TOPICAL REVIEW

A Review of Computer Vision-Based Monitoring Approaches for Construction Workers' Work-Related Behaviors

JIAQI LI [1], QI MIAO[1], ZHENG ZOU[2], HUAGUO GAO[1], LIXIAO ZHANG[2], ZHAOBO LI[3,4], AND NAN WANG[1]

[1]School of Civil Engineering, University of Science and Technology Liaoning, Anshan 114051, China
[2]College of Transportation Engineering, Dalian Maritime University, Dalian 116026, China
[3]Hohhot Science and Technology Innovation Service Center, Hohhot 010011, China
[4]School of Mechanics and Civil Engineering, China University of Mining and Technology, Xuzhou 221116, China

Corresponding author: Jiaqi Li (lijiaqi@ustl.edu.cn)

ABSTRACT Construction workers' behaviors directly affects labor productivity and their own safety, thereby influencing project quality. Recognizing and monitoring the construction-related behaviors is therefore crucial for high-quality management and orderly construction site operation. Recent strides in computer vision technology suggest its potential to replace traditional manual supervision approaches. This paper explores research on monitoring construction workers' behaviors using computer vision. Through bibliometrics and content-based analysis, the authors present the latest research in this area from three perspectives: "Detection, Localization, and Tracking for Construction Workers," "Recognition of Workers' Construction Activities," and "Occupational Health and Safety Behavior Monitoring." In terms of the literature's volume, there has been a notable increase in this field. Notably, the focus on safety-related literature is predominant, underscoring the concern for occupational health. Vision algorithms have witnessed an increase in the utilization of object detection. The ongoing and future research trajectory is anticipated to involve multi-algorithm integration and an emphasis on enhancing robustness. Then the authors summarize the review from engineering impact and technical suitability, and analyze the limitations of current research from the perspectives of technical approaches and application scenarios. Finally, it discusses future research directions in this field together with generative AI models. Furthermore, the authors hope this paper can serves as a valuable reference for both scholars and engineers.

INDEX TERMS Computer vision, construction worker, construction behavior, construction site, monitoring.

I. INTRODUCTION

The construction industry is recognized as a labor-intensive field that involves multi-tasking operations. On-site workforce management can be challenging due to its cumbersome nature, which has the potential to result in low productivity and high safety risks [1]. Possible inactive construction activities, inefficiency, and irresponsible attitudes of operators have been found to result not only in time and resource wastage

but also in economic loss for the entire project, along with a decline in safety and construction quality [2], [3].

Lately, there has been a growing emphasis on proactively evaluating the occupational health and safety challenges confronted by construction laborers, as indicated by research [4]. Furthermore, several countries have recognized the need to strengthen construction supervision regulations at the governmental policy level [5], [6].

Traditionally, professionals conduct on-site inspections and supervise and record work, which often leads to problems such as insufficient coverage, unreasonable personnel
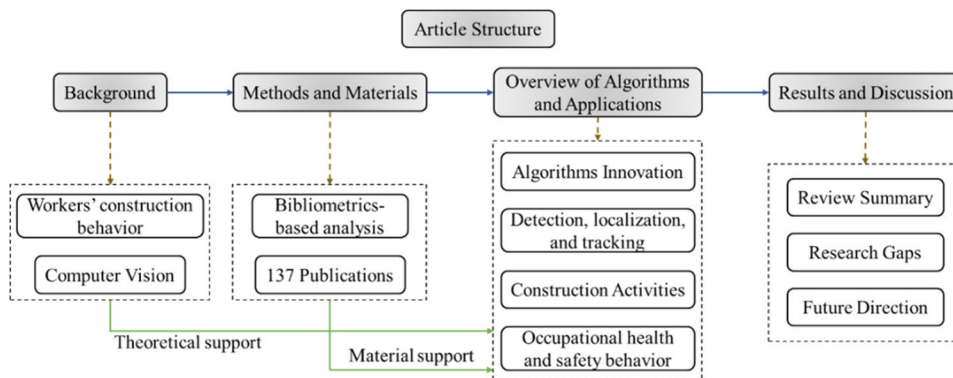
The associate editor coordinating the review of this manuscript and approving it for publication was Zhenhua Guo.

FIGURE 1. The research logic framework.

scheduling, time-consuming tasks, and subjective supervision results [7]. In recent years, many countries have embraced the concept of intelligent construction, driven by the development of information technology and the Internet of Things [8], [9], [10]. Thanks to technology and policy, several intelligent technologies, such as sensor technology, audio technology, and computer vision technology, are being employed in engineering construction supervision.

The primary types of sensors used for monitoring construction workers are position sensors [11] and acceleration sensors [12]. However, using these sensors requires them to be attached to the worker, which can affect their comfort and increase equipment maintenance costs [13]. The audio-based approach cannot identify construction tasks that produce less sound or no sound at all [14]. In the field of civil engineering, computer vision technology has become a high-precision non-contact monitoring method, which is used in matters such as cracks monitoring [15], [16], [17], strains measurement [18], [19], damage recognition [20], [21]. Vision-based approaches can also record worker activities in real-time, allowing project managers to track construction site productivity, progress, and safety risks, and make timely decisions about project progress. Using computer vision to achieve intelligent control of smart construction sites is not only the next frontier science, but also the driving force for the development of smart construction sites [22].

Nonetheless, there have been fewer studies conducted to monitor worker behavior than those assessing building structure damages, as indicated by Mostafa and Hegazy [23]. In recent years, several comprehensive reviews have been conducted, focusing on safety concerns in engineering construction. Zhang et al. [24] and Fang et al. [25] have explored the application of computer vision technology in worker safety. From an economic development perspective, Luo et al. [26]have elaborated on construction safety issues, emphasizing emerging trends like deep learning and interdisciplinary technologies. Additionally, Zeng et al.'s work has pointed artificial intelligence as directions in construction safety research [27] Construction workers' behavior plays a pivotal role in influencing construction quality [2]. Their

trajectories, work behaviors, and safety practices need to be seamlessly integrated into intelligent management systems [24], [28]. In recent years, the number of related works using computer vision technology to conduct research from the perspective of construction workers' behavior has been increasing year by year, and the algorithmic techniques are constantly updated and iterative, so a comprehensive review is needed to summarize and review the latest related literatures in the field in order to understand the current research trends and status, challenges and limitations. At the same time, we hope that this paper can give some reference to engineers to help innovation in construction management.

To achieve this objective, the remainder of the paper is organized as follows: Chapter 2 outlines the methodology employed for literature search, Chapter 3 describes the literature situation, Chapter 4 presents the computer vision techniques utilized in reviewed research, Chapter 5 provides applications scenarios of computer vision techniques to worker construction-related behaviors, Chapter 6 lists summaries from the technical and practical application perspectives, and finally, research gaps and limitations are identified, and potential future research directions are suggested.

II. RESEARCH METHODOLOGY

Figure 1 is the technical framework of this article. Initially, the authors establish the study's background in the introduction, providing the groundwork for subsequent research. Following this, a meticulous screening and analysis were conducted on 137 publications published after 2013, all of which are pertinent to the work behaviors of construction workers. Subsequently, based on the research content within these articles, the examination and advancement of knowledge are executed from three perspectives: the localization and tracking of construction workers, the recognition of construction activities, and the monitoring of occupational health and safety behaviors. These three dimensions comprehensively encompass the critical issues of workers' movement trajectories, construction activities, and safety practices, all of which exert a direct or indirect influence on project quality control and the safety of construction personnel. Building upon the insights

gleaned from the literature review, a comprehensive synthesis and discussion are presented, elucidating the current state of research and delineating prospective avenues for future investigation.

The authors conducted a search of relevant literature using Web of Science (WOS). The search string used was: (worker AND construction) AND ("computer vision" OR "vision-based" OR "deep learning" OR "image processing" OR vision). The retrieval process comprised three sequential phases. In the initial stage, utilizing the aforementioned string, 596 scholarly articles from journal publications and conference papers were obtained from the WOS core collection. In the second stage, review papers were excluded, and literature from 2013 to June 2023 was retained, leaving 496 papers. The third stage involved a careful manual search of the abstract, keywords, and body content to exclude content unrelated to "construction workers" and "machine vision," leaving 137 papers for analysis.

Figure 2 illustrates the distribution of the 137 articles analyzed in this study by their year of publication and the number of citations they received in each year. It is evident that the number of publications in this field has grown rapidly since 2015. Moreover, the number of citations reveals that the research outcomes have received significant attention from the academic community. This trend could be attributed to the growing interest in intelligent construction technology and the advancements in computer vision technology since 2010. The popularity of computer vision was further enhanced with the introduction of the Faster R-CNN algorithm in 2015. Figure 3 displays the sources of the papers, including Automation in Construction, Journal of Computing in Civil Engineering, Journal of Construction Engineering and Management, Advanced Engineering Informatics, IEEE Access, etc. These journals are reputable and may have a considerable potential to make further contributions to the field.
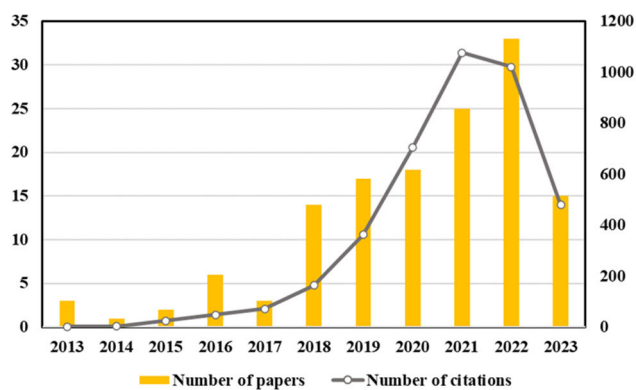

FIGURE 2. The publication frequency across various years.

III. OVERVIEW OF REVIEWED LITERATURES

VOSViewer, a bibliometric software, was utilized to cluster the reviewed papers and identify current topics of interest in the field [29]. The authors utilized VOSViewer for an in-depth analysis of the paper's keywords, which led to the creation of a heat map (depicted in Figure 4). In the process of generating this heat map, each point on the map is assigned a color based on the density of the elements in its vicinity. A higher density corresponds to a closer proximity to red, while lower density leans toward blue. The magnitude of this density is contingent upon the significance of these elements, providing a rapid overview of critical research areas [27]. It is evident that the central themes in the literature revolve around keywords such as "deep learning," "computer vision," "construction," and "workers," among others.

Subsequently, the authors analyzed the literature's authors and their institutions, and the results are visualized in Figure 5 and Figure 6. In the figures, a larger font size and nodes correspond to a higher volume of published literature. Various colors are used to denote distinct clusters, while the lines connecting the circles illustrate the interconnectedness between authors and their affiliations through cross-referencing [30].The filtering condition for this analysis was a minimum of three publications, and the size of the bubbles in the figures indicates that Hong Kong Polytechnic University [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51] and Huazhong University of Science and Technology. [32], [33], [34], [36], [38], [42], [50], [51], [52], [53], [54], [55], [56], [57], [58] are currently leading the field. Other institutions conducting research in this area include Chung Ang University [59], [60], [61], [62], [63], [64], [65], University of Illinois [66], [67], [68], [69], [70], [71], [72], [73], Dalian University of Technology [74], [75], [76], [77], [78], [79], and more. Figure 5 highlights that Li Heng's team [31], [32], [33], [34], [35], [36], [38], [39], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51] at Hong Kong Polytechnic University and Luo Hanbin's team [33], [42], [52], [53], [54], [55], [56], [57], [58] at Huazhong University of Science and Technology are currently leading the field, with similar colors representing the same research area. Additionally, Li et al. [75], [76], [78], Zhang et al. [74], [75], [76], [78], [80], and Cai and Cai [81], [82], [83], [84] have shown potential in this field.

Figure 7 illustrates the research progress in each country or region, highlighting Mainland China's current leadership in this field. This may be attributed to the rapid growth of China's construction industry since the 21st century, surpassing the United States in terms of construction output by 2012, with the gap in output widening over the following decade. However, this growth has also brought about negative consequences, including safety and environmental concerns, leading China to prioritize enhanced regulation, smart construction, and green construction. As a result, this concerted effort has catalyzed the swift advancement and implementation of computer vision technology in the construction industry. The United States, as the world's number one economy, has also made significant contributions to this field. With its technology powerhouse status, the United States provides fertile ground for the development of artificial intelligence
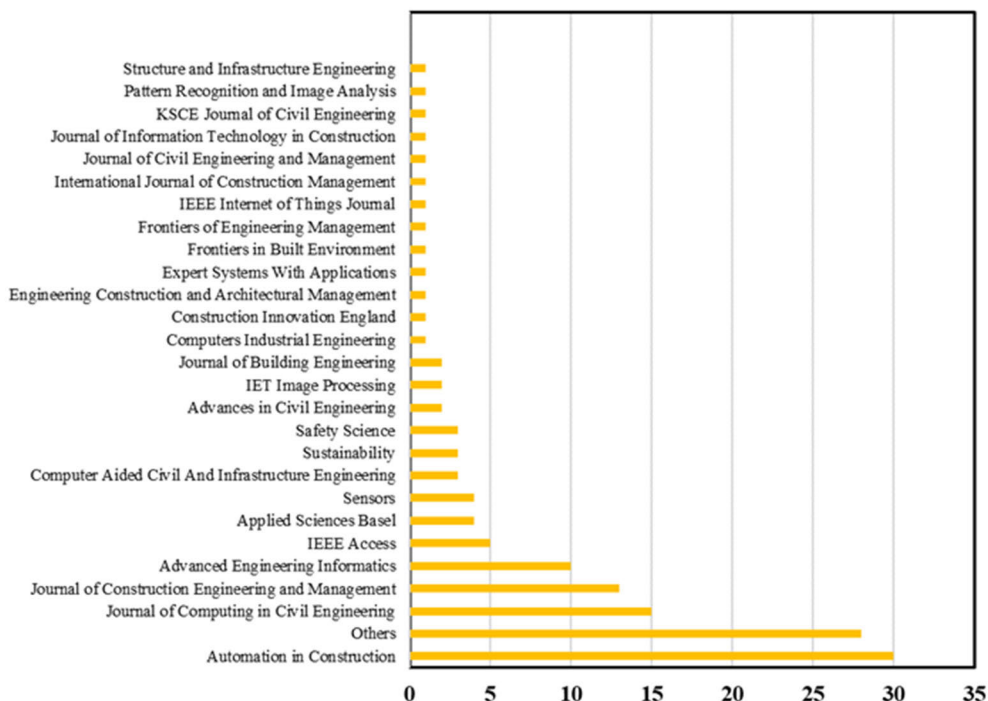
FIGURE 3. The volume of scholarly articles published in relevant academic journals.
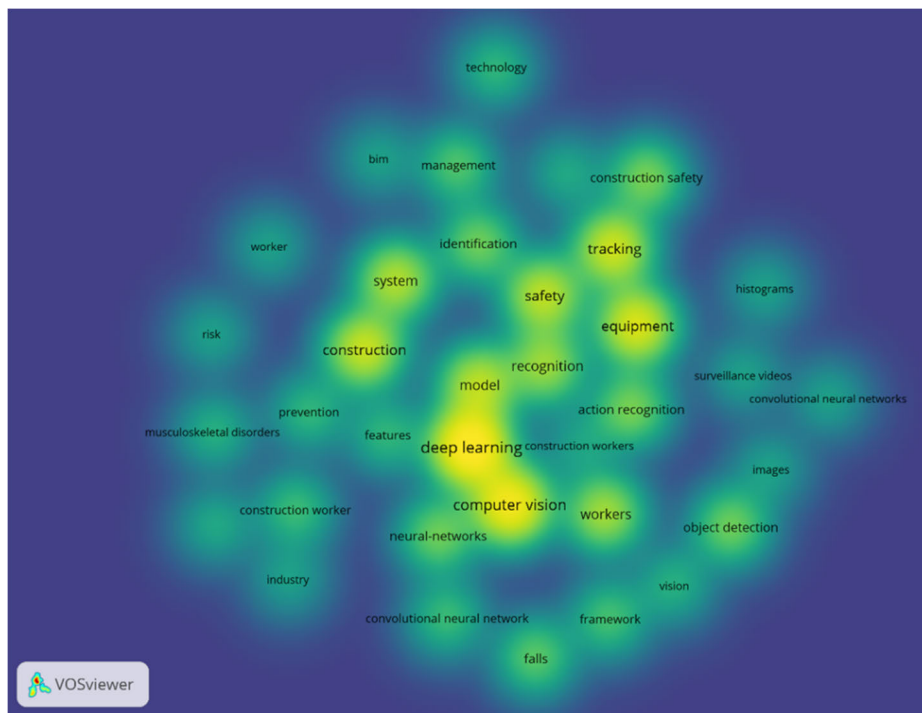

FIGURE 4. Keywords hotspot heatmap.

technology. Furthermore, South Korea, Australia, and other countries have also shown significant development in the "computer vision + worker construction behavior + construction" field.

