Review

Current Status and Future Directions of Deep Learning Applications for Safety Management in Construction

Hieu T. T. L. Pham [1], Mahdi Rafieizonooz [1], SangUk Han [1,*] and Dong-Eun Lee [2]

1 Department of Civil and Environmental Engineering, Seoul Campus, College of Engineering, Hanyang University, Seoul 04763, Korea; phamhieu1816@hanyang.ac.kr (H.T.T.L.P.); mahdirafiei@hanyang.ac.kr (M.R.)

2 Department of Architectural Engineering, School of Architectural, Civil, Environmental and Energy Engineering, Kyungpook National University, Daegu 41566, Korea; dolee@knu.ac.kr

* Correspondence: sanguk@hanyang.ac.kr

Abstract: The application of deep learning (DL) for solving construction safety issues has achieved remarkable results in recent years that are superior to traditional methods. However, there is limited literature examining the links between DL and safety management and highlighting the contributions of DL studies in practice. Thus, this study aims to synthesize the current status of DL studies on construction safety and outline practical challenges and future opportunities. A total of 66 influential construction safety articles were analyzed from a technical aspect, such as convolutional neural networks, recurrent neural networks, and general neural networks. In the context of safety management, three main research directions were identified: utilizing DL for behaviors, physical conditions, and management issues. Overall, applying DL can resolve important safety challenges with high reliability; therein the CNN-based method and behaviors were the most applied directions with percentages of 75% and 67%, respectively. Based on the review findings, three future opportunities aiming to address the corresponding limitations were proposed: expanding a comprehensive dataset, improving technical restrictions due to occlusions, and identifying individuals who performed unsafe behaviors. This review thus may allow the identification of key areas and future directions where further research efforts need to be made with priority.

Keywords: construction safety; unsafe behaviors; physical safety management; safety management issues; deep learning

1. Introduction

Construction is a large, dynamic, and complex field offering a large number of job opportunities for millions of people worldwide [1]. In addition, construction sites also contain various risks (e.g., struck-by accidents [2] and fall accidents [3]), and the accident rate continues to rise over time. According to global statistical data, the construction industry's accidental death and injury rates are three and two times higher than those of other industries, respectively [4]. The number of fatal injuries in this industry in the United States increased by 16%, from 781 in 2011 to 908 in 2014 [5], and its injuries and accidents in 2015 were 50% higher than those in any other industry [3]. These percentages reached 40% of the total accidents in Japan, 25% in the United Kingdom, and 50% in Ireland [6]. Although various countries have put effort into construction safety-related laws, regulations, and management systems over the past decades, their safety performance in construction is still unsatisfactory [7]. Thus, it is essential to apply an appropriate method to assist safety management in the construction industry.

To prevent occupational accidents, Sarkar and Maiti (2020) [8] investigated and reported several existing approaches, such as survey-based qualitative analysis, conventional

statistical analysis, and data-driven machine-learning-based analysis. By reviewing publications examining the application of machine learning (ML) approaches in accident analysis, they also illustrated that ML outperforms its traditional counterpart, owing to its several potential benefits, including the capability to deal with large dimensional data, flexibility in recreating data generation structures regardless of complexity, and predictive and interpretive potential by extracting relationships/rules among attributes in data [8]. In support of this observation, Xu and Saleh (2021) [9] argued that ML has the potential to provide new insights and opportunities to address critical challenges in safety applications. However, one of the challenges of ML is that ML problems become extremely difficult for high-dimensional data [10]. Compared to traditional ML, deep learning (DL) algorithms can deal with high-dimensional input data, and they become highly efficient in resolving the issue of data sources such as images and videos when equipped with convolutional layers [9]. Moreover, the rapid development of graphics processing units (GPUs) has dramatically improved the computing capacity for processing ML algorithms, leading to an increase in the number of DL applications [11]. Therefore, Xu and Saleh (2021) [9] emphasized that in all applications to date, DL has considerably outperformed shallow ML algorithms. In this context, researchers in the construction industry have made considerable efforts to keep up with the pace of DL applications [12]. The amount of research on DL in construction has grown exponentially over the past few years, and the applications have spread over many construction areas since their inception [13]. For example, Akinosho et al., (2020) [12] proved that DL was applied to prevalent construction challenges, such as structural health monitoring, construction site safety, building occupancy modeling, and energy demand prediction [12]. In the context of construction safety, DL has also proven its potential for safety management. DL can be used to extract different types of data such as images, videos, text, and signals to reduce construction accident cases by detecting on-site damage conditions [14], detecting unsafe behaviors [15], and analyzing construction safety documents [16].

DL is a subset of ML, and can theoretically deal with all categories of ML [9]. For example, different types of DL techniques used in real-time object detection help develop new helmet detection systems with higher accuracy and less training time [17]. Zhong et al., (2020) [18] demonstrated that DL can be used to automatically extract unstructured safety data from accident reports. As a result, managers become better positioned to make informed and timely decisions about how to ensure construction safety [18]. With these prominent and widespread applications of DL in construction safety, researchers need to understand what typical types of data can be used for different methods (e.g., convolutional neural networks, recurrent neural networks, etc.) for gaining high performance. Moreover, with the extremely rapid advancement of DL algorithms, the review of recent literature can play an important role in understanding the research status of DL studies and exploring an opportunity of its application for further enhancement of construction safety. However, there is limited literature examining the theoretical links between DL and safety management. For example, several review studies, such as [19,20], have mainly focused on construction safety without the detailed review on DL techniques. Hou et al., (2021) [21] carried out a review of the relevant papers on applications of DL for safety management in the architecture, engineering, and construction (AEC) industry; however, a comprehensive linkage between safety and DL methods (e.g., data types and quantities, DL algorithms and their performance, safety factors) was not fully investigated. Moreover, how the results of DL studies can be applied in safety management practice was not clearly presented and discussed by Hou et al., (2021) [21]. By addressing those issues, researchers and managers in the field of construction safety may better understand what type of method has achieved highly accurate results along with the type and amount of data has been used for a certain safety task, as well as the actions managers can take from the result of DL models for improving safety management. This study aims to fill these gaps by comprehensively reviewing DL studies in the construction safety area.

Specifically, this literature review is performed to (1) identify and summarize the current status of recent DL studies in the construction safety area for showing how DL could be applied in previous studies; (2) analyze the links of data type and quantity, and DL models applied and newly proposed with three main research directions of construction safety (e.g., behaviors, physical conditions, and management issues) for understanding how to apply DL models in different safety-related tasks; (3) review the contributions of DL results in safety management practice; and (4) outline practical challenges and future opportunities associated with the applications for improving and fully exploiting the DL contributions in safety. This review may thus allow the identification of key areas and future directions where further research efforts need to be made with priority. The remainder of this paper is organized as follows. The paper firstly presents the research methodology used in this review (Section 2). An overview of DL algorithms commonly used for construction safety is then presented from a technical aspect (Section 3). Subsequently, this paper summarizes the current status of safety-related papers for an in-depth understanding of DL applications for safety management (Section 4). Along with a comprehensive review, this study discusses the contributions, practical challenges, and future opportunities of applying DL approaches to practice (Sections 5 and 6). Finally, the major findings are summarized to present the significance of this study (Section 7).

2. Research Methodology

With the purpose of analyzing the current status of DL studies in safety performed to understand how well the DL methods have been applied for safety management as well as how distinct DL models could address safety issues with different specific types of data, this study adopts a content-analysis-based review method, a systematic and structured technique "for compressing many words of text into fewer content categories based on explicit coding rules" to identify key research themes for literature review [22]. Content analysis is a research tool utilized to determine the presence of certain words, themes, or concepts within several given qualitative data (i.e., text). Using content analysis, researchers can analyze and quantify the presence, meanings and relationships of certain words, themes, or concepts [23]. This method has been well-recognized and widely used for reviewing and synthesizing literature, and rationalizing outcomes in the research field of engineering/construction management [22,24–26]. The review process based on this method consists of three phases: literature search, title- and abstract-based literature selection, and full-paper-based literature selection, as described in Figure 1. In the literature search, an exhaustive search was carried out with keywords regarding DL and construction safety that aimed to find all articles related to the field of review. The title- and abstract-based literature selection was then conducted to filter papers applying DL to handle safety issues based on reading titles and abstracts. After that, an overall screening was performed in the phase of full-paper-based literature selection that aimed to identify the articles relevant only to construction safety and DL by reading the full paper. Therefore, the most significant DL studies on construction safety were collected and reviewed to guarantee the provision of fit and quality research materials for this study.

2.1. Literature Search

The first step of the review was an exhaustive search in Scopus and Google Scholar. Keywords and Boolean operators, AND and OR, were used to ensure that all relevant literature was captured from 2014 to 2021. According to Akinosho et al., (2020) [12], DL became popular with the achievements of CNNs in the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC2012), and its applications in the construction industry have achieved significance since around 2014. Thus, the chosen dates were based on the DL revolution. The search strings used were "deep learning" OR "computer vision" OR "CNN" OR "RNN" OR "neural networks" AND "construction safety" OR "construction hazard" OR "construction accident" OR "safety management". Initially, 387 documents were identified. To limit the scope of the search results, these documents were further screened

by including only journal articles published in English and the remaining 145 papers. Moreover, we chose articles with the highest level of relevance to the research scope, namely, engineering, computer science, materials science, and management. After this screening, a total of 126 documents, including articles and conference papers, were selected as the literature sample.
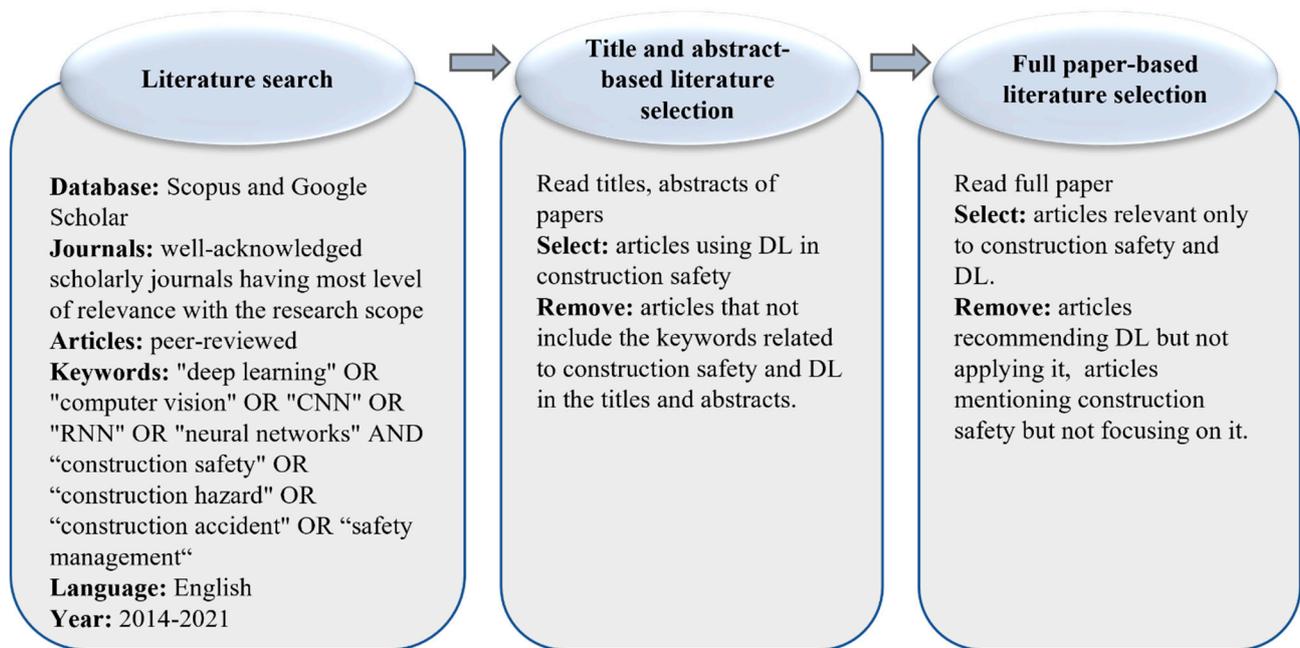
Literature search

Database: Scopus and Google Scholar
Journals: well-acknowledged scholarly journals having most level of relevance with the research scope
Articles: peer-reviewed
Keywords: "deep learning" OR "computer vision" OR "CNN" OR "RNN" OR "neural networks" AND "construction safety" OR "construction hazard" OR "construction accident" OR "safety management"
Language: English
Year: 2014-2021

Title and abstract-based literature selection

Read titles, abstracts of papers
Select: articles using DL in construction safety
Remove: articles that not include the keywords related to construction safety and DL in the titles and abstracts.

Full paper-based literature selection

Read full paper
Select: articles relevant only to construction safety and DL.
Remove: articles recommending DL but not applying it, articles mentioning construction safety but not focusing on it.

Figure 1. Research methodology.

2.2. Title- and Abstract-Based Literature Selection

This stage of document screening was conducted to identify articles relevant to construction safety and DL for further analysis. These documents from the literature search were manually screened by reading and exploring the titles and abstracts to identify and extract relevant articles. Publications that did not include keywords regarding construction safety and deep learning in titles or abstracts were screened out. The total number of documents remaining after this phase was approximately 98.

2.3. Full-Paper-Based Literature Selection

This phase aims to remove irrelevant papers by examining the contents of the articles. The remaining documents from the previous phase were screened by reading the full paper to identify articles relevant only to construction safety and DL. For example, articles (e.g., [27]) that only mentioned "deep learning" but did not focus on DL methods, were removed. Several articles, such as [28], were also removed, as they did not focus on safety in construction, although the term "construction safety" was found in its abstract. Similar articles that applied DL in manufacturing, structural assessment, and crack and defect detection were removed as they did not focus on safety issues in the construction industry. After the third screening, a total of 66 papers remained for an in-depth review and analysis.

2.4. Results

According to the final paper selection, a total of 66 papers in journals shown in Figure 2 were identified for further analysis. Figure 3 shows the number of publications by year, which proves the development of DL applications in construction safety in recent years. The number of studies using DL increased from 2018 to 2021 and is likely to continue to rise in the coming years. Figures 4–7 present an overview of the reviewed papers. In addition to extracting information related to DL models and safety factors, which is the purpose of

this study, we also present the type of data and accident types to provide a comprehensive overview of what types of accidents researchers have attempted to reduce. Overall, these figures show that the CNN-based method and behaviors were the most applied directions with percentages of 75% and 67%, respectively; images were the most used data in these models (73%), and struck-by and other general accidents were two types of accidents DL studies have focused on with the percentages of 36% and 38%, respectively.
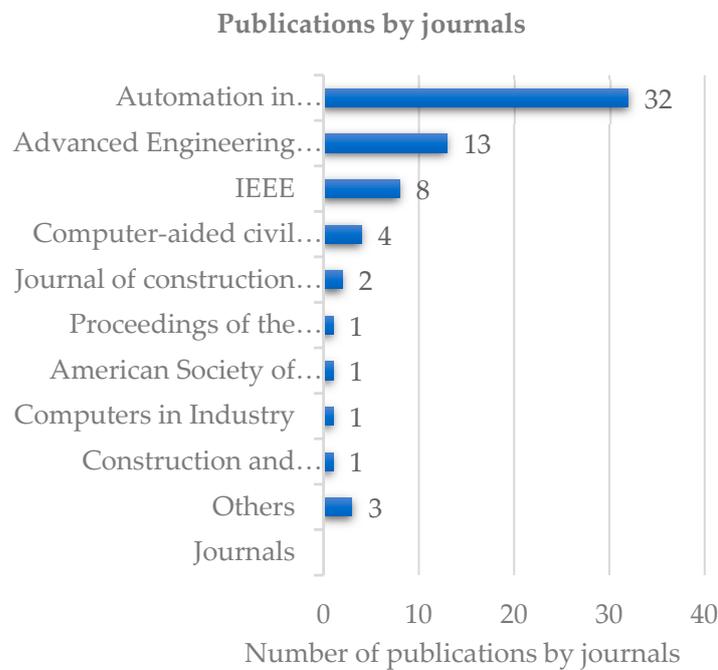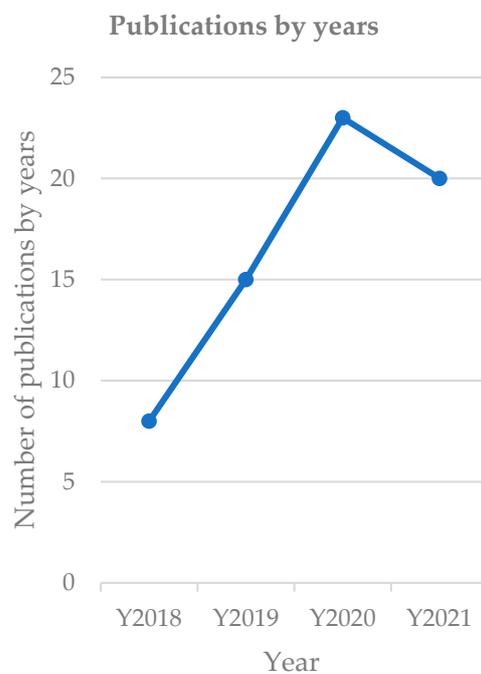
Figure 2. Publications by journals.

Figure 3. Publications by years.

