Computer Vision for Safety Management A Case Study in the Construction Industry

Nguyen Quoc Khanh
Dept. of Vingroup AI Engineer Prgram Training
VinBigData Joint Stock Company
Ha Noi, Viet Nam
orcid.org/0009-0007-4278-149X

***Abstract*—The growing deployment of computer vision in industrial processes significantly contributes to strengthening the manufacturing sector in terms of productivity and safety of the workers. Manufacturing workers are often working in hazardous environments handling different dangerous equipment putting their life on the line every day. Work accidents are reminders for which companies must make efforts to reduce its occurrence and their adverse impact on the lives of workers. The growing deployment of computer vision in industrial processes significantly contributes to strengthening the manufacturing sector in terms of productivity and safety of the workers. Manufacturing workers are often working in hazardous environments handling different dangerous equipment putting their life on the line every day. Work accidents are reminders for which companies must make efforts to reduce its occurrence and their adverse impact on the lives of workers. Due to the inherent complexities of construction sites, safety management remains a persistent challenge and a critical responsibility. The primary objective of construction safety management is to guarantee the project's successful and secure completion, prioritize the well-being of workers, and mitigate construction-related accidents and their associated costs. The advent of computer vision (CV) technology has revolutionized traditional approaches to construction safety management. This paper examines the transformative impact of CV technology on construction site safety and management efficiency from an application-oriented perspective. Initially, the paper provides an overview of the fundamental principles and methodologies underpinning CV technology, accompanied by a detailed description of the literature analysis methodology employed. Subsequently, the paper delves into the diverse applications of CV technology, encompassing real-time construction site monitoring, worker safety management, equipment behavior tracking, and material quality control. Furthermore, the paper highlights the substantial potential of CV technology in mitigating accidents, enhancing safety performance, and offers valuable insights into future research directions. In conclusion, this paper presents a comprehensive overview of the construction industry's pursuit of leveraging CV technology to elevate safety management practices, serving as both an informative and instructive resource.**

***Keywords*—Computer vision, safety management, worker behavior.**

I. Introduction

The construction industry has experienced rapid expansion due to its significant global economic influence. However, it is considered one of the most hazardous industries, accounting for approximately 20% of worldwide occupational fatalities. In 2023, The Ministry of Labor, War invalids and Social Affairs of the People's Republic of Viet Nam reported 7,394 work accidents, causing 7,553 victims. The total cost of work accidents and property damage was nearly 16,357 billion VND and more than 149,770 working days, just in the labor relations sector. [1]. According to ILO statistics, construction sites worldwide witness at least 60,000 fatalities annually, representing approximately 1/6 of all work-related deaths. In developed nations, 25%–40% of construction site fatalities are work-related. Disturbingly, in certain countries, 30% of construction workers experience back pain or other musculoskeletal disorders [2]. The construction sector is widely acknowledged as having one of the highest incidences of occupational illnesses and accidents globally. Moreover, the industry's inherent risks contribute to severe health issues among its workforce and escalate safety management costs [3]. Consequently, there is an urgent need to prioritize and enhance safety management practices throughout the engineering and construction processes.

The constantly changing nature of construction sites, including the movement of personnel and the temporary placement of various equipment, contributes to the multitude of risks that can arise [4]. Another significant factor is the frequent disregard for worker safety. The engineering community has recently acknowledged the limitations of traditional management practices, such as relying solely on safety supervisors, in effectively addressing this management challenge. As a result, scientists have been actively developing a wide array of advanced smart technologies, including wearables, cameras, robots, and smartphones, along with sophisticated computer techniques such as cloud computing, the Internet of Things, and artificial intelligence (AI). These technological advancements aim to automate and enhance safety management by replacing conventional human-based information decision-making tools [5].

Among these innovative methods, Computer Vision (CV) is rapidly gaining traction due to its exceptional capability to process complex image data. CV technology enables the monitoring of equipment and worker safety on construction sites, the recognition of visual data such as material quantities and surfaces, and the tracking of object movement and changes [6].

The increasing availability of high-resolution cameras, the exponential growth of computer database capacity, and the advancement of technologies such as deep learning, knowledge graphs, and neural networks have contributed to the growing popularity of CV technology [7]. (1) CV excels in efficiently integrating and processing vast amounts of complex data; (2) CV operates reliably in extreme temperatures, dusty environments, electromagnetic fields, explosive atmospheres, and other hazardous working conditions; (3) CV is a non-intrusive technology, a crucial aspect when applied to construction safety monitoring.

This paper leverages a literature review and analysis to explore the primary avenues for future research and assess the progress of CV technology in engineering safety management. It focuses on the various elements that CV technology utilizes as classification labels for further analysis. These elements encompass four key domains: construction site monitoring, worker behavior and activity guidance and control, safe equipment operation, and material management. This study aims to provide valuable insights and recommendations for the future application of CV technology in related domains, ultimately contributing to the rational and sustainable growth of the engineering sector.

II. Overview of Computer Vision and Deep Learning

Computer Vision (CV), a branch of Artificial Intelligence (AI), holds the potential to extract valuable insights from digital visual data, including images and videos. Situated at the crossroads of various disciplines such as Computer Science, Mathematics, Engineering, Psychology, and Physics, CV aims to replicate the functionalities of the human visual system. Drawing inspiration from the human eye, Computer Vision endeavors to replicate its visual processing capabilities. The fusion of Computer Vision and Deep Learning has yielded remarkable advancements, achieving exceptional accuracy and efficiency in visual recognition tasks. Deep Learning (DL), a subfield of Machine Learning (ML), employs sophisticated networks capable of autonomous learning from unlabeled data. These networks, central to DL, possess the ability to learn intricate patterns and representations from data without explicit supervision. A fundamental process in computer vision tasks is feature extraction from images, which involves the initial identification of edges, corners, colors, and objects.

The initial step in extracting meaningful information from raw images involves the detection of fundamental visual elements such as edges, corners, colors, and objects. Traditionally, algorithms like SIFT (Scale-Invariant Feature Transform), SURF (Speeded-Up Robust Features), and BRIEF (Binary Robust Independent Elementary Features) [78] have been employed for feature extraction from raw images. However, these conventional algorithms encounter difficulties in effectively extracting features when dealing with images of lower quality or when the number of classes for classification increases. Deep Learning (DL) approaches offer solutions to overcome these feature extraction challenges. Figure 1 illustrates the workings of deep learning algorithms in conjunction with computer vision. In 2012, a groundbreaking research paper showcased the superiority of deep learning methods over existing state-of-the-art (SOTA) techniques in image recognition at the renowned ImageNet CV competition. This pivotal event marked a turning point in the field, highlighting the substantial benefits of integrating DL with CV. This breakthrough sparked a surge of interest in Convolutional Neural Networks (CNNs), leading to the development of various "ConvNets" models. ConvNET architectures have demonstrated exceptional performance in image classification tasks, approaching near-perfect accuracy levels of almost 100%. Consequently, DL has emerged as an indispensable tool in the realm of Computer Vision.

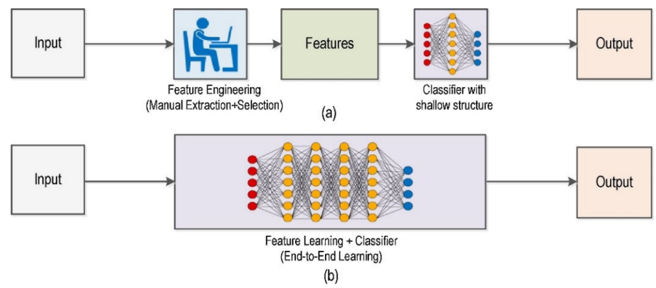

Fig. 1. . (a) Traditional Computer Vision workflow vs. (b) Deep Learning workflow. Figure from [78]

Deep Learning Methods possess the capability to autonomously acquire meaningful insights from annotated training data. This attribute renders them exceptionally valuable in Computer Vision, particularly in the utilization of Convolutional Neural Networks (CNNs) for image classification and object detection tasks.

A. Computer Vision Techniques for Constructional Intelligence

A concise discussion of five Computer Vision techniques follows.

1) Image Classification: This technique involves inputting a labeled image training dataset into a computer system for processing. The system then analyzes these images to learn and discern their visual characteristics. Expanding image classification tasks across multiple frames leads to action recognition by integrating predictions from each frame.