IV. TECHNOLOGIES

A. OVERVIEW OF COMPUTER VISION TECHNIQUES

Computer Vision, also known as Machine Vision, is a technology that uses mathematical algorithms and computers to
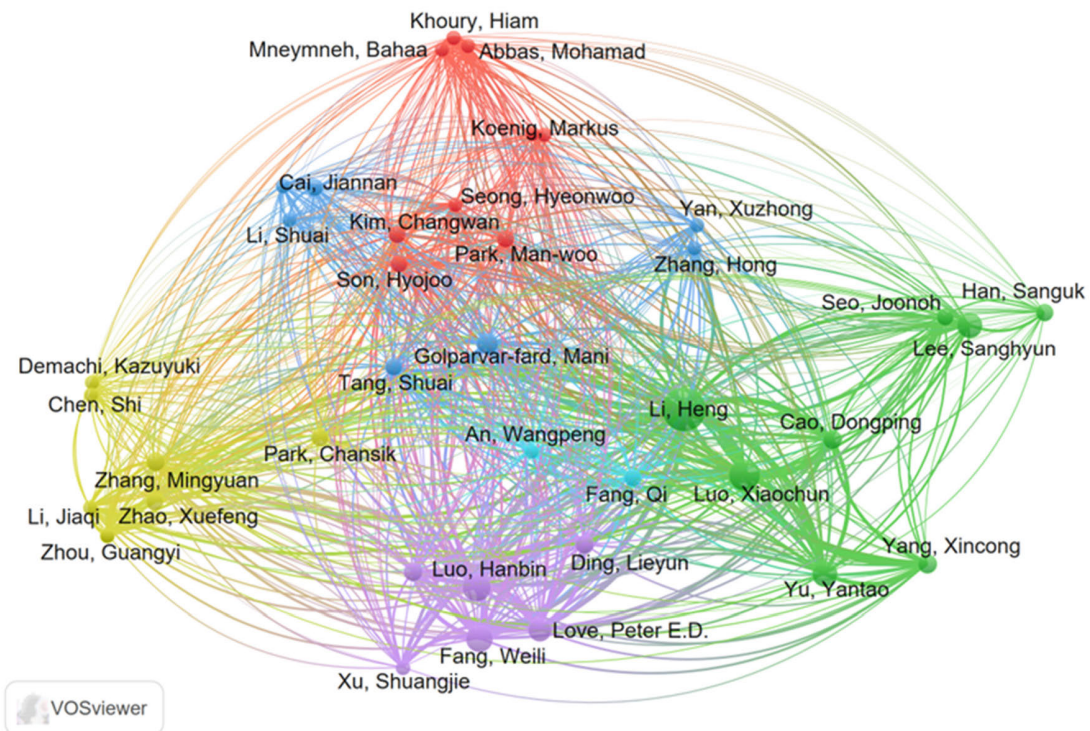
FIGURE 5. Analysis of research authors.
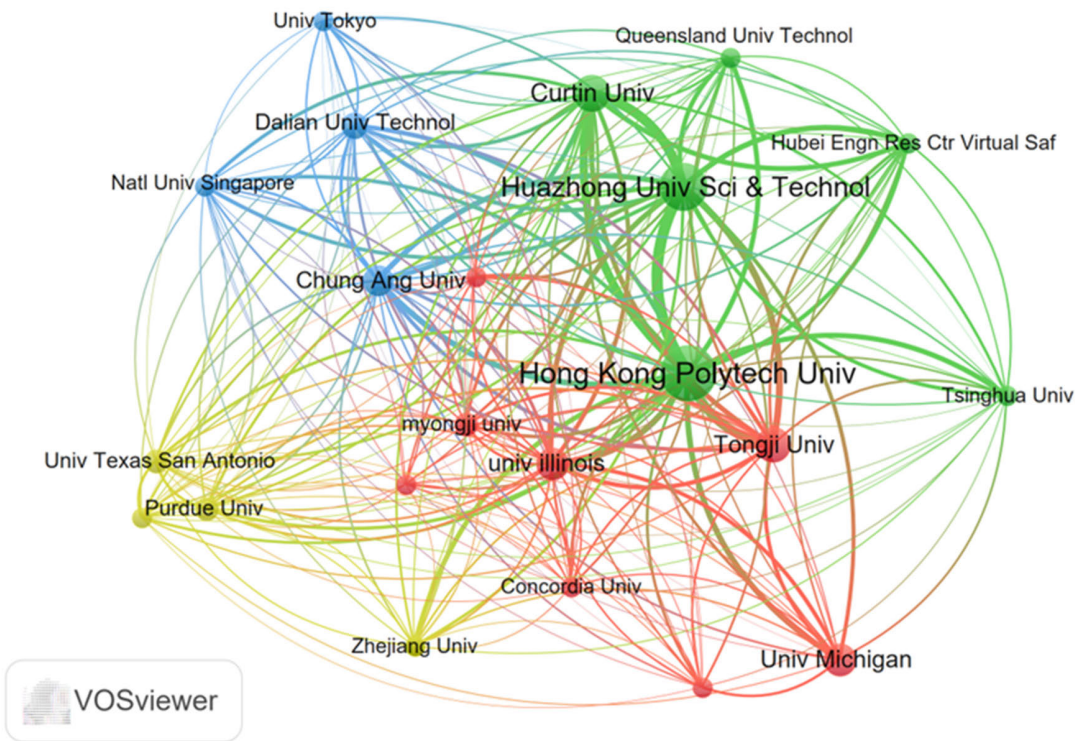

FIGURE 6. Analysis of research institutions.

automate and simulate human vision. Its primary purpose is to extract valuable information from image and video data and then analyze, process, and understand it.

Traditional computer vision methods rely mainly on image processing and pattern recognition techniques. These methods include feature extraction techniques like edge
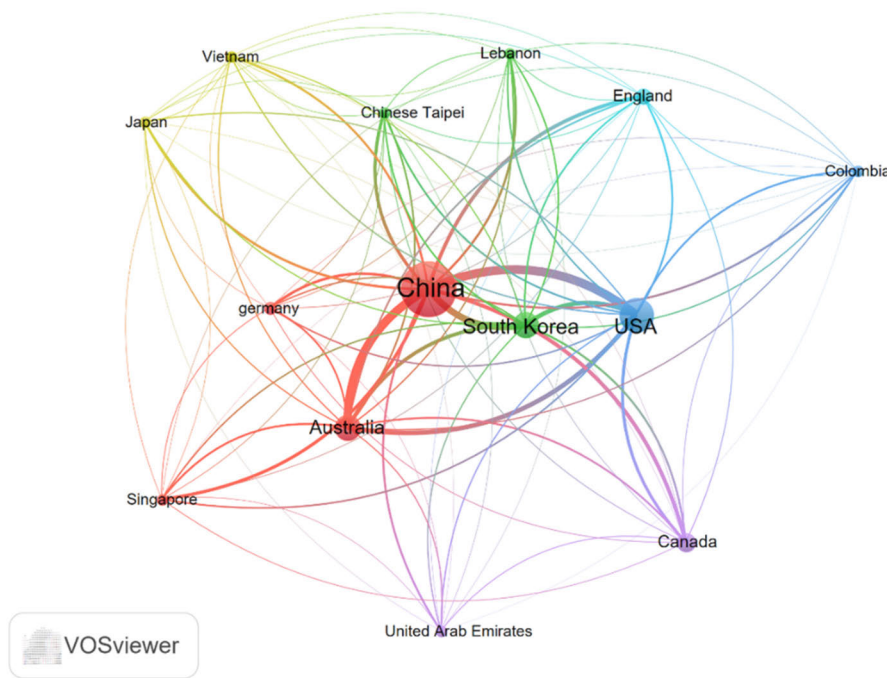
FIGURE 7. Analysis of research countries or areas.

detection [85]. Image classification methods such as support vector machines (SVM) [86] can be used to classify objects based on their features. Object tracking methods like Kalman filter [87] can track the position and motion state of objects in consecutive frames. Additionally, segmentation methods based on thresholding, edges, regions etc. [88], can divide images into different regions or parts. While these traditional methods can be useful for tasks such as image recognition, their application in complex scenarios is limited due to the need for manually designed features.

Currently, deep learning-based computer vision methods have become mainstream. Convolutional neural networks (CNN) and their derivatives, such as ResNet [89], are widely used for various tasks, including image classification, detection, and segmentation. Recurrent neural networks (RNN) [90] are used for sequential data processing, such as video, speech, and text. Object detection networks such as Faster R-CNN [91] and YOLO [92] are used to locate and measure the size of target objects in images, while segmentation networks such as Mask R-CNN [93] are used for image segmentation into different parts or regions. Generative adversarial networks (GANs) [94] are used for image generation and restoration, as they can produce high-quality images and enhance image restoration. Furthermore, long and short term memory networks (LSTM) [95] can be used for visual tasks such as video classification and object tracking.

Computer vision technologies mentioned above provide powerful support for quickly extracting information from construction sites and identifying construction site conditions and worker behavior.

B. COMPUTER VISION ALGORITHMS' INNOVATION

The construction environment is characterized by its intricate and dynamic nature, where diverse categories of heavy machinery coexist and operate in parallel, alongside considerable personnel mobility. To better adapt computer vision technology to engineering needs, researchers have improved and innovated upon existing research results in the field of computer vision, updating the technical framework. For example, the recognition of steel bar-related activities performed by construction workers was addressed by Luo et al. [52] through the proposition of a three-stream CNN which can capture static spatial features, short-term motion, and long-term motion in video clips. Yang et al. [96] proposed the Spatial Temporal Relation Transformer (STR-Transformer), which better fuses temporal and spatial features in construction video clips. They also created a video clip dataset including seven types of construction workers' behaviors. Fang et al. [56] proposed an improved Faster R-CNN that achieves better multi-scale detection in detecting and locating workers and equipment on construction sites. Liu et al. [97] proposed Multi-Domain Convolutional Neural Network to improve the tracking performance of construction workers. Huang et al. [98] achieved better construction worker hardhat wear detection performance by improving YOLO v3. Park et al. [99] introduced DIoU and NMS for YOLO v5, utilizing weight triplet attention, expansion deature-level, and soft-pool to improve the performance and enhance the ability of detecting workers at construction sites, especially the ability to detect overlapping objects in complex environments was enhanced. Other studies have also involved various computer vision

TABLE 1. Typical research for algorithms integrated.

Authors	Objectives	Algorithms
Cheng et al[111]	Tracking of construction workers and monitoring the PPE's usage	Person ReID + CNN
Wei et al[54]	Identify a person's identity from videos captured from construction sites	spatial attention network + temporal attention networks
Ting et al[33]	Predict construction workers' unsafe behaviors	SiamMask + improved Social-LSTM + PNPoly
Yan et al[35]	Assessing the risk of collisions between construction workers and trucks	Faster R-CNN + 3D bounding box reconstruction
Yang et al[112]	3D Localization of Workers in Construction Sites	Faster R-CNN + DeepSORT + ReID + Triangulation
Luo et al[41]	Recognize construction workers' activities	YOLO v3 + SORT + 3D CNN + Clustering filting
Deng et al[113]	Detect workers and predict their movement trajectories	SURF + improved GMM + HOG + SVM
Li et al[114]	Hardhat-wearing tracking detection	YOLO v5 + Strong SORT
Lee et al[115]	PPE usage detection	YOLACT employs MobineNetv3 + DeepSORT
Li et al[116]	Recognize construction workers' activities	YOLO + ST-GCN (Spatial temporal graph convolutional networks)

methods for algorithm improvements [100], [101], [102], [103], [104], [105], [106].

Establishing a highly stable vision algorithm model is partly dependent on high-quality training data. While some open-source databases like COCO and ImageNet are available, they have limited applications in the construction industry. To address this limitation, Yang et al. [96] created a video clip dataset comprising seven types of construction workers' behavior. Jixiu et al. [107] also developed an open-access dataset called GDUT-HWD, which includes 3174 images for hardhat detection. Additionally, several studies have developed new databases, including Yang et al. [108], Luo et al. [52], Tian et al. [44], and Xiong et al. [48]. These efforts have resulted in an increased availability of training data for construction-related computer vision tasks.

Section IV-A emphasizes that different classes of vision algorithms can fulfill distinct functions, and for construction sites with complex scenes, a single class of algorithm may not suffice for behavior recognition. Thus, combining multiple algorithms is necessary. For instance, Xiao et al. [109] proposed a construction worker tracking method that utilizes the features of various algorithms, such as Mask R-CNN and Kalman Filter, for tracking multiple construction workers, even in the face of challenges like occlusion and feature scale variation. Piao et al. [110] proposed a framework for computer vision methods that incorporates Dynamic Bayesian Network, Openpose, and Faster R-CNN to assess the fall risk of construction workers during dynamic construction. Roberts et al. [66] developed a computer vision-based method for construction worker activity analysis, which included YOLO v3, Alphapose, and i3D based on CNN. Their findings indicated that incorporating pose estimation enhances construction activity analysis. Fang et al. [38] integrated

Faster R-CNN for detecting construction activities, SORT (Simple Online And Realtime Tracking) for tracking individuals, and face recognition for confirming the identity of workers, to eventually identify whether there are uncertified workers engaged in irrelevant construction tasks. Li et al. [78] used YOLO v5 to detect Personal Protective Equipment (PPE) and Openpose to detect skeleton joints, incorporating the visual feature information to determine whether the PPE was correctly used, using 1D-CNN. Ding et al. [58] combined CNN and LSTM to identify construction workers' unsafe behaviors. Cai et al. [84] used Faster R-CNN to detect the head and body orientation of construction workers, and subsequently applied a multi-task learning network to assess the visual attention direction of construction workers. Fang et al. [51] combined Mask R-CNN and Cascaded Pyramid Network to localize construction workers in monocular vision, then identified their unsafe behaviors. Table 1 presents other papers that integrated multiple algorithms to achieve improved visual recognition.
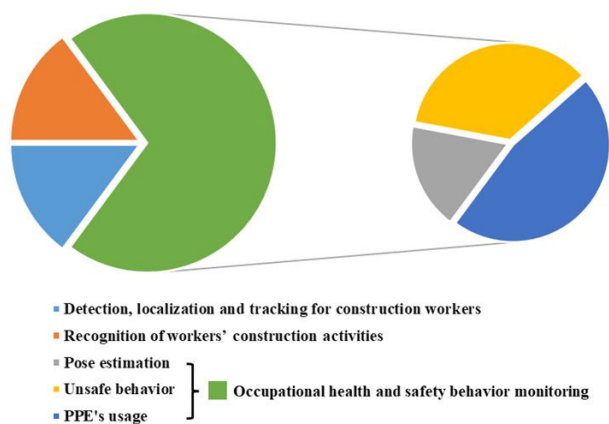
V. APPLICATIONS

This section provides an in-depth analysis of the reviewed articles from the application perspective. The literature materials selected in this paper consider all construction-related behaviors, including workers' movement trajectory, construction activities, and safety behaviors. These behaviors have indirect or direct impacts on project quality control and workers' safety. The section analyzes the literature from three perspectives: construction workers' location and tracking, workers' construction activities recognition, and monitoring of occupational health and safety behaviors.