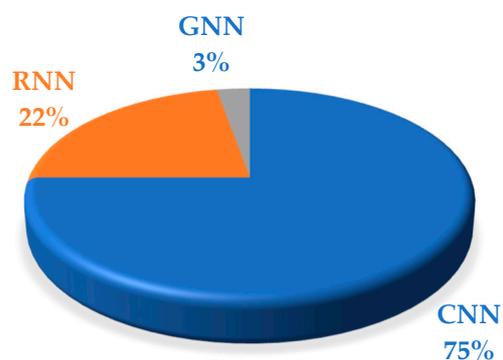**Percentage of methods in total cases**

Figure 4. Percentage of methods in total cases.

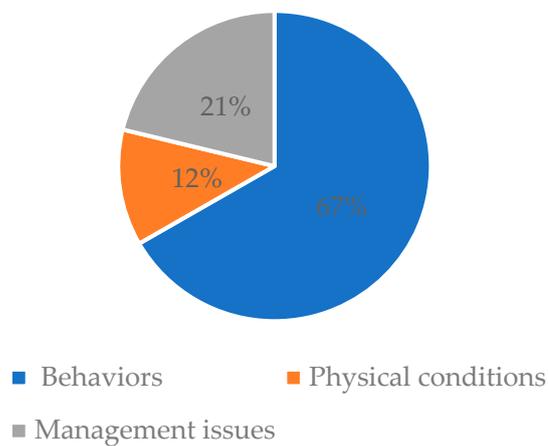**Percentage of publications by safety factors**

Figure 5. Percentage of publications by safety factors.
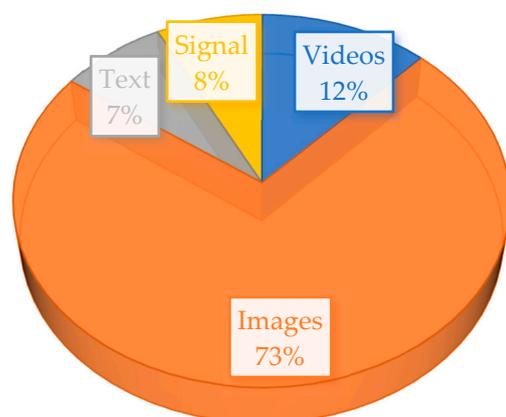
Percentages of publications by data

Figure 6. Percentages of publications by data.

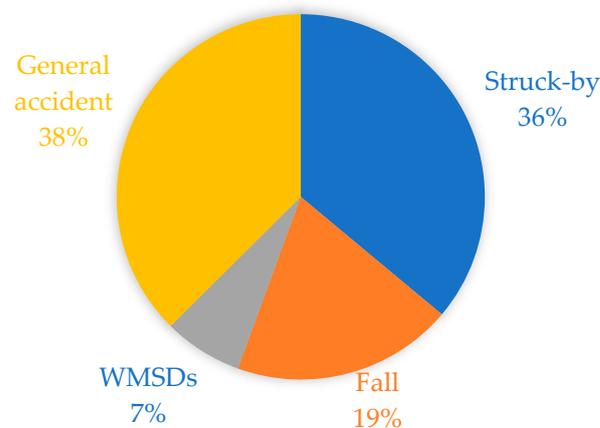Percentage of publications by accident types

Figure 7. Percentages of publications by accident types.

3. Overview of Deep Learning Architectures

DL is a set of ML algorithms that attempt to learn features at multiple levels with different levels of abstraction [29]. The grades in these learned models correspond to different levels of concepts, where the same lower-level concepts can support many higher-level concepts [29]. Thus, a DL architecture can be defined as an artificial neural network (ANN) with two or more hidden layers to enhance prediction accuracy [29,30]. Three important reasons for the popularity of DL today are the drastic increase in the abilities of chip processing (e.g., GPU units), the significant increase in the size of data used for training, and the recent algorithm advances in ML and signal/information processing studies [29,31]. These advances have enabled DL methods to exploit complex, compositional nonlinear functions, and effectively use both labeled and unlabeled data [29]. Therefore, unlike the architectures of shallow ML, DL networks are capable of processing nonlinear information [32] and provide training for both supervised and unsupervised categories [33]. With the outstanding ability in processing various types of data, including images, videos, text, speech, and signals, DL networks and techniques have been implemented widely in various fields such as image classification [34], object detection [35], object tracking [36], activity recognition [37], information extraction [38], text classification [39], and speech recognition [40].

According to Khallaf and Khallaf (2021) [13], DL is called "deep" due to the number of layers available in the network model. Generally, the DL architecture is composed of three types of layers: an input layer, hidden layers, and an output layer; the typical architecture of DL is shown in Figure 8. Data are received in an input layer, features are extracted from the datasets via hidden layers depending on the purpose of their application, and the resulting features are passed to the output layer for prediction. In the network, the output of the previous layer is used as the input of the next layer. There are different types of DL architectures [13], and for safety management, the most commonly used types of DL include convolutional neural networks (CNNs), recurrent neural networks (RNNs), and general neural networks (GNNs).

3.1. Convolutional Neural Networks

For DL, the term "deep" is derived from the many hidden layers in the ANN structure [41]. Unfortunately, this structure is receptive to translation and shift deviation, which may adversely affect the performance of classification [42]. To eliminate these drawbacks, an extended ANN version, the CNN, was developed, which can ensure spatial translation and shift invariance [43]. The CNN is a supervised DL architecture mainly used for image analysis applications [30,44,45]. Similar to the ANN, the network consists of multiple

hidden layers between an input layer and an output layer (Figure 9). However, the hidden layers comprise convolutional, pooling, and fully connected layers. The convolution filter acts as a feature extractor by learning hidden patterns from different input signals [41] and generating relevant feature maps through kernels or filters [30]. The calculation of convolution is defined as

$$O_{x,y} = \sum_{c=0}^{C_{in}-1} \sum_{j=0}^{K-1} \sum_{i=0}^{K-1} I_{x+s\times i, y+s\times j} \times w_{i,j} + b_{x,y} \tag{1}$$

where $I_{x+s\times i, y+s\times j}$ is the value of the input feature at the point of $(x + s \times i, y + s \times j)$, C_{in} is the number of input channels, K is the kernel size, s is the stride of convolutional layer, $w_{i,j}$ is the weight in the kernels, $b_{x,y}$ is the bias, and $O_{x,y}$ is the value of the output feature at the point of (x, y). This convolutional layer thus allows the detection of low-level features, such as lines and edges, as well as high-level features such as shapes and objects [46]. In this process, the convolutional layer can enhance the input data features and reduce noise [32]. The convolutional layer is likely connected to a pooling layer with a nonlinear mapping function (e.g., rectified linear unit (ReLU)) [47]. The appropriate pooling layer has a positive effect on reducing the input dimension without losing information [47]. Different types of pooling methods exist, such as global pooling, average pooling, and max pooling [30]. In particular, for extracting features from images, the performance of the maximum pooling method is better than that of average pooling [48]. Maximum pooling splits the input image into multiple rectangular regions based on the size of the filter, and its output is the maximum value for each region [49]. The output of the max pooling layer can be calculated as

$$N_{x,y}^{out} = \underset{m,n\in[0,i-1]}{Max} \left(N_{x+m,y+n}^{in} \right) \tag{2}$$

where the max pooling layers take the maximum value from the region $i \times i$ of input as the output, $N_{x+m,y+n}^{in}$ is the value of the input at the point of $(x + m, y + n)$, and $N_{x,y}^{out}$ is the value of output at the point of (x, y). This process is known as downsampling or subsampling [30]. After these layers, the fully connected layer commonly connects all neurons from the previous layer to every single neuron [32]. Thus, this layer sets a weighted sum of all the previous layer outputs to determine a specific target output [41].
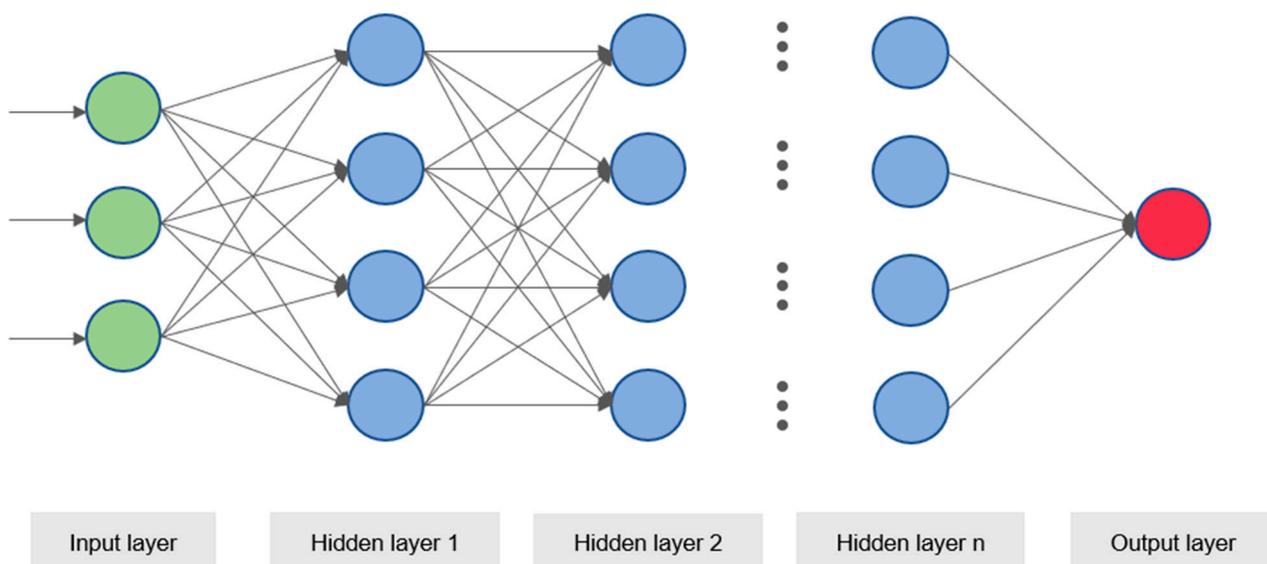

Figure 8. Typical deep learning architecture with three types of layers: an input layer, hidden layers, and an output layer.
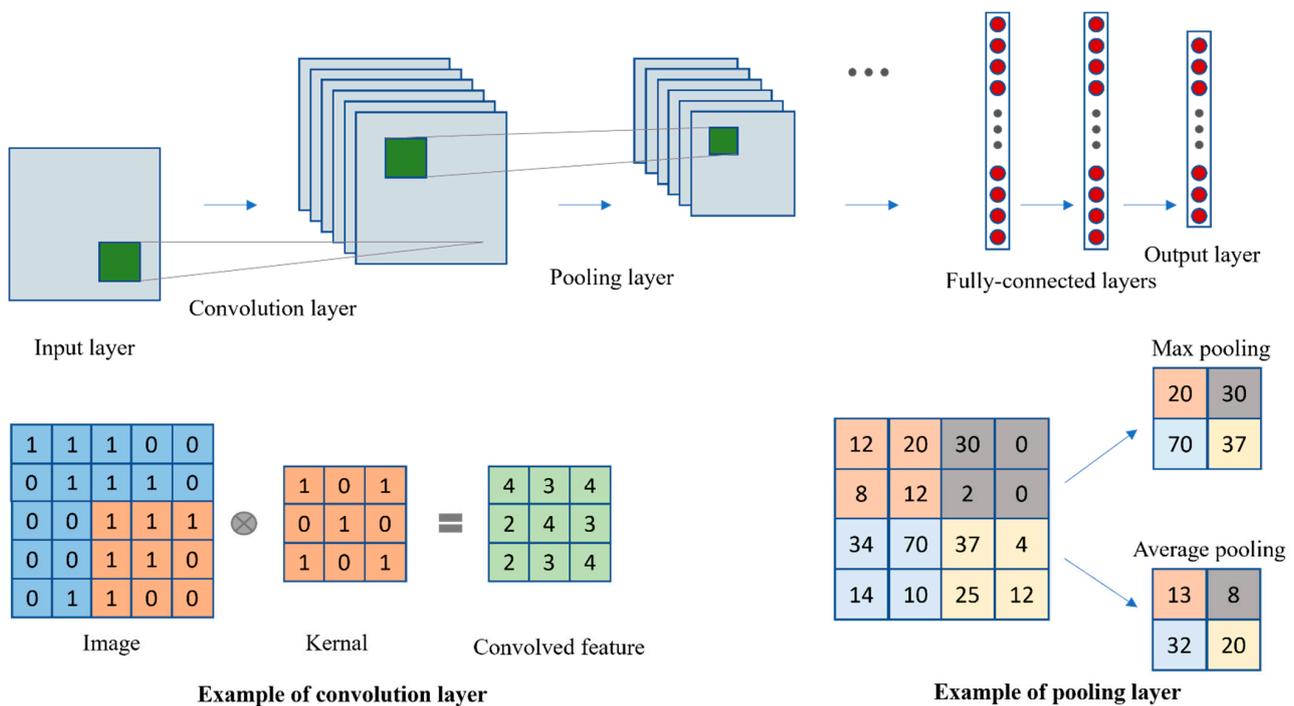
Figure 9. A typical CNN architecture.

The variations of CNN methods include region-based CNN (R-CNN), fast R-CNN, faster R-CNN, and you only look once (YOLO). As discussed above, DL methods with convolutional networks are widely used for image processing tasks. Among the various applications of CNNs, object detection frameworks combining both classification and localization to detect and draw boxes around objects in images have markedly developed in recent years [50]. According to Koirala et al., (2019) [50], early object detection frameworks based on CNN used a sliding window approach at evenly spaced locations over the image, where many patches are generated to classify each patch as containing an object or not. Thus, feeding all available patches for multiscale detection to a CNN slowed the object detection framework [50]. R-CNN replaced the sliding window method by using a group of boxes for the image and then analyzing each box if either of the boxes contained a target [51]. The entire target identification method through R-CNN uses the following three models: a linear SVM classifier for object identification, CNN employed for characteristic extraction, and a regression model required to tighten the bounding boxes [52]. Therefore, the drawbacks of R-CNN are multiple stages of training, taking up disk space and training time consuming cumbersome steps [53]. Therefore, a fast R-CNN was developed to improve the detection speed of R-CNN [50]. In place of using three different models of R-CNN, fast R-CNN [54] employs a model to extract characteristics from different regions. However, the drawback of the fast R-CNN method is that it is based on a selective search [55]; for example, 2000 sections are excerpted per image [52]. Thus, this approach may increase the running time of the fast R-CNN method [52]. In contrast, faster R-CNN creatively utilizes the convolution network to create the proposed box and shares the convolution network with the object detection network, which reduces the number of proposed frames, for example, from approximately 2000 to approximately 300 [56]. However, despite the speed of faster R-CNN-based detection model being improved compared to that of fast R-CNN, it is still too slow to apply to real-time video streaming [50]. To address this limitation, YOLO was developed to generate a one-step process involving detection and classification [57]. YOLO's idea differs from other traditional systems in that bounding box predictions and class predictions are performed simultaneously [57], making YOLO one of the fastest object detection methods [50].

3.2. Recurrent Neural Networks

The neurons of a fully connected network or a CNN are fully connected in different layers but disconnected in the same layer; each layer processes signals independently and then propagates to the next layer [48]. In this regard, this architecture cannot resolve the problem of relationships between input data [32]. RNNs can be considered as another class of DL networks that are used for sequential data for supervised and unsupervised learning [29]. An RNN can "remember" past information and utilize the knowledge learned from the past to make its present decision [58]. In RNNs, the output of the previous step is stored and utilized to calculate the current output (Figure 10), which means that the network's input contains both the data from the input layer and the output of the previous hidden layers [32]. The output of the RNN model can be calculated as

$$h_t = f(Ux_t + Wh_{t-1} + b_h) \tag{3}$$

$$O_t = softmax(Vh_t + b_o) \tag{4}$$

where U is the weights matrix of the input x_t to the hidden layers, W is the duplicated recurrent weight matrix, V represent sts the hidden to output weight matrix, f is a nonlinear activation function, and b_h and b_o are the biases added to the hidden and output layers, respectively. Thus, the RNN is extremely powerful for modeling sequence data (e.g., speech or text) [29].
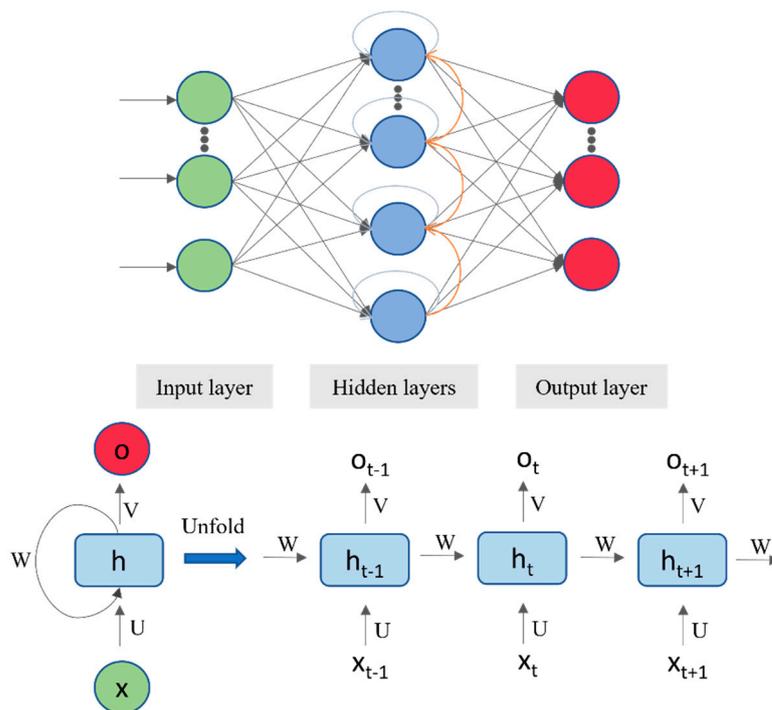
Figure 10. A typical RNN architecture.

