Article

Real-Time Early Safety Warning for Personnel Intrusion Behavior on Construction Sites Using a CNN Model

Jinyu Zhao [1,*], Yinghui Xu [1], Weina Zhu [1], Mei Liu [1] and Jing Zhao [2]

1 School of Urban Economics and Management, Beijing University of Civil Engineering and Architecture, Beijing 102616, China; yinghuixu_bucea@163.com (Y.X.); sophiezwn@163.com (W.Z.); liumei@bucea.edu.cn (M.L.)
2 School of Economics and Management, Hebei University of Architecture, Zhangjiakou 075000, China; dongjiaqi1301@162.com
* Correspondence: zhaojinyu@bucea.edu.cn

Abstract: The high number of annual safety accidents and casualties reflects the problems of slow detection of safety accidents and untimely early warnings in current construction safety management, and China urgently needs new methods and technologies to improve the safety management efficiency of the construction industry. However, there are fewer achievements in the use of new technologies for intelligent construction safety management, and most of the research focuses on intrusion detection and specific event alarms, which cannot be well implemented for systematic early warning functions. Based on the existing research and the characteristics of early warning scenarios, this study introduces the convolutional neural network (CNN) to build a video image recognition and classification model to give early safety warnings for intrusion behavior in hazard areas of construction and demonstrates the warning effect and accuracy with practical cases. First, it clarifies the early warning demand information, such as the attributes of construction personnel and hazard areas. Then, the construction model is realized by multi-scale hierarchical feature extraction mapping, the Softmax classification function, and the argmax function. Finally, from the empirical analysis, it can be seen that an early safety warning based on the CNN model has an accurate ability to identify the intrusion behavior of construction site personnel, which can reduce the probability of construction safety accidents to a certain extent, and provide enlightenment for further realization of intelligent construction sites.

Keywords: construction safety management; intrusion behavior; early warning; convolutional neural network (CNN)

1. Introduction

The construction industry absorbs the most urban labor force and has made outstanding contributions to promoting high-quality economic development in China, but its total number of accidents remains high among industries, and the effect of safety control on construction sites is not obvious. According to the Ministry of Housing and Urban-Rural Development of the People's Republic of China, there were 737 production safety accidents and 824 deaths in housing and municipal engineering in 2021, with the main types of accidents including falls, object strikes, pit collapses, and crane injuries [1]. In order to solve these accidents and reduce the large number of deaths in China's construction industry, we can draw on early safety warning technologies that have been widely used in other industries. Combining these technologies with the particular environment of construction sites to build a sound early warning system and provide timely alerts when there are signs of accidents. Therefore, it has become an urgent problem for the construction industry to effectively realize the early warning of accidents, improve the level of safety management, and reduce the incidence of accidents on construction sites.

The occurrence of construction safety accidents is mostly due to the joint action of "Men, Machine, Medium and Management" factors. Combined with the "Industrial Safety Theory" of Heinrich [2] and related research results [3,4], it can be seen that personnel who do not meet the safety requirements entering the construction hazard area is the key factor causing construction safety accidents. Here, the behavior of construction personnel to enter the construction hazard area at random is the "Personnel Invasion Behavior", which usually contains two meanings: (1) non-professional construction personnel enter the construction area containing professional sources of danger; (2) non-site construction personnel enter the construction area. Construction sites are characterized by the cross-operation of related professions, uneven quality of personnel, uncertain construction environment, and so on, so the two kinds of intrusion behaviors are prevalent in daily production activities of China's construction sites. It can be seen that the realization of early personnel intrusion warnings on construction sites is crucial for construction safety management.

Established research mainly uses personnel location information and personnel behavior and action information to achieve safety warnings in various fields [5,6]. Navon and Kolton [7] established an automatic monitoring model that can automatically divide hazard areas, identify dangerous actions, and provide protective measures. Zheng et al. [8] established a real-time observation system for dangerous omen events, which realized the function of early warning by predicting the possible location of dangerous events in real time and taking timely action after the occurrence of these events. In addition, Shuang and Zhang [9] applied machine learning techniques to determine the hierarchical relationship of fatal causes of construction site accidents, providing early warning signs of fatal and unsafe factors for those involved.

In summary, traditional safety management mainly relies on manual supervision, which is not only time-consuming and labor-intensive but also inefficient, and some potential safety hazards cannot be controlled at the source. Construction safety management should consider the use of more advanced technology to strengthen coordinated management [10] and provide timely reminders or warnings to non-construction personnel entering hazard areas and approaching hazardous sources. The existing research mainly focuses on personnel location monitoring, safety protection, and special accident warnings, etc. The realization of the early warning function is not yet comprehensive, and there are fewer studies on real-time monitoring and early warning involving personnel intrusion behavior in hazard areas on construction sites. For the above problems, this paper introduces CNN technology to carry out early safety warning research on personnel intrusion behavior after identifying hazard areas that are prone to have serious consequences and collect information on construction personnel. CNN technology has the features of image labeling, feature mapping acquisition, personnel identification, and simulation, which can make up for the deficiencies of other techniques and lay the foundation for real-time monitoring of unsafe posture.