Figure 8 illustrates the distribution of literature among various application domains, with the majority of the

research dedicated to occupational health and safety behavior monitoring, constituting over 70% of the total literature. This finding suggests that safety in construction is a significant concern among researchers worldwide. Within this domain, researchers show a higher inclination towards detecting Personal Protective Equipment (PPE) use, which constitutes 32.8% of the total reviewed literature.

A. DETECTION, LOCALIZATION AND TRACKING FOR CONSTRUCTION WORKERS

Engineering construction involves multi-tasking operations with a high turnover of personnel, and the distribution of the location and the trajectory of construction workers affect both productivity and efficiency. Thus, detecting and tracking construction workers is a crucial task in construction management.

- Detection, localization and tracking for construction workers
- Recognition of workers' construction activities
- Pose estimation
- Unsafe behavior
- PPE's usage
- Occupational health and safety behavior monitoring

FIGURE 8. Analysis of application fields.

When detecting construction workers, relying on a single algorithm, such as SVM, is often considered in early stage. For instance, Memarzadeh et al. [72] extracted HOG features and then used SVM to detect construction workers and equipment. On the other hand, Fang et al. [56] and Son et al. [62] both used the Faster R-CNN algorithm for object detection and achieved better performance. Park et al. [99] proposed SOC-YOLO to improve the detection of construction workers in complex scenes when they overlap. Achieving more accurate localization for individual workers on a temporal scale requires combining various algorithms [112], [117]. For instance, Yang et al. [112] added DeepSORT after Faster R-CNN for multi-object detection and then used the Person ReID algorithm for individual worker localization on the time scale. Some researchers have extended worker detection applications. For example, Huang et al. [118] and Mei et al. [119] used skeleton-based and object detection-based approaches, respectively, to monitor whether a worker has intruded into a prohibited area based on localization. Wei et al. [54] integrated spatial attention networks and temporal attention networks to recognize individual identities.

Cai et al. [84] used Faster R-CNN to detect the head and body orientation of construction workers, and then applied a multitask learning network to evaluate the visual attention direction of workers. Yan et al. [39] first detected workers using Faster R-CNN and then applied 3D pose estimation to locate them in 3D space.

The fundamental basis for the implementation of worker tracking lies in the localization and detection of construction personnel, facilitating an enhanced comprehension of labor allocation and mobility within construction sites. The Kalman Filtering algorithm has found wide-ranging applications in the field of object tracking, as evidenced by the reviewed literature [109], [113], [120], [121], [122]. It is common to combine Kalman Filtering with object detection algorithms. For example, Neuhausen et al. [120], [121] used YOLO v3 to detect construction workers and then tracked them using the Kalman filter, demonstrating the robustness of the Kalman filter in compensating for the limitations of YOLO v3. Yongyue et al. [122] proposed a different approach by combining Kalman Filtering with the individual pose joint point estimation algorithm, Openpose, to reduce the manual annotation workload. Similar results can be achieved by combining object detection algorithms with other techniques. Son and Kim [63] proposed the combination of YOLO v4 and Siamese Network for construction worker detection and tracking, achieving an accuracy of 0.975. Wan et al. [106] improved YOLO v5 by attention mechanism, astrous spatial pooling, and universal upsampling to track construction workers by classifying hardhats, which in turn identified unauthorized intrusions. Angah and Chen [123] integrated Mask R-CNN, Matching, and Rematching, enabling multiple construction worker tracking on the screen. Liu et al. [97] introduced a multidomain representation for CNNs to improve tracking effects in complex dynamic scenes. Furthermore, researchers have expanded tracking into 3D space based on 2D tracking. Cai and Cai [81] proposed a hybrid approach of visual tracking and radio localization to address the problem of easily losing targets in visual recognition. Lee and Park [124] employed two cameras for separate tracking and localization, followed by Entity Matching for 3D tracking.

LSTM networks can analyze input data with a time series component. To predict the action trajectories of construction workers, Cai et al. [83] proposed a context-aware LSTM-based method. Similarly, Tang et al. [73] combined LSTM with a mixture density network to achieve the prediction of construction workers' action trajectories.

Computer vision-based detection and tracking methods have the potential to effectively assist project managers in analyzing the precise location information of construction workers. As computer vision technology continues to advance, the accuracy and stability of tracking methods will gradually improve, leading to enhanced tracking capabilities of individuals in 3D space and improved prediction of action trajectories.

TABLE 2. Construction activities recognition research details.

Authors	Objectives	Algorithms
Yang et al (2016)[108]	Workers' construction activities recognition (LayBrick, Transporting, CutPlate, Drilling, TieRebar, Nailing, Plastering, Shoveling, Bolting, Welding, Sawing)	SVM
Liu et al (2016)[125]	Construction workers' actions (walking, lifting, crawling)	a silhouette-based human action recognition method
Luo et al (2018)[52]	Workers' construction activities recognition (Steel bending, Transporting, Walking)	An improved three-stream CNN
Yang (2018)[126]	Workers' construction activities recognition (LayBrick, CutPlate, Drilling, TieRebar, Nailing, Plastering, Shoveling, Bolting, Welding, Sawing)	dense trajectories + data-driven scene parsing
Fang et al (2018)[38]	Recognize workers' construction activities (Making framework, Installing windows, cutting rebar, mixing concrete, pouring concrete, operating excavator et al.) and detect non-certified work	Faster R-CNN + relationship between detected boxes + SORT + face recognition
Luo et al (2018)[43]	Recognize workers' construction activities (Placing concrete, machining or transferring formwork, building formwork, transporting goods, et al.)	Faster R-CNN + relationship between detected boxes + relevance network
Luo et al (2018)[45]	Recognize workers' construction activities (measuring, preparing, fixing formwork, connecting rebar, fixing rebar, transporting rebar, welding rebar, et al.)	Two-stream CNN
Luo et al (2019)[41]	Recognize workers' construction activities (standing, bending, squatting, transporting, pulling, leveling, et al.)	YOLO v3 + SORT + 3D CNN + Clustering algorithm
Luo et al (2019)[49]	Recognize workers' construction activities (standing and carrying, leveling, moving, leaning forward, leaning backward, sitting, standing with hands empty)	Bayesian Nonparametric Learning
Cai et al (2019)[128]	Recognize working groups and recognize group construction activities (Spotting, road paving, et al.)	Positional cues + attentional cues + LSTM
Roberts et al (2020)[66]	Construction worker activity analysis (bricklaying and plastering operations)	YOLO v3 + Alphapose + i3D based on CNN
Tang et al (2020)[67]	Safety inspection and construction activities recognition	Faster R-CNN + detected boxes + human-object interaction
Liu et al (2020)[40]	Recognize workers' construction activities (Masonry work, cart transportation, rebar work, plastering, tiling)	Sentence generation (CNN + RNN + word embedding)
Luo et al (2020)[50]	Recognize workers' construction activities in groups (Moving with hands empty, sitting and resting, transporting, placing concrete, cutting formwork, placing rebar, et al.)	YOLO v3 + SORT +KCF + C3D + CRF
Li et al (2021)[75]	Recognize workers' construction activity (assembling reinforcement) and evaluate the work productivity	CenterNet + detected boxes + productivity evaluation
Li et al (2022)[76]	Recognize workers' construction activities on a reinforcement processing area (Straighten, check, transferring, process)	Faster R-CNN + spatial relationship between detected boxes
Torabi et al (2022)[129]	Recognize workers' construction activities (Placing/fixing rebars, drilling, hammering, standing, walking, transporting, et al)	YOWO53
Bhokare et al (2022)[127]	Recognize workers' construction activities (Brick laying, Carpentry, Excavating, Concrete Screeding)	YOLO v3
Yang et al (2023)[96]	Recognize workers' construction activities (Smoke, stand, climb ladder, wear helmet, fall down, walk, talk, et al)	STR-Transformer
Li and Li (2021) [79]	Recognize construction activities of workers under occlusion (Driving truck, Transporting cement, Checking the power socket, Cleaning up the plank, Lashing rebar, Paving concrete, Installing scaffolding, Smearing plaster)	Openpose + GAN + ResNet50
Li et al[116]	Recognize workers behaviors (throwing, operating, crossing)	YOLO + ST-GCN

B. RECOGNITION OF WORKERS' CONSTRUCTION ACTIVITIES

The accurate recognition of workers' construction activities plays a crucial role in enabling project managers to effectively monitor construction progress and optimize the allocation of labor resources. It serves as a valuable tool for analyzing labor productivity. Previous studies primarily employed traditional image processing algorithms for construction activities

TABLE 3. Typical research for pose estimation.

Authors	Objectives	Algorithms	2D/3D
Khosrowpour et al(2014)[69]	Pose estimation and activity classification	3D pose estimation + activity classifier + Hidden Markov Model	3D
Zhang et al (2018)[46]	Pose estimation	multi-stage convolutional neural network	3D
Yu et al (2019)[31]	Pose estimation	3D human body detection + REBA (Rapid Entire Body Assessment)	3D
Yu et al (2019)[34]	Pose estimation + Physical fatigue assessment	3D motion capture algorithm + biomechanical analysis	3D
Yu et al (2020)[47]	Dataset creation + Pose estimation	A 3D pose dataset consists of diversified working poses + residual artificial neural network	3D
Seo et al (2021)[137]	Posture classification	Shape-based features + radial histograms of silhouettes + SVM	2D
Cai et al (2022)[82]	Pose estimation	Pose estimation + LSTM + multitask learning	2D

recognition. For instance, Liu et al. [125] utilized a silhouette-based approach, while Yang et al. [108] employed an SVM approach for classifying image features. However, with technological advancements, Yang [126] enhanced the method by incorporating data-driven scene parsing while retaining the dense trajectories.

In the last several years, the advancement of convolutional neural networks has led to reduced training costs for object detection algorithms, as well as improved detection speed and accuracy. These advancements have facilitated the application of object detection algorithms in construction activities recognition. Luo et al. [43] employed Faster R-CNN for the identification of construction workers and various entities depicted in images captured at construction sites. They employed a relevance network to recognize multiple construction activities by analyzing the spatial relationship between them. Similar studies include Fang et al. [38], who employed SORT and face recognition for the detection of non-certified work, and Li et al. [75], who used CenterNet to detect construction workers and objects to evaluate construction productivity while recognizing reinforcement assembly activities.

Extensive investigations in the domain of construction activity recognition have been conducted by the research group led by Luo Xiaochun and Li Heng, delving into the amalgamation of object detection algorithms with complementary methodologies. For instance, they combined YOLO v3 with SORT and 3D CNN to achieve spatial localization and construction activities recognition of construction workers [41]. In another study, the researchers utilized YOLO v3, SORT, KCF (Kernelized Correlation Filter), C3D, and CRF (Conditional Random Field) algorithms in multiple steps, enabling the recognition of construction activities performed by workers in groups [50]. Considering the spatio-temporal properties of construction activities, Li et al. [76] directly employed the Faster R-CNN algorithm to recognize construction activities in images, simultaneously detecting construction workers and determining the activities

performed by different operators based on the spatial relationship between individuals and activities. Bhokare et al. [127] also proposed a method based on YOLO v3 for direct recognition of construction activities.

To effectively integrate construction activities with spatio-temporal information, utilizing video clips for recognition is a more promising approach. Previous studies have proposed the use of multistream convolutional neural networks [45], [52]. In recent years, research methods based on videos have become more diverse. Roberts et al. [66] incorporated pose estimation algorithms into video-based construction activities recognition. Similarly, Cai et al. [128] introduced an attention direction estimation method to identify groups of construction workers and subsequently classified their activities using LSTM. Li et al [116] identified three activities of construction workers through YOLO and ST-GCN, which are throwing, operating and crossing. Torabi et al. [129] and Yang et al. [96] employed YOWO53 and Transformer, respectively, demonstrating high recognition accuracy on today's advanced computing hardware. Li and Li [79] applied Openpose and GAN to estimate the complete skeleton joints of construction workers under occlusion, then used ResNet to recognize construction activities.

Table 2 presents the comprehensive details of the reviewed literature in this section. The literature encompasses a broad spectrum of application scenarios, extending beyond the physical activities of construction workers to encompass various processes within civil engineering construction. These technologies hold significant potential for facilitating intelligent construction and enhancing project management.

C. OCCUPATIONAL HEALTH AND SAFETY BEHAVIOR MONITORING FOR CONSTRUCTION WORKERS

1) POSE ESTIMATION

The analysis of construction workers' body posture enables the understanding of their fatigue level, safety condition, and construction behavior. While algorithms like Openpose [130]

and Alphapose [131] offer rapid and accurate estimation of 2D skeleton joint points, it is important to note that 2D pose estimation may not fully capture the nuanced movements of construction workers. Consequently, researchers in the engineering field have placed greater emphasis on pose estimation in 3D space.

Earlier studies relied on 3D depth cameras as vision algorithms were not as advanced. Han et al. [68] employed a depth camera to acquire data and reconstruct the 3D coordinates of joint points belonging to construction workers. This approach facilitated the detection of perilous actions performed during the process of ascending ladders. Khosrowpour et al. [69] also employed a depth camera to acquire 3D joint point data and performed activity classification. Liu et al. [132] utilized a stereo video camera to capture 3D skeleton joint points. In an early stage approach, Seo et al. [133] transformed motion capture data obtained from vision-based methods for biomechanical analysis. In another study, Seo et al. [37] employed shape- and radial histogram-based features to estimate the posture of construction workers. Additionally, Seo et al. [134] compared multiple pose estimation methods and concluded that the sensor-based approach yielded the least error.

In past few years, the rise of deep learning methods has facilitated direct vision-based pose estimation. Zhang et al. [46] employed a multi-stage convolutional neural network to estimate the 3D pose of construction workers using monocular vision. Yu et al. [36] utilized a vision-based approach to obtain the 3D spatial pose of construction workers and further collected ergonomic information through an intelligent insole, enabling the analysis of their physical condition. In another study, Yu et al. [47] curated a construction posture dataset encompassing various postures to facilitate 3D posture estimation. Chu et al. [135] developed an ergonomic 3D posture assessment framework for construction workers, which integrated tracking algorithms, 2D detection, and 3D body generation. Kim et al. [136] constructed a synthetic dataset to enhance 3D pose estimation, while Tian et al. [44] compiled a construction worker motion dataset that can contribute to 3D pose recognition. Table 3 provides a list of typical studies in pose estimation.

2) UNSAFE BEHAVIOR RECOGNITION

During construction tasks, it is crucial to monitor the actions of construction workers as they may engage in hazardous movements or enter dangerous areas due to various factors. Such behaviors pose significant safety risks and require vigilant monitoring.

Computer vision techniques facilitated in extracting postural features related to unsafe behaviors of construction workers and assessing safety by analyzing their spatial location in relation to other entities. A research team from Huazhong University of Science and Technology developed a safety monitoring framework during the Wuhan Metro construction project, which effectively detects and predicts unsafe behaviors, including instances of construction workers illegally

walking on structural supports [32], [33], [53], [57]. The risk of falls is a major concern for construction workers, and Piao et al. [110] successfully employed Dynamic Bayesian Network, Openpose, and Faster R-CNN to detect fall risks step by step. During hoisting operations, workers face the risk of object strikes in dangerous areas. Chian et al. [138] utilized CenterNet to detect and track construction workers entering these hazardous zones. During the Covid-19 pandemic, Chian et al. [139] also employed CenterNet to monitor safe distances between construction workers. Ladder climbing poses a potential fall risk, which has been addressed by Han et al. [70], Anjum et al. [59], Ding et al. [58], and Chen et al. [140] using various computer vision methodologies. Furthermore, several computer vision-based studies have contributed to the recognition of unsafe behaviors, encompassing safety risk determination [71], [141], fall detection [142], and identification of unsafe actions [48], [65], [143], [144].