Despite the promising performance of RNN, vanishing gradient is a significant problem in the conventional RNN because it makes the gradient easily vanish (e.g., the previous information is lost through multiple layers), and the model learning process becomes much more difficult [59]. One solution to solve this problem is to use long short-term memory (LSTM) networks, which can store sequences for a long time, as well as using gated recurrent units (GRUs) [60,61]. The LSTM algorithm combines a memory block with three gates: input, output, and forget gates [41]. The input gate determines what new information is saved and updated in the cell state, the output gate determines what information is utilized based on the cell state, and the forget gate is used to delete the unimportant

information from the cell state. Thus, the difference from RNN is that LSTM can determine what information is useful through the cell, which can avoid the disappearance of the gradient to some extent [48]. The learning capacity of the LSTM cell is also superior to that of a conventional recurrent cell [62]. However, additional parameters increase the computational burden [62]. To reduce the number of parameters, the GRU combines the input and forget gates of the LSTM model into an update gate, and the output gate in the LSTM model is called a reset gate [63]. Thus, the GRU is an extension of LSTM, which achieves a performance comparable to that of LSTM but uses fewer parameters and makes training faster [64].

3.3. General Neural Networks

In addition to the two common methods of DL (i.e., CNN and RNN), bidirectional encoder representations from transformers (BERT) [39] (Figure 11) and other deep learning models for natural language processing (NLP) (Figure 12) and computer vision (CV) [65] (Figure 13) have also been applied in safety management. Unlike recent language representation models, BERT is designed to pretrain deep bidirectional representations from unlabeled text by jointly conditioning on both left and right contexts in all layers [66]. BERT's execution for tasks consists of two phases: pretraining for language understanding and fine-tuning for a specific task such as text classification and text summarization [67]. A pretrained language model can be defined as a black box containing previous knowledge of natural language [68]. The BERT by Devlin et al., (2018) [66] used encoders in a transformer as a substructure for pretraining models for NLP tasks. Specifically, the BERT-based model is pretrained using two unsupervised tasks: (1) the masked language model (LM) predicts some randomly masked tokens in the input to train the bidirectional encoder and (2) next sentence prediction (NSP) predicts the following sentence of the input sentence to understand sentence relationships, so the pretrained BERT model can be more suitable for other NLP applications [69]. BERT can be fine-tuned using a dense layer of neural networks for different classification tasks [68]. The advantages of BERT include its ability to address contextual information extraction owing to its bidirectional ability and faster training capabilities [67]. With the above characteristics, the BERT model demonstrated state-of-the-art performance in many NLP tasks [70]. BERT is known to achieve exceptional results in 11 natural language understanding (NLU) tasks [66]. However, BERT still has specific drawbacks, including the use of BERT-large, made up of 24-layered transformer encoder blocks, and producing a total of 340 million parameters, which may tend to be computationally expensive [67].
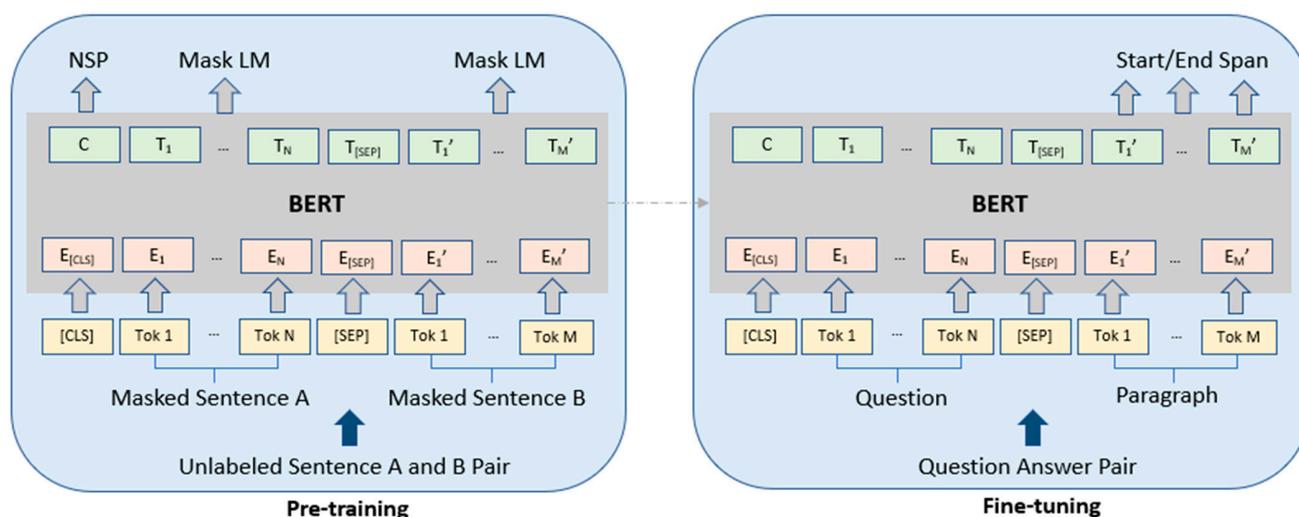

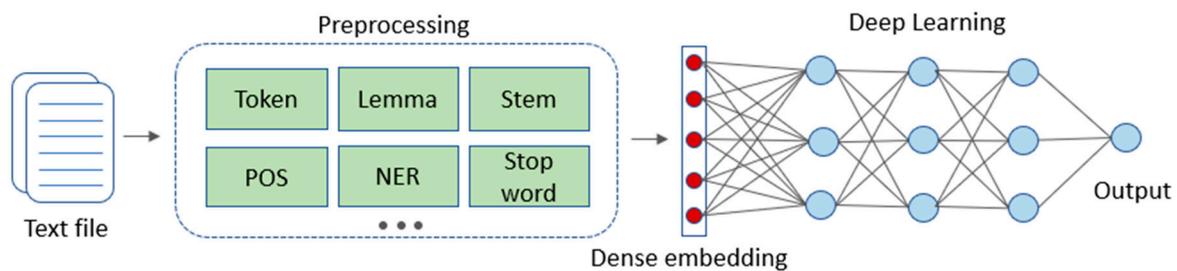**Figure 11.** An example of BERT architecture.

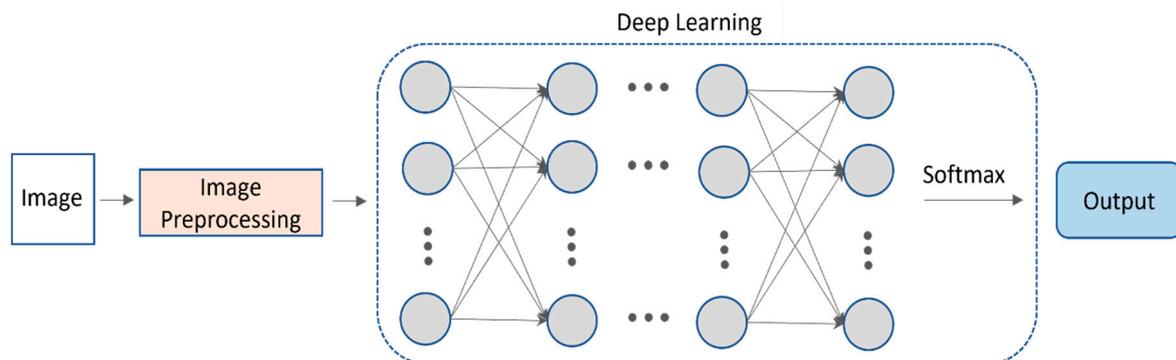**Figure 12.** General NLP model based on DL.

Figure 13. General CV model based on DL.

4. Deep Learning Applications for Construction Safety Management

According to Reason's model [71], on-site safety management is the last layer of management for preventing accidents and requires considerable emphasis. In this context, we focus on construction safety aspects based on a safety management system (SMS). An SMS integrates activities and functions to identify accidents and manage risks in the workplace [72]. Construction safety management can be divided into preconstruction and construction phases [73]. In the preconstruction phase, the potential safety accidents are normally identified based on the experience of safety officers or project managers and eliminated through safety training and safety planning [74]. During construction, hazards are prevented by monitoring workers and the environment at construction sites [75]. Therefore, in general, a safety management system approach focuses on three main aspects: behaviors, physical conditions, and management issues [76,77]. Figure 5 shows the percentage of publications based on these safety factors. The common types of behaviors on construction sites identified [77,78] are (1) pose and gesture, (2) action, (3) interaction, (4) activity, and (5) personal protection equipment (PPE) and safety compliance. We then presented factors that influence the physical conditions on construction sites [76], including (1) site condition (SC), (2) work environment (WE), and (3) site layout (SL). Finally, management issues were discussed [79,80] based on the following subcategories: (1) safety management plan, (2) accident investigation and analysis, and (3) hazard identification and risk management. The general applications of DL in construction safety are shown in Figure 14.

4.1. Behaviors

Unsafe worker behavior is a significant cause of workplace accidents [81]. It has been proven that 88% accidents are caused by workers' unsafe behavior [82]. According to Fam et al., (2012) [83], unsafe behavior occurs when an employee fails to respect safety rules, standards, instructions, procedures, and specified project criteria. In general, unsafe behaviors are factors related to workers' awareness, unsafe actions, and noncompliance attitudes that cause dangerous consequences (e.g., injury). Due to the varying levels of abstraction and complexity of human behaviors, Edwards et al., (2016) [78] proposed a five-level classification system for workers' behaviors, which included pose, gesture,

action, interaction, and activity. Likewise, Guo et al., (2021) [77] proposed a six-level hierarchical framework of safety behavior with the contribution of the safety compliance factor. According to a series of these studies and based on the applications of DL in construction safety, the unsafe behaviors causing accidents in construction are categorized as (1) pose and gesture, (2) action, (3) interaction, (4) activity, and (5) personal protection equipment (PPE) and safety compliance. Table 1 summarizes the DL studies on behaviors in the construction industry.
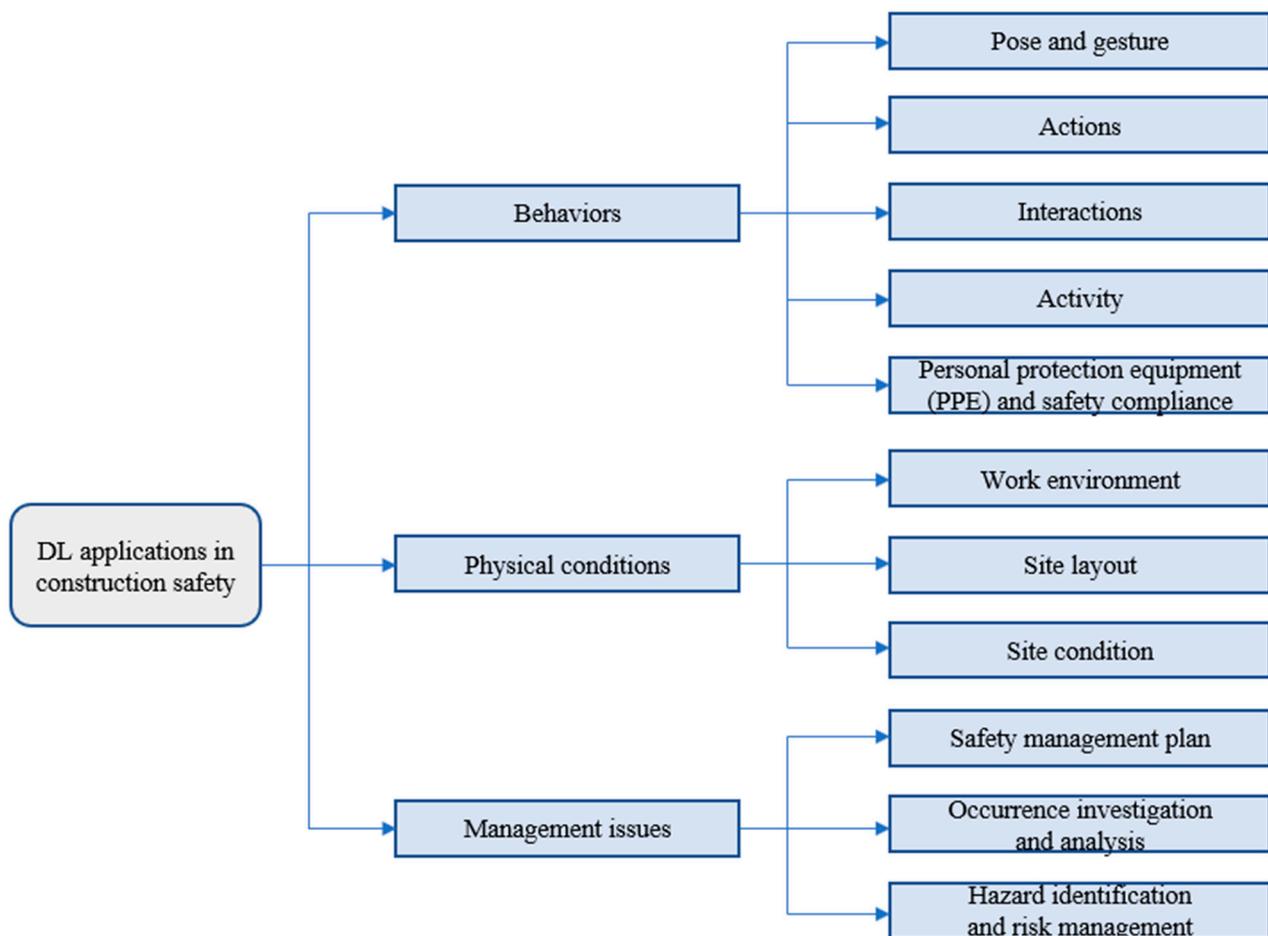

Figure 14. DL applications in construction safety.

Table 1. Construction safety studies about behaviors.

Categories	Type of Data	Numbers of Data Training–Validation–Testing	Method	Accuracy Value	Object/Action	Accident Type	References
Pose and gesture	Images	4483–641–1281	CNN	Accuracy: 0.93	Excavator's pose	(Struck-by)	[84]
	Images	N/A–N/A–3241	CNN	Accuracy: 0.94 Precision: 0.96 Recall: 0.98	Worker's standing, walking, squatting, sitting, or bending.	(Struck-by)	[85]
	Images	2116–235–1008	RNN	Accuracy: 0.96	Ergonomic postures	WMSDs	[86]
	Images	10,000–N/A–N/A	CNN	Accuracy: 0.91	Workers' and excavators' status	Struck-by	[87]
	Images	N/A	CNN	Accuracy: 0.83	Workers' standing still, bending, ladder-climbing/stepping/standing	Fall	[88]
	Videos	N/A	CNN + RNN	F1-score: 0.83	Ergonomic postures	WMSDs	[89]
	Signal	2196 (training and testing)	RNN	Accuracy: 0.99 F1-score: 0.99	Ergonomic postures	WMSDs	[90]
	Signal	32,396 (60%–N/A–40%)	RNN	Accuracy: 0.95	Workers' standing, bending, squatting, walking, twisting, kneeling, and using stairs	WMSDs	[91]
	Images	N/A	CNN	Accuracy: 0.96	Ergonomic postures	WMSDs	[92]
Action	Videos	160–N/A–40	CNN + RNN	Accuracy: 0.92	Ladder-climbing actions	Fall	[15]
	Images	1461–N/A–450	CNN	Precision: 0.75 Recall: 0.9	Worker traversing supports	Fall	[93]
Interaction	Videos	10 (80%–10%–10%)	RNN	N/A	Worker–equipment interactions	Struck-by	[94]
	Videos	5 (70%–10%–20%)	RNN	Accuracy: 0.9	Excavators and dump truck interactions during earthmoving tasks	Struck-by	[95]
	Images	2169 (training and validation)–241	CNN	Precision: 0.87	Worker–equipment interactions	Struck-by	[96]
	Images	3652–N/A–913	CNN	Precision: 0.66 Recall: 0.65	Worker–tool interactions	General accident	[97]
	Images	4114–N/A–398	CNN	Precision: 0.91	Worker–equipment interactions	Struck-by	[98]
	Images	6000–N/A–N/A	CNN	Precision: 0.96 Recall: 0.93	Worker–excavator interactions	(Struck-by)	[99]
	Images	523,966–N/A–50,000	CNN	Accuracy: 0.96 Precision: 0.98 Recall: 0.98 F1-score: 0.98	Worker–equipment interactions	Struck-by	[2]
	Images	N/A	CNN	Recall: 0.5	Components' or crews' relationships	General accident	[100]
	Videos	8000–N/A–2000	RNN	Accuracy: 0.95	Worker–equipment interactions	Struck-by	[37]
	Videos	210–70–84	CNN + RNN	Accuracy: 0.93	Hand signals for instructing tower crane operations	(Struck-by)	[101]
	Images	N/A	CNN	Precision: 1 Recall: 0.82	Worker–equipment interactions	Struck-by	[102]