2) Object Detection: Object Detection focuses on pinpointing and classifying objects within an image, encompassing the generation of bounding boxes. Object detection involves identifying and labeling objects within an image. The methodology employed introduces variations in this technique. Applying object detection to each frame of a video sequence results in an object tracking task [79].

3) Object Tracking: This technique pertains to the monitoring of one or more moving objects within a defined scene. Object tracking is a technique used to follow the movement of one or multiple objects within a specific scene. It has been traditionally employed to observe and analyze real-world interactions, with applications extending to video frames.

4) Semantic Segmentation: As a crucial aspect of Computer Vision, segmentation involves partitioning an image into distinct pixel groups that can be labeled and categorized. Semantic segmentation, a crucial aspect of computer vision, divides an image into meaningful segments. More specifically, semantic segmentation aims to decipher the role of each pixel within an image.

5) Instance Segmentation: This technique classifies all instances of various classes. For example, in a complex scene containing numerous overlapping objects and backgrounds, this technique enables the classification of all objects, the identification of their disparities, and an understanding of their interrelationships. These aspects require comprehensive evaluation. Object detection involves identifying and labeling objects within an image.

III. Computer Vision based framework for the monitoring of Safety at construction

While related, safety and compliance represent distinct concepts within informatics. The Occupational Safety and Health Administration (OSHA), a division of the United States Department of Labor, emphasizes compliance as a crucial element in mitigating workplace injuries and accidents, as stated in [80]. Previous research endeavors have delved into understanding the causal factors underlying occupational accidents within the construction industry. Notably, in 2015, J. Seo et al. [81] introduced a framework designed to enhance health and safety monitoring for construction workers. Examining prior incidents and their contributing factors underscores the need for heightened emphasis on safety and compliance protocols. By implementing robust safety and compliance measures, the potential for such accidents can be significantly reduced, as highlighted in [82] and [83].

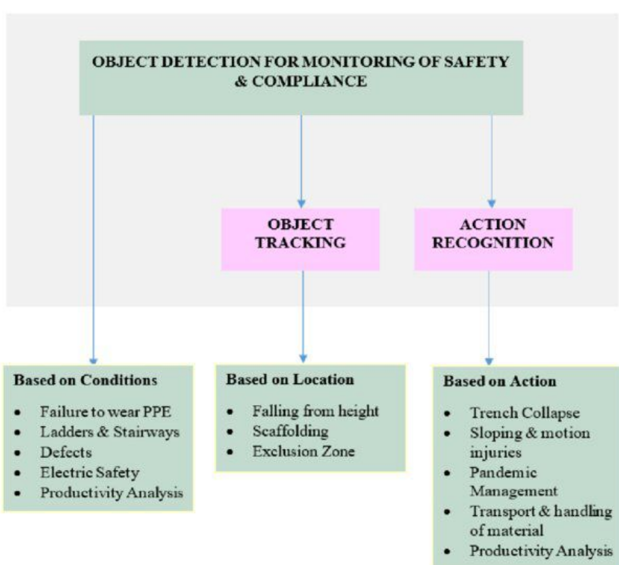

Fig. 2. Framework for the Safety and Compliance Management System

Figure 2 illustrates a framework for a Safety and Compliance Management System aimed at fostering safer construction environments. The initial step, as depicted in Figure 2, involves the compilation of an image dataset acquired through image sensing devices. This dataset leverages three distinct computer vision (CV) techniques: (I) Object detection; (II) Object tracking; and (III) Action recognition, all tailored to identify unsafe and non-compliant conditions. Object detection serves as a foundation for identifying condition- or situation-based risks, paving the way for subsequent object tracking and action recognition methodologies. Object detection serves as a foundation for identifying condition- or situation-based risks, paving the way for subsequent object tracking and action recognition methodologies. Object tracking algorithms facilitate the monitoring of location-based unsafe and non-compliant activities. Employing sequential images, action recognition techniques analyze the actions of construction-related entities (e.g., workers and equipment) to Employing sequential images, action recognition techniques analyze the actions of construction-related entities (e.g., workers and equipment) to identify unsafe practices that deviate from established safety and compliance protocols.

1) Personal Protective Equipment (PPE): Personal protective equipment also termed to as "PPE", is an equipment for the protection against personal injury and illnesses at the workplace. These illnesses and injuries are the outcome of various workplace hazards, for example, radiological, chemical, electrical hazards etc. [84][85]. In Figure 3, various types of PPE are presented

Fig. 3. Types of PPE [84][85]

IV. Research Method

This study aims to shed light on the current state of CV technology utilization in engineering safety management through a comprehensive analysis of existing literature. The research employs content analysis, a widely used methodology in social science research. Content analysis is a specialized research technique that utilizes a systematic, objective, and quantitative approach to analyze the content of literature. Its primary objective is to uncover hidden connections within the literature and generate well-informed predictions about future trends. Keyword clustering, based on bibliometric analysis, is used to map research hotspots in the field of construction safety related to CV. This approach ensures neutrality in categorizing the applications of CV technology in the subsequent section and lays the foundation for the discussion that follows.

A. Literature search and selection

The majority of the literature reviewed in this paper is sourced from the Core Collection database in Web of Science (WOS). WOS hosts some of the most prestigious academic literature systems in science and technology, including SCI, SSCI, and others [8]. It encompasses the most reputable and influential academic journals globally and serves as a valuable resource for tracking the research interests of international scientists. WOS is particularly renowned for its robust bibliometric capabilities, enabling efficient and effective literature measurement and the identification of relevant, high-quality research.

This study employs an advanced search strategy that utilizes Boolean operators and synonyms to ensure the accurate and comprehensive retrieval of relevant literature. The search query used is as follows: “TS=(CV OR based vision); TS=(construction OR architecture); TS=(safety management OR risk management); #3 AND #2 AND #1”. This search strategy yielded 164 articles relevant to the research topic, including journal articles, conference papers, review articles, and early access publications.

The articles span from 1997 to 2023, and no further time filtering was applied due to the limited number of articles available.

B. Literature analysis

1) Analysis based WOS: WOS provides statistical analysis of the selected articles. Figure 4 illustrates the annual growth in the number of publications focusing on CV in construction safety. The graph highlights a significant increase in the utilization of CV technology for safety management in the engineering industry after 2018. This surge coincides with the rise in popularity of technologies such as neural networks and deep learning, suggesting that advancements in computer algorithms have paved the way for the widespread adoption and progress of CV applications.

Figure 5 reveals a striking number of publications (71) originating from China, underscoring the country’s research focus on this topic. It’s important to note that the figure displayed in Figure 5 for China is 69 due to the exclusion of two articles from Taiwan Province. However, the adjusted number, including Taiwan Province, is 71. This near 50% share of publications highlights the growing emphasis Chinese researchers place on worker safety. In 2018, China reported 734 construction-related accidents and 840 fatalities, while in 2020, there were 689 accidents and 794 fatalities.

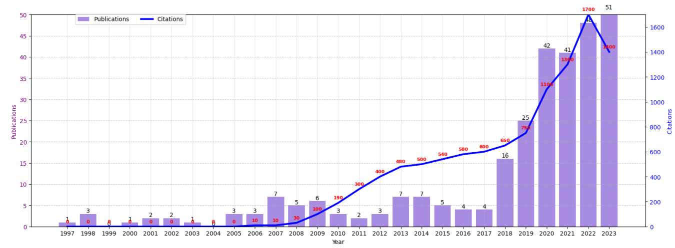

Fig. 4. Times cited and publications over time.

Fig. 5. Illustration depicting the geographical distribution of research publications.

2) Analysis based CiteSpace: Leveraging CiteSpace, a specialized tool for visual analysis of research literature, this study conducted a keyword co-occurrence network analysis on a dataset of 164 articles. CiteSpace, grounded in graph theory-based spectral clustering algorithms, offers inherent advantages over traditional co-reference networks that rely on linking relationships for node clustering instead of attributes. The credibility of clustering results derived from CiteSpace is typically evaluated using two primary metrics: Silhouette (-S value), representing the average profile of clusters, and Modularity, indicating the clustering module’s value (-vQalue). Generally, an S value greater than 0.5 suggests reasonable clustering, exceeding 0.7 indicates satisfactory clustering, while a modularity value of 0.7 is considered reliable. In this analysis, as depicted in Figure 6, the clustering results yielded a Q value of 0.5117 and an S value of 0.7899, signifying the trustworthiness of the outcome. The results, visualized in Figure 6, provide insights into prevailing research hotspots within the field.