2. Literature Review

2.1. Overview of Early Warning Approaches for Intrusion Behavior

The idea of early warning arose from the prediction of economic fluctuations [11]. Gu [12] proposed three stages of early warning through his study of economic warnings, namely defining the meaning of warning, finding the source of warning, analyzing warning signs, and predicting the level of warning, where warning refers to the monitoring objects, the source of warning is the source of danger, the warning signs are the precursors of the occurrence of police situations, and the level of warning is the predicted hazard level.

The study of early safety warnings for personnel intrusion behavior on construction sites focuses on two aspects: one is to identify the construction hazard area, and the other is to provide real-time warnings and deal with the intrusion. For the identification of construction hazard areas, in addition to research on classification [13] and evaluation [14], real-time positioning technology can be used to collect the location trajectory of construction personnel and achieve static or dynamic identification and automatic classification of

construction hazard area [15]. A real-time hazard area automatic identification model based on BIM and real-time positioning technology can identify potential hazard areas for construction personnel paths [16]. In terms of intrusion warnings, the identification of personnel intrusion behavior is the first step in early safety warnings. Behavior-based safety (BBS) is often used to identify unsafe worker behavior, but traditional BBS management systems rely on empirical manual procedures that do not effectively capture complex intrusion-related information.

At present, methods to identify and warn about personnel intrusion behavior from a technological perspective show greater advantages. Some studies use positioning technologies, such as ultra-wide band (UWB) [17] and real-time location sensing [15,18] to track the location of workers and machinery to identify and warn of personnel intrusion behavior at construction sites. Computer-based location tracking technology integrated with BBS [19] and BIM [20] is effective in automatically identifying construction personnel entering hazard areas, providing timely warnings, and capturing worker responses. However, the positioning system is used to capture the temporal and spatial trajectories of construction personnel and machinery and cannot identify intrusions in dynamic environments that evolve over time. In addition, wearable sensor technology requires high costs, affects worker operations [21], and is relatively few used in intrusion behavior recognition.

With the development of artificial intelligence, machine vision technology, which does not require contact with the observed object, is gradually being used in hazardous working environments that are not suitable for manual work or in construction situations where manual vision has difficulty meeting the requirements. The application of this technology in construction personnel intrusion is mainly through the collection of on-site video data, and then machine vision algorithms, such as ResnNet [22] and R-CNN [23], to obtain information about the location of construction site personnel, which is also known as target detection. Gao [4] adopted a moving target detection algorithm to detect workers entering the dangerous area in real time and to discriminate intrusion behavior based on their behavioral characteristics. Wei et al. [24] utilized a spatiotemporal attention network to remove redundant information from the video, thus achieving accurate and automatic identification of workers from the construction scene. Lie et al. [25] developed a novel deep-learning scheme that actively recognizes construction worker violations through the aggregation of CNN and long-term and short-term memory networks, allowing the extraction of image features without the involvement of cumbersome parameters. Each of these detection models has its own strengths, but most of them still need to improve their real-time performance, accuracy, and simplicity of implementation steps. Wang [26] combined BIM technology with machine vision technology to construct an early warning model for construction personnel intrusion in the hazard areas of a building construction site.

Existing research is focused on enabling proactive and real-time intrusion warnings. Intrusion sensing and alerting technologies have the potential to improve worker safety in construction site work zones by providing warnings when a hazardous situation exists [27]. When equipment is too close to the unknown or other equipment, radio frequency (RF) remote sensing and actuating technology can alert moving workers and equipment operators in active and real-time mode [28]. Modern proactive safety systems provide advanced real-time tracking of on-site workers, which alerts them to real-time audio warnings if they get too close to danger [29]. Visualization techniques can enhance safety management by assisting with safety training, job hazard area (JHA) identification, on-site safety monitoring, and warnings [21]. The intrusion detection system uses Bluetooth low energy (BLE) beacons to visualize intrusion in hazard areas of construction sites by combining spatiotemporal trajectories of workers captured by the system with BIM [30]. BLE real-time location systems (RTLS) can filter out unnecessary alarms and generate timely vibration-tactile alarms, providing easily perceived proximity safety alarms for construction site workers [31]. A haptic-based warning mechanism has been introduced to identify intrusion

hazards in construction work zones. Tactile signals in this mechanism can increase worker awareness of the dangers of intrusion [32].

From the existing literature, it can be seen that most studies have used localization techniques to obtain location information for people and machinery, and there are few early warning studies that use machine vision techniques, such as CNN, to monitor workers' intrusion into hazard areas in real time. Although intrusion detection techniques have been widely studied and applied, they can only support in-event inspection and post-event control. Therefore, further in-depth research is still needed to achieve proactive measures to improve the efficiency of construction safety management through early warning systems.