Table 1. *Cont.*

Categories	Type of Data	Numbers of Data Training–Validation–Testing	Method	Accuracy Value	Object/Action	Accident Type	References
Activity	Images	96–N/A–N/A	CNN	Precision: 0.52 Recall: 0.45 F1-score: 0.48	Mixed activities of workers and equipment	General accident	[35]
	Images	703–235–N/A	CNN	Accuracy: 0.86	Scaffolding activity	Fall	[103]
	Videos	7–N/A–3	CNN + RNN	mAP: 0.73	Earthmoving activity	(Struck-by)	[36]
	Images	N/A	CNN	Accuracy: 0.84	Concrete pouring	(General accident)	[104]
Safety compliance	Images	944–240–288	CNN	mAP: 0.72	PPE (hard hat, vest)	Fall and struck-by	[105]
	Images	81,000–N/A–19,000	CNN	Precision: 0.96 Recall: 0.95	PPE (hard hat)	Fall and struck-by	[106]
	Images	6029–N/A–6000	CNN	Precision: 0.94 Recall: 0.83	PPE (hard hat, glasses, dust mask, safety belt)	Fall and struck-by	[107]
	Images	1587–N/A–1587	CNN	mAP: 0.84	PPE (hard hat)	(General accident)	[108]
	Images	2583–N/A–726	CNN	Accuracy: 0.9	PPE (hard hat, vest)	Fall and struck-by	[109]
	Images	N/A	CNN	Precision: 0.9 Recall: 0.93	PPE (hard hat, harness, anchorage)	Fall	[110]
	Images	693–N/A–130	CNN	Precision: 0.99 Recall: 0.95	PPE (harness)	Fall	[111]
	Images	8000–N/A–N/A	CNN	Precision: 0.83 Recall: 0.83	Noncertified work of workers	(General accident)	[112]
	Images	1366–N/A–N/A	CNN	mAP: 0.55	PPE (hard hat)	Struck-by	[113]
	Images	7000–N/A–200	CNN	Precision: 0.91 Recall: 0.9	PPE (hard hat)	(General accident)	[114]
	Images	64,115–2693–N/A	CNN	mAP: 0.86	PPE (hard hat)	Fall and struck-by	[115]
	Images	1587–N/A−1587	CNN	mAP: 0.87	PPE (hard hat)	(General accident)	[116]
	Images	100,000–N/A–N/A	CNN	mAP: 0.89	PPE (hard hat, vest)	Struck-by	[117]
	Images	9800–N/A–9000	CNN	Accuracy: 0.94 Precision: 0.96 Recall: 0.96	PPE (hard hat)	Struck-by	[118]
	Images	13,000–N/A–1300	CNN	mAP: 0.93	PPE (hard hat)	Fall and struck-by	[119]
	Images	20,554–N/A–1501	CNN	mAP: 0.94	PPE (hard hat)	(General accident)	[120]
	Images	5000–N/A–1000	CNN	mAP: 0.96	PPE (hard hat)	(General accident)	[121]
	Images	N/A	CNN	mAP: 0.58	PPE (hard hat)	Fall	[122]

Note: The mAP represents mean average precision, and WMSDs represent work-related musculoskeletal disorders. The accident types in parentheses were judged by the authors' assessment and not specified in the paper.

4.1.1. Pose and Gesture

Posture-related safety risks have been a significant concern in construction projects that need to be addressed [90]. Pose and gesture are defined as the spatial arrangement of a human body at a single temporal instance, or a temporal pose series or action primitives on a subaction scale [77]. The worker's safety risk level can be assessed based on the worker's current posture by calculating the similarity of the workers' posture to the identified hazardous postures [88]. Several methods can be employed to represent human posture: images, text descriptions, or skeleton data [88]. The goal of human pose estimation is to specify the position of human joints from images or skeleton data provided using motion-capturing hardware [123]. Text description is a user-friendly way to facilitate human understanding, but it removes the objective and quantitative features of the postures [88]. Based on this, researchers have utilized DL methods for detecting unsafe postures using different types of data (e.g., videos [92], images [86], and signals [91]).

DL has been widely and successfully applied for detecting unsafe workers' postures with different typical statuses, including standing still, climbing down, standing on the ladder, and bending. For example, based on posture-location fusion evaluation, Chen et al., (2019) [88] proposed deep CNN architectures to extract human skeletons from sensor images for evaluating the ladder-climbing posture of construction workers with an accuracy of 83%. Likewise, Son et al., (2019) [85] illustrated the ability to accurately and rapidly detect workers in construction sites under different poses in image by using a CNN-based model with an accuracy of 94.3%. In addition, with the development of RNN-based DL, Kim and Cho (2020) [91] obtained the best-performing accuracy of 82.39% from the model using the LSTM network when compared with other conventional ML algorithms in the motion recognition of workers, including standing, bending, squatting, etc. Among these gestures, recent research has focused on ergonomic posture, which poses the highest risk for musculoskeletal disorders (MSDs). For example, Yang et al. (2020) [90] investigated the feasibility of identifying varying physical loading conditions by analyzing the lower body movements of workers while moving concrete bricks. With a high accuracy of 98.6%, the findings contribute to the literature on classifying ergonomically at-risk workers and preventing work-related musculoskeletal disorders (WMSDs) in physically demanding occupations, thus enhancing the health and safety of the construction workplace. Similarly, Zhao and Obonyo (2020) [89] and Yu et al. (2019) [92] proposed a DL-based ergonomic assessment tool to provide automatic and detailed ergonomic assessments of workers based on images.

According to Luo et al., (2020) [84], similar to human poses, the posture of construction machines can be represented by key points. Thus, in addition to the worker's postures, DL was also applied to detect the poses of construction machines. For example, Luo et al., (2020) [84] developed a CNN-based model to automatically estimate the poses of excavators in images captured at construction sites. The experimental results demonstrated the promising performance of the proposed methodology framework for automatically evaluating different full-body poses of construction equipment with high accuracy and fast speed. Likewise, Luo et al., (2020) [87] developed a real-time smart surveillance system based on the YOLOv2 detection approach that can detect people and the status of excavators in hazardous areas. The results proved that the developed systems could provide immediate feedback concerning unsafe behavior and thus enable appropriate actions to be taken to prevent reoccurrence.

4.1.2. Action

Falls are highly frequent accidents in the construction industry, and occupational injuries and fatalities caused by falls from height pose a severe public problem worldwide [124]. According to prevention strategies for falling accidents in construction proposed by Huang and Hinze (2003) [3] and Chi et al., (2005) [125], fatal occupational falls on-site were closely associated with serious on-site risk factors, including poor work practices and bodily actions. Thus, it is essential to achieve and improve unsafe action

recognition to ensure the safety of construction. In the study by Guo et al., (2021) [77], action is defined as a series of gestures that form a contextual event, or more specifically, action in construction is a single activity executed by a subject, such as ladder-climbing, walking, and running. In particular, the actions' pattern and pace vary from individual to individual as well as from time to time [126]. Thus, it can be determined that different action categories can have similar postures, and one action category can have a variety of postures [127]. According to Gong et al., (2011) [127], action is classified as either action at a single moment as depicted in an image or action in a time period as shown in a sequence of images. Based on this, studies have used DL to recognize actions on construction sites from images/videos. Ding et al., (2018) [15] developed a new hybrid DL model that integrates a CNN and LSTM to automatically recognize workers' unsafe actions from videos. By extracting the visual features from videos using a CNN model and sequencing the learning features using LSTM models, the results revealed that the accuracy of the model exceeded the current state-of-the-art descriptor-based methods for the detection of safe/unsafe actions conducted by workers on-site. Likewise, an automatic computer-vision approach that utilizes an R-CNN-based model was proposed by Fang et al. (2019) [93] to detect individuals traversing structural supports from photographs during construction. By automatically identifying the presence of people and recognizing the relationship between people and concrete/steel supports, the results demonstrated that the developed model could accurately detect people traversing concrete/steel supports during construction; thus, the proposed approach could be used by site managers to automatically identify unsafe behavior and provide feedback to individual workers about their likelihood of falling from heights.

4.1.3. Interaction

In several cases, whether an action is safe depends on the status of other objects [77]. As a proof of this concept, Zhang et al., (2020) [99] proved that constant interaction and the state of random movement increase the risks of worker injury [99]. One of the accidents caused by inappropriate interactions between entities on construction sites is struck-by accidents, which led to 804 fatalities from 2011 to 2015 [37]. Therefore, to recognize unsafe behavior, current researchers not only recognize involved objects (e.g., workers, crane, and load) in terms of their identity, location, and movement direction, but more importantly, attempt to understand the interactions between these objects. Interaction is a pairwise or reciprocal action committed by two or more entities. In the concept of construction safety, entities can be defined as human (workers, managers, etc.) or objects (excavators, dump trucks, etc.). Each entity has a single action that reflects its state compared to the other entity. For example, earthmoving activities involve interactions between dump trucks and excavators.

Recognizing ongoing activities and related working groups is crucial as it allows the comprehension of jobsite context, which in turn enables the interpretation of worker intentions, their movement prediction, and the detection of inappropriate interactions that are counterproductive and may cause harmful consequences [37]. To consider the applications of DL in the interaction assessment of on-site entities, there are three different interaction types: human-to-human interaction, human-to-object interaction, and object-to-object interaction. Human-to-human interaction is an action committed by two people or groups of people (workers and managers), human-to-object interaction is an action committed directly by people to an object or multiple objects, and object-to-object interaction is an action committed by two objects or groups of objects. The interaction between construction workers and equipment is a crucial reason for on-site safety hazards [96]. Therefore, the risks posed by this interaction have received significant attention in current DL studies. For example, various studies have identified and evaluated the spatial relationship between construction workers and equipment to prevent struck-by hazards from images based on DL algorithms such as faster R-CNN [97,99,102] and YOLO [2]. Moreover, by extracting information from images, studies proposed CNN-based models for not only automatically

predicting potential safety hazards by detecting construction workers and equipment and identifying hazardous zones [96], but also tracking and analyzing spatial-temporal interactions on construction sites for real-time detection [98]. Likewise, to demonstrate that the sequence-to-sequence method could better predict trajectories and avoid error accumulation compared to conventional predictions, Cai et al. (2020) [94] and Cai et al. (2019) [37] proposed an LSTM method using construction videos that integrates both personal movement and workplace contextual information (e.g., movements of neighboring entities, workgroup information, and potential destination information). Studies have also focused on monitoring the equipment's interactions and crew relationships using DL methods. For example, based on data of historical motion from camera videos and activity attributes, Luo et al., (2021) [95] proposed an RNN framework, called GRU, for predicting future construction excavator and truck poses and monitoring when either one-to-one or group interactions of construction machines exist during earthmoving tasks. Similarly, Xiong et al., (2019) [100] developed an automated hazard identification system (AHIS) based on the CNN method to detect visual relationships between objects, including site components or crews. The results demonstrated that the proposed visual relationship detection method had the potential to enrich the semantic representation of operation facts, which could lead to better automation in construction hazard detection.

4.1.4. Activity

The information on basic actions may not be sufficient for safety analysis and schedule assessment; therefore, in recent years, researchers have attempted to recognize actions with a higher level of abstraction and complexity [77]. Guo et al., (2021) [77] showed that various on-site human activities are characterized by a complex spatial and temporal composition of objects and actions. According to the definition proposed by Turaga et al., (2008) [128], activity is a complex series of actions performed by several people who could interact with each other in a constrained manner over longer durations compared to action. Therefore, activity in construction safety can be defined as a group of actions and/or interactions that are executed to describe high-level work such as roofing, formwork, and scaffolding activities. Each action and interaction can be considered as a subactivity event in such scenarios [78]. In the context of construction safety, DL has been applied in activity recognition with different events such as scaffolding activity [103], earthmoving activity (27), and concrete pouring activity [104].

Scaffolding-related falls are an important potential threat at the job site, causing a significant number of accidents annually [129]. According to Khan et al., (2021) [103], the fatality rate due to falls from scaffolds, ladders, working platforms, and roof edges, was 60%. Therefore, the detection of unsafe activities during scaffolding activities has received attention from researchers. For example, in a study conducted by Khan et al., (2021) [103], a deep neural network, mask R-CNN, was proposed for monitoring mobile scaffold safety and detecting workers' unsafe behaviors from image dataset, including 703 training and 235 validation data with an overall accuracy of 0.86. DL was also applied to monitor other construction activities. By using the temporal and spatial CNN for recognizing basic actions during concrete pouring tasks, a hierarchical statistical method proposed by Luo et al., (2019) [104] proved the ability to recognize workers' activities with an average accuracy of 0.84. Similarly, Lin et al., (2021) [36] analyzed consecutive image sequences to automatically identify irregular operations during earthmoving work and its visualization. Therein, faster R-CNN was adapted with transfer learning to detect workers and pieces of construction equipment on the jobsite, and a hybrid model integrating CNN and LSTM was employed for action recognition. The results illustrated that the proposed framework could aid field managers in efficiently identifying potential abnormal activities, providing opportunities for further investigations and appropriate adjustments.

4.1.5. PPE and Safety Compliance

Safety rules are intended to outline safety guidelines for people and activities occurring in the workplace to ensure construction safety. Safety compliance involves following these rules in construction, adhering to safety procedures, and carrying out work safely. One of the regulations on construction sites is the use of protective equipment. Personal protective equipment, also termed as "PPE", is equipment designed to protect people against personal injury while performing tasks at the workplace. PPE includes helmets for avoiding head injuries, hand gloves for hand protection, safety glasses for eye protection, vests, boots, harnesses, and respirators [130]. A survey conducted by the US Bureau of Labor Statistics (BLS) suggested that 84% of workers who had suffered head injuries were not wearing head protection equipment [131]. Fang et al. (2018) showed that 75.1% of decedents from fall from height did not use personal fall arrest systems (PFAS) [110]. The "fatal four" (i.e., fall, struck-by object, electrocution, and caught-in/between) accounted for nearly 60% of all fatalities in construction in 2017, and the majority of these fatalities could have been prevented by wearing appropriate PPE [109]. However, there are often cases in which construction workers ignore regulations [113], and not all construction workers are aware of the importance of wearing hard hats [106]. In practice, many workers tend to take off their hard hats because of religious values [132] or discomfort due to weight and to cool off at high temperatures [106]. In addition, some frequent accidents are closely related to workers who are not certified to perform specific tasks. To support this observation, Fang et al. (2018) [112] showed that fewer accidents occur when workers are qualified and their qualifications are appropriately certified.

Previous studies have utilized DL methods to detect behaviors that do not follow construction safety rules, thereby preventing serious injuries. As discussed above, one of the most significant actions in noncompliance with construction safety regulations is the failure to wear appropriate PPE. In this regard, detecting workers with non-PPE has received considerable attention in recent studies. For example, by extracting information from images, various researchers have proposed PPE detection algorithms to identify the proper use of hard hats on human objects using DL methods such as faster R-CNN [106,111,114], YOLO [105,107,115,117,119–122], and CNN-based algorithms [109,110,113,118]. In addition, according to Wu et al. (2019) [108], the colors of hard hats can signify different roles on construction sites, providing an accessible way to improve construction safety management. Thus, in addition to detecting hard hats, researchers identified their corresponding colors that can achieve a mean average precision (mAP) of at least 0.84 [108,116]. Moreover, accidents are less likely when workers are qualified and their qualifications are properly certified [133]. Hence, DL was also applied to check whether a site worker is working within the constraints of their certification [112]. A faster R-CNN model was used to detect common objects based on the latest face detection and face recognition methods. The experimental results demonstrated the reliability and accuracy of the DL-based method to detect workers carrying out work for which they are not certified to facilitate safety inspections and monitoring.

4.2. Physical Conditions

According to the accident causation model [82], unsafe conditions and unsafe actions are considered as two direct causes of accidents. Therefore, safety performance can be improved if one can moderate people's unsafe behavior and improve their work conditions [134]. According to Li et al., (2018) [25], a hazardous working environment is a workplace with unusual hazards that violate the prevailing safety standards, thus being considered unsuitable for work [25]. In the context of construction safety, unsafe conditions can include poor lighting, temporary structure instability, unsecured equipment, etc., which can cause unfortunate accidents at construction sites. According to the extant literature [76], the common types of physical conditions identified include: (1) site condition (SC), (2) work environment (WE), and (3) site layout (SL). These conditions were also research directions of previous DL studies, and a summary of these studies is presented in Table 2.

Table 2. Construction safety studies about physical conditions.

Categories	Type of Data	Numbers of Data Training–Validation–Testing	Method	Accuracy Value	Object/Action	Accident Type	References
Work environment (WE)	Images	4000–N/A–667	CNN	Accuracy: 0.97	Guardrail	Fall	[14]
	Images	N/A	CNN	Accuracy: 0.9	Crane cracks	(General accident)	[135]
	Signal	N/A	CNN	N/A	Diaphragm wall deformation	(General accident)	[136]
	Signal	55–N/A–15	RNN	N/A	Brake pedal aperture for automatic wheel loader	(General accident)	[137]
	Images	10,000–N/A–N/A	CNN	Accuracy: 0.95 Precision: 0.76	Construction machines at nighttime	Struck-by	[138]
Site layout (SL)	Images	240 (90%–N/A–10%)	CNN	mAP: 0.99	Dense multiple construction vehicles	(General accident)	[139]
	Images	N/A	GNN	Accuracy: 0.95	Safety-rule violations of complex construction scenes	(General accident)	[65]
Site condition (SC)	Signal	600 (80%–N/A–20%)	RNN	Accuracy: 0.99	Prediction of water inflow into drill and blast tunnels	(General accident)	[140]

Note: The mAP represents mean average precision. The accident types in parentheses were judged by the authors' assessment and not specified in the paper.