Construction sites often involve the simultaneous operation of various large machinery and random stacking of materials, increasing the risk of collisions between construction workers and these objects. The application of computer vision methods for assessing collision risks between workers and machinery or materials plays a vital role in ensuring the safety of construction workers. In earlier studies, Kim et al. [145] employed Gaussian mixture models to identify collision risks between construction workers and mechanical equipment. Fang et al. [51] utilized Mask R-CNN and Cascaded Pyramid Network to locate the relationships between construction workers, equipment, and materials using monocular vision. Zhang et al. [74] applied Faster R-CNN to detect workers and machines at construction sites, obtaining their location coordinates, and determining collision risks through fuzzy inference. Yan et al. [35] employed the Faster R-CNN algorithm to detect key points of construction trucks, followed by 3D reconstruction to establish distance recognition in the 3D scene. Shin and Kim [146] integrated YOLO v3 and the Openpose algorithm to identify collision risks between construction workers and trucks. Zhang and Ge [77] employed the latest Transformer technology to identify dynamic collision risks between construction workers and tower cranes. Additionally, other studies have contributed to collision risk identification [64], [147], [148] and fall risk identification [149]. Table 4 presents some of the typical research literature in this field.

3) PPE'S USAGE INSPECTION

Safety accidents are prevalent in engineering construction, with a significant number of casualties resulting from fall accidents and object strikes, as reported by the Ministry of Urban and Rural Construction of the People's Republic of China and the U.S. Bureau of Labor Statistics [152], [153], [154]. To mitigate the occurrence of such casualties, various countries have implemented mandatory policies and regulations regarding personal protective equipment (PPE). For instance, the State Administration of Quality Supervision,

TABLE 4. Typical research for unsafe behavior recognition.

Authors	Objectives	Algorithms
Fang et al (2019)[53]	Construction workers' unsafe behavior recognition (walking through structural supports)	Mask RCNN + Overlapping detection module
Ting et al (2021)[33]	Predict construction workers' unsafe behaviors (walking through structural supports)	SiamMask + improved Social-LSTM + PNPoly
Mei et al (2023)[119]	Construction workers' unsafe behavior recognition (enter dangerous area)	YOLO v5
Ding et al[58]	Construction worker unsafe behavior recognition (during ladder climbing)	CNN + LSTM
Hu et al[150]	Construction worker unsafe behaviors when working at height（lying, relying, jumping, throwing, with no helmet）	Faster R-CNN
Kim et al[145]	Evaluate risk of collision between construction workers and machinery	GMM + Kalman filtering + Fuzzy Inference
Zhang et al (2020)[74]	Evaluate risk of collision between construction workers and machinery	Faster R-CNN + Fuzzy Inference
Yan et al (2020)[35]	Evaluate risk of collision between construction workers and trucks	Faster RCNN + 3D bounding box reconstruction
Shin et al (2022)[146]	Evaluate risk of collision between construction workers and trucks	YOLO v3 + Openose + Perception-Based Safety Ellipse
Zhang and Ge (2022)[77]	Evaluate risk of collision between construction workers and crane	FairMOT + Transformer
Lee et al[151]	Check whether the behavior complies with safety regulations (Self-checking for PPE, responding to the warning alarm, checking the joint, step board inspection, and hooking the safety hook)	Skeleton-based method + spatio-temporal graph convolutional network

Inspection and Quarantine of China and the U.S. Occupational Safety and Health Administration have specified the use of hardhats during construction operations [155], [156], and the requirement for safety harnesses when working at heights. With the advancements in artificial intelligence technology, numerous computer vision-based studies have emerged to address the detection of PPE usage.

Before the rapid advancement of deep learning techniques, traditional methods for processing image features and machine learning algorithms were more commonly employed. For instance, Park et al. [157] utilized image processing methods to extract the spatial relationship between construction workers and hardhats. They subsequently employed SVM to detect and match this feature, determining whether construction workers were wearing hardhats or not. Mneymneh et al. [158], [159], [160] extracted various image features such as SURF and employed template matching and cascade classifiers for hardhat detection. In order to achieve better performance in safety vest detection, Seong et al. [161] compared various features and classifier algorithms. The results demonstrated that combining the C4.5 classifier with YCbCr and SVM classifier yielded superior outcomes. Similar to previous application areas, deep learning methods are rapidly replacing traditional techniques. Weili et al. [55] employed Faster R-CNN to detect the usage of safety harnesses by construction workers at heights. In another study conducted by the same group, Fang et al. [42] employed

Faster R-CNN for far-field monitoring at construction sites to detect hardhat wearing. The utilization of the Faster R-CNN algorithm has significantly enhanced image recognition accuracy [162]. Fu et al. [163] employed this algorithm for hardhat wearing detection, while Kamoona et al. [164] applied it to detect high-visibility vests.

As deep learning algorithms continue to advance, Delhi et al. [165] applied YOLO v3 for detecting the usage of hardhats and jackets among construction workers. YOLO v3 is an algorithm that has made significant strides in the field of object detection, leading to extensive research employing or enhancing this algorithm to detect the personal protective equipment (PPE) of construction workers [98], [102], [166], [167]. Among them, Nath et al. [166] developed three distinct models based on the YOLO v3 algorithm. In their second model, the algorithm simultaneously detected individual workers and verified PPE compliance using a single CNN framework, achieving a mean average precision (mAP) of 72.3%. Additionally, apart from v3, the method has been continuously updated to incorporate algorithms such as YOLO v4 and v5, which exhibit even better performance in hardhat detection [100], [101], [104], [168], [169], [170], [171], [172], [173]. Nguyen et al. [101] created a dataset comprising 11,978 images and tested various versions of YOLO v5. Experimental results indicated that the enhanced YOLO v5s demonstrated the best detection performance, achieving a precision of 0.74 on their custom dataset.

TABLE 5. Typical research for ppe's usage inspection.

Authors	Objectives	Algorithms	Performance
Park et al (2015)[157]	Monitoring hardhats' usage	using geometric and spatial relationship between individuals and hardhats + SVM	the precision for issuing the safety alerts reached 94.3%
Fang et al (2018)[42]	Detecting non-hardhat use	Faster R-CNN	> 90%
Kamoona et al (2019)[164]	Detect high-visibility vests' usage	Faster R-CNN + Anomaly Detection	98% F1 score
Nath et al (2020)[166]	Detect construction workers' PPE's usage (hardhat, vest)	YOLO v3	72.3% mAP
Nguyen et al (2023)[101]	Detect construction workers' PPE's usage (hardhats, masks, googles, gloves, suit, shoes)	Improved YOLO v5	0.74 Precision; 0.66 Recall; 0.70 F1 Score
Wang et al (2022)[104]	Monitoring hardhats' usage	Improved YOLO v5	91.6 % accuracy; 29 FPS detection speed
Li et al (2020)[185]	Monitoring hardhats' usage	SSD-MobileNet	95% precision; 77% recall
Yue et al (2022)[105]	Monitoring hardhats' usage	improved boosted random ferns	92.74% accuracy
Gu et al (2021)[183]	Monitoring hardhats' usage	YOLO v4 + Openpose	95.1% Precision
Cheng et al (2022)[111]	Detect PPE's usage (hardhat, vest)	Person ReID + PPE detection	87.33% Accuracy
Iannizzotto et al (2021)[182]	Detect PPE's usage (helmet, mask, vest)	YOLO-based DNN + Fuzzy Logic	>85%
Li et al (2022)[78]	PPE's standardized usage inspection (hardhat, safety harness)	YOLO v5 + Openpose + 1D-CNN	0.9467 Accuracy
Chen et al (2022)[184]	PPE's proper use detection (hardhat, glasses, dust mask)	NLP + YOLO v4 + graph representation	98.54% Precision; 88.79% Recall
Li et al (2023)[114]	PPE's usage detection	YOLO v5 +Strong SORT	95.4% mAP
Chern et al[179]	Detect PPE's usage (strap, hardhat, harness, hook)	Semantic / Depth estimation model	78.5% Accuracy / 86.22% Accuracy
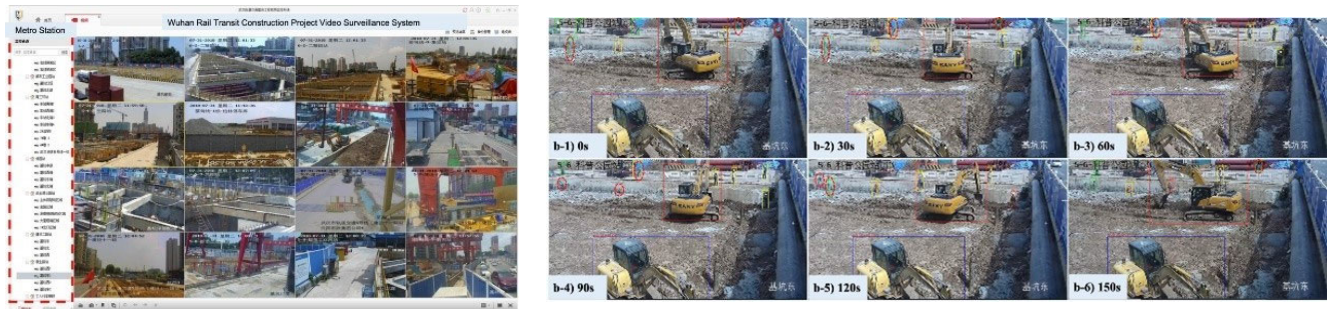
FIGURE 9. Wuhan metro monitoring system [187].

To enhance detection accuracy, researchers have explored various algorithms [61], [105], [174], [175], [176], [177], [178], [179] and integrated multiple algorithms to address the PPE monitoring challenge [60], [78], [111], [114], [115], [180], [181], [182], [183], [184]. Some scholars have also developed open-source PPE usage datasets for the broader academic community [107]. Xu et al. [175] proposed a novel detection strategy called the matching-recheck strategy and evaluated safety harness detection using the newly introduced Efficient YOLO v5 on a custom dataset, with 94% mAP.

Chen et al. [184] utilized YOLO v3 and Openpose to extract features related to PPE and individual joints. They employed Scene graph representation to describe the spatial relationship between joints and PPE, enabling the detection of multiple PPE items such as hardhats, dust masks, safety glasses, and safety belts. Li et al. [78] introduced the hierarchy of control and emphasized the importance of ensuring correct usage of PPE from an administrative control perspective. By combining the features extracted by YOLO v5 and Openpose, they employed 1D-CNN for classification and detection of

incorrect usage of hardhats and safety harnesses. Table 5 provides further details on some typical studies.

D. SUMMARY

This section presents an overview of computer vision applications for recognizing construction worker behavior on construction sites, primarily focusing on occupational health and safety risks. These articles represent significant advancements in utilizing computer vision technology to enhance both productivity and safety in engineering construction. Compared to other sensing and monitoring technologies, computer vision technology offers immense potential for development, cost-effectiveness, and a wide range of applications. According to survey results, CNNs emerge as the most widely employed technique among researchers, leveraging their ability to rapidly extract features compared to traditional machine learning algorithms. With the continuous evolution of CNNs, object detection technology has been extensively adopted, demonstrating strong adaptability in monitoring worker construction behavior. It has achieved breakthroughs in detection performance and real-time monitoring, enabling the tracking of construction workers' location, activities, and safety risks. As object detection technology continues to advance and Transformer technology gains prominence, computer vision technology holds the promise of making even greater contributions to the field's development.

VI. DISCUSSION

This article presents the latest research outcomes concerning the utilization of computer vision techniques for the recognition of construction workers' behavior. Considering prevailing trends and the global emphasis on the high-quality development of infrastructure, we anticipate that this research avenue will continue to garner considerable enthusiasm in the ensuing years. Within this section, we introduce the impact of relevant technologies on construction, undertaking a comparative analysis of the appropriateness of various algorithmic approaches. Subsequently, we summarize the current research gaps and delve into possible future research directions. Notably, we incorporate insights from generative AI models, a highly prevalent trend this year, providing an additional perspective and offering recommendations for researchers engaged in related pursuits.

A. IMPACT ON ENGINEERING AND CONSTRUCTION

Artificial intelligence (AI) has spearheaded the evolution of intelligent construction, enhancing the efficiency of construction operations. Some practitioners assert that the application of AI technology can yield cost savings of up to 20% in building construction [186]. Over recent years, with the continual refinement of algorithms and methods, a growing number of outcomes are finding practical applications in engineering. China, boasting the largest economy in the construction industry, has witnessed a growth in engineering applications. Noteworthy instances include the successful implementation of results from Luo Hanbin's team at Huazhong University

of Science and Technology in the Wuhan metro project [53], [56], [57], [58]. The established quality and safety management platform comprehensively monitors on-site construction conditions, provides real-time updates on quality and safety risks, and guides dynamic adjustments to the construction process [187] (refer to Figure 9). Additionally, Li et al. [75], [76] and Yang et al.'s outcomes [80] were integrated into a super high-rise construction project in the Donggang Business District of Dalian City in collaboration with China Construction Eighth Engineering Division. This integration led to a reduction in project managers' workload, indirectly contributing to a zero-accident occurrence throughout the project. While some results cited in this paper may not explicitly showcase the integration of applications in real projects, the data collected originates from actual construction sites and holds the potential for further development into tangible products [43], [84], [99]. Technology companies like Baidu and HikVision have launched commercial solutions for smart construction sites, aligning with the computer vision trend. In contrast to university scholars' emphasis on methodological innovation, commercial solutions prioritize functionality. In the current stage, the swift progress of AI technology has influenced public decision-making [8], [9], [10], and the concept of smart sites is expected to persist as a focal point in the engineering and construction field. Anticipating increased investment in research and development, it is foreseeable that novel methodologies will expedite technological integration with commercial solutions, further enriching the body of knowledge.

B. TECHNICAL SUITABILITY

One of the challenges hindering the widespread adoption of computer vision technology in the engineering construction field is the method's applicability. The construction site environment is intricate, exhibiting significant variations from one project to another. Factors such as lighting conditions and occlusion pose substantial impacts on the algorithm's recognition outcomes. Consequently, the algorithm's robustness becomes a critical consideration.

In the realm of computer vision, earlier algorithms predominantly relied on image processing and machine learning techniques, exemplified by methods like HOG features [72] and support vector machine (SVM) [108], [157]. However, these algorithms suffer from sluggish image feature extraction, inability to handle extensive data training, and exhibit low robustness and limited engineering applicability.

CNN, featuring weight sharing and automated feature extraction, exhibits the capacity to train effectively on extensive datasets. However, when the dataset primarily represents a singular scene composition, its applicability to other projects may be compromised [35], [50], [80]. In such instances, remedying this limitation necessitates secondary training. Conversely, if the dataset encompasses diverse lighting conditions, CNNs demonstrate efficacy in addressing such challenges. Notably, methods utilizing Faster R-CNN have yielded satisfactory results in scenarios involving varied

lighting conditions [42], [76]. The emergence of larger-scale datasets related to construction behavior further substantiates the success of algorithms like Faster R-CNN and YOLO [44], [48], [52], [107], [108]. In terms of methodology, there are also ways to address applicability issues such as occlusion [79]. As algorithms mature, the focus on improving robustness becomes pertinent. For instance, Park et al. [99] enhanced YOLO v5 with multiple strategies, enabling the detection of construction obscured and overlapped workers even in low light environments (Figure 10). Anticipating continued algorithmic and methodological advancements, coupled with ongoing dataset updates, the resolve of addressing the technique's robustness appears promising.
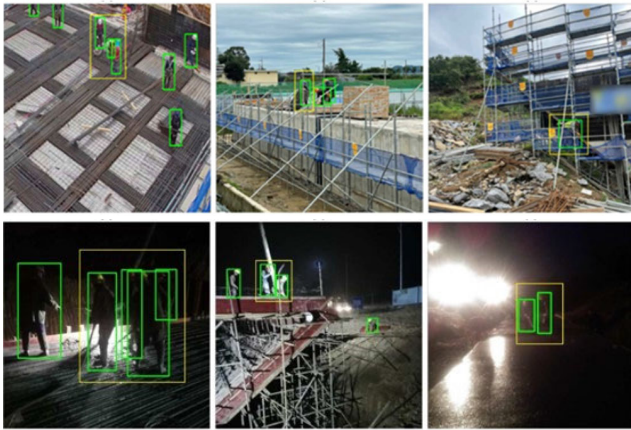
FIGURE 10. Detection results from SOC-YOLO model [99].