Figure 6 presents the outcomes of a keyword cluster analysis encompassing 10 distinct clusters. The keywords are sequentially numbered from 0 to 9, with lower numerical values indicating a higher number of keywords within each cluster. Each cluster comprises a collection of semantically related

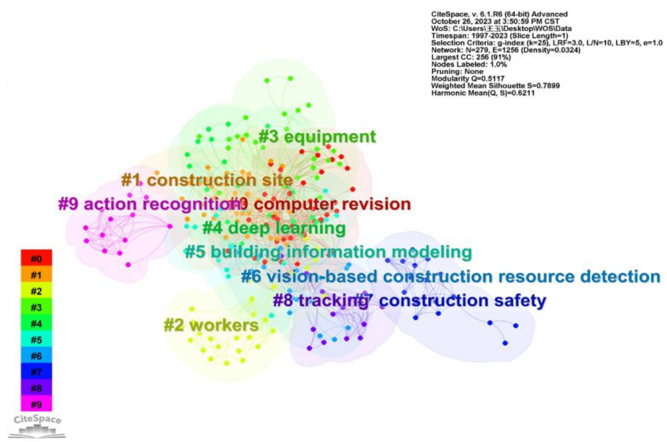

Fig. 6. Visualization of keyword co-occurrence network analysis using CiteSpace.

terms. Through a comprehensive analysis of these ten keyword clusters, this research identifies four primary categories that encompass the objects of Computer Vision (CV) in the realm of engineering safety management: construction sites, workers, equipment, and materials. These categories provide a foundational framework for the subsequent discussions presented in this study.

The temporal emergence of research topics closely aligned with the study's theme is illustrated through the co-occurrence network presented in Figure 3. Notably, terms such as "neural networks," "deep learning," and "convolutional neural networks" emerged prominently in 2018. This period signifies a pivotal juncture in the field's evolution, marked by the emergence and subsequent convergence of these keywords, further corroborating the findings derived from the earlier analysis based on the Web of Science (WOS).

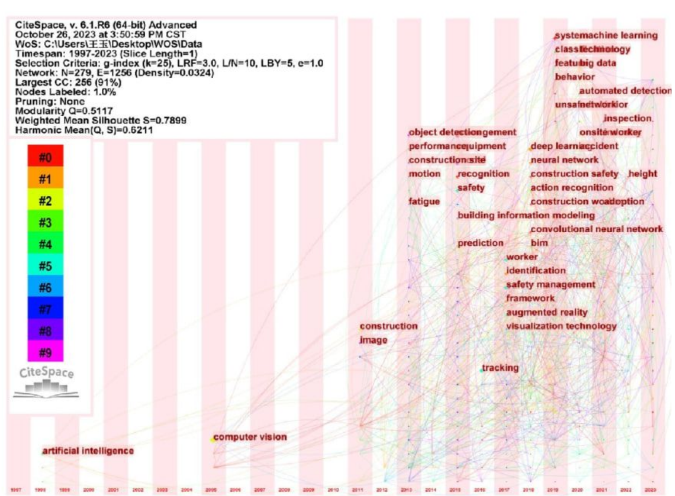

Fig. 7. Temporal representation of prominent keyword appearances based on CiteSpace analysis.

V. Literature Review Based on Content Analysis

From a health and safety perspective, there are three levels of computer vision development and application: L1 recognition and tracking, L2 assessment, and L3 prediction, as shown in Table 1.

TABLE I
COMPUTER VISION DEVELOPMENT FRAMEWORK

Level of Development	Functions	Main Research Question
L1: Recognition and Tracking	L1.1 Object Recognition	What is an Object?
	L1.2 Object Tracking	Where is an Object?
	L1.3 Action Recognition	What is an Object Doing?
L2: Assessment	L2.1 Assess the object	Is the object a hazard?
	L2.2 Assess the behavior	Is the action unsafe?
	L2.3 Assess the condition	Is the working condition (scenario) safe or unsafe?
L3: Prediction	L3.1 Predict behavior	How will the object behave?
	L3.2 Predict the incident	Will the next incident occur?
	L3.3 Early warning	What are the leading indicators of the next incident?

A. *Monitoring of Construction Sites*

Within the context of construction site management, one prominent application of CV lies in safety monitoring. This entails the continuous observation of the worksite using CV technology to detect potential hazards such as fire, smoke, or structural instability, enabling timely activation of emergency response protocols. (2) Identification of hazardous areas: CV facilitates the identification and demarcation of high-risk zones or areas containing hazardous materials within a construction site, thereby enhancing worker awareness and promoting precautionary measures. (3) Traffic monitoring: To ensure traffic safety, particularly on roads and bridges within and surrounding project sites, CV-powered systems monitor traffic flow and vehicle movement patterns.

In a recent study focusing on safety monitoring, Yan et al. [9] proposed a comprehensive conceptual framework for managing construction accidents. Their approach leveraged CV technology and other methodologies to effectively manage the aftermath of a building collapse incident in Shanghai by providing comprehensive situational awareness. Similarly, Ahn et al. [10] introduced a novel methodology for improving fire detection and mitigating accidents and property damage. Their approach involved developing a CV system integrated with indoor closed-circuit television (CCTV) technology, achieving a remarkable accuracy rate of 91% and demonstrating the ability to detect fires in indoor environments within a one-second timeframe.

The specific hazards present on a construction site are inherently variable and context-dependent. However, a technology-driven approach to identifying and assessing high-risk areas can significantly mitigate the occurrence of accidents. In line with this, Pushkar et al. [10] proposed the utilization of image-based 3D site reconstruction as a means to pinpoint the most hazardous locations, including areas where workers are exposed to potential risks, such as machinery and equipment operation zones, as well as material storage areas. Real-time image processing techniques, employing multiple cameras and CV algorithms, are utilized to monitor the safety of these designated areas. Furthermore, Luo et al. [12] developed an intelligent video surveillance system that leverages CV to differentiate between stationary and moving personnel within various hazardous zones characterized by varying risk levels.

Given the frequent utilization of heavy vehicles and mechanical equipment on construction sites, effective traffic accident management is paramount for enhancing overall safety. Zhu et al. [13] developed a CV-based early warning system designed for real-time monitoring of traffic conditions at construction site crossings. This cost-effective system effectively handles complex traffic scenarios involving buses, transport vehicles, workers, and pedestrians.

In conclusion, the integration of CV technology for tasks such as construction site safety monitoring, hazardous area identification, and traffic management holds immense potential for enhancing construction safety, reducing accidents, and improving project quality. Consequently, these advancements play a pivotal role in ensuring the successful execution of construction projects and safeguarding worker well-being.

B. *Identification and Tracking of Workers*

The application of CV technology in safeguarding construction workers can be categorized into three primary domains: (1) Training and education programs designed to enhance worker safety awareness and knowledge. (2) Identification of potentially hazardous worker behavior. (3) Real-time tracking and identification of worker locations.

In the past decade, researchers have made significant efforts to recognize various project-related objects using computer vision. As shown in Table 2, the focus of research has been on recognizing and tracking the workforce and on-site equipment. More recently, specific efforts have been made to apply object recognition to construction health and safety.

Research findings indicate a strong correlation between the occurrence of accidents and two key factors: inadequate training and limited comprehension among construction workers [14]. Numerous research endeavors have focused on leveraging technological advancements to enhance training effectiveness, with augmented reality (AR) emerging as a prominent technology in this domain. The fundamental aspect of these advancements involves the implementation of CV [14]. Kivrak et al. [15] developed an innovative safety education methodology utilizing augmented reality glasses to provide an engaging and immersive learning experience, enabling users to visualize safety animations. In a 2018 study, Eiris et al. developed a safety training system platform incorporating augmented 360-degree panoramic reality (PARS) [16]. Testing conducted on a group of 30 participants revealed a significant improvement in their ability to identify all four designated hazard samples. The findings of these studies suggest that CV technology is poised to facilitate the development of advanced training platforms capable of enhancing safety education and training, thereby improving workers' ability to identify and address potential hazards. Future worker training initiatives aim to incorporate cutting-edge technologies to foster engagement through immersive experiences.