4.2.1. Work Environment (WE)

The nature of the construction working environment poses both health and safety risks to workers. According to a report by the Occupational Safety and Health Administration (OSHA), approximately 40% of all construction fatalities are caused by falls from heights, followed by struck-by objects, electrocution, and caught-in/between [141]. To support this, Kolar et al. (2018) [14] showed that "fall protection, construction" was at the top of the list of the most frequently violated OSHA standards. In addition, the results from the study of Arditi et al., (2007) [142] indicated that the safety risks at nighttime could be five times higher than those in the day time due to several significant factors, including the lower illumination conditions and the fatigue of workers and machine operators. Therefore, managing, monitoring, and improving the work environment, including guarding systems, structural defects, functional defects, lighting, and noise, etc., play an important role in reducing accidents at construction sites. Passive falling prevention approaches, such as guardrails, warning lines, and fall arrest systems, often act as on-site measures for reducing the risk of falling [14].

With the development of DL, researchers have developed models for monitoring construction safety under different work environments. For example, Kolar et al. (2018) [14] developed a safety guardrail detection model based on a CNN to check whether the guardrail system is set up appropriately. The results showed that the proposed model could obtain a high accuracy of 0.97, so their model has the potential to improve construction site situations. Similarly, studies have also demonstrated that the CNN-based model can reduce the number of injuries and fatalities by detecting structural defects such as crane cracks [135] and concrete diaphragm wall (CDW) deflections [136]. In addition, by considering the poor lighting conditions that can affect the visibility of monitoring construction safety, Xiao et al., (2021) [138] proposed a vision-based method for automatically tracking construction machines at night by integrating DL illumination enhancement. The results showed that with a multiple-object tracking accuracy (MOTA) of 0.95 and a multiple-object tracking precision (MTOP) of 0.76, the proposed methodology could also be used to help accomplish automated monitoring tasks during construction at nighttime to improve safety performance.

4.2.2. Site Layout (SL)

Construction is characterized by its dynamics, such as multiple construction workers, diverse types of equipment and materials, and continuously changing working environments [19]. Quickly changing and complex workplace conditions were identified as the direct cause of more than 30% of construction accidents [143]. Therefore, proper site layout management, including arrangement, storage, and positioning of agents (e.g., construction vehicles, heavy machines, materials, etc.), is an urgent requirement to avoid hazardous issues such as site congestion and failure to properly locate utilities. However, activities involving multiple pieces of equipment and workers taking place often in a unique, complex, and dynamic environment always create challenges for monitoring proper site layout. Thus, the development of DL has proven the ability to assist in effectively managing safe layout in construction sites. For example, Wang et al. (2019) [65] used a DL-based approach for automatic safety assessment based on object relationships learned from labeled images of complex construction scenes with safety rule violations. Similarly, Guo et al., (2020) [139] proposed a CNN-based end-to-end approach for precisely detecting dense multiple construction vehicles using images from unmanned aerial vehicle (UAV). The results illustrated that the proposed method was of great significance to ensure the safety of construction sites by accurately identifying many dense vehicles with an AP of 0.99.

4.2.3. Site Condition (SC)

Site conditions, including weather, temperature, and geographical conditions, considerably affect safety during the construction process. Awolusi et al. (2018) [144] showed that both health and safety risks of workers are posed by the construction work environment. This is partly because most of the activities are performed outdoors, significantly exposing workers to weather elements [144]. In addition, Mahmoodzadeh et al., (2021) [140] proved that other natural environmental conditions, such as groundwater inflows during tunnel construction, were among the most common and challenging issues faced by constructors and designers in karst regions. The sudden and unexpected significant water inflow at the heading often damages construction machinery and leads to worker fatalities [140]. For example, a large-scale water inflow accident occurred in the Yesanguan tunnel of the Yichang–Wanzhou railway in China on 5 August, 2007 [145]. Therefore, applying DL to the prediction of the influence of natural conditions has made important contributions to safety management. For example, by proposing an LSTM-based prediction model, Mahmoodzadeh et al., (2021) [140] proved that their proposed model could predict water inflow into tunnels with higher accuracy than other ML techniques; thus, this model could ensure safety and help with scheduling during the underground construction process.

4.3. Management Issues

Safety management, a method of applying on-site safety policies, procedures, and practices convolving a construction project, is one of the most frequently used techniques to regulate construction activities and control risks [146]. Various studies related to construction safety confirmed that most accidents at construction sites could have been reduced and prevented by establishing a proper and consistent safety management process or program of planning, education/training, and inspection [147]. In general, common safety management activities in the construction industry include monitoring, controlling safety rules, planning, training, and managing the practice process to ensure safety at the construction site. According to the extant literature [79,80] and based on the context of considering DL applications on construction safety, the categories of safety management identified include (1) safety management plan, (2) accident investigation and analysis, and (3) hazard identification and risk management. Table 3 lists previous studies regarding applications of DL in handling safety management issues in the construction industry.

Table 3. Construction safety studies about management issues.

Categories	Type of Data	Numbers of Data Training–Validation–Testing	Method	Accuracy Value	Object/Action	Accident Type	References
Safety management plan	Images	10,000–N/A–1500	CNN	Accuracy: 0.95	Workers and excavators	(General accident)	[148]
	Images	19,404–4000–18,264	CNN	mAP: 0.55	Moving object detection (workers and equipment)	Struck-by	[149]
	Images	2324–26–231	CNN	Recall: 0.86 mAP: 0.83	Construction equipment	Struck-by	[150]
	Images	34,510 (66%–17%–17%)	CNN + RNN	Precision: 0.99 Recall: 1 F1-score: 0.99	Construction activity scenes	(General accident)	[151]
	Videos	4–N/A–8	CNN + RNN	Accuracy: 0.79	Recognizing people's identity	(General accident)	[152]
	Images	2094–523–654	CNN	mAP: 0.91	Construction equipment	(Struck-by)	[153]
Accident investigation and analysis	Text	95–N/A–50	RNN	F1-score: 0.84	Information extraction from accident reports	General accident	[38]
	Text	2624–N/A–657	GNN	Accuracy: 0.87 Precision: 0.51 Recall: 0.54	Text classification of near-misses safety reports	General accident	[39]
	Text	3000	CNN	Precision: 0.8 Recall: 0.68 F1-score: 0.71	Hazard record analysis	General accident	[154]
	Text	90,000 (90%–N/A–10%)	RNN	F1-score: 0.87	Automatically learning injury precursors	General accident	[16]
	Text	2000	CNN	Precision: 0.65 Recall: 0.61	Classifying and visualizing accident narratives	General accident	[18]
	Images	2000–N/A–N/A	CNN	Precision: 0.89 Recall: 0.93	A gate scenario and an earthmoving scenario	General accident	[155]
Hazard identification and risk management	Images	40,000	CNN	N/A	Identifying hazards	General accident	[156]
	Images	6000–N/A–1000	CNN	Accuracy: 0.93	Worker localization and hazard detection	General accident	[157]

Note: The mAP represents mean average precision. The accident types in parentheses were judged by the authors' assessment and not specified in the paper.

4.3.1. Safety Management Plan

With the presence of cost and time pressures and the frequent need to perform unplanned work (e.g., rework), people tend to take risks to make their work more efficient [158–160]. The upshot of this case is that people tend to commit unsafe actions, especially when they know they are not being supervised [152]. Therefore, safety management plans regarding publishing safety policies, objectives, and requirements; proposing plans; making decisions; and monitoring safety play an important role. The purpose of health and safety monitoring is to ensure effective measurement and management of construction workers' safety practices against existing safety plans and standards [19]. Visual

information related to construction activity scenes is becoming increasingly important for construction management [161,162]. The scene of construction activity in images can be defined as an integral overview of the activity in pictures that synchronously contain objects (e.g., workers, equipment, and materials), their interrelationships (e.g., cooperation between objects or coexistence of objects), and other vital scenario elements (e.g., earthmoving and concrete pouring) [151]. Thus, with the development of DL, automatically manifested construction activity scenes [151] provide managers with information for making decisions and safety management plans [148].

Recent research has focused on providing site managers the status of construction sites by detecting construction objects to assist in planning safety management at construction sites. For example, various studies proposed DL models such as faster R-CNN [148,149], YOLO [150], LSTM [151], and the CNN-based method [153] to provide supervisors with more insight into the real-time status of large-scale construction jobsites so they could assist supervisors in inspecting construction safety and processes [148,150,153]. In addition, as discussed above, workers sometimes have the proclivity to commit unsafe actions, especially when they know they are not being supervised [152], so it is important to provide direct feedback to people committing unsafe actions so that they can modify their future behavior. In a notable study by Wei et al., (2019) [152], a novel DL approach was developed to automatically determine a person's identity, which can be utilized by site managers to automatically recognize individuals engaging in unsafe behavior; therefore, it can be used to provide immediate feedback about their actions and possible consequences.

4.3.2. Accident Investigation and Analysis

Accidents and incidents should be analyzed for better implementation and continuous improvement of safety management systems [79]. Collecting and organizing accident reports, regulations, and laws, and then presenting them publicly, are considered good practices for improving the safety management of construction sites [38]. Safety reports are an extremely valuable information source that can be used by site managers to learn about the conditions and events contributing to the occurrence of accidents [158,163]. Therefore, it enhances managers' safety awareness and urges them to prevent accidents or related construction work issues [38]. Nowadays, using DL, accident documents are processed to provide useful information for safety management under two main tasks: information extraction and text classification. Information extraction is the task of finding structured information from unstructured or semistructured text [164], which is essential for handling continuously growing data published on the online, especially in the Big Data era [165]. For example, Feng and Chen (2021) [38] adopted the BiLSTM-CRF model to automatically extract information from accident reports, so this model could help to raise workers' security awareness and prevent hazards and accidents. Similarly, Baker et al., (2020) [16] compared two state-of-the-art DL architectures, CNN and hierarchical attention networks (HAN) based on GRU, to automatically learn injury precursors from raw construction accident reports. The results illustrated that HAN outperformed CNN almost everywhere with a mean performance of 0.87; thus, the HAN model can extract useful information, which not only allows the exploration of empirical relationships for postanalysis and project statistics, but can also be used proactively during typical work planning, job risk analyses, prejob meetings, and audits. Another application of DL is text classification, which is a fundamental task in the natural language processing area where one needs to assign one or multiple predefined labels to a text sequence [166]. For example, previous studies proposed DL-based models to classify and analyze the narrative surrounding accidents and to better understand their causal nature from accident reports [18,39,154]. In addition, Xiao et al., (2021) [155] proposed a DL-based method for the collection and automatic generation of video highlights from construction videos. The proposed CNN-based approach was validated through two case studies: a gate scenario and an earthmoving scenario. With a score of 0.89 for precision and 0.93 for recall, the proposed model proved that it could offer potential benefits to construction management in

terms of significant reduction in video storage space and efficient indexing of construction video footage, which was beneficial for project management tasks such as safety control.

4.3.3. Hazard Identification and Risk Management

Dynamic and complex construction environments have caused significant risks during construction. Unfortunately, studies across the world have reported that a substantial portion (approximately 50%) of hazards remain unrecognized [167–169]. These unrecognized hazards expose construction workers to unanticipated risks and potential injuries [168]. Therefore, identifying hazards and managing risks play an important role in construction safety management. DL has been used to identify risks with notable achievements. For example, Fang et al., (2020) [156] integrated computer vision algorithms with ontology models to develop a knowledge graph that can automatically and accurately recognize hazards while complying with safety regulations, even when they are subjected to change. Therein, mask R-CNN was adopted in their research for entity detection. The results showed that the proposed approach could successfully detect falls from height (FFH) hazards in varying contexts from images. Similarly, a mask RCNN-based framework was developed by Jeelani et al., (2021) [157] for an automated system that detects hazardous conditions and objects in real-time with over 93% accuracy; therefore, this model can assist workers and safety managers in identifying risks in complex and dynamic construction environments.

5. Overall Research Trends in Safety Management: Summary of Contributions and Limitations

In this study, three safety management factors, including behaviors, physical conditions, and management issues, were identified in the context of applying DL models to construction safety. This section provides an overview of the research trends from technical and managerial aspects (e.g., data types, algorithms, and safety issues) (Figures 15 and 16). Table 4 shows the accuracy of the studies using DL for construction safety. Overall, a CNN is the most commonly used method applied in these studies from the major data of the images, and unsafe behaviors is the main research direction with high performance, gaining a variety of contributions to safety management.
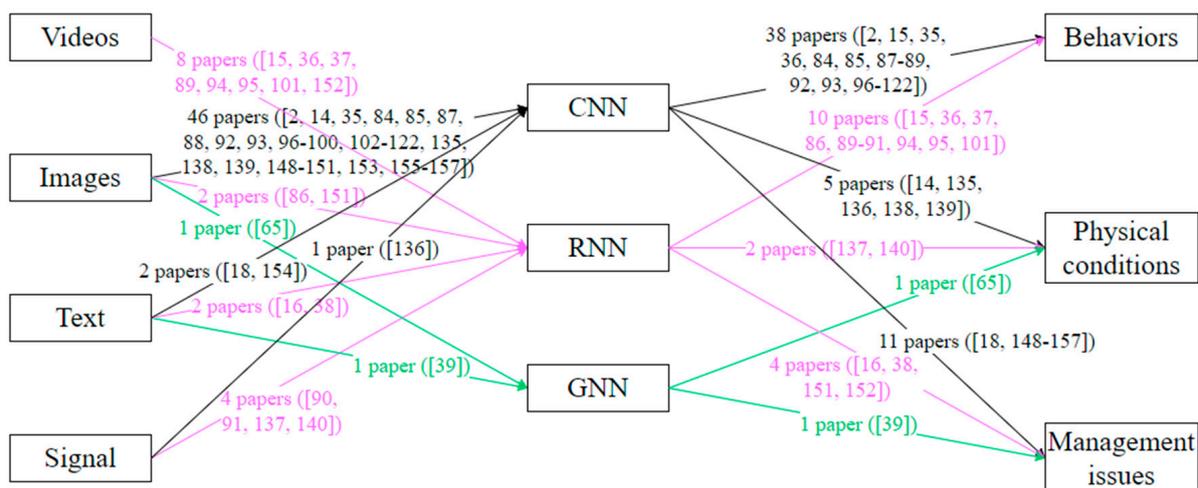

Figure 15. The link between types of data, methods, and safety factors.
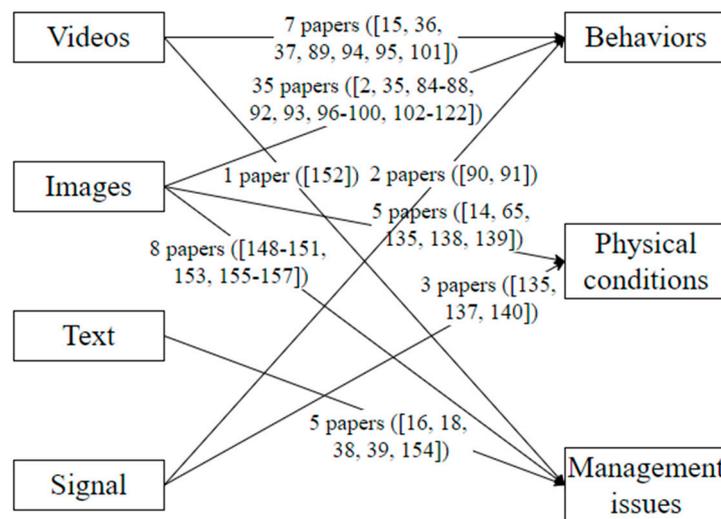
Figure 16. The link between types of data and safety factors.

Table 4. Accuracy of DL studies.

Safety Factors	Methods	Accuracy Value	(Min–Max) Value	Mean Value	References
Behaviors	CNN	Accuracy	0.83–0.96	0.91	[2,15,84,85,87,88,92,101,103,104,109,118,148]
		Precision	0.52–0.99	0.88	[2,35,85,93,96–99,102,106,107,110–112,114,118]
		Recall	0.45–0.98	0.84	[2,35,85,93,97,99,100,102,106,107,110–112,114,118]
		F1-score	0.48–0.98	0.76	[2,35,89]
		mAP	0.55–0.96	0.81	[36,105,108,113,115–117,119–122]
	RNN	Accuracy	0.90–0.99	0.95	[15,37,86,90,91,95]
		F1-score	0.83–0.99	0.91	[89,90]
		mAP		0.73	[36]
Physical conditions	CNN	Accuracy	0.9–0.97	0.94	[14,135,138]
		Precision		0.76	[138]
		mAP		0.99	[139]
	RNN	Accuracy		0.99	[140]
	GNN	Accuracy		0.95	[65]
Management issues	CNN	Accuracy	0.79–0.95	0.89	[148,152,157]
		Precision	0.65–0.99	0.83	[18,151,154,155]
		Recall	0.61–1.0	0.82	[18,150,151,154,155]
		F1-score	0.71–0.99	0.85	[151,154]
		mAP	0.55–0.91	0.76	[149,150,153]
	RNN	Accuracy		0.79	[152]
		Precision		0.99	[151]
		Recall		1	[151]
		F1-score	0.84–0.99	0.90	[16,38,151]
	GNN	Accuracy		0.87	[39]
		Precision		0.51	[39]
		Recall		0.54	[39]

mAP: mean average precision.