C. RESEARCH GAPS AND POSSIBLE FUTURE RESEARCH DIRECTIONS

1) TECHNICAL APPROACHES

Considering the profound influence of visual algorithm performance on the precision of behavior recognition, enhancing the accuracy and speed of visual algorithms remains a pivotal trajectory for future development in this domain. In the realm of CNN-based methods, accuracy enhancements can be achieved by modifying the network structure, while computational speed can be augmented through alterations in convolution techniques or the implementation of alternative methods, such as refining the loss function.

While CNN-based methods excel in local feature extraction, they often exhibit robustness issues to data noise and deformation. Presently, Transformer technology, grounded in a self-attention mechanism, demonstrates superior modeling ability and enhanced robustness. Its application has begun to permeate the field of engineering and construction. Nevertheless, the Transformer model's intricate structure and numerous parameters pose challenges. Hence, a plausible future direction involves developing a lightweight Transformer model tailored for the construction domain. Potential measures encompass the incorporation of lightweight

attention modules, utilization of convolution or pooling operations to reduce sequence length, among other strategies.

The integration of various computer vision algorithms can effectively tackle intricate behavior recognition tasks, constituting a methodologically sound approach. However, to align with the practical demands of engineering applications, future research could explore the establishment of a unified specification for operational environments and normalized data formats applicable to diverse computer vision algorithms. Alternatively, the development of novel algorithms or monitoring strategies geared towards multifunctional real-time monitoring could be pursued.

2) APPLICATION SCENARIOS

From the reviewed articles, the predominant focus has been on addressing occupational health and safety risks within the construction context. There is a noticeable dearth of studies pertaining to the tracking of construction workers and the recognition of construction activities. While understanding the location distribution and movement trajectory of workers proves valuable in visualizing the labor force's distribution status and aiding resource allocation, vision-based tracking encounters challenges in preventing the loss of targets across spatio-temporal scales. This predicament has led to the increased attention given to ReID (Re-Identification) algorithms in recent years. Future research directions could involve the development and exploration of ReID algorithms tailored for construction workers or the pursuit of alternative methods with comparable functionality.

Construction activities recognition can intuitively reflect the productivity status of the site, and should be given enough attention in future research, which can establish a large-scale construction activities database for the development of computer vision-based approaches.

Engineering construction sites present intricate scenarios, with environmental variations observed across different locations. Some research is carried out in laboratory settings for image acquisition and testing [78], while others collect data from specific engineering projects [76]. Consequently, the performance and effectiveness of these approaches cannot be assured when applied to different sites. To illustrate, considering Personal Protective Equipment (PPE) usage detection, numerous studies report accuracy rates surpassing 90%. However, these achievements are often based on self-constructed datasets for testing rather than utilizing open-source public datasets.

Future research endeavors should focus on constructing extensive public datasets encompassing diverse facets of worker construction behavior. This comprehensive dataset should span static images and dynamic videos, encompassing a spectrum of construction scenarios across various project types. It is strongly advised that researchers employ these public datasets during the development of algorithmic frameworks. Doing so would significantly contribute to the broader comprehension of algorithms among researchers, fostering collaboration and knowledge exchange.

3) GENERATIVE AI'S RESPONSE

In 2023, generative AI has emerged as a prominent field within AI, with ChatGPT [188] and NewBing [189] standing out as two highly popular generative AI models. ChatGPT, developed by OpenAI, is a conversational AI model built upon the GPT 3.5 series models, which have been trained using reinforcement learning techniques. NewBing, on the other hand, is an interactive tool offered by Microsoft that incorporates ChatGPT technology to provide more innovative insights and suggestions.

To gather Q&A information related to generative AI, Dr. Yang Zhaozheng from the University of Strasbourg adopted both NewBing and ChatGPT, directing specific inquiries to each model. It should be noted that the statements presented by Generative AI models do not reflect the perspectives of Dr. Yang or any of the contributors to this article.

Here are some suggestions for future research directions given by generative AI models:

a. Enhance object detection and recognition algorithms for construction workers in complex indoor environments.

b. Improvement is sought in accuracy and robustness despite variations in lighting conditions, occlusions, camera movements, and background noises.

c. Combining computer vision with technologies like deep learning, IoT, augmented reality, and digital twin for better performance in progress monitoring systems.

d. Progress monitoring will not only cover physical aspects but also psychological, social, and environmental factors affecting workers' well-being and behavior.

e. Standard benchmarks and evaluation metrics will be established for comparing different computer vision-based progress monitoring methods and systems.

f. Integration of multi-sensor data, such as wearable sensors or GPS, with visual data for a more comprehensive understanding of workers' behavior.

g. Development of worker-centered monitoring approaches while respecting privacy and autonomy.

h. Exploration of robust monitoring systems that can adapt to changing environments, including lighting conditions and weather.

i. Implementation of real-time monitoring and intervention systems to provide immediate feedback to workers and prevent potential hazards.

Moreover, generative AI models offer substantial potential benefits. Excellent design capabilities are exemplified by Cadence's Allegro X AI technology. This generative AI model facilitates the optimization of the system design process. Furthermore, the generative AI models' capacity holds promise for creating algorithms to recognize the construction workers' behaviors. When this technology matures, the AI model can seamlessly execute automated testing of product features, including compatibility testing for software applications. Additionally, in the realm of construction management, where decision-making is integral, generative AI models can prove invaluable by providing targeted advice.

VII. CONCLUSION

This paper presents a comprehensive review of the literature pertaining to the study of monitoring construction workers' behavior using computer vision technology. Initially, the researchers conducted a keyword search on the WOS platform, resulting in 596 papers from the past decade. Through manual inspection, 137 papers were selected for further analysis. The distribution of publication years indicates that the field is undergoing rapid development, with ongoing updates to technical methods. The authors employed VOSViewer software to identify research hotspots and highlight notable researchers and institutions in the field.

The paper provides an overview of computer vision techniques and technological advancements in worker behavior recognition. The reviewed literature is categorized into three sections based on application scenarios: " Detection, localization and tracking for construction workers," "Recognition of workers' construction activities," and" Occupational health and safety behavior monitoring." The abundance of literature demonstrates the preference of current researchers for employing computer vision to address safety concerns in engineering construction. The analysis reveals a trend wherein vision-based approaches begin to be adopted in real-world engineering, impacting decision-making processes in civil engineering. Initially, across different application scenarios, early research inclined towards addressing challenges through conventional image processing methods and machine-learning classification. However, as CNN and object detection techniques matured, vision methods exhibited enhanced robustness, adept at handling behavior recognition issues even in intricate scenarios, leading to the swift displacement of traditional methods. Recent years have witnessed an increase in the integration of multiple algorithms to address real-world challenges, marking a prevalent trend. Additionally, the emergence of transformer technology holds the promise of transcending existing algorithmic limitations in this domain.

The authors conclude by summarizing research gaps and outlining future directions. Further optimization is required in terms of accuracy and speed, especially when integrating multiple algorithms, necessitating the establishment of a unified framework platform and integration strategy. In terms of application scenarios, improvements can be made to enhance scene applicability and diversify data sources. Furthermore, insights from the generative AI model underscore the importance of considering the privacy and autonomy of construction workers, suggesting their involvement in monitoring projects. This paper serves as a reference for engineers and researchers, aiming to contribute to the field of engineering and construction while benefiting fellow scholars' scientific studies.

REFERENCES

[1] Y. Zhang, T. Wang, and K. Yuen, "Construction site information decentralized management using blockchain and smart contracts," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 37, no. 11, pp. 1450–1467, Sep. 2022, doi: 10.1111/mice.12804.

[2] M. Yihua and X. Tuo, "Research of 4M1E's effect on engineering quality based on structural equation model," *Syst. Eng. Proc.*, vol. 1, pp. 213–220, Jan. 2011, doi: 10.1016/j.sepro.2011.08.034.

[3] H. Zhou, Y. Zhao, Q. Shen, L. Yang, and H. Cai, "Risk assessment and management via multi-source information fusion for undersea tunnel construction," *Autom. Construct.*, vol. 111, Mar. 2020, Art. no. 103050, doi: 10.1016/j.autcon.2019.103050.

[4] C. Sheehan, R. Donohue, T. Shea, B. Cooper, and H. D. Cieri, "Leading and lagging indicators of occupational health and safety: The moderating role of safety leadership," *Accident Anal. Prevention*, vol. 92, pp. 130–138, Jul. 2016, doi: 10.1016/j.aap.2016.03.018.

[5] *Opinions on Promoting the Sustainable and Healthy Development of the Construction Industry—General Office of the State Council*. Accessed: Apr. 20, 2023. [Online]. Available: http://www.gov.cn/zhengce/zhengceku/2017-02/24/content_5170625.htm

[6] *Bringing Innovation to the Worksite with' Smart Construction—The Government of Japan—JapanGov*. Accessed: Apr. 20, 2023. [Online]. Available: https://www.japan.go.jp/tomodachi/2017/autumn2017/power_of_innovation.html

[7] T. Dawood, Z. Zhu, and T. Zayed, "Computer vision-based model for moisture marks detection and recognition in subway networks," *J. Comput. Civil Eng.*, vol. 32, no. 2, Mar. 2018, Art. no. 04017079, doi: 10.1061/(ASCE)CP.1943-5487.0000728.

[8] L. D. Xu, E. L. Xu, and L. Li, "Industry 4.0: State of the art and future trends," *Int. J. Prod. Res.*, vol. 56, no. 8, pp. 2941–2962, Mar. 2018, doi: 10.1080/00207543.2018.1444806.

[9] GOV.UK. *Construction 2025: Strategy*. Accessed: Apr. 20, 2023. [Online]. Available: https://www.gov.uk/government/publications/construction-2025-strategy

[10] *Notice on the Issuance of the '14th Five-Year Plan' for the Development of the Construction Industry*. Accessed: Apr. 20, 2023. [Online]. Available: http://www.gov.cn/zhengce/zhengceku/2022-01/27/content_5670687.htm

[11] A. Montaser and O. Moselhi, "RFID indoor location identification for construction projects," *Autom. Construct.*, vol. 39, pp. 167–179, Apr. 2014, doi: 10.1016/j.autcon.2013.06.012.

[12] N. D. Nath, R. Akhavian, and A. H. Behzadan, "Ergonomic analysis of construction worker's body postures using wearable mobile sensors," *Appl. Ergonom.*, vol. 62, pp. 107–117, Jul. 2017, doi: 10.1016/j.apergo.2017.02.007.

[13] L. Joshua and K. Varghese, "Accelerometer-based activity recognition in construction," *J. Comput. Civil Eng.*, vol. 25, no. 5, pp. 370–379, Sep. 2011, doi: 10.1061/(asce)cp.1943-5487.0000097.

[14] Y.-C. Lee, M. Scarpiniti, and A. Uncini, "Advanced sound classifiers and performance analyses for accurate audio-based construction project monitoring," *J. Comput. Civil Eng.*, vol. 34, no. 5, Sep. 2020, Art. no. 04020030, doi: 10.1061/(ASCE)CP.1943-5487.0000911.

[15] Y. Zhang, Y.-Q. Ni, X. Jia, and Y.-W. Wang, "Identification of concrete surface damage based on probabilistic deep learning of images," *Autom. Construct.*, vol. 156, Dec. 2023, Art. no. 105141, doi: 10.1016/j.autcon.2023.105141.

[16] Y. Tang, Z. Huang, Z. Chen, M. Chen, H. Zhou, H. Zhang, and J. Sun, "Novel visual crack width measurement based on backbone double-scale features for improved detection automation," *Eng. Struct.*, vol. 274, Jan. 2023, Art. no. 115158, doi: 10.1016/j.engstruct.2022.115158.

[17] Z. Wu, Y. Tang, B. Hong, B. Liang, and Y. Liu, "Enhanced precision in dam crack width measurement: Leveraging advanced lightweight network identification for pixel-level accuracy," *Int. J. Intell. Syst.*, vol. 2023, pp. 1–16, Sep. 2023, doi: 10.1155/2023/9940881.

[18] B. Xie, L. Zhang, M. Ding, W. Li, and X. Zhao, "Strain measurement based on cooperative operation with different smartphones," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 38, no. 2, pp. 241–252, Sep. 2022, doi: 10.1111/mice.12919.

[19] Y. Tang, M. Zhu, Z. Chen, C. Wu, B. Chen, C. Li, and L. Li, "Seismic performance evaluation of recycled aggregate concrete-filled steel tubular columns with field strain detected via a novel mark-free vision method," *Structures*, vol. 37, pp. 426–441, Mar. 2022, doi: 10.1016/j.istruc.2021.12.055.

[20] Y. Zhang, X. Zhao, and P. Liu, "Multi-point displacement monitoring based on full convolutional neural network and smartphone," *IEEE Access*, vol. 7, pp. 139628–139634, 2019, doi: 10.1109/ACCESS.2019.2943599.

[21] Y. Zhang and K.-V. Yuen, "Bolt damage identification based on orientation-aware center point estimation network," *Struct. Health Monitor.*, vol. 21, no. 2, pp. 438–450, Mar. 2021, doi: 10.1177/14759217211004243.

[22] W. Ma, "Technical framework of energy-saving construction management of intelligent building based on computer vision algorithm," *Soft Comput.*, pp. 1–11, May 2023, doi: 10.1007/s00500-023-08424-1.

[23] K. Mostafa and T. Hegazy, "Review of image-based analysis and applications in construction," *Autom. Construct.*, vol. 122, Feb. 2021, Art. no. 103516, doi: 10.1016/j.autcon.2020.103516.

[24] M. Zhang, R. Shi, and Z. Yang, "A critical review of vision-based occupational health and safety monitoring of construction site workers," *Saf. Sci.*, vol. 126, Jun. 2020, Art. no. 104658, doi: 10.1016/j.ssci.2020.104658.

[25] W. Fang, P. E. D. Love, H. Luo, and L. Ding, "Computer vision for behaviour-based safety in construction: A review and future directions," *Adv. Eng. Informat.*, vol. 43, Jan. 2020, Art. no. 100980, doi: 10.1016/j.aei.2019.100980.

[26] F. Luo, R. Y. M. Li, M. J. C. Crabbe, and R. Pu, "Economic development and construction safety research: A bibliometrics approach," *Saf. Sci.*, vol. 145, Jan. 2022, Art. no. 105519, doi: 10.1016/j.ssci.2021.105519.

[27] L. Zeng and R. Y. M. Li, "Construction safety and health hazard awareness in web of science and Weibo between 1991 and 2021," *Saf. Sci.*, vol. 152, Aug. 2022, Art. no. 105790, doi: 10.1016/j.ssci.2022.105790.

[28] D. Fang, Y. Huang, H. Guo, and H. W. Lim, "LCB approach for construction safety," *Saf. Sci.*, vol. 128, Aug. 2020, Art. no. 104761, doi: 10.1016/j.ssci.2020.104761.

[29] U. A. Bukar, M. S. Sayeed, S. F. A. Razak, S. Yogarayan, O. A. Amodu, and R. A. R. Mahmood, "A method for analyzing text using VOSviewer," *MethodsX*, vol. 11, Dec. 2023, Art. no. 102339, doi: 10.1016/j.mex.2023.102339.