TABLE II
RECOGNIZABLE AND TRACKABLE OBJECTS

Recognizable and Trackable Objects	**References**
Concrete columns	(Zhu et al. 2010)
Dump trucks	(Rezazadeh Azar and McCabe 2011)
Excavators and dump trucks	(Golparvar-Fard et al. 2012)
Hydraulic excavators	(Azar and McCabe 2012)
Tower cranes	(Yang et al. 2012)
Workers	(Park and Brilakis 2012; Zhu et al. 2016)
Excavators, loaders, bulldozers, rollers, and backhoes	(Tajeen and Zhu 2014)
Dump trucks, excavators, loaders, concrete mixer trucks, and road rollers	(Kim et al. 2017)
Trade recognition	(Fang et al. 2018)
Concrete mixer trucks	(Kim and Kim 2018)
Safety harnesses	(Fang et al. 2018)
Safety guardrails	(Kolar et al. 2018)
Hardhats	(Fang et al. 2018)

The identification and monitoring of construction worker behavior constitute a crucial aspect of mitigating safety incidents. However, the comprehensive implementation of this task has been challenging due to its demanding and intricate nature. Working at heights has long been recognized as a hazardous occupation. In practice, a significant proportion of workers exhibit a lack of awareness or intentional disregard for the use of essential safety equipment, such as helmets and safety harnesses. To address these concerns, Fang et al. [17] proposed a safety management approach that combines convolutional neural networks and CV. This approach enables the prompt implementation of interventions and preventive measures upon the detection of specific behaviors, leading to a substantial reduction in fall incidents within the context of working at heights. In their study, Hayat et at [18] presented a novel framework employing computer vision technology for the automated inspection of safety helmets, achieving a 92.44% accuracy rate even in challenging conditions such as low light or the presence of small objects. Another study by Ding et al. [19] introduced a hybrid deep learning model combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) to automatically categorize worker actions across different postures.

In their work, Alateeq et al. [20] researchers have developed a method using closed-circuit television (CCTV) footage for worker recognition, integrating deep learning and computer vision techniques to improve the identification and classification of construction worker activities and safety compliance. These advancements highlight the rapid progress of computer vision in worker behavior monitoring, particularly when combined with wearable devices for enhanced convenience. Integrating intelligent methodologies like deep learning and advanced camera technologies like depth cameras, binocular cameras, and optical cameras can contribute to a more precise and automated approach to managing worker behavior [21].

While safety management systems for worker location tracking on construction sites often rely on Internet of Things (IoT) systems and sensor devices like Radio Frequency Identification (RFID) to represent worker location information virtually, computer vision technology offers enhanced stability, convenience, and real-time tracking capabilities. Park et al. [22] have introduced a computer-based visual system to identify and track construction workers within video frames, enabling the detection of worker positions and the determination of spatial coordinates for materials, buildings, and equipment at the construction site. Park et al. [23] has explored strategies for tracking workers in densely populated outdoor environments by combining camera technology with computer vision methods. Integrating computer vision with virtual reality and human-computer interaction has also been proposed by Park et al. [24] to combine tracking and worker localization processes. An online visual tracking approach using Multi-Domain Convolutional Neural Networks (MD-CNN) has demonstrated satisfactory tracking performance even in complex construction environments by Lin et al. [25].

Traditionally, ensuring worker safety on construction sites has been labor-intensive, requiring managers and engineers to physically inspect the site for potential hazards. Computer vision offers a solution by automating worker position tracking and behavior monitoring through image analysis.

C. Equipment Monitoring

Machinery-related accidents, particularly those involving large vehicles like cranes and excavators, are a significant concern on construction sites. Vision-based techniques for monitoring machinery location, detecting movements, and providing guidance during operations have been a focal point in construction automation research.

Azar et al. [26] proposed a novel computer vision methodology using an active zoom camera to monitor excavator and dump truck operations, achieving a recognition accuracy between 80% and 90%. This study highlights the benefits of active camera systems for such monitoring tasks. Xiao et al. [27] introduced a tracking methodology for monitoring multiple machinery and equipment units using hash features extracted from images as an appearance model. This approach, combining computer vision and deep learning, effectively tracked multiple units operating simultaneously in challenging conditions like poor lighting and occlusion.

Extensive research has explored the identification of attitudes in construction machinery using various methods, including IMU sensors, with a particular focus on machinery collision avoidance as a crucial aspect of safety management. Zhang et al. [28] proposed an assessment framework for enhancing collision avoidance safety using computer vision and fuzzy reasoning techniques, identifying machinery and worker proximity and crowding as primary contributors to collision accidents. To improve system efficiency, a faster R-CNN algorithm was incorporated to establish a quantitative relationship in identifying these factors. Zhang et al [29] developed a framework for collision avoidance safety using computer vision (CV) and fuzzy reasoning techniques. They found that proximity and crowding of machinery and workers were the leading causes of collision accidents. The authors integrated a faster R-CNN algorithm to improve system efficiency by quantifying these factors.

The successful execution of earthworks relies on the skilled operation of large machinery and the management of complex geological conditions, which can pose hazards related to material conveyance. A novel vision technique using two depth cameras was introduced to accurately assess the spatiotemporal deformation of soil during earthmoving excavation by Tsuchiya et al. [30]. The study also revised the existing model for predicting bucket resistance during excavation, incorporating the new measurements. While algorithmic advancements could further enhance the method's precision, this study holds significant implications for using vision technology in accurate assessments. Naghshbandi et al. [31] examined the application of computer vision (CV) technology for enhancing safety measures during earthmoving activities, focusing on automated safety and target monitoring. The utilization of multi-camera calibration, data fusion, and attitude estimation techniques contributed to improved detection accuracy and reduced processing complexity.

Integrating computer vision technology into the regulation of construction machinery and vehicles has the potential to mitigate hazards associated with unsafe areas and blind spots, improving operational management in challenging environments like earthworks.

D. Materials and Resources Management

In construction safety management, computer vision (CV) is primarily used for identifying, classifying, and tracking materials and resources like temporary scaffolding and formwork, with a focus on monitoring material quality to ensure safety standards. The labor-intensive nature of construction sites requires significant workforce, machinery, materials, and temporary provisions, and inadequate management of these components can negatively impact construction schedules and safety protocols [32].

In their investigation, Dimitrov et al. [33] proposed a methodology for classifying materials using CV and support vector machine classifiers The researchers collected a dataset consisting of over 20 commonly used construction materials, with 150 images obtained for each category. Notably, the images were carefully conditioned to ensure optimal lighting conditions at the site, resulting in enhanced accuracy of the classification process. Hence, Mahami et al. [34] developed and presented a visual material classification approach based on deep learning. A total of 1231 images from diverse construction sites were gathered to train a substantial dataset. The study successfully attained a detection accuracy of 97.35% for various building materials, even in intricate surroundings.

The field of vision-based inspection for assessing the quality of building materials and structural integrity, with a particular focus on crack detection, has experienced significant advancements in recent years. In their study, Dinh et al. [35] integrated inspection robots and CV techniques to facilitate the automatic extraction of visible cracks from photo. This was achieved by the utilisation of image binarisation thresholding, employing non-parametric peak detection algorithms. According to Koch et al. [36], CV has made significant advancements in automating the detection of cracks, joint damage, and cavities in various infrastructural elements such as tunnels, asphalt pavements, and pipelines. The authors emphasise the significance of this visual quantification approach in ensuring the safety and quality assurance of these materials. In their recent review, Deng et al. [37] have showcased the application of CV in crack detection, along with the integration of artificial intelligence techniques. These techniques have witnessed significant advancements, transitioning from manual image processing techniques (IPTs) that relied on low-level features, to more sophisticated feature learning methods. As a result, the current approaches offer comprehensive and precise quantification outcomes.

VI. Challenges of Computer Vision in Construction

While numerous attempts have been made to automate hazard identification on construction sites using computer vision, a fully automated system remains elusive. Although we can identify individuals not wearing personal protective equipment (PPE), determining their identity and verifying the proper use of PPE, such as ensuring safety harness hooks are secured to rails, remains a challenge. However, deep learning offers potential for real-time hazard identification through data analytics. Let's delve into the potential applications and technical hurdles of deep learning algorithms, especially convolutional neural networks (CNNs), in safety management. (1) identifying individuals who fail to comply with PPE requirements; and (2) determining if PPE is used correctly (for example, if a safety harness hook is attached to a rail). Deep learning, however, has the potential to provide data analytics for real-time, automated hazard identification. Next, we explore the potential applications and technical challenges of deep learning algorithms, particularly CNNs, in safety management.