5.1. Recognition of Unsafe Behavior

The advancement of DL has opened up significant opportunities for examining unsafe behaviors in construction. Among the five categories of behaviors that DL has focused on, construction workers (21 of 44 papers) and PPE (17 of 44 papers) are the main objects of interest. For objects, different algorithms (e.g., object detection algorithms [35], object tracking [36], and activity recognition [37]) have demonstrated good performance in detecting and tracking workers. For example, by using DL-based object detection architectures, previous studies detected workers and PPE successfully with an accuracy exceeding 0.90 [85,87,107,111,118]. In addition, recognizing equipment operations (e.g., dump trucks and excavators) has also attracted much attention from researchers for mainly examining the interaction between entities. For example, researchers proposed DL-based models to monitor and analyze the interaction between workers and equipment with an accuracy range of 0.65 to 1.00 [2,97–99,102].

As various DL methods that use a CNN, RNN, and GNN have been applied, different formats of data (e.g., videos, images, and signals) have been used to detect those representing unsafe behaviors in the data. In particular, detection and tracking of unsafe behaviors were performed mainly using videos and images (85%). The reason for this phenomenon is partly because collecting videos and images at construction sites is easier and more common than other types of data (e.g., signals). According to Daniel and Chen (2003) [170], along with digital camcorders, video conferencing, digitized movies, and video emails that are making their way into everyday life, it is almost certain that the use of video data will multiply by multiple times in the coming years. Moreover, nowadays, there are various publicly available data sources such as Microsoft's Common Objects in Context (MS COCO) [171], ImageNet [172], Pascal VOC [173], etc., which researchers can easily access.

From an algorithmic perspective, recent neural networks, especially CNNs, have achieved considerable success in various areas, including image/video understanding, processing, compression, etc. [174]. The trained CNN can be used to handle classification, recognition, and prediction tasks on test data with highly efficient adaptability [174]. Therefore, CNN was dominantly applied in detecting unsafe behaviors using image data sources (34 of 44 papers). For videos and other sequence data such as signals (e.g., time-series data), RNNs, designed for sequence learning [175], were also used with high performance. For example, various studies have utilized RNN models to detect unsafe behaviors from videos with an accuracy exceeding 0.9 [15,37,95].

5.2. Physical Condition Identification

Previous research on unsafe physical conditions have focused mainly on structural defects and site layout status at construction sites. The main objects of interest in such research include structures (e.g., guardrails and diaphragm walls) and equipment (e.g., cranes, wheel loaders, and construction vehicles). For example, various studies proposed CNN-based methods to detect structural defects such as guardrail defects [14], crane cracks [135], and diaphragm wall deformations [136] from images and signals with an accuracy of up to 0.97. In addition, to evaluate whether the site layout is appropriate, entities in the construction sites need to be detected precisely. Therefore, in a construction environment involving a wide range of heavy equipment (e.g., tower cranes, dump trucks, and excavators), recognizing equipment operations has also attracted much attention from researchers (50% of the total number of papers regarding physical conditions), and these DL studies can gain accuracy of over 0.9 [135,137–139].

For detecting unsafe physical conditions, image was the most used type of data in DL models (62.5% of total papers). By applying the image classification task, the status of physical conditions regarding guardrails [14], the surface of crane cracks [135], and dense multiple construction vehicles [139] were detected and located to ensure safety at construction sites. Moreover, because of the growing interest in CNNs, the most common tool used for image analysis and image classification [34], they have been applied the most

in handling issues related to physical conditions. In addition, to predict other physical conditions such as diaphragm wall deformation [136] and water flow [140] during the underground construction process, time-series data was used to describe properties related to deformation and inflow over time. For this application, an RNN is commonly used to deal with such issues with a high accuracy, reaching 0.99 [136,140].

5.3. Safety Management

DL has been used effectively to support construction safety management. By using image datasets, the CNN method was utilized the most (8 of 14 papers) to provide managers with real-time statuses of large-scale construction jobsites, so this can assist in improving their decision-making regarding safety and planning [148–151]. These studies mainly focused on workers (six of eight papers) and equipment (e.g., pump trucks, excavators, rollers, and tower cranes) (seven of eight papers). For example, various studies applied CNN models to detect workers and construction equipment from images with an accuracy range of 0.55–1.0. In addition, to minimize safety risks in construction, data are recorded in various formats (e.g., video, photographs, and safety reports), which researchers have used to monitor safety [134]. Thus, various studies have used videos and images (64%) and accident reports (36%) to aid the investigation and analysis of risks at construction sites. For example, previous studies applied DL models for NLP tasks (e.g., text classification and information extraction) with high accuracy, ranging from 0.54 to 0.87 [16,18,38,39,154].

Besides recognizing individuals committing unsafe actions from images, identifying the person's identity also plays an important role in supporting safety management. Once a person's identity can be determined, site managers can provide specific feedback regarding their unsafe behaviors [152]. However, very little research has focused on this issue (one of 66 papers). In a notable study conducted by Wei et al. (2019) [152], a DL model was applied to determine a person's identity by computing the c between the identity feature with previously saved features of other people's identities; however, this study reported practical limitations such as the limited number of activities (e.g., people walking), and the possibility of delay in recognizing a person's identity in real-time because of the computation requirements placed on the attention network to extract representations from videos.

5.4. The Summary of Contributions and Limitations of Deep Learning on Safety Management

This study reviewed the contributions and limitations specified in previous papers and reports the key contributions with limitations, as summarized and outlined in Table 5. In terms of contributions, by detecting unsafe physical conditions, construction workers and equipment, as well as their behaviors, the multiple contributions of DL models include monitoring safety and proactively preventing hazards, evaluating proactive safety risk levels, strategizing effective training solutions, designing effective hazard recognition and management practices, and applying operator assistance systems in construction machinery to achieve active safety. The investigation and analysis of safety reports can not only be used proactively during typical work planning, job hazard analyses, prejob meetings, and audits, but also raise the safety awareness of workers and professionals. However, applying DL in construction safety still has challenges such as the limitation of the dataset, the influence of performance due to the presence of occlusions, blurriness, and background patches, and the lack of consideration of an individual's identity during action recognition.

Table 5. The summary of contributions and limitations of DL in safety management.

Purposes	Contributions	Limitations	References
Detecting workers and equipment, estimating, recognizing, and analyzing their behaviors.	DL models can support monitoring safety and proactively preventing hazards by sending early warning information combined with the on-site alarm equipment to the management staff so they can provide instant feedback concerning unsafe behavior, and appropriate actions can be put in place to prevent reoccurrence.	The training dataset was limited.	[2,15,35,84,91,93,94,96,98,99,103, 111,148,149,151,152,156,157]
		The accuracy of the method is affected by the presence of occlusions, confusion with background patches, poor illumination, and blurriness.	[84,87,97,105,107,110,148,155,156]
		They do not associate any personal identification with the output for verification.	[105,106,118,156,157]
		Not mentioned.	[37,86,95,101,102,108,109,112–117,119–122,150,153]
	Proactive and automatic safety risk level can be analyzed and evaluated for making decisions on risk management.	The dataset was limited.	[36,89,94,99]
		Cases of the on-site experiment failed due to visual obstacles.	[92]
		Not mentioned.	[85,86,88,90,104]
	DL models support strategizing effective training solutions and designing effective hazard recognition and management practices.	The dataset was limited.	[93,99,100]
		The accuracy of the method is affected by the presence of occlusions.	[93]
		Individual workers need to be identified.	[157]
	The proposed method can be applied to operator assistance systems in construction machinery to achieve active safety.	The dataset was limited.	[2]
Detecting unsafe physical conditions.	DL models can support monitoring safety and early warning, so managers can provide the appropriate solutions to prevent or control risks.	The dataset was limited.	[14,140]
		Occlusion was not addressed.	[14]
		Not mentioned.	[65,135–139]
	Before the predicted deformation reaches the threshold limit, control strategies can be implemented to avoid excessive deformation and the corresponding risks to the engineering project and surrounding environment.	Not mentioned.	[136]
Investigating and analyzing safety reports.	The results can be used proactively during typical work planning, job hazard analyses, prejob meetings, and audits.	Not mentioned	[16,39]
	DL models raise the security awareness of workers and professionals to better understand and prevent hazards and accidents, and aid in educating workers about "what not to do" and "what to do".	The dataset was limited.	[18,38,154]
		Not mentioned.	[16,39]

6. Future Research Directions

Despite recent technical advances in DL, there are still challenges in its practical applications. Based on the limitations identified and summarized, directions for future research are discussed to resolve these issues and further expand their applications. These directions include (1) expanding a comprehensive dataset, (2) improving technical restrictions due to occlusions, and (3) identifying individual who performed unsafe behaviors.

6.1. Expanding a Comprehensive Dataset

In a dynamic and complex construction environment involving many human resources, diverse types of equipment, as well as many types of actions of humans and equipment, larger and more comprehensive datasets are important for improving the performance of DL models. According to Ding et al., (2018) [15], some worker actions could

not be recognized due to the training sample size and the limited number of unsafe actions considered. Therefore, with larger datasets, the model may further improve and provide more accurate results. However, there is currently no comprehensive and common dataset publicly available, not only for specific tasks such as object detection, pose detection, and activity recognition but also for a variety of construction sites, different viewpoints, lighting, and occlusion conditions. Although several studies such as Xuehui et al., (2021) [149] presented the Moving Objects in Construction Sites (MOCS) image dataset for detecting objects at construction sites, its use may be limited by the size and type of the dataset. Therefore, further research is needed to generate and share a comprehensive dataset for the research community. Potential solutions may include generating publicly available datasets by developing a DL-based methodology to automatically create safety reports in natural language based on construction site imagery and using models to collect and amalgamate reports across the industry through continuous updates as new data arrive. In a study on DL in generating radiology reports, Monshi et al., (2020) [176] reported that CNNs used for image analysis could be integrated alongside RNNs for NLP and natural language generation (NLG), generating radiology coherent paragraphs in the medical field. Thus, once this DL application is applied in construction, creating automatic safety reports based on construction site images increases the number of datasets. A platform then needs to be built for public access so that researchers can easily share and upload datasets.

6.2. Improving Technical Restrictions Due to Occlusions

In dynamic and continuously changing construction environments, as images and videos data are mostly used, DL models have faced challenges such as occlusion [84], poor illumination and blurriness [105], and background clutter [97]. For example, Fang et al., (2019) [93] reported that the DL model could not detect all people traversing structural supports due to the presence of occlusions. However, previous studies often ignored occlusions by assuming no occlusion (e.g., the guardrail is always visible for detection in [14]). To handle these issues, potential solutions may include the following. First, a method is needed to search and identify the optimum placement of cameras (e.g., position and distance of a camera, the effect of occlusion, and lighting conditions) where full or maximum coverage of resources (e.g., workers, materials, and machines) can be achieved. Second, to handle the self-occlusion of projected objects in a 2D vision, reconstructing the 3D bounding boxes of these objects can be conducted using DL models to estimate depth and reconstruct depth scenes as a global 3D model from monocular images. Finally, another method for coping with occlusions is to combine vision-based approaches with sensor-based methods (e.g., the global positioning system), which can provide the location and motion of objects.

6.3. Identifying Individuals Who Performed Unsafe Behaviors

Providing feedback to individuals regarding the likelihood of their unsafe actions can result in immediate behavior modification and targeted safety training [93]. Therefore, in addition to identifying unsafe actions at construction sites, it is necessary to identify who performed these unsafe actions. Based on this, site managers can automatically identify unsafe behavior in real time and provide feedback to individuals about their unsafe behaviors. However, previous studies have not focused on the identification of workers (e.g., [156,157]). To achieve this goal, several solutions can be used in the future. First, sensors can identify a person's identity and location [177]. Thus, future research can combine the results of an individual's identity from sensors and action monitoring of the DL model to identify those who do not perform unsafe actions. Second, this issue can be addressed by developing a DL approach to identify individuals from videos by integrating temporal and spatial information. Wei et al., (2019) [152] provided an example of this approach. This DL approach focuses on using the spatial attention network for extracting spatial feature maps, temporal attention networks for extracting temporal information, and computing the distance between features to recognize a person's identity. In addition, a

person's identity can also be recognized by face recognition models based on a CNN [178], so future research can combine face recognition and action recognition to identity workers performing unsafe behaviors.

7. Conclusions

This study synthesized and reviewed the current DL studies applied to safety management in the construction industry. It was found that DL studies had paid attention to three main research directions, including behaviors, physical conditions, and management issues. By providing detailed summaries of DL applications in each category, this paper aims to support researchers and managers in the field of construction safety with a specific overview regarding what type of method has achieved highly accurate results, along with the type and amount of data that has been used for a certain safety task, as well as the actions managers can take from the result of DL models for improving safety management. In general, detecting unsafe behaviors was the main research direction of previous studies (67%) with high performance, which has contributed to safety management in the construction industry. Moreover, the results indicated that CNN modeling was the most common method used in these studies (75%) and achieved high accuracy, which could reach up to ~1.0, from the primary data of images (73%). In addition to providing the overall trends of DL applications, this literature also presents limitations and future directions for applying DL in construction safety. In a dynamic and complex construction environment involving many human and equipment resources, expanding larger and more comprehensive datasets is important for improving the DL model performance. In addition, the presence of occlusions causing challenges for DL studies using image and video data should be addressed in future studies. Another direction is to identify individuals who performed unsafe behaviors for immediate behavior modification and targeted safety training. DL is an emerging area of construction safety and is still developing, so outlining key challenges and corresponding proposal research can aid in developing DL applications in the future. We expect that this paper will provide not only new lines of advanced methods for researchers working on safety management but also opportunities to apply DL in practice.

References

1. Rostami, A.; Sommerville, J.; Wong, L.; Lee, C. Risk management implementation in small and medium enterprises in the UK construction industry. *Eng. Constr. Archit. Manag.* **2015**, *22*, 91–107. [CrossRef]
2. Son, H.; Kim, C. Integrated worker detection and tracking for the safe operation of construction machinery. *Autom. Constr.* **2021**, *126*, 103670. [CrossRef]
3. Huang, X.; Hinze, J. Analysis of construction worker fall accidents. *J. Constr. Eng. Manag.* **2003**, *129*, 262–271. [CrossRef]
4. Sousa, V.; Almeida, N.M.; Dias, L.A. Risk-based management of occupational safety and health in the construction industry–Part 1: Background knowledge. *Saf. Sci.* **2014**, *66*, 75–86. [CrossRef]
5. 2003–2014 Census of Fatal Occupational Injuries. Available online: http://www.bls.gov/ (accessed on 15 March 2015).
6. Agwu, M.O.; Olele, H.E. Fatalities in the Nigerian construction industry: A case of poor safety culture. *J. Econ. Manag. Trade* **2014**, *4*, 431–452. [CrossRef]

7.   Fang, D.; Huang, Y.; Guo, H.; Lim, H.W. LCB approach for construction safety. *Saf. Sci.* **2020**, *128*, 104761. [CrossRef]

8.   Sarkar, S.; Maiti, J. Machine learning in occupational accident analysis: A review using science mapping approach with citation network analysis. *Saf. Sci.* **2020**, *131*, 104900. [CrossRef]

9.   Xu, Z.; Saleh, J.H. Machine learning for reliability engineering and safety applications: Review of current status and future opportunities. *Reliab. Eng. Syst. Saf.* **2021**, 107530. [CrossRef]

10.   Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.

11.   Chen, H.; Chen, A.; Xu, L.; Xie, H.; Qiao, H.; Lin, Q.; Cai, K. A deep learning CNN architecture applied in smart near-infrared analysis of water pollution for agricultural irrigation resources. *Agric. Water Manag.* **2020**, *240*, 106303. [CrossRef]

12.   Akinosho, T.D.; Oyedele, L.O.; Bilal, M.; Ajayi, A.O.; Delgado, M.D.; Akinade, O.O.; Ahmed, A.A. Deep learning in the construction industry: A review of present status and future innovations. *J. Build. Eng.* **2020**, *32*, 101827. [CrossRef]

13.   Khallaf, R.; Khallaf, M. Classification and analysis of deep learning applications in construction: A systematic literature review. *Autom. Constr.* **2021**, *129*, 103760. [CrossRef]

14.   Kolar, Z.; Chen, H.; Luo, X. Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images. *Autom. Constr.* **2018**, *89*, 58–70. [CrossRef]

15.   Ding, L.; Fang, W.; Luo, H.; Love, P.E.; Zhong, B.; Ouyang, X. A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Autom. Constr.* **2018**, *86*, 118–124. [CrossRef]

16.   Baker, H.; Hallowell, M.R.; Tixier, A.J.-P. Automatically learning construction injury precursors from text. *Autom. Constr.* **2020**, *118*, 103145. [CrossRef]

17.   Arya, K.M.; Ajith, K.K. A Review on Deep Learning Based Helmet Detection. In Proceedings of the International Conference on Systems, Energy & Environment (ICSEE) 2021, Kannur, India, 22–23 January 2021; Government College of Engineering Kannu: Kerala, India, 2021.