2.2. Overview of the Convolutional Neural Network Model

A convolutional neural network (CNN) was proposed in 1962 by Hubel and other biologists after studying neuronal cells in the visual cortex of cats [33]. The network has a convolutional layer to form a feature map, a pooling layer to downscale the feature map data, and a full connection layer to summarize the input feature information of the pooling layer, and it runs in a stepwise fashion. These features can greatly reduce the complexity of the network and the computational complexity of the training process and improve the accuracy and efficiency of feature recognition, which are mainly used in risk intelligence detection [34] and machine vision detection [35,36]. The CNN is one of the most popular deep neural networks and performs very well on pattern recognition and machine learning problems, especially in applications dealing with image data [37]. The CNN model was built to evaluate its performance on image recognition and detection datasets, and the results showed good computational accuracy [38,39]. CNN also has great development prospects in the field of construction safety.

Some scholars have already begun to use machine learning techniques to analyze civil engineering activities. Exploring intelligent detection of concrete damage reduction by building a CNN model, thus reducing the frequency of construction site safety accidents caused by damage to concrete components [40]. Kalman filtering is used to locate workers and construction machinery in real time and predict their movement trends to prevent collisions between workers and machinery [41]. A CNN-based safety fence detection model can improve the efficiency of safety inspection at construction sites [42]. Data learning and mining of typical safety hazards at construction sites through CNN models, while using machines to automatically identify typical hazards, could achieve the goal of improving the efficiency of construction site safety management [43]. A deep learning model that integrates the advantages of CNN and RNN is capable of capturing powerful acoustic features from time-frequency spectrograms and utilizing acoustic modal techniques to accurately identify the start and end times of individual construction activities for efficient automated monitoring of building construction activities [44].

Target detection technology with multi-target classification and localization as the main task is the basis of intrusion recognition implementation, and CNN has gradually become an important part of target detection. The R-CNN (regions with CNN features) algorithm is proposed for the first time to realize the application of deep learning in the field of target detection [45]. With the continuous improvement of technology, Faster R-CNN has greatly improved in accuracy and detection speed and has also become the final version of the family of target detection structures such as R-CNN, SPP-Net, and Fast R-CNN [46]. Since computer vision pattern recognition methods for identifying unsafe behaviors on construction sites rely on manual computation of complex image features, some studies have borrowed from other fields and used machine vision algorithms, such as CNN, to obtain construction site coordinate information for locating construction site personnel. CNN-based deep learning methods can effectively detect and monitor workers' behavior in complex environments [47,48]. A hybrid deep learning model has integrated CNN and long short-term memory (LSTM) and can automatically identify unsafe operations of workers [49]. A construction activity image recognition system incorporating deep learning can utilize the powerful image extraction capability of CNN to automatically feed

feature data into the network for full-connectivity training, thereby effectively improving the efficiency of construction management [25].

According to the literature mentioned above, the existing research has produced a series of results in terms of construction site personnel localization, personnel attributes, and construction site component recognition, indicating that it is feasible to apply CNN to construction site safety management. However, the research mainly focuses on personnel safety protection equipment wearing recognition, personnel simple behavior recognition, component breakage recognition, etc., which can only solve specific safety management problems. For complex engineering projects, there are still restrictions on safety warnings for construction site personnel. Therefore, it is worth further research on combining early warning theory and CNN to establish a construction hazard area intrusion warning system.

3. Methodology

Adopting CNN as the realization method of the early warning function in this study, the early warning demand information should be determined first, i.e., the "hazard area" information and "personnel attribute information" required by the early warning function should be determined. Then, it is necessary to implement the function, i.e., selecting the appropriate algorithm to build the model, and then use the collected data for model pre-training and model validation.

3.1. Process of Personnel Intrusion Behavior Early Warning

The implementation of the early safety warning function for personnel intrusion in the construction hazard area can be described as follows: first, identify the construction personnel in the monitoring video at the edge of the hazard area and calculate the personnel coordinates through the background coordinate system. Then, the relative position of the personnel and hazard area is determined based on the coordinates, and if the relative distance is greater than the safety threshold set in advance and the personnel is not a construction professional in the area, a strong reminder is made, and the safety management personnel are notified. The work process is shown in Figure 1 below:
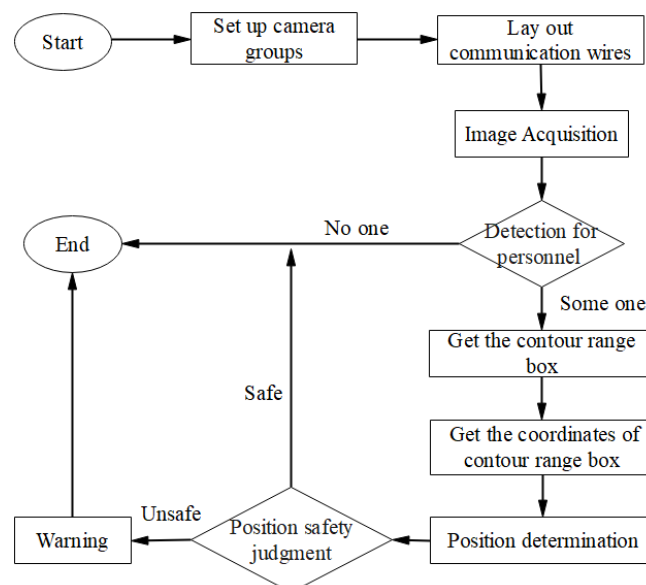

Figure 1. Early Safety Warning Process of Invasion in Construction Hazard Area.