A. Applying computer vision in practice

A prerequisite for effective hazard identification (minimizing misdetections and ensuring accuracy) and successful computer vision application is access to a comprehensive, high-quality image database encompassing various types to train CNNs. However, the lack of a sufficiently large database poses a significant obstacle to leveraging computer vision for efficient safety monitoring. Unlike publicly available datasets in computer science, such as ImageNet and Microsoft® Common Objects in Context (COCO), construction datasets require unique characteristics. These include considerations for spatial conflicts, cluttered backgrounds, occlusions, variations in poses and scales, and the dynamic, ever-changing nature of construction environments. Consequently, many potential hazards remain hidden from plain sight (e.g., potential collisions with plant or equipment), and structural quality assessment becomes challenging. This difficulty arises from the limitations of deep learning models in predicting previously unobserved objects and extracting concealed information through computer vision.

The scarcity of datasets has compelled researchers to rely on relatively small image samples for their experimental hazard identification work. This reliance on limited data has made it challenging to compare and contrast reported evaluation metrics like precision, recall, and accuracy with other approaches. The variability in the quality of datasets used for training and testing further complicates the determination of the validity and reliability of results reported in existing literature. Therefore, there is a pressing need for robust and objective evaluation criteria to compare and contrast various computer vision approaches advocated for safety management.

B. Technical challenges

As previously mentioned, computer vision studies have often relied on small databases and supervised learning approaches to identify unsafe behavior. This approach leads to weak generalizations due to two main factors: (a) the assumption that training and testing databases share the same distribution; and (b) the limitations of training machine learning models on small datasets, which restricts inter and intra-class variability. Consequently, this hinders their ability to accurately recognize unsafe behavior and generalize to different datasets [42]. However, techniques like transfer learning and data augmentation (e.g., cropping, flipping, and random rotation) can mitigate the accuracy and reliability issues associated with small training datasets.

Deep learning models excel at learning correlations between input and output features, but they struggle to establish causality. Understanding the interplay between human behavior and the corresponding work environment is crucial to contextualize information surrounding hazards on construction sites. For instance, behavior-based safety (BBS) observes and identifies unsafe actions, providing direct feedback to individuals who commit such actions to encourage safer future behavior [43,44,45,46].

While deep learning demonstrates proficiency in hazard recognition, it's essential to acknowledge that these approaches are often task-specific. This limitation poses a significant challenge as no single approach can effectively identify a wide range of unsafe behaviors, making the practical implementation of computer vision costly and time-consuming. Therefore, developing new algorithms and training them to detect a broad spectrum of common unsafe behaviors and conditions encountered on construction sites is crucial.

Deep learning models often operate as 'black boxes', lacking transparency in their decision-making processes [47,48,49]. While efforts have been made to address this opacity by visualizing the contributions of individual nodes within complex networks involving millions of parameters, achieving complete transparency in deep learning remains an unresolved challenge [48]. The inability to pinpoint the exact features extracted and learned by these nodes makes it difficult to understand the detection process and identify parameters requiring adjustment for accurate hazard detection, hindering the justification for adopting deep learning [48].

VII. Areas for future research in deep learning and computer vision

To overcome these challenges and ensure the effective and efficient application of computer vision in safety monitoring, we propose potential research areas aligned with the framework we designed and developed.

A. Combining deep learning and computer vision with digital technologies

Current computer vision approaches in construction often exhibit low levels of information utilization, necessitating higher accuracy levels for hazard detection. As safety regulations become increasingly complex, interdependent, and stringent due to statutory requirements, existing computer vision approaches may struggle to keep pace. If these approaches fail to adapt to evolving regulatory demands and the nuances of construction, they risk becoming obsolete.

1) Ontology and computer vision: To address the challenges posed by evolving safety regulations, we propose integrating ontology with computer vision approaches. Ontology provides a formal, conceptualized framework of knowledge, offering a simplified representation of a domain by describing objects, concepts, and their relationships [50]. Its purpose is to enable computer applications to represent and reason about knowledge efficiently. When combined with computer vision, objects can be automatically detected, and attributes like classes and geometry can be extracted from images. This integration paves the way for a semantic computer vision framework comprising four key procedures: (1) ontological modeling of hazards (e.g., unsafe behavior and plant status); (2) entity and attribute detection using computer vision; (3) extraction of spatial and temporal semantic relationships from video data; and (4) data reasoning for hazard identification.

We propose that a semantic model integrating ontology and computer vision, even with limited data, can be effective for hazard identification using deep learning. This approach leverages not only accurate object detection but also the spatial-temporal relationships between objects for hazard reasoning. Several studies have demonstrated the satisfactory performance of existing computer vision-based approaches in detecting a variety of objects [51], suggesting that this proposed semantic approach, which doesn't require a specific training database, holds significant potential.

2) Group with as-built visual data, as-planned model and IoT: A deep learning visual analytics system is proposed to assess project performance. This system would leverage a three-dimensional (3D) semantic reconstruction model [52,53], built from as-built and as-planned data, and utilize computer vision techniques.

The system would enable the automatic and scalable generation of high-quality 3D as-built models. This would be achieved by processing large volumes of images and videos from diverse sources and aligning them with the as-planned model.

Furthermore, it would enhance situational awareness for safety purposes, streamline claim analysis, and aid in accident investigations. Field snapshots and videos, aligned with the as-planned model, could be used for annotation, reporting, documentation, and communication. Computer vision's role in this system would extend beyond mere object extraction and attribute identification (e.g., distance, class).

It would also encompass the real-time construction of an as-built 3D semantic model. Concurrently, sensor data (e.g., locations) from installed components could be extracted and integrated with the computer vision detection results via the IoT. This integration would facilitate the storage and continuous real-time updating of data within a 3D model of the constructed asset. Consequently, potential hazards, including structural defects or failures, could be identified within the model.

3) As-built visual data, AR/VR, and building information models: AR applications have found widespread use in construction, augmenting both virtual and real-world environments [54].

- These applications have been developed to facilitate information retrieval during construction and facility management, particularly for safety purposes.
- They enable the visualization of underground utilities, enhance visual perception for excavation safety and subsurface utility inspection, and provide real-time 3D operational instructions overlaid on the actual site to assist with assembly and other complex tasks.
- Real-time 3D operational instructions, overlaid on the actual site, can be obtained to aid in assembly and other intricate operations.

Existing AR capabilities and mobile devices (e.g., Apple's ARKit for iOS, Google's ARCore for Android) are sufficiently advanced to support the identified visualization applications [55]. However, there is a need to enable resource-constrained mobile devices to support the deep learning-powered analytics engine.

B. Insight from multiple data fusion

Multiple data fusion involves integrating data from various sources to generate information that is more consistent, accurate, and useful than what a single source could provide[56]. In construction safety, data can originate from a wide array of sources, including safety reports and non-visual sensors.

1) Utilizing video streams and multi-model fusion: While deep learning has traditionally been employed in supervised settings with labeled data, there has been a notable shift towards unsupervised models to enhance object detection speed, accuracy, and reliability.

Building upon advancements in deep learning and computer vision within computer science[57,58,59,60], a self-taught deep learning-based unsupervised learning approach is proposed for hazard identification using video streaming. In this approach, video streams would serve as inputs to a deep learning model equipped with a self-taught learning mechanism. This mechanism, with adjustable parameters, would enable continuous self-training using video streams, producing output frames for testing.

Given the abundance of readily available remote sensing image data resources, which have been utilized in various applications[61-65], we propose their fusion for automated hazard identification in construction. For instance, by fusing thermal and 2D images, unsafe behaviors such as smoking in construction can be detected. This is possible because the temperature of a cigarette exceeds that of its immediate surroundings, making it easily distinguishable from non-hotspots using thermal imaging [66].

