.; Elghaish, F.; Baduge, S. Excavator 3D Pose Estimation Using Deep Learning and Hybrid Datasets. *Adv. Eng. Inform.* **2023**, *55*, 101875. [CrossRef]
66. Wen, L.; Kim, D.; Liu, M.; Lee, S. 3D Excavator Pose Estimation Using Projection-Based Pose Optimization for Contact-Driven Hazard Monitoring. *J. Comput. Civ. Eng.* **2023**, *37*, 04022048. [CrossRef]
67. Tian, Z.; Yu, Y.; Xu, F.; Zhang, Z. Dynamic Hazardous Proximity Zone Design for Excavator Based on 3D Mechanical Arm Pose Estimation via Computer Vision. *J. Constr. Eng. Manag.* **2023**, *149*, 4023048. [CrossRef]
68. Zhao, J.; Cao, Y.; Xiang, Y. Pose Estimation Method for Construction Machine Based on Improved AlphaPose Model. *Eng. Constr. Archit. Manag.* **2022**. [CrossRef]
69. Pedro, A.; Chien, P.H.; Park, C.S. Towards a Competency-Based Vision for Construction Safety Education. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *143*, 012051. [CrossRef]
70. Zhu, Z.; Park, M.-W.; Koch, C.; Soltani, M.; Hammad, A.; Davari, K. Predicting Movements of Onsite Workers and Mobile Equipment for Enhancing Construction Site Safety. *Autom. Constr.* **2016**, *68*, 95–101. [CrossRef]
71. Konstantinou, E.; Lasenby, J.; Brilakis, I. Adaptive Computer Vision-Based 2D Tracking of Workers in Complex Environments. *Autom. Constr.* **2019**, *103*, 168–184. [CrossRef]
72. Kim, H.; Lee, J.K.; Shin, J.; Choi, J. Visual Language Approach to Representing KBimCode-Based Korea Building Code Sentences for Automated Rule Checking. *J. Comput. Des. Eng.* **2019**, *6*, 143–148. [CrossRef]
73. Hussain, R.; Lee, D.Y.; Pham, H.C.; Park, C.S. Safety Regulation Classification System to Support BIM Based Safety Management. In Proceedings of the ISARC 2017—Proceedings of the 34th International Symposium on Automation and Robotics in Construction, Taiwan, China, 28 June–1 July 2017.
74. Wolfswinkel, J.F.; Furtmueller, E.; Wilderom, C.P.M. Using Grounded Theory as a Method for Rigorously Reviewing Literature. *Eur. J. Inf. Syst.* **2013**, *22*, 45–55. [CrossRef]
75. Lee, D.; Khan, N.; Park, C. Rigorous Analysis of Safety Rules for Vision Intelligence-Based Monitoring at Construction Jobsites. *Int. J. Constr. Manag.* **2021**, *23*, 1768–1778. [CrossRef]
76. Khan, N.; Ali, A.K.; Tran, S.V.T.; Lee, D.; Park, C. Visual Language-Aided Construction Fire Safety Planning Approach in Building Information Modeling. *Appl. Sci.* **2020**, *10*, 1704. [CrossRef]
77. OSHA. *Commonly Used Statistics | Occupational Safety and Health Administration*; OSHA: Dar es Salaam, Tanzania, 2017.
78. Collins, R.; Zhang, S.; Kim, K.; Teizer, J. Integration of Safety Risk Factors in BIM for Scaffolding Construction. In Proceedings of the 2014 International Conference on Computing in Civil and Building Engineering, Orlando, FL, USA, 23–25 June 2014; pp. 307–314.
79. Jeremiah Wooten OSHA Reveals Top 10 Health & Safety Violations for 2019. Available online: <https://inspectioneering.com/blog/2019-12-18/8922/oshas-top-ten-safety-and-health-violations-for-2019> (accessed on 13 May 2020).