(1) Start the workstation system. After setting up the camera group and installing the communication wires, the operator can open the camera group and the video identification terminal of the workstation through remote control at the workstation and transmit the real-time video acquired by the camera group in the hazard area of the construction site to the terminal in the form of frame images.

(2) Realize real-time detection of construction site personnel. First, CNN obtained the characteristic map of construction site workers. Then, the classification function of the Softmax function is used to separate the construction personnel in any video frame image, so as to achieve the purpose of detecting the construction site personnel in the video.

(3) Real-time measurement of personnel coordinates. After the coordinate recognition of the human body and the drawing of the human body range frame, the coordinate regression function is used to realize the coordinate output of the human body range frame.

(4) Realize early warnings according to detection results. When the body and construction of the danger zone at the edge of the distance is less than the safe distance of construction site rules, under the condition of appropriate comprehensive safety, the personnel access the properties and reminders or warnings for location information are given, and the image is saved automatically. The early safety warning procedure is performed in the next frame image in a row to complete construction hazard area identification.

The "composite classification model" determines whether the construction workers enter the hazard area and then provides the corresponding level of warning according to the attribute information of the workers and the category of the hazard area.

(1) Ordinary reminder: When personnel enter the hazard area conforming to their own attribute information;

(2) Emergency reminder: When personnel enter the hazard area that does not conform to the information of their own attributes, they will make an emergency reminder and send the information to the security officer for disposal.

3.2. Identification of Hazard Areas at Construction Sites

3.2.1. Division of Hazard Areas

At present, most of the construction site hazard areas are divided into pits, holes, edges, machinery, and other sources of danger as the origin to expand a certain distance, and then the closed area is drawn with this distance as radius or side length. The classification of hazard areas in this case is based on the classification of construction safety accidents in previous years and the proportion of accidents, but it is difficult to cope with the complex construction site environment by relying on the several existing types of fixed accidents as a support.

By analyzing the construction sites for each type of professional work, this paper found that the daily working scenes and common hazards of civil construction personnel, machinery operators, and electric power engineering personnel are greatly different, so unified management of different types of work often leads to inefficiency. Due to the differentiation of safety education, personnel can be safer in their own professional area than in the construction area of other types, which is why accidents are mostly concentrated in the work area of non-specialized trades. It has also been documented that a large proportion of construction professionals are involved in safety accidents, usually in non-professional work areas [50]. Therefore, this paper will divide the hazard area based on the working area of each type of construction personnel and combine it with the rules for defining the unsafe environment of the construction site, to achieve regional specialization.

(1) Hazard areas in civil construction. The area may contain hazards such as edges, opening holes, etc. The rules for defining this large category of hazard area draw on previous research, expanding the edge of the construction area by 1 m and using this expansion line as the boundary point for the civil construction hazard area [51].

(2) Hazard area for the movement of machinery. Danger sources include working cranes arms, moving machinery and equipment, etc. Mark the path of the machinery and equipment activity and extend the outermost edge by 1 m as the boundary of the machinery movement hazard area.

(3)　Hazard area for power leakage. The main sources are wires and cables. In the direction perpendicular to the main body of the wire, the projection of the wire and cable is expanded by 1 m, which defines the boundary of the leakage hazard area.

3.2.2. Modeling of Construction Site

According to the above method of dividing the hazard area of the construction site, the construction site is simulated, as shown in Figure 2.
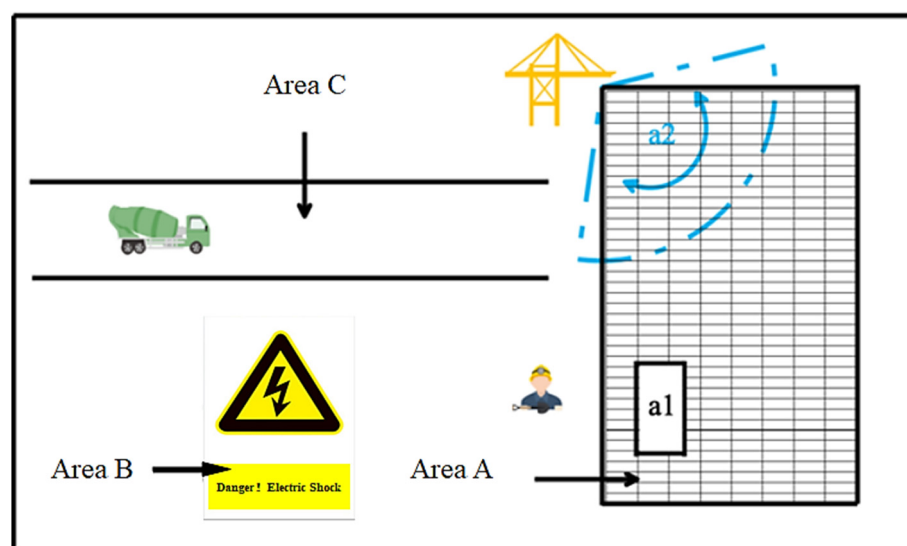

Figure 2. Construction Site Simulation.