18.   Zhong, B.; Pan, X.; Love, P.E.; Ding, L.; Fang, W. Deep learning and network analysis: Classifying and visualizing accident narratives in construction. *Autom. Constr.* **2020**, *113*, 103089. [CrossRef]

19.   Seo, J.; Han, S.; Lee, S.; Kim, H. Computer vision techniques for construction safety and health monitoring. *Adv. Eng. Inform.* **2015**, *29*, 239–251. [CrossRef]

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

21.   Hou, L.; Chen, H.; Zhang, G.K.; Wang, X. Deep Learning-Based Applications for Safety Management in the AEC Industry: A Review. *Appl. Sci.* **2021**, *11*, 821. [CrossRef]

22.   Mok, K.Y.; Shen, G.Q.; Yang, J. Stakeholder management studies in mega construction projects: A review and future directions. *Int. J. Proj. Manag.* **2015**, *33*, 446–457. [CrossRef]

23.   Content Analysis. Available online: https://www.publichealth.columbia.edu/research/population-health-methods/content-analysis (accessed on 30 September 2021).

24.   Zhang, M.; Shi, R.; Yang, Z. A critical review of vision-based occupational health and safety monitoring of construction site workers. *Saf. Sci.* **2020**, *126*, 104658. [CrossRef]

25.   Li, X.; Yi, W.; Chi, H.L.; Wang, X.; Chan, A.P. A critical review of virtual and augmented reality (VR/AR) applications in construction safety. *Autom. Constr.* **2018**, *86*, 150–162. [CrossRef]

26.   Liang, X.; Shen, G.Q.; Bu, S. Multiagent systems in construction: A ten-year review. *J. Comput. Civ. Eng.* **2016**, *30*, 04016016. [CrossRef]

27.   Zhong, B.; Li, H.; Luo, H.; Zhou, J.; Fang, W.; Xing, X. Ontology-based semantic modeling of knowledge in construction: Classification and identification of hazards implied in images. *J. Constr. Eng. Manag.* **2020**, *146*, 04020013. [CrossRef]

28.   Zhong, B.; Xing, X.; Love, P.; Wang, X.; Luo, H. Convolutional neural network: Deep learning-based classification of building quality problems. *Adv. Eng. Inform.* **2019**, *40*, 46–57. [CrossRef]

29.   Deng, L.; Yu, D. Deep learning: Methods and applications. *Found. Trends Signal Process.* **2014**, *7*, 197–387. [CrossRef]

30.   Shamshirband, S.; Fathi, M.; Dehzangi, A.; Chronopoulos, A.T.; Alinejad-Rokny, H. A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. *J. Biomed. Inform.* **2020**, *113*, 103627. [CrossRef] [PubMed]

31.   Guo, Y.; Liu, Y.; Oerlemans, A.; Lao, S.; Wu, S.; Lew, M.S. Deep learning for visual understanding: A review. *Neurocomputing* **2016**, *187*, 27–48. [CrossRef]

32.   Hao, X.; Zhang, G.; Ma, S. Deep learning. *Int. J. Semant. Comput.* **2016**, *10*, 417–439. [CrossRef]

33.   Vateekul, P.; Koomsubha, T. A study of sentiment analysis using deep learning techniques on Thai Twitter data. In Proceedings of the 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), Khon Kaen, Thailand, 13–15 July 2016; pp. 1–6.

34.   Mikołajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the 2018 International Interdisciplinary PhD Workshop (IIPhDW), Swinoujscie, Poland, 9–12 May 2018; pp. 117–122.

35.   Kim, K.; Kim, S.; Shchur, D. A UAS-based work zone safety monitoring system by integrating internal traffic control plan (ITCP) and automated object detection in game engine environment. *Autom. Constr.* **2021**, *128*, 103736. [CrossRef]

36.   Lin, Z.-H.; Chen, A.Y.; Hsieh, S.-H. Temporal image analytics for abnormal construction activity identification. *Autom. Constr.* **2021**, *124*, 103572. [CrossRef]

37. Cai, J.; Zhang, Y.; Cai, H. Two-step long short-term memory method for identifying construction activities through positional and attentional cues. *Autom. Constr.* **2019**, *106*, 102886. [CrossRef]

38. Feng, D.; Chen, H. A small samples training framework for deep Learning-based automatic information extraction: Case study of construction accident news reports analysis. *Adv. Eng. Inform.* **2021**, *47*, 101256. [CrossRef]

39. Fang, W.; Luo, H.; Xu, S.; Love, P.E.; Lu, Z.; Ye, C. Automated text classification of near-misses from safety reports: An improved deep learning approach. *Adv. Eng. Inform.* **2020**, *44*, 101060. [CrossRef]

40. Graves, A.; Mohamed, A.-R.; Hinton, G. Speech recognition with deep recurrent neural networks. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 6645–6649.

41. Faust, O.; Hagiwara, Y.; Hong, T.J.; Lih, O.S.; Acharya, U.R. Deep learning for healthcare applications based on physiological signals: A review. *Comput. Methods Programs Biomed.* **2018**, *161*, 1–13. [CrossRef] [PubMed]

42. Fukushima, K.; Miyake, S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and Cooperation in Neural Nets*; Springer: Berlin/Heidelberg, Germany, 1982; pp. 267–285.

43. LeCun, Y.; Bengio, Y. Convolutional networks for images, speech, and time series. *Handb. Brain Theory Neural Netw.* **1995**, *3361*, 1995.

44. Lawrence, S.; Giles, C.L.; Tsoi, A.C.; Back, A.D. Face recognition: A convolutional neural-network approach. *IEEE Trans. Neural Netw.* **1997**, *8*, 98–113. [CrossRef] [PubMed]

45. Chen, S.; Xu, H.; Liu, D.; Hu, B.; Wang, H. A vision of IoT: Applications, challenges, and opportunities with china perspective. *IEEE Internet Things J.* **2014**, *1*, 349–359. [CrossRef]

46. Han, J.; Zhang, D.; Cheng, G.; Liu, N.; Xu, D. Advanced deep-learning techniques for salient and category-specific object detection: A survey. *IEEE Signal Process. Mag.* **2018**, *35*, 84–100. [CrossRef]

47. Wang, D.; Guo, Q.; Song, Y.; Gao, S.; Li, Y. Application of multiscale learning neural network based on CNN in bearing fault diagnosis. *J. Signal Process. Syst.* **2019**, *91*, 1205–1217. [CrossRef]

48. Mu, R.; Zeng, X. A review of deep learning research. *KSII Trans. Internet Inf. Syst. (TIIS)* **2019**, *13*, 1738–1764.

49. Lu, J.; Tan, L.; Jiang, H. Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification. *Agriculture* **2021**, *11*, 707. [CrossRef]

50. Koirala, A.; Walsh, K.B.; Wang, Z.; McCarthy, C. Deep learning–Method overview and review of use for fruit detection and yield estimation. *Comput. Electron. Agric.* **2019**, *162*, 219–234. [CrossRef]

51. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.

52. Adarsh, P.; Rathi, P.; Kumar, M. YOLO v3-Tiny: Object Detection and Recognition using one stage improved model. In Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 6–7 March 2020; pp. 687–694.

53. Liu, B.; Zhao, W.; Sun, Q. Study of object detection based on Faster R-CNN. In Proceedings of the 2017 Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; pp. 6233–6236.

54. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.

55. Uijlings, J.R.; Van De Sande, K.E.; Gevers, T.; Smeulders, A.W. Selective search for object recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171. [CrossRef]

56. Salvador, A.; Giró-i-Nieto, X.; Marqués, F.; Satoh, S.i. Faster r-cnn features for instance search. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 9–16.

57. Huang, R.; Pedoeem, J.; Chen, C. YOLO-LITE: A real-time object detection algorithm optimized for non-GPU computers. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 2503–2510.

58. Chen, C.; Qin, C.; Qiu, H.; Tarroni, G.; Duan, J.; Bai, W.; Rueckert, D. Deep learning for cardiac image segmentation: A review. *Front. Cardiovasc. Med.* **2020**, *7*, 25. [CrossRef] [PubMed]

59. Abdulkadir, S.J.; Alhussian, H.; Nazmi, M.; Elsheikh, A.A. Long short term memory recurrent network for standard and poor's 500 index modelling. *Int. J. Eng. Technol* **2018**, *7*, 25–29. [CrossRef]

60. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

61. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**, arXiv:1412.3555.

62. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **2019**, *31*, 1235–1270. [CrossRef] [PubMed]

63. Kuan, L.; Yan, Z.; Xin, W.; Yan, C.; Xiangkun, P.; Wenxue, S.; Zhe, J.; Yong, Z.; Nan, X.; Xin, Z. Short-term electricity load forecasting method based on multilayered self-normalizing GRU network. In Proceedings of the 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing, China, 26–28 November 2017; pp. 1–5.

64. Althelaya, K.A.; El-Alfy, E.-S.M.; Mohammed, S. Stock market forecast using multivariate analysis with bidirectional and stacked (LSTM, GRU). In Proceedings of the 2018 21st Saudi Computer Society National Computer Conference (NCC), Riyadh, Saudi Arabia, 25–26 April 2018; pp. 1–7.

65. Wang, Y.; Liao, P.-C.; Zhang, C.; Ren, Y.; Sun, X.; Tang, P. Crowdsourced reliable labeling of safety-rule violations on images of complex construction scenes for advanced vision-based workplace safety. *Adv. Eng. Inform.* **2019**, *42*, 101001. [CrossRef]
66. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
67. Acheampong, F.A.; Nunoo-Mensah, H.; Chen, W. Transformer models for text-based emotion detection: A review of BERT-based approaches. *Artif. Intell. Rev.* **2021**, *54*, 5789–5829. [CrossRef]
68. Alatawi, H.S.; Alhothali, A.M.; Moria, K.M. Detecting white supremacist hate speech using domain specific word embedding with deep learning and BERT. *IEEE Access* **2021**, *9*, 106363–106374. [CrossRef]
69. Tang, M.; Gandhi, P.; Kabir, M.A.; Zou, C.; Blakey, J.; Luo, X. Progress notes classification and keyword extraction using attention-based deep learning models with BERT. *arXiv* **2019**, arXiv:1910.05786.
70. Tenney, I.; Das, D.; Pavlick, E. BERT rediscovers the classical NLP pipeline. *arXiv* **2019**, arXiv:1905.05950.
71. Reason, J. Human error: Models and management. *BMJ* **2000**, *320*, 768–770. [CrossRef] [PubMed]
72. Guo, B.H.; Yiu, T.W. Developing leading indicators to monitor the safety conditions of construction projects. *J. Manag. Eng.* **2016**, *32*, 04015016. [CrossRef]
73. Zhang, L.; Wu, X.; Skibniewski, M.J.; Zhong, J.; Lu, Y. Bayesian-network-based safety risk analysis in construction projects. *Reliab. Eng. Syst. Saf.* **2014**, *131*, 29–39. [CrossRef]
74. Guo, H.; Yu, Y.; Skitmore, M. Visualization technology-based construction safety management: A review. *Autom. Constr.* **2017**, *73*, 135–144. [CrossRef]
75. Carter, G.; Smith, S.D. Safety hazard identification on construction projects. *J. Constr. Eng. Manag.* **2006**, *132*, 197–205. [CrossRef]
76. Zou, P.X.; Sunindijo, R.Y. *Strategic Safety Management in Construction and Engineering*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
77. Guo, B.H.; Zou, Y.; Fang, Y.; Goh, Y.M.; Zou, P.X. Computer vision technologies for safety science and management in construction: A critical review and future research directions. *Saf. Sci.* **2021**, *135*, 105130. [CrossRef]
78. Edwards, M.; Deng, J.; Xie, X. From pose to activity: Surveying datasets and introducing CONVERSE. *Comput. Vis. Image Underst.* **2016**, *144*, 73–105. [CrossRef]
79. Jazayeri, E.; Dadi, G.B. Construction safety management systems and methods of safety performance measurement: A review. *J. Saf. Eng.* **2017**, *6*, 15–28.
80. Petersen, D. *Techniques of Safety Management: A Systems Approach*; American Society of Safety Engineers: Park Ridge, IL, USA, 1989.
81. Welford, A.T. *Fundamentals of Skill*; Methuen: London, UK, 1968.
82. Heinrich, H.W. *Industrial Accident Prevention. A Scientific Approach*, 2nd ed.; McGraw-Hill Book Company, Inc.: New York, NY, USA, 1941.
83. Fam, I.M.; Nikoomaram, H.; Soltanian, A. Comparative analysis of creative and classic training methods in health, safety and environment (HSE) participation improvement. *J. Loss Prev. Process Ind.* **2012**, *25*, 250–253. [CrossRef]
84. Luo, H.; Wang, M.; Wong, P.K.-Y.; Cheng, J.C. Full body pose estimation of construction equipment using computer vision and deep learning techniques. *Autom. Constr.* **2020**, *110*, 103016. [CrossRef]
85. Son, H.; Choi, H.; Seong, H.; Kim, C. Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks. *Autom. Constr.* **2019**, *99*, 27–38. [CrossRef]
86. Lee, H.; Yang, K.; Kim, N.; Ahn, C.R. Detecting excessive load-carrying tasks using a deep learning network with a Gramian Angular Field. *Autom. Constr.* **2020**, *120*, 103390. [CrossRef]
87. Luo, H.; Liu, J.; Fang, W.; Love, P.E.; Yu, Q.; Lu, Z. Real-time smart video surveillance to manage safety: A case study of a transport mega-project. *Adv. Eng. Inform.* **2020**, *45*, 101100. [CrossRef]
88. Chen, H.; Luo, X.; Zheng, Z.; Ke, J. A proactive workers' safety risk evaluation framework based on position and posture data fusion. *Autom. Constr.* **2019**, *98*, 275–288. [CrossRef]
89. Zhao, J.; Obonyo, E. Convolutional long short-term memory model for recognizing construction workers' postures from wearable inertial measurement units. *Adv. Eng. Inform.* **2020**, *46*, 101177. [CrossRef]
90. Yang, K.; Ahn, C.R.; Kim, H. Deep learning-based classification of work-related physical load levels in construction. *Adv. Eng. Inform.* **2020**, *45*, 101104. [CrossRef]
91. Kim, K.; Cho, Y.K. Effective inertial sensor quantity and locations on a body for deep learning-based worker's motion recognition. *Autom. Constr.* **2020**, *113*, 103126. [CrossRef]
92. Yu, Y.; Yang, X.; Li, H.; Luo, X.; Guo, H.; Fang, Q. Joint-level vision-based ergonomic assessment tool for construction workers. *J. Constr. Eng. Manag.* **2019**, *145*, 04019025. [CrossRef]
93. Fang, W.; Zhong, B.; Zhao, N.; Love, P.E.; Luo, H.; Xue, J.; Xu, S. A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network. *Adv. Eng. Inform.* **2019**, *39*, 170–177. [CrossRef]
94. Cai, J.; Zhang, Y.; Yang, L.; Cai, H.; Li, S. A context-augmented deep learning approach for worker trajectory prediction on unstructured and dynamic construction sites. *Adv. Eng. Inform.* **2020**, *46*, 101173. [CrossRef]
95. Luo, H.; Wang, M.; Wong, P.K.-Y.; Tang, J.; Cheng, J.C. Construction machine pose prediction considering historical motions and activity attributes using gated recurrent unit (GRU). *Autom. Constr.* **2021**, *121*, 103444. [CrossRef]

96. Wang, M.; Wong, P.; Luo, H.; Kumar, S.; Delhi, V.; Cheng, J. Predicting safety hazards among construction workers and equipment using computer vision and deep learning techniques. In Proceedings of the International Symposium on Automation and Robotics in Construction, Banff Alberta, AB, Canada, 21–24 May 2019; IAARC Publications: Banff Alberta, AB, Canada, 2019; pp. 399–406.

97. Tang, S.; Roberts, D.; Golparvar-Fard, M. Human-object interaction recognition for automatic construction site safety inspection. *Autom. Constr.* **2020**, *120*, 103356. [CrossRef]

98. Kim, D.; Liu, M.; Lee, S.; Kamat, V.R. Remote proximity monitoring between mobile construction resources using camera-mounted UAVs. *Autom. Constr.* **2019**, *99*, 168–182. [CrossRef]

99. Zhang, M.; Cao, Z.; Yang, Z.; Zhao, X. Utilizing computer vision and fuzzy inference to evaluate level of collision safety for workers and equipment in a dynamic environment. *J. Constr. Eng. Manag.* **2020**, *146*, 04020051. [CrossRef]

100. Xiong, R.; Song, Y.; Li, H.; Wang, Y. Onsite video mining for construction hazards identification with visual relationships. *Adv. Eng. Inform.* **2019**, *42*, 100966. [CrossRef]

101. Wang, X.; Zhu, Z. Vision-based hand signal recognition in construction: A feasibility study. *Autom. Constr.* **2021**, *125*, 103625. [CrossRef]

102. Yan, X.; Zhang, H.; Li, H. Computer vision-based recognition of 3D relationship between construction entities for monitoring struck-by accidents. *Comput.-Aided Civ. Infrastruct. Eng.* **2020**, *35*, 1023–1038. [CrossRef]

103. Khan, N.; Saleem, M.R.; Lee, D.; Park, M.-W.; Park, C. Utilizing safety rule correlation for mobile scaffolds monitoring leveraging deep convolution neural networks. *Comput. Ind.* **2021**, *129*, 103448. [CrossRef]

104. Luo, X.; Li, H.; Yang, X.; Yu, Y.; Cao, D. Capturing and understanding workers' activities in far-field surveillance videos with deep action recognition and Bayesian nonparametric learning. *Comput.-Aided Civ. Infrastruct. Eng.* **2019**, *34*, 333–351. [CrossRef]

105. Nath, N.D.; Behzadan, A.H.; Paal, S.G. Deep learning for site safety: Real-time detection of personal protective equipment. *Autom. Constr.* **2020**, *112*, 103085. [CrossRef]

106. Fang, Q.; Li, H.; Luo, X.; Ding, L.; Luo, H.; Rose, T.M.; An, W. Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Autom. Constr.* **2018**, *85*, 1–9. [CrossRef]

107. Chen, S.; Demachi, K. Towards on-site hazards identification of improper use of personal protective equipment using deep learning-based geometric relationships and hierarchical scene graph. *Autom. Constr.* **2021**, *125*, 103619. [CrossRef]

108. Wu, J.; Cai, N.; Chen, W.; Wang, H.; Wang, G. Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset. *Autom. Constr.* **2019**, *106*, 102894. [CrossRef]

109. Nath, N.D.; Behzadan, A.H. Deep Learning Detection of Personal Protective Equipment to Maintain Safety Compliance on Construction Sites. In Proceedings of the Construction Research Congress 2020: Computer Applications, Tempe, AZ, USA, 8–10 March 2020; pp. 181–190.