80. Hoła, A.; Sawicki, M.; Szóstak, M. Methodology of Classifying the Causes of Occupational Accidents Involving Construction Scaffolding Using Pareto-Lorenz Analysis. *Appl. Sci.* **2018**, *8*, 48. [CrossRef]
81. Cho, C.; Kim, K.; Asce, A.M.; Park, J.; Asce, A.M.; Cho, Y.K. Data-Driven Monitoring System for Preventing the Collapse of Scaffolding Structures. *J. Constr. Eng. Manag.* **2018**, *144*, e1535. [CrossRef]
82. Sakhakarmi, S.; Park, J.; Asce, A.M.; Cho, C. Enhanced Machine Learning Classification Accuracy for Scaffolding Safety Using Increased Features. *J. Constr. Eng. Manag.* **2019**, *145*, 04018133. [CrossRef]
83. Xu, Y.; Tuttas, S.; Hoegner, L.; Stilla, U. Reconstruction of Scaffolds from a Photogrammetric Point Cloud of Construction Sites Using a Novel 3D Local Feature Descriptor. *Autom. Constr.* **2018**, *85*, 76–95. [CrossRef]
84. Jung, Y.; Oh, H.; Jeong, M.M. An Approach to Automated Detection of Structural Failure Using Chronological Image Analysis in Temporary Structures. *Int. J. Constr. Manag.* **2018**, *19*, 178–185. [CrossRef]

85. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2015, Boston, MA, USA, 7–12 June 2015; pp. 779–788. [\[CrossRef\]](#)
86. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [\[CrossRef\]](#)
87. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
88. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
89. YOLOv5—Ultralytics | Revolutionizing the World of Vision AI. Available online: <https://ultralytics.com/yolov5> (accessed on 19 July 2023).
90. Google TensorFlow Lite Converter. Available online: <https://www.tensorflow.org/lite/convert> (accessed on 19 April 2021).
91. Zhang, Y. TensorFlow Lite Object Detection Android Demo · GitHub. Available online: https://github.com/tensorflow/examples/tree/master/lite/examples/object_detection/android (accessed on 19 April 2021).
92. Le, Q.T.; Pedro, A.; Lim, C.R.; Park, H.T.; Park, C.S.; Kim, H.K. A Framework for Using Mobile Based Virtual Reality and Augmented Reality for Experiential Construction Safety Education. *Int. J. Eng. Educ.* **2015**, *31*, 713–725. [\[CrossRef\]](#)
93. Omale Reuben Peters, O.O. Health Risks and Safety of Construction Site Workers in Akure, Nigeria. *J. Arts Soc. Sci.* **2013**, *13*, 75–94.
94. Lipscomb, H.J.; Glazner, J.E.; Bondy, J.; Guarini, K.; Lezotte, D. Injuries from Slips and Trips in Construction. *Appl. Ergon.* **2006**, *37*, 267–274. [\[CrossRef\]](#)
95. QEB. *Slips, Trips and Falls on the Level*; QBE Insurance Group: Sydney, Australia, 2015.
96. Lim, T.-K.; Park, S.-M.; Lee, H.-C.; Lee, D.-E. Artificial Neural Network–Based Slip-Trip Classifier Using Smart Sensor for Construction Workplace. *J. Constr. Eng. Manag.* **2016**, *142*, 04015065. [\[CrossRef\]](#)
97. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. Scaled-YOLOv4: Scaling Cross Stage Partial Network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13029–13038.
98. Solawetz, J. How to Train a Scaled-YOLOv4 Object Detection Model. 2021. Available online: <https://blog.paperspace.com/how-to-train-scaled-yolov4-object-detection/> (accessed on 13 May 2020).
99. Van Etten, A. Satellite Imagery Multiscale Rapid Detection with Windowed Networks. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision, WACV 2019, Waikoloa Village, HI, USA, 7–11 January 2019; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2019; pp. 735–743.
100. Riggs, J. HSE Statistics Should Remind Us about the Dangers of Working at Height | Planning, BIM & Construction Today. Available online: <https://www.pbctoday.co.uk/news/health-safety-news/hse-statistics-should-remind-us-about-the-dangers-of-working-at-height/39615/> (accessed on 13 May 2020).
101. Yuan, X.; Anumba, C.J.; Parfitt, M.K. Cyber-Physical Systems for Temporary Structure Monitoring. *Autom. Constr.* **2016**, *66*, 1–14. [\[CrossRef\]](#)
102. Cho, C.; Sakhakarmi, S.; Kim, K.; Park, J.W. Scaffolding Modelling for Real-Time Monitoring Using a Strain Sensing Approach. In Proceedings of the ISARC 2018—35th International Symposium on Automation and Robotics in Construction and International AEC/FM Hackathon: The Future of Building Things, Berlin, Germany, 20–25 July 2018; International Association for Automation and Robotics in Construction (I.A.A.R.C): Edinburgh, UK, 2018.