Area A indicates the hazard area for civil construction, where "a1" indicates a hazard area, such as edges and opening holes, and "a2" indicates a hazard area under crane arms. Area B is the hazard area for power leaks. Area C is the hazard area for mechanical movement. This study mainly focuses on civil construction workers (C), machine operators (M) and electrical technicians (E) who tend to have high casualty rates at construction sites.

3.3. Quantification of Construction Personnel Information
3.3.1. Collection of Personnel Attribute Information

Construction projects consist of multiple divisional projects, and each divisional project requires a variety of professional construction personnel; hence, each complete construction project contains a variety of different attributes of workers. The different worker attributes are mainly represented by key information, such as the type of work, equipment, and years of experience. The collection of this type of information is achieved through the "Construction Site Personnel Management System" developed with the fifth author's participation [52]. The system is built based on BIM-RFID technology, which can realize real-time detection of personnel attribute information on the construction site.

The construction site personnel management system under BIM-RFID technology includes: RFID chip system, BIM security area system, site management system, intelligent storage system, and intelligent gate system. First, during the process of construction personnel wearing built-in RFID chip helmets entering the site from the induction channel, the information of construction personnel including basic information of workers, work type, project, and necessary safety equipment is entered. The RFID chip containing personnel information is programmed using Python, and then, based on the BIM platform, the construction personnel's attribute information is monitored. The recorded work type information can distinguish the professional construction area of the personnel so as to detect whether the workers stay in the non-specialized construction area.

The personnel attributes contain three pieces of information: job type, coordinates, and time. As mentioned above, letters C, M, and E represent their type of work (T). The coordinates

of personnel are represented by (x, y). The time of personnel presence is indicated by t. The attribute information of individual personnel can be expressed as $(T, (x, y), t)$.

3.3.2. Quantification of Personnel Behavior

The vast majority of construction site accidents occur in the hazard area, so it is meaningful to know exactly if a construction worker is in the hazard area. By adding the position input function of "Personnel Identification Box" (as shown in Figure 3) to the construction personnel classification model, the real-time coordinates (x, y) of personnel are obtained, and then the boundary coordinate information of personnel and hazard areas can be used to realize the quantitative analysis of the safety of construction site personnel.

$$(x, y) = \left(\frac{x_1 + x_2 + x_3 + x_4}{4}, \frac{y_1 + y_2 + y_3 + y_4}{4} \right) \tag{1}$$
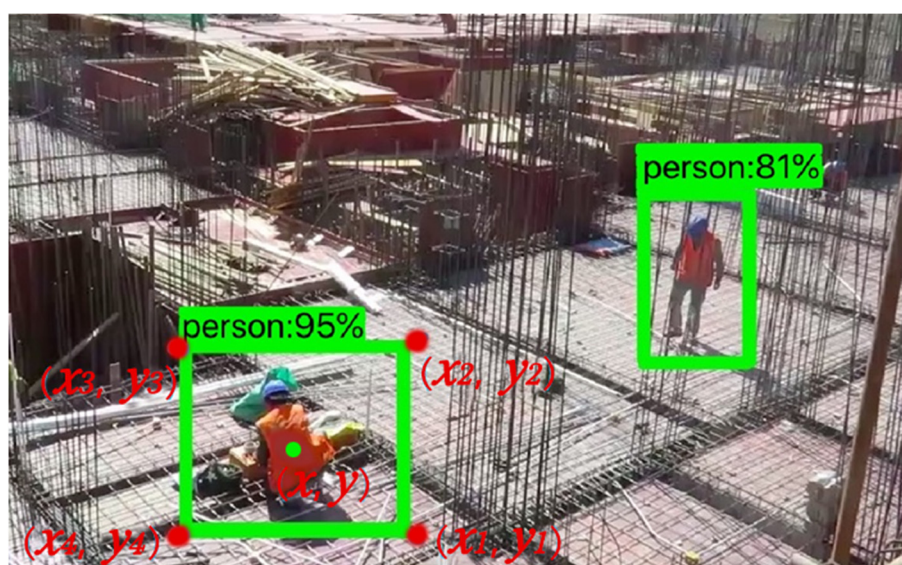

Figure 3. Principle of Personnel Coordinate Determination.