[30] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, Aug. 2010, doi: 10.1007/s11192-009-0146-3.

[31] Y. Yu, X. Yang, H. Li, X. Luo, H. Guo, and Q. Fang, "Joint-level vision-based ergonomic assessment tool for construction workers," *J. Construct. Eng. Manage.*, vol. 145, no. 5, May 2019, Art. no. 04019025, doi: 10.1061/(ASCE)CO.1943-7862.0001647.

[32] H. Wu, B. Zhong, H. Li, P. Love, X. Pan, and N. Zhao, "Combining computer vision with semantic reasoning for on-site safety management in construction," *J. Building Eng.*, vol. 42, Oct. 2021, Art. no. 103036, doi: 10.1016/j.jobe.2021.103036.

[33] T. Kong, W. Fang, P. E. D. Love, H. Luo, S. Xu, and H. Li, "Computer vision and long short-term memory: Learning to predict unsafe behaviour in construction," *Adv. Eng. Informat.*, vol. 50, Oct. 2021, Art. no. 101400, doi: 10.1016/j.aei.2021.101400.

[34] Y. Yu, H. Li, X. Yang, L. Kong, X. Luo, and A. Y. L. Wong, "An automatic and non-invasive physical fatigue assessment method for construction workers," *Autom. Construct.*, vol. 103, pp. 1–12, Jul. 2019, doi: 10.1016/j.autcon.2019.02.020.

[35] X. Yan, H. Zhang, and H. Li, "Computer vision-based recognition of 3D relationship between construction entities for monitoring struck-by accidents," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 35, no. 9, pp. 1023–1038, Sep. 2020, doi: 10.1111/mice.12536.

[36] Y. Yu, H. Li, W. Umer, C. Dong, X. Yang, M. Skitmore, and A. Y. L. Wong, "Automatic biomechanical workload estimation for construction workers by computer vision and smart insoles," *J. Comput. Civil Eng.*, vol. 33, no. 3, May 2019, Art. no. 04019010, doi: 10.1061/(ASCE)CP.1943-5487.0000827.

[37] J. Seo, K. Yin, and S. Lee, "Automated postural ergonomic assessment using a computer vision-based posture classification," in *Proc. Construct. Res. Congr.*, 2016, p. 818, doi: 10.1061/9780784479827.082.

[38] Q. Fang, H. Li, X. Luo, L. Ding, T. M. Rose, W. An, and Y. Yu, "A deep learning-based method for detecting non-certified work on construction sites," *Adv. Eng. Informat.*, vol. 35, pp. 56–68, Jan. 2018, doi: 10.1016/j.aei.2018.01.001.

[39] X. Yan, H. Zhang, and H. Li, "Estimating worker-centric 3D spatial crowdedness for construction safety management using a single 2D camera," *J. Comput. Civil Eng.*, vol. 33, no. 5, Sep. 2019, Art. no. 04019030, doi: 10.1061/(ASCE)CP.1943-5487.0000844.

[40] H. Liu, G. Wang, T. Huang, P. He, M. Skitmore, and X. Luo, "Manifesting construction activity scenes via image captioning," *Autom. Construct.*, vol. 119, Nov. 2020, Art. no. 103334, doi: 10.1016/j.autcon.2020.103334.

[41] X. Luo, H. Li, H. Wang, Z. Wu, F. Dai, and D. Cao, "Vision-based detection and visualization of dynamic workspaces," *Autom. Construct.*, vol. 104, pp. 1–13, Aug. 2019, doi: 10.1016/j.autcon.2019.04.001.

[42] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, T. M. Rose, and W. An, "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos," *Autom. Construct.*, vol. 85, pp. 1–9, Jan. 2018, doi: 10.1016/j.autcon.2017.09.018.

[43] X. Luo, H. Li, D. Cao, F. Dai, J. Seo, and S. Lee, "Recognizing diverse construction activities in site images via relevance networks of construction-related objects detected by convolutional neural networks," *J. Comput. Civil Eng.*, vol. 32, no. 3, May 2018, Art. no. 04018012, doi: 10.1061/(ASCE)CP.1943-5487.0000756.

[44] Y. Tian, H. Li, H. Cui, and J. Chen, "Construction motion data library: An integrated motion dataset for on-site activity recognition," *Sci. Data*, vol. 9, no. 1, pp. 1–15, Nov. 2022, doi: 10.1038/s41597-022-01841-1.

[45] X. Luo, H. Li, D. Cao, Y. Yu, X. Yang, and T. Huang, "Towards efficient and objective work sampling: Recognizing workers' activities in site surveillance videos with two-stream convolutional networks," *Autom. Construct.*, vol. 94, pp. 360–370, Oct. 2018, doi: 10.1016/j.autcon.2018.07.011.

[46] H. Zhang, X. Yan, and H. Li, "Ergonomic posture recognition using 3D view-invariant features from single ordinary camera," *Autom. Construct.*, vol. 94, pp. 1–10, Oct. 2018, doi: 10.1016/j.autcon.2018.05.033.

[47] Y. Yu, H. Li, J. Cao, and X. Luo, "Three-dimensional working pose estimation in industrial scenarios with monocular camera," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1740–1748, Feb. 2021, doi: 10.1109/JIOT.2020.3014930.

[48] R. Xiong, Y. Song, H. Li, and Y. Wang, "Onsite video mining for construction hazards identification with visual relationships," *Adv. Eng. Informat.*, vol. 42, Oct. 2019, Art. no. 100966, doi: 10.1016/j.aei.2019.100966.

[49] X. Luo, H. Li, X. Yang, Y. Yu, and D. Cao, "Capturing and understanding workers' activities in far-field surveillance videos with deep action recognition and Bayesian nonparametric learning," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 34, no. 4, pp. 333–351, Oct. 2018, doi: 10.1111/mice.12419.

[50] X. Luo, H. Li, Y. Yu, C. Zhou, and D. Cao, "Combining deep features and activity context to improve recognition of activities of workers in groups," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 35, no. 9, pp. 965–978, Feb. 2020, doi: 10.1111/mice.12538.

[51] Q. Fang, H. Li, X. Luo, C. Li, and W. An, "A sematic and prior-knowledge-aided monocular localization method for construction-related entities," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 35, no. 9, pp. 979–996, Mar. 2020, doi: 10.1111/mice.12541.

[52] H. Luo, C. Xiong, W. Fang, P. E. D. Love, B. Zhang, and X. Ouyang, "Convolutional neural networks: Computer vision-based workforce activity assessment in construction," *Autom. Construct.*, vol. 94, pp. 282–289, Oct. 2018, doi: 10.1016/j.autcon.2018.06.007.

[53] W. Fang, B. Zhong, N. Zhao, P. E. D. Love, H. Luo, J. Xue, and S. Xu, "A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network," *Adv. Eng. Informat.*, vol. 39, pp. 170–177, Jan. 2019, doi: 10.1016/j.aei.2018.12.005.

[54] R. Wei, P. E. D. Love, W. Fang, H. Luo, and S. Xu, "Recognizing people's identity in construction sites with computer vision: A spatial and temporal attention pooling network," *Adv. Eng. Informat.*, vol. 42, Oct. 2019, Art. no. 100981, doi: 10.1016/j.aei.2019.100981.

[55] W. Fang, L. Ding, H. Luo, and P. E. D. Love, "Falls from heights: A computer vision-based approach for safety harness detection," *Autom. Construct.*, vol. 91, pp. 53–61, Jul. 2018, doi: 10.1016/j.autcon.2018.02.018.

[56] W. Fang, L. Ding, B. Zhong, P. E. D. Love, and H. Luo, "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach," *Adv. Eng. Informat.*, vol. 37, pp. 139–149, Aug. 2018, doi: 10.1016/j.aei.2018.05.003.

[57] J. Liu, W. Fang, P. E. D. Love, T. Hartmann, H. Luo, and L. Wang, "Detection and location of unsafe behaviour in digital images: A visual grounding approach," *Adv. Eng. Informat.*, vol. 53, Aug. 2022, Art. no. 101688, doi: 10.1016/j.aei.2022.101688.

[58] L. Ding, W. Fang, H. Luo, P. E. D. Love, B. Zhong, and X. Ouyang, "A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory," *Autom. Construct.*, vol. 86, pp. 118–124, Feb. 2018, doi: 10.1016/j.autcon.2017.11.002.

[59] S. Anjum, N. Khan, R. Khalid, M. Khan, D. Lee, and C. Park, "Fall prevention from ladders utilizing a deep learning-based height assessment method," *IEEE Access*, vol. 10, pp. 36725–36742, 2022, doi: 10.1109/ACCESS.2022.3164676.

[60] M. Khan, R. Khalid, S. Anjum, S. V.-T. Tran, and C. Park, "Fall prevention from scaffolding using computer vision and IoT-based monitoring," *J. Construct. Eng. Manage.*, vol. 148, no. 7, pp. 1–15, Jul. 2022, doi: 10.1061/%28ASCE%29CO.1943-7862.0002278.

[61] J. Lim, D. G. Jung, C. Park, and D. Y. Kim, "Computer vision process development regarding worker's safety harness and hook to prevent fall accidents: Focused on system scaffolds in South Korea," *Adv. Civil Eng.*, vol. 2022, pp. 1–12, Jul. 2022, doi: 10.1155/2022/4678479.

[62] H. Son, H. Choi, H. Seong, and C. Kim, "Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks," *Autom. Construct.*, vol. 99, pp. 27–38, Mar. 2019, doi: 10.1016/j.autcon.2018.11.033.

[63] H. Son and C. Kim, "Integrated worker detection and tracking for the safe operation of construction machinery," *Autom. Construct.*, vol. 126, Jun. 2021, Art. no. 103670, doi: 10.1016/j.autcon.2021.103670.

[64] H. Son, H. Seong, H. Choi, and C. Kim, "Real-time vision-based warning system for prevention of collisions between workers and heavy equipment," *J. Comput. Civil Eng.*, vol. 33, no. 5, p. 04019029, Sep. 2019, doi: 10.1061/(ASCE)CP.1943-5487.0000845.

[65] N. Khan, M. R. Saleem, D. Lee, M.-W. Park, and C. Park, "Utilizing safety rule correlation for mobile scaffolds monitoring leveraging deep convolution neural networks," *Comput. Ind.*, vol. 129, Aug. 2021, Art. no. 103448, doi: 10.1016/j.compind.2021.103448.

[66] D. Roberts, W. T. Calderon, S. Tang, and M. Golparvar-Fard, "Vision-based construction worker activity analysis informed by body posture," *J. Comput. Civil Eng.*, vol. 34, no. 4, p. 04020017, Jul. 2020, doi: 10.1061/(ASCE)CP.1943-5487.0000898.

[67] S. Tang, D. Roberts, and M. Golparvar-Fard, "Human-object interaction recognition for automatic construction site safety inspection," *Autom. Construct.*, vol. 120, Dec. 2020, Art. no. 103356, doi: 10.1016/j.autcon.2020.103356.

[68] S. Han and S. Lee, "A vision-based motion capture and recognition framework for behavior-based safety management," *Autom. Construct.*, vol. 35, pp. 131–141, Nov. 2013, doi: 10.1016/j.autcon.2013.05.001.

[69] A. Khosrowpour, J. C. Niebles, and M. Golparvar-Fard, "Vision-based workface assessment using depth images for activity analysis of interior construction operations," *Autom. Construct.*, vol. 48, pp. 74–87, Dec. 2014, doi: 10.1016/j.autcon.2014.08.003.

[70] S. Han, S. Lee, and F. Peña-Mora, "Vision-based detection of unsafe actions of a construction worker: Case study of ladder climbing," *J. Comput. Civil Eng.*, vol. 27, no. 6, pp. 635–644, Nov. 2013, doi: 10.1061/(asce)cp.1943-5487.0000279.

[71] S. Tang and M. Golparvar-Fard, "Machine learning-based risk analysis for construction worker safety from ubiquitous site photos and videos," *J. Comput. Civil Eng.*, vol. 35, no. 6, Nov. 2021, Art. no. 04021020, doi: 10.1061/(ASCE)CP.1943-5487.0000979.

[72] M. Memarzadeh, M. Golparvar-Fard, and J. C. Niebles, "Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors," *Autom. Construct.*, vol. 32, pp. 24–37, Jul. 2013, doi: 10.1016/j.autcon.2012.12.002.

[73] S. Tang, M. Golparvar-Fard, M. Naphade, and M. M. Gopalakrishna, "Video-based motion trajectory forecasting method for proactive construction safety monitoring systems," *J. Comput. Civil Eng.*, vol. 34, no. 6, Nov. 2020, Art. no. 04020041, doi: 10.1061/(ASCE)CP.1943-5487.0000923.

[74] M. Zhang, Z. Cao, Z. Yang, and X. Zhao, "Utilizing computer vision and fuzzy inference to evaluate level of collision safety for workers and equipment in a dynamic environment," *J. Construct. Eng. Manage.*, vol. 146, no. 6, Jun. 2020, Art. no. 04020051, doi: 10.1061/(ASCE)CO.1943-7862.0001802.

[75] J. Li, X. Zhao, G. Zhou, M. Zhang, D. Li, and Y. Zhou, "Evaluating the work productivity of assembling reinforcement through the objects detected by deep learning," *Sensors*, vol. 21, no. 16, p. 5598, Aug. 2021, doi: 10.3390/s21165598.

[76] J. Li, G. Zhou, D. Li, M. Zhang, and X. Zhao, "Recognizing workers' construction activities on a reinforcement processing area through the position relationship of objects detected by faster R-CNN," *Eng., Construct. Architectural Manage.*, vol. 30, no. 4, pp. 1657–1678, Jan. 2022, doi: 10.1108/ECAM-04-2021-0312.

[77] M. Zhang and S. Ge, "Vision and trajectory-based dynamic collision prewarning mechanism for tower cranes," *J. Construct. Eng. Manage.*, vol. 148, no. 7, Jul. 2022, Art. no. 04022057, doi: 10.1061/(ASCE)CO.1943-7862.0002309.

[78] J. Li, X. Zhao, G. Zhou, and M. Zhang, "Standardized use inspection of workers' personal protective equipment based on deep learning," *Saf. Sci.*, vol. 150, Jun. 2022, Art. no. 105689, doi: 10.1016/j.ssci.2022.105689.

[79] Z. Li and D. Li, "Action recognition of construction workers under occlusion," *J. Building Eng.*, vol. 45, Jan. 2022, Art. no. 103352, doi: 10.1016/j.jobe.2021.103352.

[80] Z. Yang, Y. Yuan, M. Zhang, X. Zhao, Y. Zhang, and B. Tian, "Safety distance identification for crane drivers based on mask R-CNN," *Sensors*, vol. 19, no. 12, p. 2789, Jun. 2019, doi: 10.3390/s19122789.

[81] J. Cai and H. Cai, "Robust hybrid approach of vision-based tracking and radio-based identification and localization for 3D tracking of multiple construction workers," *J. Comput. Civil Eng.*, vol. 34, no. 4, Jul. 2020, Art. no. 04020021, doi: 10.1061/(ASCE)CP.1943-5487.0000901.

[82] J. Cai, X. Li, X. Liang, W. Wei, and S. Li, "Construction worker ergonomic assessment via LSTM-based multi-task learning framework," in *Proc. Construct. Res. Congr.*, 2022, p. 224, doi: 10.1061/9780784483961.023.

[83] J. Cai, Y. Zhang, L. Yang, H. Cai, and S. Li, "A context-augmented deep learning approach for worker trajectory prediction on unstructured and dynamic construction sites," *Adv. Eng. Informat.*, vol. 46, Oct. 2020, Art. no. 101173, doi: 10.1016/j.aei.2020.101173.