2) Alignment between computer vision and text reports: Combining text and image data allows a deep learning model to reason about and comprehend the nature of risk. This combination offers two research avenues:

- First, leveraging reports to enhance the accuracy of unsafe behavior identification. This involves using deep learning and computer vision to extract and encode images from feature representations (e.g., important regions).
- Second, automating the generation of safety reports from images. Current on-site safety inspections heavily rely on manual record-keeping of observed hazards, which are later transferred into computer systems for report generation[66,67].

3) Alignment between computer vision and non-visual sensor data: Sensor data from various locations on a construction site can be used for hazard detection. Different sensor types, such as location sensors (e.g., Radio Frequency Identification, Global Positioning Systems), have been deployed to collect safety data[69-71]. By fusing images with data from multiple non-visual sensors, the range of hazards detectable by deep learning-based computer vision can be expanded.

- **Unsafe behavior:** This includes identifying unsafe behaviors. Sensors like location and identity sensors can capture an individual's coordinates and identity[74,75]. Computer vision can then extract information such as object classes, activities, attributes, and 3D coordinates of objects[66,72,73].
- **Unsafe plant:** Sensors can be employed to collect mechanical data, such as bending moment and angle, from a crane during lifting operations. This data, in conjunction with computer vision, enables the identification of contextual factors and plant conditions that contribute to unsafe situations.
- **Structural defects:** Researchers have utilized sensor data for structural health monitoring to pinpoint defect origins[76, 77]. Additionally, computer vision and deep learning have been employed to visually recognize defects. However, these methods haven't effectively determined the internal composition quality of civil infrastructure like tunnels and bridges. Ground-penetrating radar (GPR), a geophysical technique using electromagnetic radiation in the UHF/VHF frequencies, detects reflected signals from subsurface structures. We propose developing a computed tomography (CT) system that integrates radar images and deep learning-based computer vision to extract information and diagnose the internal structural quality.

VIII. Application of Computer Vision in Occupational Safety in the VinBigData Ecosystem

VinBigData Joint Stock Company was established on the basis of a number of scientific research achievements of the Big Data Research Institute of Vingroup (established in August 2018) in the fields of Data Science and Artificial Intelligence (especially in image and language processing).

With mastery of core technology and large-scale multi-industry databases, VinBigdata focuses on developing highly applicable products and solutions, increasing user experience, improving the quality of life and optimizing production and business activities.

In this article, the Vizone ecosystem, an ecosystem of intelligent image analysis solutions, will be discussed.

Vizone is an ecosystem of image analysis solutions applying Computer Vision (CV) and Artificial Intelligence (AI) technology, including: Facial recognition, security and safety monitoring, Customer behavior analysis, Electronic customer identification and Identity card recognition... The ecosystem synthesizes comprehensive technology solutions with the goal of bringing a safe and convenient life to everyone.

Fig. 8. Vizone ecosystem overview

A. Vizone Secure - Smart surveillance camera solution

Vizone Secure is a comprehensive smart surveillance camera solution for businesses. Developed from Computer Vision (CV) and Artificial Intelligence (AI) technology, Vizone Secure can operate effectively in any environment with high accuracy, meeting diverse needs and scales, helping to optimize operating processes.

Outstanding features

- 99% Accurate facial recognition
- 0.3s Instant recognition speed
- 90% Object detection (people, vehicles,...)

Object recognition

- Pet detection
- Detect abandoned objects in hallways, in elevators
- Foreign object recognition
- Detect illegal parking in no-parking zones
- Detect objects overloaded compared to regulations

Characteristic & behavior analysis

- Emotion analysis (happy/sad/satisfied)
- Detect people wearing masks
- Analyze facial attributes, gender, age...
- Detection carrying knife/sword/dangerous behavior

Statistics feature

- Count people
- Heat map
- Count vehicles
- Count objects, objects on request

Face recognition

- Identify strangers, regular customers
- Identify strangers in the day
- Identify correct customers (with photos in the system)
- Identify and warn of blacklists available in the data

Vizone Secure's automatic monitoring features support businesses not only in detecting dangerous behaviors but also in predicting and preventing accidents, creating a safer working environment:

1. Detecting dangerous behavior

- The Vizone Secure system is capable of analyzing worker behavior in real time, helping to detect dangerous behaviors such as carrying weapons (knives, swords) or performing incorrect operations.
- Computer vision integrated with surveillance cameras can detect behaviors that do not comply with labor safety regulations, such as not using protective equipment or standing in dangerous areas.
- Automatically recognize and warn when detecting unusual behaviors, helping managers quickly intervene and prevent accidents.

2. Monitor the use of personal protective equipment (PPE)

- Computer vision helps identify and monitor whether workers comply with wearing protective equipment such as helmets, gloves, reflective vests or masks in the work area.
- The system will detect and issue warnings if it detects workers not using protective equipment as prescribed, helping businesses improve safety compliance in factories or construction sites.

3. Monitoring dangerous areas and restricted access

- Vizone Secure helps monitor dangerous areas, restricted access areas such as no-parking zones, areas with dangerous machinery.
- The system is capable of detecting unauthorized access to these areas and immediately sending out warning signals, helping to prevent unnecessary accidents.
- Identifying and analyzing the location of employees in real time helps manage access to safe areas, ensuring there are no unauthorized intrusions.

4. Analyzing employee emotions and behaviors

- Computer vision technology can also be used to analyze employee emotions and behaviors in the work environment, thereby recognizing signs of fatigue, stress or other emotional expressions that can affect performance and safety.
- Through these analyses, businesses can come up with timely solutions to improve working conditions and minimize work accidents caused by psychological or health factors.

5. Optimizing the safety inspection process

- Vizone Secure's intelligent monitoring system enables the automation of safety inspections in businesses. Inspection activities are recorded and reported automatically, making it easy for managers to monitor and improve work safety measures.
- Integration with a data analysis system helps businesses identify areas or behaviors that pose potential risks, thereby providing effective improvement solutions.

6. Measure and report safety data

- Vizone Secure provides detailed statistics on safety violations, incidents, and high-risk behaviors. This data is collected and analyzed automatically, supporting managers in making decisions to improve safety policies.
- Thanks to its powerful data analysis capabilities, the system can report positive or negative changes in safety levels, thereby helping businesses adjust their occupational safety strategies accordingly.

B. Vizone Access - Face Access Control Solution

Vizone Access is a face access control solution that applies Artificial Intelligence (AI) and Deep Learning technology, supporting timekeeping, check-in/out, monitoring access at facilities, recognizing VIP customers, controlling employees, and performing specialized services.

Outstanding features

- **99.9%** Outstanding accuracy
- **<1s** Fast business processing speed
- **24/7** Stable operation
- **Face recognition**

- Face recognition in special cases (wearing glasses, wearing masks,...)

- **Fake detection**
- **Easy integration into existing systems**
- **Support and management of report settings**

Applications of Vizone Access in labor safety: Vizone Access is not only used for access control but also optimally supports labor safety in the workplace. With real-time facial recognition capabilities, the system can prevent unauthorized individuals from entering dangerous areas, while ensuring that employees comply with safety regulations, such as wearing masks or goggles. Tamper detection reduces the risk of unauthorized access, while easy integration into existing surveillance systems ensures effective deployment in manufacturing and construction environments.

IX. Conclusion

This paper employs CiteSpace for bibliometric analysis to provide a comprehensive overview of computer vision (CV) technology's extensive research applications in construction safety management, focusing on site environments, workers, machinery, and materials, highlighting CV's importance as a cutting-edge technology for enhancing safety management effectiveness.

(1) Computer vision technology facilitates real-time, continuous monitoring of complex construction sites, enabling the collection of real-time data on worker behavior, machinery operation, material quantity and quality, and overall site safety, allowing for the timely identification of risk factors and implementation of corrective actions.

(2) Computer vision technology enables risk prevention and intelligent decision-making by leveraging accumulated image data to create a comprehensive database, which, when combined with intelligent algorithms, facilitates the generation of strategies and systems for risk identification, assessment, and decision-making, ultimately enhancing the efficiency of construction safety management.

Furthermore, this paper outlines several key trends in the application of computer vision across various aspects of construction safety management:

(1) Increased intelligence and automation will enable systems to automatically detect and respond to unsafe behaviors, hazardous areas, and equipment failures, minimizing the need for human intervention and enhancing the efficiency and accuracy of safety monitoring.