In Equation (1), (x, y) is the body center coordinates of the personnel on the elevation perpendicular to the ground; x_1, x_2, x_3, x_4 are the horizontal values of the four corners of the personnel identification box; y_1, y_2, y_3, y_4 are the vertical values of the four corners of the personnel identification box.

After obtaining the coordinates of the personnel, the linear distance (s) between the worker and the edge of the hazard area is calculated. The calculation formula is shown in Equation (2):

$$s = |\delta_a - \delta_b| \tag{2}$$

where "s" denotes the linear distance between the worker and the hazard area. "δ_a" denotes the horizontal coordinates of the worker on the elevation perpendicular to the edge of the hazard area, and "δ_b" denotes the horizontal coordinates of the edge of the hazard area on the elevation. If the linear distance (s) > the safety threshold, then no voice alert is given; if, on the contrary, the linear distance (s) ≤ the safety threshold, perform voice alerts.

3.4. CNN Model of Intrusions in Hazard Area

3.4.1. Personnel Image Acquisition and Marking

Under the premise of conforming to the "Technical Code for Video Surveillance on Construction Site", the video camera groups that can obtain images from all angles of the construction hazard area are arranged. There have been studies on the layout of camera sets at construction sites, which are divided into two categories: flat view perspective and

top view perspective [36]. With more accurate acquisition of human real-time behavior as the starting point, the layout of camera sets is selected from a flat view perspective.

After setting up the camera set, the image collection of the workers on the construction site was carried out. Through the data communication line of the construction site that has been set up, the captured images are transmitted to the image processing terminal, and the pre-prepared LabelMe marking software is used to mark the collected pictures containing the daily activities of the workers on the construction site; that is, the "objects" to be recognized by CNN in this training. Finally, the marked field photos are unequally divided into two parts, and the larger number of photo sets are divided into training sets, while the smaller portion is divided into verification sets.

3.4.2. Personnel Image Feature Mapping Acquisition

Feature mapping is the key to successful identification of construction workers in images later and is a prerequisite for early warning of construction site workers' intrusive behavior. For a captured construction site image, a fast local Laplace filter can be built to enhance the features of each pixel in the image [53]. Before performing a fast local Laplace filter transformation containing images of construction workers, if the image of each layer is set to be 1/4 of the image of the previous layer, the image resolution of each layer is halved compared to the image of the previous layer, and a Gaussian pyramid on the construction site image is built (see Figure 4 below).
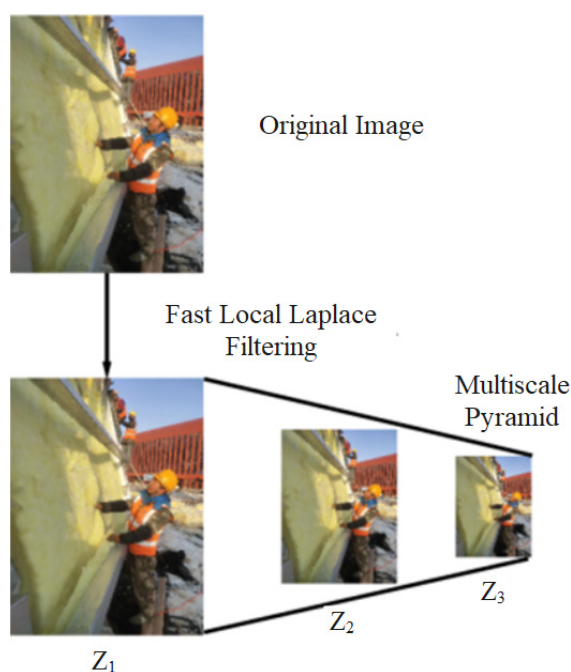

Figure 4. Schematic Diagram of Multi-scale Pyramid Construction Site Image.

In the above figure, the image "Z_1" has the same resolution as the original image, and the images "Z_2", "Z_3" ... "Z_n" ($n \geq 3$) are in turn half the resolution of the previous image. This process, on the one hand, can improve the accuracy of image recognition at a later stage; on the other hand, after the "filtering" process, an image is theoretically used as n images, which reduces the pressure of image acquisition at an earlier stage.

After the images are filtered by the ladder of the "multi-scale pyramid", the distance between each image in the "pyramid" has the properties of ($A = 0$) and ($\sigma = 1$) (A is the mean value, σ is the standard deviation). The feature recognition model is f_x, and a_x is any parameter of the function f_x, then the feature recognition model is considered to be the

CNN model at each resolution, and all model parameters within the model are shared at each scale by the principle of neuronal operation:

$$\theta_x = \theta_0, x \in \{1, 2, \dots, N\} \tag{3}$$

θ_0 is initial parameter.