103. KOSHA. *Korean Occupational Safety and Health Agency Report*; KOSHA: Ulsan, Republic of Korea, 2019; pp. 1–230.
104. Darko, A.; Chan, A.P.C.; Yang, Y.; Tetteh, M.O. Building Information Modeling (BIM)-Based Modular Integrated Construction Risk Management—Critical Survey and Future Needs. *Comput. Ind.* **2020**, *123*, 103327. [\[CrossRef\]](#)
105. Park, C.S.; Kim, H.J. A Framework for Construction Safety Management and Visualization System. *Autom. Constr.* **2013**, *33*, 95–103. [\[CrossRef\]](#)
106. Ding, L.; Fang, W.; Luo, H.; Love, P.E.D.; Zhong, B.; Ouyang, X. A Deep Hybrid Learning Model to Detect Unsafe Behavior: Integrating Convolution Neural Networks and Long Short-Term Memory. *Autom. Constr.* **2018**, *86*, 118–124. [\[CrossRef\]](#)
107. Papazoglou, I.A.; Aneziris, O.; Bellamy, L.; Ale, B.J.M.; Oh, J.I.H. Uncertainty Assessment in the Quantification of Risk Rates of Occupational Accidents. *Risk Anal.* **2015**, *35*, 1536–1561. [\[CrossRef\]](#) [\[PubMed\]](#)
108. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [\[CrossRef\]](#)
109. Gómez-de-Gabriel, J.M.; Fernández-Madrugal, J.A.; López-Arquillos, A.; Rubio-Romero, J.C. Monitoring Harness Use in Construction with BLE Beacons. *Measurement* **2019**, *131*, 329–340. [\[CrossRef\]](#)
110. KOSHS. Preventing Falls in Construction. Available online: <https://kiprc.uky.edu/sites/default/files/2022-01/construction-falls-2016.pdf> (accessed on 26 April 2021).
111. Umer, W.; Li, H.; Lu, W.; Szeto, G.P.Y.; Wong, A.Y.L. Development of a Tool to Monitor Static Balance of Construction Workers for Proactive Fall Safety Management. *Autom. Constr.* **2018**, *94*, 438–448. [\[CrossRef\]](#)
112. Leamon, T.B.; Murphy, P.L. Occupational Slips and Falls: More than a Trivial Problem. *Ergonomics* **1995**, *38*, 487–498. [\[CrossRef\]](#)
113. Kumar, A. A Modern Pothole Detection Technique Using Deep Learning. In Proceedings of the 2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, 28–29 February 2020.

114. Dai, J. Real-Time and Accurate Object Detection on Edge Device with TensorFlow Lite. *J. Phys. Conf. Ser.* **2020**, *1651*, 12114. [[CrossRef](#)]
115. Galvez, R.L.; Bandala, A.A.; Dadios, E.P.; Vicerra, R.R.P.; Maningo, J.M.Z. Object Detection Using Convolutional Neural Networks. In Proceedings of the IEEE Region 10 Annual International Conference, Proceedings/TENCON, Kochi, India, 17–20 October 2019; pp. 2023–2027. [[CrossRef](#)]
116. Chen, Y.; Li, W.; Sakaridis, C.; Dai, D.; Van Gool, L. Domain Adaptive Faster R-CNN for Object Detection in the Wild. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3339–3348.
117. Kawazoe, Y.; Shimamoto, K.; Yamaguchi, R.; Shintani-Domoto, Y.; Uozaki, H.; Fukayama, M.; Ohe, K. Faster R-CNN-Based Glomerular Detection in Multistained Human Whole Slide Images. *J. Imaging.* **2018**, *4*, 91. [[CrossRef](#)]
118. Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Chen, J.; Liu, X.; Pietikäinen, M. Deep Learning for Generic Object Detection: A Survey. *Int. J. Comput. Vis.* **2020**, *128*, 261–318. [[CrossRef](#)]
119. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; IEEE Computer Society: Washington, DC, USA, 2015; pp. 1–9.
120. Su, J. How to Train a Custom TensorFlow Lite Object Detection Model. Available online: <https://blog.roboflow.com/how-to-train-a-tensorflow-lite-object-detection-model/> (accessed on 17 April 2021).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.