[84] J. Cai, L. Yang, Y. Zhang, S. Li, and H. Cai, "Multitask learning method for detecting the visual focus of attention of construction workers," *J. Construct. Eng. Manage.*, vol. 147, no. 7, Jul. 2021, Art. no. 04021063, doi: 10.1061/(ASCE)CO.1943-7862.0002071.

[85] J. Canny, "A computational approach to edge detection," in *Readings in Computer Vision*, M. A. Fischler and O. Firschein, Eds. San Francisco, CA, USA: Morgan Kaufmann, 1987, pp. 184–203, doi: 10.1016/B978-0-08-051581-6.50024-6.

[86] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[87] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, Mar. 1960, doi: 10.1115/1.3662552.

[88] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004, doi: 10.1023/B%3AVISI.0000022288.19776.77.

[89] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.

[90] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," 2014, *arXiv:1409.2329*.

[91] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," 2015, *arXiv:1506.01497*.

[92] *GitHub—Ultralytics/YOLOv5: YOLOv5 in PyTorch > ONNX > CoreML > TFLite*. Accessed: Apr. 21, 2023. [Online]. Available: https://github.com/ultralytics/yolov5

[93] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.

[94] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2018, *arXiv:1611.07004*.

[95] MIT Press. *Long Short-Term Memory | Neural Computation*. Accessed: Apr. 24, 2023. [Online]. Available: https://direct.mit.edu/neco/article-abstract/9/8/1735/6109/Long-Short-Term-Memory?redirectedFrom=full text

[96] M. Yang, C. Wu, Y. Guo, R. Jiang, F. Zhou, J. Zhang, and Z. Yang, "Transformer-based deep learning model and video dataset for unsafe action identification in construction projects," *Autom. Construct.*, vol. 146, Feb. 2023, Art. no. 104703, doi: 10.1016/j.autcon.2022.104703.

[97] W. Liu, Y. Shao, S. Zhai, Z. Yang, and P. Chen, "Computer vision-based tracking of workers in construction sites based on MDNet," *IEICE Trans. Inf. Syst.*, vol. 106, no. 5, pp. 653–661, 2023, doi: 10.1587/transinf.2022DLP0045.

[98] L. Huang, Q. Fu, M. He, D. Jiang, and Z. Hao, "Detection algorithm of safety helmet wearing based on deep learning," *Concurrency Comput., Pract. Exper.*, vol. 33, no. 13, p. e6234, Jul. 2021, doi: 10.1002/cpe.6234.

[99] M. Park, D. Q. Tran, J. Bak, and S. Park, "Small and overlapping worker detection at construction sites," *Autom. Construct.*, vol. 151, Jul. 2023, Art. no. 104856, doi: 10.1016/j.autcon.2023.104856.

[100] K. Han and X. Zeng, "Deep learning-based workers safety helmet wearing detection on construction sites using multi-scale features," *IEEE Access*, vol. 10, pp. 718–729, 2022, doi: 10.1109/ACCESS.2021.3138407.

[101] N. Ngoc-Thoan, D.-Q.-T. Bui, C. N. N. Tran, and D.-H. Tran, "Improved detection network model based on YOLOv5 for warning safety in construction sites," *Int. J. Construct. Manage.*, pp. 1–11, Feb. 2023, doi: 10.1080/15623599.2023.2171836.

[102] H. Peng and Z. Zhang, "Helmet wearing recognition of construction workers using convolutional neural network," *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–8, Apr. 2022, doi: 10.1155/2022/4739897.

[103] J. Jain, R. Parekh, J. Parekh, S. Shah, and P. Kanani, "Helmet detection and license plate extraction using machine learning and computer vision," in *Proc. Int. Conf. Cognition Recongion*, 2023, pp. 258–268, doi: 10.1007/978-3-031-22405-8_20.

[104] L. Wang, Y. Cao, S. Wang, X. Song, S. Zhang, J. Zhang, and J. Niu, "Investigation into recognition algorithm of helmet violation based on YOLOv5-CBAM-DCN," *IEEE Access*, vol. 10, pp. 60622–60632, 2022, doi: 10.1109/ACCESS.2022.3180796.

[105] S. Yue, Q. Zhang, D. Shao, Y. Fan, and J. Bai, "Safety helmet wearing status detection based on improved boosted random ferns," *Multimedia Tools Appl.*, vol. 81, no. 12, pp. 16783–16796, May 2022, doi: 10.1007/s11042-022-12014-y.

[106] H.-P. Wan, W.-J. Zhang, H.-B. Ge, Y. Luo, and M. D. Todd, "Improved vision-based method for detection of unauthorized intrusion by construction sites workers," *J. Construct. Eng. Manag.*, vol. 149, no. 7, Jul. 2023, Art. no. 04023040, doi: 10.1061/JCEMD4.COENG-13294.

[107] J. Wu, N. Cai, W. Chen, H. Wang, and G. Wang, "Automatic detection of hardhats Worn by construction personnel: A deep learning approach and benchmark dataset," *Autom. Construct.*, vol. 106, Oct. 2019, Art. no. 102894, doi: 10.1016/j.autcon.2019.102894.

[108] J. Yang, Z. Shi, and Z. Wu, "Vision-based action recognition of construction workers using dense trajectories," *Adv. Eng. Informat.*, vol. 30, no. 3, pp. 327–336, Aug. 2016, doi: 10.1016/j.aei.2016.04.009.

[109] B. Xiao, H. Xiao, J. Wang, and Y. Chen, "Vision-based method for tracking workers by integrating deep learning instance segmentation in off-site construction," *Autom. Construct.*, vol. 136, Apr. 2022, Art. no. 104148, doi: 10.1016/j.autcon.2022.104148.

[110] Y. Piao, W. Xu, T.-K. Wang, and J.-H. Chen, "Dynamic fall risk assessment framework for construction workers based on dynamic Bayesian network and computer vision," *J. Construct. Eng. Manage.*, vol. 147, no. 12, Dec. 2021, Art. no. 04021171, doi: 10.1061/(ASCE)CO.1943-7862.0002200.

[111] J. P. Cheng, P. K.-Y. Wong, H. Luo, M. Wang, and P. H. Leung, "Vision-based monitoring of site safety compliance based on worker re-identification and personal protective equipment classification," *Autom. Construct.*, vol. 139, Jul. 2022, Art. no. 104312, doi: 10.1016/j.autcon.2022.104312.

[112] Q. Zhang, Z. Wang, B. Yang, K. Lei, B. Zhang, and B. Liu, "Reidentification-based automated matching for 3D localization of workers in construction sites," *J. Comput. Civil Eng.*, vol. 35, no. 6, Nov. 2021, Art. no. 04021019, doi: 10.1061/(asce)cp.1943-5487.0000975.

[113] H. Deng, Z. Ou, and Y. Deng, "Multi-angle fusion-based safety status analysis of construction workers," *Int. J. Environ. Res. Public Health*, vol. 18, no. 22, p. 11815, Nov. 2021, doi: 10.3390/ijerph182211815.

[114] F. Li, Y. Chen, M. Hu, M. Luo, and G. Wang, "Helmet-wearing tracking detection based on StrongSORT," *Sensors*, vol. 23, no. 3, p. 1682, Feb. 2023, doi: 10.3390/s23031682.

[115] Y.-R. Lee, S.-H. Jung, K.-S. Kang, H.-C. Ryu, and H.-G. Ryu, "Deep learning-based framework for monitoring wearing personal protective equipment on construction sites," *J. Comput. Design Eng.*, vol. 10, no. 2, pp. 905–917, Mar. 2023, doi: 10.1093/jcde/qwad019.

[116] P. Li, F. Wu, S. Xue, and L. Guo, "Study on the interaction behaviors identification of construction workers based on ST-GCN and YOLO," *Sensors*, vol. 23, no. 14, p. 6318, Jul. 2023, doi: 10.3390/s23146318.

[117] I. Jeelani, K. Asadi, H. Ramshankar, K. Han, and A. Albert, "Real-time vision-based worker localization & hazard detection for construction," *Autom. Construct.*, vol. 121, Jan. 2021, Art. no. 103448, doi: 10.1016/j.autcon.2020.103448.

[118] H. Huang, H. Hu, F. Xu, Z. Zhang, and Y. Tao, "Skeleton-based automatic assessment and prediction of intrusion risk in construction hazardous areas," *Saf. Sci.*, vol. 164, Aug. 2023, Art. no. 106150, doi: 10.1016/j.ssci.2023.106150.

[119] X. Mei, X. Zhou, F. Xu, and Z. Zhang, "Human intrusion detection in static hazardous areas at construction sites: Deep learning-based method," *J. Construct. Eng. Manage.*, vol. 149, no. 1, Jan. 2023, Art. no. 04022142, doi: 10.1061/(ASCE)CO.1943-7862.0002409.

[120] M. Neuhausen, D. Pawlowski, and M. König, "Comparing classical and modern machine learning techniques for monitoring pedestrian workers in top-view construction site video sequences," *Appl. Sci.*, vol. 10, no. 23, p. 8466, Nov. 2020, doi: 10.3390/app10238466.

[121] M. Neuhausen, P. Herbers, and M. König, "Using synthetic data to improve and evaluate the tracking performance of construction workers on site," *Appl. Sci.*, vol. 10, no. 14, p. 4948, Jul. 2020, doi: 10.3390/app10144948.

[122] L. Yongyue, Z. Zhou, and Y. Wang, "A tracking method of multi-workers onsite with Kalman filter and OpenPose," in *Proc. ICCREM*, 2021, p. 280, doi: 10.1061/9780784483848.031.

[123] O. Angah and A. Y. Chen, "Tracking multiple construction workers through deep learning and the gradient based method with re-matching based on multi-object tracking accuracy," *Autom. Construct.*, vol. 119, Nov. 2020, Art. no. 103308, doi: 10.1016/j.autcon.2020.103308.

[124] Y.-J. Lee and M.-W. Park, "3D tracking of multiple onsite workers based on stereo vision," *Autom. Construct.*, vol. 98, pp. 146–159, Feb. 2019, doi: 10.1016/j.autcon.2018.11.017.

[125] M. Liu, D. Hong, S. Han, and S. Lee, "Silhouette-based on-site human action recognition in single-view video," in *Proc. Construct. Res. Congr.*, 2016, p. 959, doi: 10.1061/9780784479827.096.

[126] J. Yang, "Enhancing action recognition of construction workers using data-driven scene parsing," *J. CIVIL Eng. Manage.*, vol. 24, no. 7, pp. 568–580, Nov. 2018, doi: 10.3846/jcem.2018.6133.

[127] S. Bhokare, L. Goyal, R. Ren, and J. Zhang, "Smart construction scheduling monitoring using YOLOv3-based activity detection and classification," *J. Inf. Technol. Construct.*, vol. 27, pp. 240–252, Mar. 2022, doi: 10.36680/j.itcon.2022.012.

[128] J. Cai, Y. Zhang, and H. Cai, "Two-step long short-term memory method for identifying construction activities through positional and attentional cues," *Autom. Construct.*, vol. 106, Oct. 2019, Art. no. 102886, doi: 10.1016/j.autcon.2019.102886.

[129] G. Torabi, A. Hammad, and N. Bouguila, "Two-dimensional and three-dimensional CNN-based simultaneous detection and activity classification of construction workers," *J. Comput. Civil Eng.*, vol. 36, no. 4, Jul. 2022, Art. no. 04022009, doi: 10.1061/(ASCE)CP.1943-5487.0001024.

[130] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," 2016, *arXiv:1611.08050*.

[131] H.-S. Fang, J. Li, H. Tang, C. Xu, H. Zhu, Y. Xiu, Y.-L. Li, and C. Lu, "AlphaPose: Whole-body regional multi-person pose estimation and tracking in real-time," 2022, *arXiv:2211.03375*.

[132] M. Liu, S. Han, and S. Lee, "Tracking-based 3D human skeleton extraction from stereo video camera toward an on-site safety and ergonomic analysis," *Construct. Innov.*, vol. 16, no. 3, pp. 348–367, Jul. 2016, doi: 10.1108/CI-10-2015-0054.

[133] J. Seo, R. Starbuck, S. Han, S. Lee, and T. J. Armstrong, "Motion data-driven biomechanical analysis during construction tasks on sites," *J. Comput. Civil Eng.*, vol. 29, no. 4, Jul. 2015, Art. no. B4014005, doi: 10.1061/(ASCE)CP.1943-5487.0000400.

[134] J. Seo, A. Alwasel, S. Lee, E. M. Abdel-Rahman, and C. Haas, "A comparative study of in-field motion capture approaches for body kinematics measurement in construction," *Robotica*, vol. 37, no. 5, pp. 1–19, Dec. 2017, doi: 10.1017/S0263574717000571.

[135] W. Chu, S. Han, X. Luo, and Z. Zhu, "Monocular vision-based framework for biomechanical analysis or ergonomic posture assessment in modular construction," *J. Comput. Civil Eng.*, vol. 34, no. 4, Jul. 2020, Art. no. 04020018, doi: 10.1061/(ASCE)CP.1943-5487.0000897.

[136] J. Kim, K. Daeho, J. Shah, and S. Lee, "Synthetic training image dataset for vision-based 3D pose estimation of construction workers," in *Proc. Construct. Res. Congr.*, 2022, p. 262, doi: 10.1061/9780784483961.027.

[137] J. Seo and S. Lee, "Automated postural ergonomic risk assessment using vision-based posture classification," *Autom. Construct.*, vol. 128, Aug. 2021, Art. no. 103725, doi: 10.1016/j.autcon.2021.103725.

[138] E. Y. T. Chian, Y. M. Goh, J. Tian, and B. H. W. Guo, "Dynamic identification of crane load fall zone: A computer vision approach," *Saf. Sci.*, vol. 156, Dec. 2022, Art. no. 105904, doi: 10.1016/j.ssci.2022.105904.

[139] Y. M. Goh, J. Tian, and E. Y. T. Chian, "Management of safe distancing on construction sites during COVID-19: A smart real-time monitoring system," *Comput. Ind. Eng.*, vol. 163, Jan. 2022, Art. no. 107847, doi: 10.1016/j.cie.2021.107847.

[140] Z. Chen, L. Wu, H. He, Z. Jiao, and L. Wu, "Vision-based skeleton motion phase to evaluate working behavior: Case study of ladder climbing safety," *Hum.-Centric Comput. Inf. Sci.*, vol. 12, pp. 1–18, Jan. 2022, doi: 10.22967/HCIS.2022.12.001.

[141] S. Wu, L. Hou, G. Zhang, and H. Chen, "Real-time mixed reality-based visual warning for construction workforce safety," *Autom. Construct.*, vol. 139, Jul. 2022, Art. no. 104252, doi: 10.1016/j.autcon.2022.104252.

[142] T. N. Dang, T. K. Le, T. P. Hong, and V. B. Nguyen, "Fast and accurate fall detection and warning system using image processing technology," in *Proc. Int. Conf. Adv. Technol. Commun. (ATC)*, Oct. 2021, pp. 207–210, doi: 10.1109/ATC52653.2021.9598204.

[143] P. D. Hung and N. T. Su, "Unsafe construction behavior classification using deep convolutional neural network," *Pattern Recognit. Image Anal.*, vol. 31, no. 2, pp. 271–284, Apr. 2021, doi: 10.1134/S1054661821020073.

[144] P. Zhai, J. Wang, and L. Zhang, "Extracting worker unsafe behaviors from construction images using image captioning with deep learning-based attention mechanism," *J. Construct. Eng. Manag.*, vol. 149, no. 2, Feb. 2023, Art. no. 04022164, doi: 10.1061/JCEMD4.COENG-12096.

[145] H. Kim, K. Kim, and H. Kim, "Vision-based object-centric safety assessment using fuzzy inference: Monitoring struck-by accidents with moving objects," *J. Comput. Civil Eng.*, vol. 30, no. 4, Jul. 2016, Art. no. 04015075, doi: 10.1061/(ASCE)CP.1943-5487.0000562.