In the range x, for any multiscale feature recognition model containing M phases as f_x, there exists:

$$f_x(X_x; \theta_x) = W_M H_{M-1} \tag{4}$$

W_m is the weight matrix of stage M, H_{M-1} is the output of stage $M-1$, $H_0 = X_x$. The output of the intermediate hidden stage m is:

$$H_m = pool(tanh(W_m H_{m-1} + b_l)) \tag{5}$$

The *pool* function represents the pooling operation, *tanh* is the activation function, W_m and b_l are weight matrices, a vector of parametric parameters, defined θ_x:

$$\theta_x = aW_m + bb_l, (a \in R, \ b \in R) \tag{6}$$

Lastly, linearizing and unifying the feature maps obtained from each CNN model generates a three-dimensional feature matrix F.

$$F = [f_1, u(f_2), \dots, u(f_N)] \tag{7}$$

u stands for sampling function. $N = 3$, as the "pyramid" in the passage has three levels.

In conclusion, the feature extraction process of video acquisition images on construction sites is obtained, as shown in Figure 5.
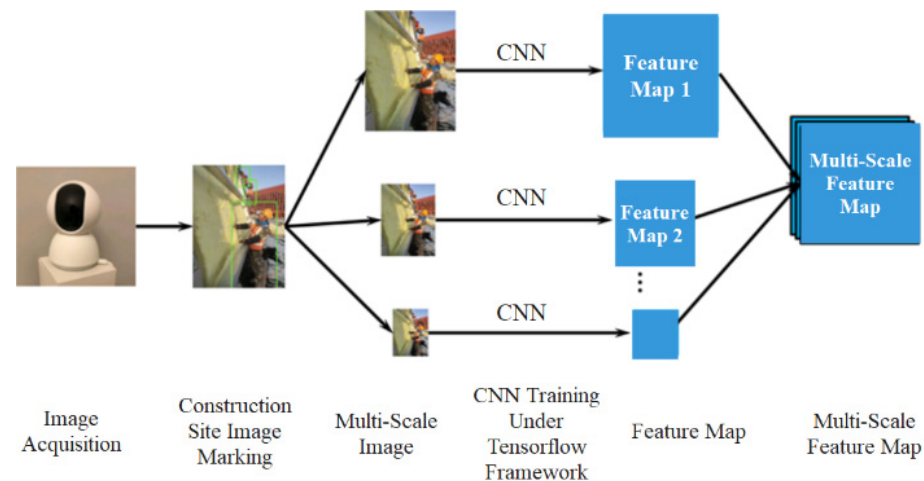

Figure 5. Multi-scale Layered Feature Extraction Process with Construction Personnel Image.

After labeling the acquired images of the construction site, multi-scale division of the image is achieved using a fast Local laplican filter, which is able to characterize the image at multiple scales and geometrically increase the image at the same time. Finally, the TensorFlow learning framework is used to train the features of the images at each scale to obtain a complete feature map, which is a generalized pixel-wise representation of construction personnel.

As can be seen from the multi-scale layered feature extraction process in the above figure, the more scales in Equation (4) $f_x = (X_x; \theta_x)$, the more capable the feature map obtained by this process will be to represent all the features of the actual effects of the construction site workers in the video.

3.4.3. Personnel Identification and Classification Model Construction

The "map" is used to classify each pixel in the image to be recognized, and the category to which most pixels in a particular image belong is calculated to realize the detection of construction site workers (as shown in Figure 6).
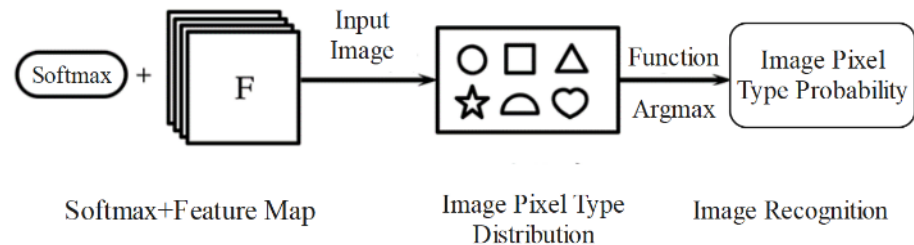

Figure 6. Identification Model of Construction Site Personnel.

"F" is a description of the characteristics of the workers at the construction site capture image obtained through a "Gaussian pyramid" transformation. The Softmax function is used to determine the pixel class distribution of the construction site surveillance images input to the detection model. Finally, through the linear regression effect of the argmax function, an overview and probability of the class to which the pixels in each region of the image belong is output to achieve accurate recognition of each part of the specific image.

The predicted distribution of each pixel of the image obtained by the classifier Softmax classifier is set to $g_{i,a}$. Then, for an image containing m pixels ($m > 0$), the distribution of the target classification to which each pixel belongs conforms to the following equation:

$$g_{m,a} = \frac{1}{X(m)} \sum_{i \in m} g_{i,a} \qquad (8)$$

where $X(m)$ is the entire set of pixels within that image, and the argmax function is processed to obtain the predicted class of each pixel in that particular image as follows:

$$P_K = \mathop{argmax}_{a \in classes} g_{m,a} \qquad (9)$$

P_K is the probability of the predicted outcome of element k in the image with respect to a particular class. The prediction results for each region of the image are combined to achieve recognition of the image.