(2) Deep integration with AI technology will further enhance machines' capabilities, enabling them to better comprehend complex construction scenarios and behaviors, thereby improving their ability to identify and predict safety issues.

(3) Multi-sensor fusion, integrating data from multiple sensors such as depth cameras, optical cameras, and radar, will provide a more comprehensive view of construction sites, enhancing the understanding of the construction environment and associated risks.

(4) Computer vision systems are poised to increasingly leverage cloud computing and big data analytics for processing and storing vast amounts of image and video data, enabling long-term data analysis, trend detection, and decision support in safety management.

(5) Integration with smart personal protective equipment (PPE), such as smart helmets or goggles, will provide workers with enhanced safety measures by monitoring their physiological status, delivering real-time alerts, and improving control over unsafe behaviors.

This integration of computer vision with smart PPE will enhance worker safety by enabling real-time monitoring of physiological status, providing alerts, and improving control over unsafe behaviors.

X. Acknowledgement

I would also like to express my gratitude to VinBigdata Big Data Research Institute for their valuable contributions through discussions and guidance throughout the project.

In addition, I would like to acknowledge the use of ChatGPT, Axriv, ASCE Library, Science Direct, Taylor & Francis for their great support in conducting this research.

References

[1] The Ministry of Labor, War invalids and Social Affairs of the People's Republic of Viet Nam: Work accident situation 2023
[2] International Labour Office 2005 Facts about safety at work.
[3] Wang M, Wong P, Luo H, Kumar S, Delhi V and Cheng J 2019 Proc. Inter. Symposium on Automation and Robotics in Construction 36 399-406.
[4] Guo B H, Zou Y, Fang Y, Goh Y M and Zou P X 2021 Safety Sci. 135 105130.
[5] Arshad S, Akinade O, Bilal M and Bello S 2023 J. Build. Eng. 107049.
[6] Xu S, Wang J, Shou W, Ngo T, Sadick A M and Wang X 2021 Archi. Comput. Methods Eng. 28 3383-3397.
[7] Fang W, Love P E, Luo H and Ding L 2020 J. Advanced Engineering Informatics 43 100980.
[8] Liu W, Meng Q, Li Z and Hu X 2021 Build. 11(9) 409.
[9] Yan X and Kim Y C 2018 A conceptual framework of ITSMCA for a building collapse accident Eng. Constr. Archit. Ma. 25(6) 721-737.
[10] Ahn Y, Choi H and Kim B.S 2023 Development of early fire detection model for buildings using computer vision-based CCTV J. Build. Eng. 65 105647.
[11] Pushkar A, Senthilvel M and Varghese K 2018 Automated progress monitoring of masonry activity using photogrammetric point cloud ISARC 35 1-7.
[12] Luo H, Wang M, Wong P K Y and Cheng J C 2020 Full body pose estimation of construction equipment using computer vision and deep learning techniques AUTCON 110 103016.
[13] Zhu D, Wen H and Deng Y 2020 Pro-active warning system for the crossroads at construction sites based on computer vision Eng. Constr. Archit. Ma. 27(5) 1145-1168.
[14] Wilkins J R 2011 Construction workers' perceptions of health and safety training programmes Constr. Ma. Econom. 29(10) 1017-1026.
[15] Jeelani I, Albert A and Han K 2020 Improving Safety Performance in Construction Using Eye-Tracking, Visual Data Analytics, and Virtual Reality In Construction Research Congress 2020 American Society Civil Eng pp 395-404.
[16] Eiris R, Gheisari M and Esmaeili B 2018 PARS: Using augmented 360-degree panoramas of reality for construction safety training Int. J. Environ. Health. Res. 15(11) 2452.
[17] Fang W, Ding L, Luo H and Love P E 2018 AUTCON 91 53-61.
[18] Hayat A and Morgado-Dias F 2022 J. Deep learning-based automatic safety helmet detection system for construction safety. Applied Sciences. 12(16) 8268.
[19] Ding L, Fang W, Luo H, Love P E, Zhong B and Ouyang X 2018 J. A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. AUTCON. 86 118-124.

[20] Alateeq M M, Fathimathul Rajeena P P and Ali M A 2023 Construction Site Hazards Identification Using Deep Learning and Computer Vision Sustainability 15(3) 2358.

[21] Yan X and Kim Y C 2018 A conceptual framework of ITSMCA for a building collapse accident Eng. Constr. Archit. Ma. 25(6) 721-737.

[22] Park M W and Brilakis I 2012 AUTCON 28 15-25.

[23] Park M.W, Palinginis E and Brilakis I 2012 Detection of construction workers in video frames for automatic initialization of vision trackers In Construction Research Congress 2012: Construction Challenges in a Flat World pp 940-949.

[24] Park M W and Brilakis I 2016 AUTCON 72 129-142.

[25] Liu W, Shao Y, Zhai S, Yang Z and Chen P 2023 IEICE Transactions on Information and Systems 106(5) 653-661.

[26] Azar E.R 2016 Construction equipment identification using marker-based recognition and an active zoom camera J. Constr. Eng. Ma. 30(3) 04015033.

[27] Xiao B and Kang S C 2021 Vision-based method integrating deep learning detection for tracking multiple construction machines J. Constr. Eng. Ma. 35(2) 04020071.

[28] Zhang M, Cao Z, Yang Z and Zhao X 2020 J. Constr. Eng. Ma. 146(6) 04020051.

[29] Zhang M and Ge S 2022 Vision and trajectory–Based dynamic collision prewarning mechanism for tower cranes J. Constr. Eng. Ma. 148(7) 04022057.

[30] Tsuchiya K and Ishigami G 2020 Vision-based measurement of spatio-temporal deformation of excavated soil for the estimation of bucket resistive force J. Terramechanics 90 11-21.

[31] Naghshbandi S N, Varga L and Hu Y 2021 J. Technologies for safe and resilient earthmoving operations: A systematic literature review AUTCON 125 103632.

[32] Goodrum P M, Zhai D and Yasin M F 2009 J. Relationship between changes in material technology and construction productivity J. Constr. Eng. Ma. 135(4) 278-287.

[33] Dimitrov A and Golparvar-Fard M 2014 J. Adv. Eng. Inform. 28(1) 37-49.

[34] Mahami H, Ghassemi N, Darbandy M T, Shoeibi A, Hussain S, Nasirzadeh F, Alizadehsani R, Nahavandi D, Khosravi A and Nahavandi S 2020 J. arXiv preprint arXiv 16344.

[35] Dinh T H, Ha Q P and La H M 2016 Computer vision based method for concrete crack detection. Int. Control Automation Robotics & Vision (ICARCV) 1-6.

[36] Koch C, Georgieva K, Kasireddy V, Akinci B and Fieguth P 2015. Adv. Eng. Inform. 29(2) 196-210.

[37] Deng J, Singh A, Zhou Y, Lu Y and Lee V C S 2022 Constr Build Mater. 356 129238.

[38] S. Shetye, S. Shetty, S. Shinde, C. Madhu and A. Mathur, "Computer Vision for Industrial Safety and Productivity," 2023 International Conference on Communication System, Computing and IT Applications (CSCITA), Mumbai, India, 2023, pp. 117-120, doi: 10.1109/CSCITA55725.2023.10104764.

[39] Wang, Yu. (2024). Research on computer vision application for safety management in construction. Theoretical and Natural Science. 30. 232-242. 10.54254/2753-8818/30/20241124.

[40] Fang, Weili & Ding, Lieyun & Luo, Hanbin & Li, Heng & Peña-Mora, Feniosky & Zhong, Botao & Zhou, Cheng. (2020). Computer vision applications in construction safety assurance. Automation in Construction. 110. 103013. 10.1016/j.autcon.2019.103013.

[41] Guo, B.; Zou, Y.; Chen, L. A Review of the Applications of Computer Vision to Construction Health and Safety. 2018.

[42] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, C. Bregler, Efficient object localization using convolutional networks, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 648–656.

[43] Y. Yu, H. Guo, Q. Ding, H. Li, M. Skitmore, An experimental study of real-time identification of construction workers' unsafe behaviors, Autom. Constr. 82 (2017) 193–206

[44] R.M. Choudhry, Behavior-based safety on construction sites: a case study, Accid. Anal. Prev. 70 (2014) 14–23

[45] H. Luo, W. Fang, P.E.D. Love, L. Ding, Computer vision for behaviour-based safety in construction: A review and future directions, Adv. Eng. Inform. 43 (2020) 100980.