110. Fang, Q.; Li, H.; Luo, X.; Ding, L.; Luo, H.; Li, C. Computer vision aided inspection on falling prevention measures for steeplejacks in an aerial environment. *Autom. Constr.* **2018**, *93*, 148–164. [CrossRef]

111. Fang, W.; Ding, L.; Luo, H.; Love, P.E. Falls from heights: A computer vision-based approach for safety harness detection. *Autom. Constr.* **2018**, *91*, 53–61. [CrossRef]

112. Fang, Q.; Li, H.; Luo, X.; Ding, L.; Rose, T.M.; An, W.; Yu, Y. A deep learning-based method for detecting non-certified work on construction sites. *Adv. Eng. Inform.* **2018**, *35*, 56–68. [CrossRef]

113. Xie, Z.; Liu, H.; Li, Z.; He, Y. A convolutional neural network based approach towards real-time hard hat detection. In Proceedings of the 2018 IEEE International Conference on Progress in Informatics and Computing (PIC), Suzhou, China, 14–16 December 2018; pp. 430–434.

114. Gu, Y.; Xu, S.; Wang, Y.; Shi, L. An advanced deep learning approach for safety helmet wearing detection. In Proceedings of the 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA, 14–17 July 2019; pp. 669–674.

115. Cao, W.; Zhang, J.; Cai, C.; Chen, Q.; Zhao, Y.; Lou, Y.; Jiang, W.; Gui, G. CNN-based intelligent safety surveillance in green IoT applications. *China Commun.* **2021**, *18*, 108–119. [CrossRef]

116. Zhang, C.; Tian, Z.; Song, J.; Zheng, Y.; Xu, B. Construction worker hardhat-wearing detection based on an improved BiFPN. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 8600–8607.

117. Zhao, Y.; Chen, Q.; Cao, W.; Yang, J.; Xiong, J.; Gui, G. Deep learning for risk detection and trajectory tracking at construction sites. *IEEE Access* **2019**, *7*, 30905–30912. [CrossRef]

118. Shen, J.; Xiong, X.; Li, Y.; He, W.; Li, P.; Zheng, X. Detecting safety helmet wearing on construction sites with bounding-box regression and deep transfer learning. *Comput.-Aided Civ. Infrastruct. Eng.* **2021**, *36*, 180–196. [CrossRef]

119. Huang, L.; Fu, Q.; He, M.; Jiang, D.; Hao, Z. Detection algorithm of safety helmet wearing based on deep learning. *Concurr. Comput. Pract. Exp.* **2021**, *2020*, e6234. [CrossRef]

120. Hu, J.; Gao, X.; Wu, H.; Gao, S. Detection of workers without the helments in videos based on YOLO V3. In Proceedings of the 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Suzhou, China, 19–21 October 2019; pp. 1–4.

121. Tan, S.; Lu, G.; Jiang, Z.; Huang, L. Improved YOLOv5 Network Model and Application in Safety Helmet Detection. In Proceedings of the 2021 IEEE International Conference on Intelligence and Safety for Robotics (ISR), Tokoname, Japan, 4–6 March 2021; pp. 330–333.
122. Lee, M.-F.R.; Chien, T.-W. Intelligent Robot for Worker Safety Surveillance: Deep Learning Perception and Visual Navigation. In Proceedings of the 2020 International Conference on Advanced Robotics and Intelligent Systems (ARIS), Taipei, Taiwan, 19–21 August 2020; pp. 1–6.
123. Kitsikidis, A.; Dimitropoulos, K.; Douka, S.; Grammalidis, N. Dance analysis using multiple kinect sensors. In Proceedings of the 2014 International Conference on Computer Vision Theory and Applications (VISAPP), Lisbon, Portugal, 5–8 January 2014; pp. 789–795.
124. Hong, Z.; Gui, F. Analysis on human unsafe acts contributing to falling accidents in construction industry. In Proceedings of the International Conference on Applied Human Factors and Ergonomics, Los Angeles, CA, USA, 17–21 July 2017; pp. 178–185.
125. Chi, C.-F.; Chang, T.-C.; Ting, H.-I. Accident patterns and prevention measures for fatal occupational falls in the construction industry. *Appl. Ergon.* **2005**, *36*, 391–400. [CrossRef]
126. Han, S.; Lee, S.; Peña-Mora, F. Comparative study of motion features for similarity-based modeling and classification of unsafe actions in construction. *J. Comput. Civ. Eng.* **2014**, *28*, A4014005. [CrossRef]
127. Gong, J.; Caldas, C.H.; Gordon, C. Learning and classifying actions of construction workers and equipment using Bag-of-Video-Feature-Words and Bayesian network models. *Adv. Eng. Inform.* **2011**, *25*, 771–782. [CrossRef]
128. Turaga, P.; Chellappa, R.; Subrahmanian, V.S.; Udrea, O. Machine recognition of human activities: A survey. *IEEE Trans. Circuits Syst. Video Technol.* **2008**, *18*, 1473–1488. [CrossRef]
129. Haupt, T.C. *The Performance Approach to Construction Worker Safety and Health*; University of Florida: Gainesville, FL, USA, 2001.
130. Nain, M.; Sharma, S.; Chaurasia, S. Safety and Compliance Management System Using Computer Vision and Deep Learning. In Proceedings of the IOP Conference Series: Materials Science and Engineering, Jaipur, India, 22–23 December 2020; p. 012013.
131. Park, M.-W.; Elsafty, N.; Zhu, Z. Hardhat-wearing detection for enhancing on-site safety of construction workers. *J. Constr. Eng. Manag.* **2015**, *141*, 04015024. [CrossRef]
132. Enshassi, A.; Mayer, P.E.; Mohamed, S.; El-Masri, F. Perception of construction managers towards safety in Palestine. *Int. J. Constr. Manag.* **2007**, *7*, 41–51. [CrossRef]
133. Törner, M.; Pousette, A. Safety in construction–a comprehensive description of the characteristics of high safety standards in construction work, from the combined perspective of supervisors and experienced workers. *J. Saf. Res.* **2009**, *40*, 399–409. [CrossRef] [PubMed]
134. Fang, W.; Ding, L.; Love, P.E.; Luo, H.; Li, H.; Pena-Mora, F.; Zhong, B.; Zhou, C. Computer vision applications in construction safety assurance. *Autom. Constr.* **2020**, *110*, 103013. [CrossRef]
135. Chen, Q.; Zhang, X.-X.; Chen, Y.; Jiang, W.; Gui, G.; Sari, H. Deep Learning-Based Automatic Safety Detection System for Crack Detection. In Proceedings of the 2020 7th International Conference on Dependable Systems and Their Applications (DSA), Xi'an, China, 28–29 November 2020; pp. 190–194.
136. Zhao, H.-J.; Liu, W.; Shi, P.-X.; Du, J.-T.; Chen, X.-M. Spatiotemporal deep learning approach on estimation of diaphragm wall deformation induced by excavation. *Acta Geotech.* **2021**, *16*, 3631–3645. [CrossRef]
137. Shi, J.; Sun, D.; Hu, M.; Liu, S.; Kan, Y.; Chen, R.; Ma, K. Prediction of brake pedal aperture for automatic wheel loader based on deep learning. *Autom. Constr.* **2020**, *119*, 103313. [CrossRef]
138. Xiao, B.; Lin, Q.; Chen, Y. A vision-based method for automatic tracking of construction machines at nighttime based on deep learning illumination enhancement. *Autom. Constr.* **2021**, *127*, 103721. [CrossRef]
139. Guo, Y.; Xu, Y.; Li, S. Dense construction vehicle detection based on orientation-aware feature fusion convolutional neural network. *Autom. Constr.* **2020**, *112*, 103124. [CrossRef]
140. Mahmoodzadeh, A.; Mohammadi, M.; Noori, K.M.G.; Khishe, M.; Ibrahim, H.H.; Ali, H.F.H.; Abdulhamid, S.N. Presenting the best prediction model of water inflow into drill and blast tunnels among several machine learning techniques. *Autom. Constr.* **2021**, *127*, 103719. [CrossRef]
141. Fatal Occupational Injuries for Selected Events or Exposures. Available online: https://www.bls.gov/news.release/cfoi.t02.htm (accessed on 10 October 2017).
142. Arditi, D.; Lee, D.-E.; Polat, G. Fatal accidents in nighttime vs. daytime highway construction work zones. *J. Saf. Res.* **2007**, *38*, 399–405. [CrossRef] [PubMed]
143. Luo, X.; Li, H.; Dai, F.; Cao, D.; Yang, X.; Guo, H. Hierarchical bayesian model of worker response to proximity warnings of construction safety hazards: Toward constant review of safety risk control measures. *J. Constr. Eng. Manag.* **2017**, *143*, 04017006. [CrossRef]
144. Awolusi, I.; Marks, E.; Hallowell, M. Wearable technology for personalized construction safety monitoring and trending: Review of applicable devices. *Autom. Constr.* **2018**, *85*, 96–106. [CrossRef]
145. Jin, X.; Li, Y.; Luo, Y.; Liu, H. Prediction of city tunnel water inflow and its influence on overlain lakes in karst valley. *Environ. Earth Sci.* **2016**, *75*, 1–15. [CrossRef]
146. Zhou, Y.; Ding, L.; Chen, L. Application of 4D visualization technology for safety management in metro construction. *Autom. Constr.* **2013**, *34*, 25–36. [CrossRef]

147. Park, C.-S.; Kim, H.-J. A framework for construction safety management and visualization system. *Autom. Constr.* **2013**, *33*, 95–103. [CrossRef]

148. Fang, W.; Ding, L.; Zhong, B.; Love, P.E.; Luo, H. Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach. *Adv. Eng. Inform.* **2018**, *37*, 139–149. [CrossRef]

149. Xuehui, A.; Li, Z.; Zuguang, L.; Chengzhi, W.; Pengfei, L.; Zhiwei, L. Dataset and benchmark for detecting moving objects in construction sites. *Autom. Constr.* **2021**, *122*, 103482. [CrossRef]

150. Zeng, T.; Wang, J.; Cui, B.; Wang, X.; Wang, D.; Zhang, Y. The equipment detection and localization of large-scale construction jobsite by far-field construction surveillance video based on improving YOLOv3 and grey wolf optimizer improving extreme learning machine. *Constr. Build. Mater.* **2021**, *291*, 123268. [CrossRef]

151. Liu, H.; Wang, G.; Huang, T.; He, P.; Skitmore, M.; Luo, X. Manifesting construction activity scenes via image captioning. *Autom. Constr.* **2020**, *119*, 103334. [CrossRef]

152. Wei, R.; Love, P.E.; Fang, W.; Luo, H.; Xu, S. Recognizing people's identity in construction sites with computer vision: A spatial and temporal attention pooling network. *Adv. Eng. Inform.* **2019**, *42*, 100981. [CrossRef]

153. Arabi, S.; Haghighat, A.; Sharma, A. A deep-learning-based computer vision solution for construction vehicle detection. *Comput.-Aided Civ. Infrastruct. Eng.* **2020**, *35*, 753–767. [CrossRef]

154. Zhong, B.; Pan, X.; Love, P.E.; Sun, J.; Tao, C. Hazard analysis: A deep learning and text mining framework for accident prevention. *Adv. Eng. Inform.* **2020**, *46*, 101152. [CrossRef]

155. Xiao, B.; Yin, X.; Kang, S.-C. Vision-based method of automatically detecting construction video highlights by integrating machine tracking and CNN feature extraction. *Autom. Constr.* **2021**, *129*, 103817. [CrossRef]

156. Fang, W.; Ma, L.; Love, P.E.; Luo, H.; Ding, L.; Zhou, A. Knowledge graph for identifying hazards on construction sites: Integrating computer vision with ontology. *Autom. Constr.* **2020**, *119*, 103310. [CrossRef]

157. Jeelani, I.; Asadi, K.; Ramshankar, H.; Han, K.; Albert, A. Real-time vision-based worker localization & hazard detection for construction. *Autom. Constr.* **2021**, *121*, 103448.

158. Love, P.E.; Smith, J.; Teo, P. Putting into practice error management theory: Unlearning and learning to manage action errors in construction. *Appl. Ergon.* **2018**, *69*, 104–111. [CrossRef]

159. Love, P.E.; Teo, P.; Ackermann, F.; Smith, J.; Alexander, J.; Palaneeswaran, E.; Morrison, J. Reduce rework, improve safety: An empirical inquiry into the precursors to error in construction. *Prod. Plan. Control* **2018**, *29*, 353–366. [CrossRef]

160. Love, P.E.; Teo, P.; Morrison, J. Unearthing the nature and interplay of quality and safety in construction projects: An empirical study. *Saf. Sci.* **2018**, *103*, 270–279. [CrossRef]

161. Gouett, M.C.; Haas, C.T.; Goodrum, P.M.; Caldas, C.H. Activity analysis for direct-work rate improvement in construction. *J. Constr. Eng. Manag.* **2011**, *137*, 1117–1124. [CrossRef]

162. Khosrowpour, A.; Niebles, J.C.; Golparvar-Fard, M. Vision-based workface assessment using depth images for activity analysis of interior construction operations. *Autom. Constr.* **2014**, *48*, 74–87. [CrossRef]

163. Goh, Y.M.; Ubeynarayana, C. Construction accident narrative classification: An evaluation of text mining techniques. *Accid. Anal. Prev.* **2017**, *108*, 122–130. [CrossRef] [PubMed]

164. Jiang, J. Information extraction from text. In *Mining Text Data*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 11–41.

165. Liu, S.; Li, Y.; Fan, B. Hierarchical RNN for few-shot information extraction learning. In Proceedings of the International Conference of Pioneering Computer Scientists, Engineers and Educators, Zhengzhou, China, 21–23 September 2018; Springer: Singapore, 2018; pp. 227–239.

166. Guo, L.; Zhang, D.; Wang, L.; Wang, H.; Cui, B. CRAN: A hybrid CNN-RNN attention-based model for text classification. In Proceedings of the International Conference on Conceptual Modeling, Xi'an, China, 22–25 October 2018; Springer: Cham, Switzerland, 2018; pp. 571–585.

167. Bahn, S. Workplace hazard identification and management: The case of an underground mining operation. *Saf. Sci.* **2013**, *57*, 129–137. [CrossRef]

168. Albert, A.; Hallowell, M.R.; Kleiner, B.M. Enhancing construction hazard recognition and communication with energy-based cognitive mnemonics and safety meeting maturity model: Multiple baseline study. *J. Constr. Eng. Manag.* **2014**, *140*, 04013042. [CrossRef]

169. Jeelani, I.; Han, K.; Albert, A. Development of immersive personalized training environment for construction workers. *Comput. Civ. Eng.* **2017**, *2017*, 407–415.

170. Daniel, G.; Chen, M. Video Visualization. In Proceedings of the IEEE Visualization, Washington, DC, USA, 19–24 October 2003.

171. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision 2014, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 740–755. [CrossRef]

172. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]

173. Everingham, M.; Eslami, S.A.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes challenge: A retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [CrossRef]

174. Ma, S.; Zhang, X.; Jia, C.; Zhao, Z.; Wang, S.; Wang, S. Image and video compression with neural networks: A review. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 1683–1698. [CrossRef]

175. Hu, Y.; Lu, X. Learning spatial-temporal features for video copy detection by the combination of CNN and RNN. *J. Vis. Commun. Image Represent.* **2018**, *55*, 21–29. [CrossRef]

176. Monshi, M.M.A.; Poon, J.; Chung, V. Deep learning in generating radiology reports: A survey. *Artif. Intell. Med.* **2020**, *106*, 101878. [CrossRef]

177. Srivastava, M.; Muntz, R.; Potkonjak, M. Smart kindergarten: Sensor-based wireless networks for smart developmental problem-solving environments. In Proceedings of the 7th Annual International Conference on Mobile Computing and Networking, Rome, Italy, 16–21 July 2001; pp. 132–138.

178. Lu, X.; Yang, Y.; Zhang, W.; Wang, Q.; Wang, Y. Face verification with multi-task and multi-scale feature fusion. *Entropy* **2017**, *19*, 228. [CrossRef]