[146] Y.-S. Shin and J. Kim, "A vision-based collision monitoring system for proximity of construction workers to trucks enhanced by posture-dependent perception and truck Bodies' occupied space," *Sustainability*, vol. 14, no. 13, p. 7934, Jun. 2022, doi: 10.3390/su14137934.

[147] M. Neuhausen, P. Herbers, and M. König, "Synthetic data for evaluating the visual tracking of construction workers," *Construct. Res. Congr.*, pp. 354–371, 2020, doi: 10.1061/9780784482865.038.

[148] Q. Hu, Y. Bai, L. He, Q. Cai, S. Tang, G. Ma, J. Tan, and B. Liang, "Intelligent framework for worker-machine safety assessment," *J. Construct. Eng. Manage.*, vol. 146, no. 5, May 2020, Art. no. 04020045, doi: 10.1061/(ASCE)CO.1943-7862.0001801.

[149] B. Yang, B. Zhang, Q. Zhang, Z. Wang, M. Dong, and T. Fang, "Automatic detection of falling hazard from surveillance videos based on computer vision and building information modeling," *Struct. Infrastruct. Eng.*, vol. 18, no. 7, pp. 1049–1063, Feb. 2022, doi: 10.1080/15732479.2022.2039217.

[150] Q. Hu, Y. Bai, L. He, J. Huang, H. Wang, and G. Cheng, "Workers' unsafe actions when working at heights: Detecting from images," *Sustainability*, vol. 14, no. 10, p. 6126, May 2022, doi: 10.3390/su14106126.

[151] B. Lee, S. Hong, and H. Kim, "Determination of workers' compliance to safety regulations using a spatio-temporal graph convolution network," *Adv. Eng. Informat.*, vol. 56, Apr. 2023, Art. no. 101942, doi: 10.1016/j.aei.2023.101942.

[152] *Ministry of Housing and Urban-Rural Development of the People's Republic of China*. Accessed: May 9, 2023. [Online]. Available: https://www.mohurd.gov.cn/

[153] *Industries at a Glance: Construction: NAICS 23: U.S. Bureau of Labor Statistics*. Accessed: May 9, 2023. [Online]. Available: https://www.bls.gov/iag/tgs/iag23.htm

[154] S. Konda, H. M. Tiesman, and A. A. Reichard, "Fatal traumatic brain injuries in the construction industry, 2003–2010," *Amer. J. Ind. Med.*, vol. 59, no. 3, pp. 212–220, Mar. 2016, doi: 10.1002/ajim.22557.

[155] *Code of Practice for Selection of Personal Protective Equipment*, Standard GB/T 11651-2008. Accessed: May 9, 2023. [Online]. Available: https://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=0307C61B4CCE89BE4316BCE2A36C57DC

[156] *1926.100—Head Protection. Occupational Safety and Health Administration*. Accessed: May 9, 2023. [Online]. Available: https://www.osha.gov/laws-regs/regulations/standardnumber/1926/1926.100

[157] M.-W. Park, N. Elsafty, and Z. Zhu, "Hardhat-wearing detection for enhancing on-site safety of construction workers," *J. Construct. Eng. Manage.*, vol. 141, no. 9, Sep. 2015, Art. no. 04015024, doi: 10.1061/(ASCE)CO.1943-7862.0000974.

[158] B. E. Mneymneh, M. Abbas, and H. Khoury, "Evaluation of computer vision techniques for automated hardhat detection in indoor construction safety applications," *Frontiers Eng. Manag.*, vol. 5, no. 2, pp. 227–239, Jun. 2018, doi: 10.15302/J-FEM-2018071.

[159] B. E. Mneymneh, M. Abbas, and H. Khoury, "Automated hardhat detection for construction safety applications," *Proc. Eng.*, vol. 196, pp. 895–902, Jan. 2017, doi: 10.1016/j.proeng.2017.08.022.

[160] B. E. Mneymneh, M. Abbas, and H. Khoury, "Vision-based framework for intelligent monitoring of hardhat wearing on construction sites," *J. Comput. Civil Eng.*, vol. 33, no. 2, Mar. 2019, Art. no. 04018066, doi: 10.1061/(ASCE)CP.1943-5487.0000813.

[161] H. Seong, H. Son, and C. Kim, "A comparative study of machine learning classification for color-based safety vest detection on construction-site images," *KSCE J. Civil Eng.*, vol. 22, no. 11, pp. 4254–4262, Sep. 2018, doi: 10.1007/s12205-017-1730-3.

[162] Y. Gu, S. Xu, Y. Wang, and L. Shi, "An advanced deep learning approach for safety helmet wearing detection," in *Proc. Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jul. 2019, pp. 669–674, doi: 10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00128.

[163] J. Fu, Y. Chen, and S. Chen, "Design and implementation of vision based safety detection algorithm for personnel in construction site," in *Proc. Int. Conf. Electr., Control, Autom. Robot. (ECAR)*. Lancaster, U.K.: Destech Publications, 2018, pp. 470–476. Accessed: May 7, 2023. [Online]. Available: https://www.webofscience.com/wos/woscc/full-record/WOS:000453087200083

[164] A. M. Kamoona, A. K. Gostar, R. Tennakoon, A. Bab-Hadiashar, D. Accadia, J. Thorpe, and R. Hoseinnezhad, "Random finite set-based anomaly detection for safety monitoring in construction sites," *IEEE Access*, vol. 7, pp. 105710–105720, 2019, doi: 10.1109/ACCESS.2019.2932137.

[165] V. S. K. Delhi, R. Sankarlal, and A. Thomas, "Detection of personal protective equipment (PPE) compliance on construction site using computer vision based deep learning techniques," *Frontiers Built Environ.*, vol. 6, pp. 1–10, Sep. 2020, doi: 10.3389/fbuil.2020.00136.

[166] N. D. Nath, A. H. Behzadan, and S. G. Paal, "Deep learning for site safety: Real-time detection of personal protective equipment," *Autom. Construct.*, vol. 112, Apr. 2020, Art. no. 103085, doi: 10.1016/j.autcon.2020.103085.

[167] H. Wu and Z. Liao, "Design and implementation of safety helmet detection system based on computer vision," in *Proc. Int. Conf. Comput. Vis., Appl., Design (CVAD)*, vol. 12155, Dec. 2021, pp. 69–73, doi: 10.1117/12.2626829.

[168] A. Hayat and F. Morgado-Dias, "Deep learning-based automatic safety helmet detection system for construction safety," *Appl. Sci.*, vol. 12, no. 16, p. 8268, Aug. 2022, doi: 10.3390/app12168268.

[169] M. Z. Shanti, C.-S. Cho, B. G. de Soto, Y.-J. Byon, C. Y. Yeun, and T. Y. Kim, "Real-time monitoring of work-at-height safety hazards in construction sites using drones and deep learning," *J. Saf. Res.*, vol. 83, pp. 364–370, Dec. 2022, doi: 10.1016/j.jsr.2022.09.011.

[170] M. M. Alateeq, R. P. P. Fathimathul, and M. A. S. Ali, "Construction site hazards identification using deep learning and computer vision," *Sustainability*, vol. 15, no. 3, p. 2358, Jan. 2023, doi: 10.3390/su15032358.

[171] M. Ferdous and S. M. M. Ahsan, "PPE detector: A YOLO-based architecture to detect personal protective equipment (PPE) for construction sites," *PeerJ Comput. Sci.*, vol. 8, p. e999, Jun. 2022, doi: 10.7717/peerj-cs.999.

[172] L. Zeng, X. Duan, Y. Pan, and M. Deng, "Research on the algorithm of helmet-wearing detection based on the optimized YOLOv4," *Vis. Comput.*, vol. 39, no. 5, pp. 2165–2175, May 2022, doi: 10.1007/s00371-022-02471-9.

[173] M. Nain, S. Sharma, and S. Chaurasia, "Authentication control system for the efficient detection of hard-hats using deep learning algorithms," *J. Discrete Math. Sci. Cryptogr.*, vol. 24, no. 8, pp. 2291–2306, Nov. 2021, doi: 10.1080/09720529.2021.2011109.

[174] M. Z. Shanti, C.-S. Cho, Y.-J. Byon, C. Y. Yeun, T.-Y. Kim, S.-K. Kim, and A. Altunaiji, "A novel implementation of an AI-based smart construction safety inspection protocol in the UAE," *IEEE Access*, vol. 9, pp. 166603–166616, 2021, doi: 10.1109/ACCESS.2021.3135662.

[175] Z. Xu, J. Huang, and K. Huang, "A novel computer vision-based approach for monitoring safety harness use in construction," *IET Image Process.*, vol. 17, no. 4, pp. 1071–1085, Nov. 2022, doi: 10.1049/ipr2.12696.

[176] H. Liang and S. Seo, "Automatic detection of construction workers' helmet wear based on lightweight deep learning," *Appl. Sci.*, vol. 12, no. 20, p. 10369, Oct. 2022, doi: 10.3390/app122010369.

[177] J. Lee and S. Lee, "Construction site safety management: A computer vision and deep learning approach," *Sensors*, vol. 23, no. 2, p. 944, Jan. 2023, doi: 10.3390/s23020944.

[178] L. Wu, N. Cai, Z. Liu, A. Yuan, and H. Wang, "A one-stage deep learning framework for automatic detection of safety harnesses in high-altitude operations," *Signal, Image Video Process.*, vol. 17, no. 1, pp. 75–82, Apr. 2022, doi: 10.1007/s11760-022-02205-3.

[179] W.-C. Chern, J. Hyeon, T. V. Nguyen, V. K. Asari, and H. Kim, "Context-aware safety assessment system for far-field monitoring," *Autom. Construct.*, vol. 149, May 2023, Art. no. 104779, doi: 10.1016/j.autcon.2023.104779.

[180] G. Yan, Q. Sun, J. Huang, and Y. Chen, "Helmet detection based on deep learning and random forest on UAV for power construction safety," *J. Adv. Comput. Intell. Intell. Informat.*, vol. 25, no. 1, pp. 40–49, Jan. 2021, doi: 10.20965/jaciii.2021.p0040.

[181] S. Chen and K. Demachi, "Towards on-site hazards identification of improper use of personal protective equipment using deep learning-based geometric relationships and hierarchical scene graph," *Autom. Construct.*, vol. 125, May 2021, Art. no. 103619, doi: 10.1016/j.autcon.2021.103619.

[182] G. Iannizzotto, L. Lo Bello, and G. Patti, "Personal protection equipment detection system for embedded devices based on DNN and fuzzy logic," *Expert Syst. Appl.*, vol. 184, Dec. 2021, Art. no. 115447, doi: 10.1016/j.eswa.2021.115447.

[183] Y. Gu, Y. Wang, L. Shi, N. Li, L. Zhuang, and S. Xu, "Automatic detection of safety helmet wearing based on head region location," *IET Image Process.*, vol. 15, no. 11, pp. 2441–2453, Sep. 2021, doi: 10.1049/ipr2.12231.

[184] S. Chen, K. Demachi, and F. Dong, "Graph-based linguistic and visual information integration for on-site occupational hazards identification," *Autom. Construct.*, vol. 137, May 2022, Art. no. 104191, doi: 10.1016/j.autcon.2022.104191.

[185] Y. Li, H. Wei, Z. Han, J. Huang, and W. Wang, "Deep learning-based safety helmet detection in engineering management based on convolutional neural networks," *Adv. Civil Eng.*, vol. 2020, pp. 1–10, Sep. 2020, doi: 10.1155/2020/9703560.

[186] *The Benefits of AI in Construction.* Accessed: Nov. 4, 2023. [Online]. Available: https://constructible.trimble.com/construction-industry/the-benefits-of-ai-in-construction

[187] H. Luo, J. Liu, W. Fang, P. E. D. Love, Q. Yu, and Z. Lu, "Real-time smart video surveillance to manage safety: A case study of a transport mega-project," *Adv. Eng. Informat.*, vol. 45, Aug. 2020, Art. no. 101100, doi: 10.1016/j.aei.2020.101100.

[188] M. Aljanabi, M. Yaseen, A. Ali, and S. Abed, "ChatGpt: Open possibilities," *Iraqi J. Comput. Sci. Math.*, vol. 4, no. 1, pp. 62–64, Jan. 2023, doi: 10.52866/20IJCSM.2023.01.01.0018.

[189] *Your AI-Powered Copilot for the Web.* Accessed: May 19, 2023. [Online]. Available: https://www.microsoft.com/en-us/bing?form=MA13FJ
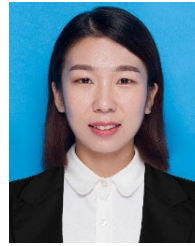
JIAQI LI was born in Anshan, Liaoning, China, in 1991. He received the B.S. degree in civil engineering from Liaoning Technical University, the M.S. degree in structural engineering from Shenyang Jianzhu University, and the Ph.D. degree in structural engineering from the Dalian University of Technology, in 2022.

Since 2022, he has been a Lecturer with the School of Civil Engineering, University of Science and Technology Liaoning. He is currently in charge of three research projects and has published eight articles as well as one patent to date. His research interests include artificial intelligence-based construction monitoring, deep learning-based building structure damage detection, and earthquake prevention for building structures.

QI MIAO was born in Anyang, Henan, China, in 2000. She received the B.S. degree in civil engineering from the Luoyang Institute of Science and Technology, in 2022. She is currently pursuing the M.S. degree in civil engineering with the University of Science and Technology Liaoning.

Her research interest includes artificial intelligence-based construction safety monitoring.

LIXIAO ZHANG was born in Hebei, China, in 1993. She received the B.S. degree in civil engineering from Northwest A&F University, in 2016, and the Ph.D. degree in structural engineering from the Dalian University of Technology, in 2022.

Since 2022, she has been a Lecturer with the College of Transportation Engineering, Dalian Maritime University. Her research interests include bridge structure health monitoring and structural damage visualization and diagnosis.

ZHENG ZOU was born in Dalian, Liaoning, China, in 1994. She received the B.S. degree in civil engineering and the Ph.D. degree in structural engineering from the Dalian University of Technology, in 2016 and 2022, respectively.

Since 2022, she has been a Lecturer with the College of Transportation Engineering, Dalian Maritime University. Her research interests include deep learning-based structure health monitoring and computer vision-based ancient building inspection.

ZHAOBO LI was born in Inner Mongolia, China, in 1989. He received the B.S. degree in engineering management from Hulunbuir University, in 2014, and the M.S. degree in structural engineering from Shenyang Jianzhu University, in 2017. He is currently pursuing the Ph.D. degree in structural engineering with the China University of Mining and Technology.

He is a Senior Engineer of the Hohhot Science and Technology Innovation Service Center. He has published more than ten academic papers in related fields and obtained more than 70 national patents, including 15 invention patents, and more than 20 national computer software copyrights. His research interest includes intelligent inspection of building curtain wall.

HUAGUO GAO was born in 1979. He received the B.S. and M.S. degrees in civil engineering from the Shenyang University of Technology. He is currently pursuing the Ph.D. degree with the Institute of Engineering Mechanics, China Earthquake Administration.

Since 2006, he has been engaged in research and teaching. He is a Professor with the School of Civil Engineering, University of Science and Technology Liaoning. His research interests include seismic damage mechanism of engineering structures, new building materials, and reinforcement of critical structures.

Prof. Gao is a member of the Board of Directors of the Liaoning Provincial Civil Engineering and Architectural Society.

NAN WANG was born in Anshan, Liaoning, China, in 1990. She received the B.S. and M.S. degrees from the University of Science and Technology Liaoning and the Ph.D. degree in civil engineering from the Nanjing University of Aeronautics and Astronautics, in 2022.

Since 2022, she has been a Lecturer with the School of Civil Engineering, University of Science and Technology Liaoning. Her research interests include magnesium cementitious materials and concrete durability.

• • •