[46] S. Guo, P. Zhang, L. Ding, Time-statistical laws of workers' unsafe behavior in the construction industry: a case study, Physica A: Statistical Mechanics and Its Applications 515 (2019) 419–429

[47] W. Samek, T. Wiegand, K.-R. Müller, Detection of construction workers under varying poses and changing the background in image sequences via very deep residual networks.

[48] Marcus, G. Deep learning: A critical appraisal. (2018) arXiv preprint arXiv:1801. 00631.

[49] M.T. Ribeiro, S. Singh, C. Guestrin, "Why should I trust you?": Explaining the predictions of any classifier, (2016), https://doi.org/10.18653/v1/n16-3020 Conference of the North American Chapter of the Associate for Computational Linguistics: Demonstractions

[50] T . Gruber, A translation appraoch to portable ontology specifications, Knowl. Acquis. 5 (2) (1993) 199–220

[51] W. Fang, L. Ding, B. Zhong, P.E.D. Love, H. Luo, Automated detection of workers and heavy equipment on construction sites: a convolutional neural network approach, Adv. Eng. Inform. 37 (2018) 139–149.

[52] P. Tang, D. Huber, B. Akinci, R. Lipman, A. Lytle, Automatic reconstruction of asbuilt building information models from laser-scanned point clouds: a review of related techniques, Autom. Constr. 19 (2010) 829–843

[53] R. Lu, I. Brilakis, Digital twinning of existing reinforced concrete bridges from labeled point clusters, Autom. Constr. 105 (2019).

[54] M. Chu, J. Matthews, P.E.D. Love, Integrating mobile building information modeling and augmented reality systems: An experimental study, Autom. Constr. 85 (2018) 305–316

[55] M. J. Kim, X. Wang, P.E.D. Love, H. Li, S. C. Kang, Virtual reality for the built environment: a critical review of recent advances, Journal of Information Technology in Construction. 18 (2), 279–305

[56] J. Llinas, D.L. Hall, An introduction to multisensor data fusion, in The Proceedings of the IEEE (1997), pp. 6–23.

[57] H. Hu, N. Phan, J. Geller, H. Vo, B. Manasi, X. Huang, S. Di Lorio, T. Dinh, S.A. Chun, Deep self-taught learning for detecting drug abuse risk behavior in tweets, Computational Data and Social Networks, Springer International Publishing, Cham, 2018, pp. 330–342

[58] P. Kang, D. Kim, S. Cho, Semi-supervised support vector regression based on selftraining with label uncertainty: An application to virtual metrology in semiconductor manufacturing, Expert Syst. Appl. 51 (2016) 85–106.

[59] X. Tang, F. Guo, J. Shen, T. Du, Facial landmark detection by semi-supervised deep learning, Neurocomputing 297 (2018) 22–32 https://doi.org/10.1016/j.neucom. 2018.01.080

[60] P.K. Rhee, E. Erdenee, S.D. Kyun, M.U. Ahmed, S. Jin, Active and semi-supervised learning for object detection with imperfect data, Cogn. Syst. Res. 45 (2017) 109–123 https://doi.org/10.1016/j.cogsys.2017.05.006

[61] Aparna, Y. Bhatia, R. Rai, V. Gupta, N. Aggarwal, A. Akula, Convolutional neural networks based potholes detection using thermal imaging, Journal of King Saud University - Computer and Information Sciences, (2019).

[62] E. Shi, Q. Li, D. Gu, Z. Zhao, A Method of Weather Radar Echo Extrapolation Based on Convolutional Neural Networks, MultiMedia Modeling, Springer International Publishing, Cham, 2018, pp. 16–28

[63] P. Ding, Y. Zhang, W.-J. Deng, P. Jia, A. Kuijper, A light and faster regional convolutional neural network for object detection in optical remote sensing images, ISPRS J. Photogramm. Remote Sens. 1

[64] N. Qin, X. Hu, H. Dai, Deep fusion of multi-view and multimodal representation of ALS point cloud for 3D terrain scene recognition, ISPRS J. Photogramm. Remote Sens. 143 (2018) 205–212

[65] V.N. Nguyen, R. Jenssen, D. Roverso, Automatic autonomous vision-based power line inspection: a review of current status and the potential role of deep learning, Int. J. Electr. Power Energy Syst. 99 (2018) 107–120

[66] H. Luo, C. Xiong, W. Fang, P.E.D. Love, B. Zhang, X. Ouyang, Convolutional neural networks: computer vision-based workforce activity assessment in construction, Autom. Constr. 94 (2018) 282–289

[67] M.Z. Hossain, F. Sohel, M.F. Shiratuddin, H. Laga, A comprehensive survey of deep learning for image captioning, ACM Comput. Surv. 51 (2019) 1–36

[68] U. Nadeem, S.A.A. Shah, F. Sohel, R. Togneri, M. Bennamoun, Deep learning for scene understanding, Handbook of Deep Learning Applications, Springer International Publishing, Cham, 2019, pp. 21–51

[69] H. Li, M. Lu, S.-C. Hsu, M. Gray, T. Huang, Proactive behavior-based safety management for construction safety improvement, Saf. Sci. 75 (2015) 107–117

[70] W. Lee, K.-Y. Lin, E. Seto, G.C. Migliaccio, Wearable sensors for monitoring on-duty and off-duty worker physiological status and activities in construction, Automconstr. 83 (2017) 341–353
[71] S. Dong, H. Li, Q. Yin, Building information modeling in combination with real-time location systems and sensors for safety performance enhancement, Saf. Sci. 102 (2018) 226–237
[72] W. Fang, L. Ding, B. Zhong, P.E.D. Love, H. Luo, Automated detection of workers and heavy equipment on construction sites: a convolutional neural network approach, Adv. Eng. Inform. 37 (2018) 139–149
[73] X. Luo, H. Li, D. Cao, Y. Yu, X. Yang, T. Huang, Towards efficient and objective work sampling: recognizing workers' activities in site surveillance videos with twostream convolutional networks, Autom. Constr. 94 (2018) 360–370
[74] T. Cheng, J. Teizer, G.C. Migliaccio, U.C. Gatti, Automated task-level activity analysis through fusion of real-time location sensors and worker's thoracic posture data, Autom. Constr. 29 (2013) 24–39
[75] A. Kelm, L. Laußat, A. Meins-Becker, D. Platz, M.J. Khazaee, A.M. Costin, M. Helmus, J. Teizer, Mobile passive radio frequency identification (RFID) portal for automated and rapid control of personal protective equipment (PPE) on construction sites, Autom. Constr. 36 (2013) 38–52
[76] D. Yang, J. Wang, D. Li, K.S.C. Kuang, Fatigue crack monitoring using plastic optical fibre sensor, Procedia Structural Integrity 5 (2017) 1168–1175
[77] X. Jin, S. Yuan, J. Chen, On crack propagation monitoring by using reflection spectra of AFBG and UFBG sensors, Sensors Actuators A Phys. 285 (2019) 491–500
[78] Mahony N, Campbell S, Carvalho A, Harapanahalli S, Hernandez G V, Krpalkova L, Riordan D, Walsh J 2010 Binary robust independent elementary features dblp, , deep learning vs. Traditional computer vision in advances in computer vision proceedings of the 2019 computer vision conference (cvc). Springer nature switzerland ag, pp. 128-144
[79] Website Source: https://www.analyticsvidhya.com/blog/2020/05/social-distancing-detection-tool-deeplearning (last updated on 15-Oct-2020)
[80] Website: https://www.osha.gov (last updated on 15-Oct-2020)
[81] Seo J et al. 2015 Computer vision techniques for construction safety and health monitoring, adv. Eng. Informat. http://dx.doi.org/10.1016/j.aei.2015.02.001
[82] Ringen K, Seegal J and England A 1995 Safety and health in the construction industry, ann. Rev. Public health 16 (1) pp 165–188
[83] Reese C D and Eidson J V 2006 Handbook of osha construction safety and health, crc press
[84] Website: https://www.osha.gov/personal-protective-equipment (last updated on 15-Oct-2020)
[85] Osha, worker safety series: construction Website: https://www.osha.gov/publications/ osha3252/3252.html (last updated on 15-Oct-2020)