4. Results

4.1. Test and Analysis Based on the CNN Model

In order to test whether the developed model meets the requirements, 850 images were separated from the video captured from this construction site. Meanwhile, in order to reduce the amount of test video processing and increase the number of samples, this paper selects 4000 images of construction site personnel that meet the requirements from the open-source MOCS (Moving Objects in Construction Sites) dataset and SODA (Site Object Detection Dataset) dataset to obtain a total of 4850 images as sample data for feature extraction and training of intrusion behavior. The constructed sample dataset is divided unequally, where 80% is divided into training sets and 20% is divided into test sets. After labeling the samples, the training is started using the TensorFlow learning framework on the basis of the downloaded weight data. The loss function used in this training is the Cross-entropy Loss, and the training error and test error are maintained at 0.03–0.06 after 200 rounds of training, which shows that the fitting of the model can meet the requirements. Specific test results are introduced into the confusion matrix for presentation, as shown in Table 1.

Table 1. Statistics of Image Recognition Results Collected at Construction Site.

Confusion Matrix		Predicted Value		Precision	Recall	F1 Score
		Positive	Negative			
True value	**Positive**	2809	423	0.8691	0.9314	0.8992
	Negative	207	1411			

The test results show that the model designed in this paper for personnel intrusion behavior has a Precision of 87%, Recall of 93%, and F1 score of 90%. It shows that the model has high accuracy of detection results, comprehensive detection targets, and high comprehensive performance, which can effectively identify the intrusion behavior of construction personnel. Meanwhile, the statistics of multiple tests show that the time to both collect and detect one frame is less than 0.05 s, which can provide accurate early safety warning for construction site personnel in real time.

4.2. Early Warning of Intrusion Behavior Based on CNN

In this paper, the hazard area adjacent to the building in the construction site is used as the object for the study of personnel intrusion behavior. The image was first collected and processed at the edge of the building, then the background of the image was modeled and the coordinate system was established (the origin of the coordinate system is located in the upper-left corner of the image, the horizontal direction to the right of the point is the positive direction of the X axis, the vertical direction of the point is the positive direction of the Y axis, and the scale was 8). The processing diagram is shown in Figure 7. According to the safety warning principle set above, the program will respond if intrusion of personnel into the hazard area is found. Then, it will issue a reminder or alarm according to the personnel attribute and automatically save the recognition image.
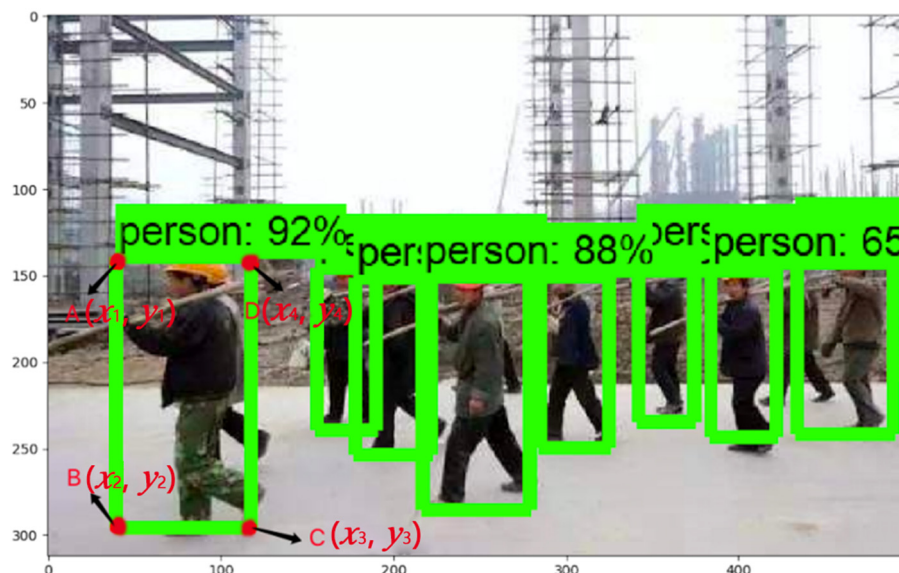
Figure 7. Automatic Identification of Hazard Area by Workers at A Construction Site.

The green box that frames each construction worker in Figure 7 is the "Personnel Identification Box" mentioned above. "Person: 92%" indicates that the objects in the box have a 92% probability of being workers. A, B, C, and D represent the four angular coordinates of the "Personnel Identification Box" in the context of this facade. For convenience, the leftmost edge of Figure 7 is assumed to be the edge of the hazard area.

As can be seen from Figure 8, the abscissa of the four corners of the leftmost person in Figure 7 are 3.20, 3.80, 3.98 and 5.27, respectively. Using Equation (1) and scale conversion, the person's center abscissa is $\left(\frac{3.20+3.80+3.98+5.27}{4}\right)/8 = 0.475$.

Figure 8. Coordinate Output of the Leftmost Construction Personnel Identification Box.

At this point, the straight-line distance between personnel and the hazard area is $\lceil 0.475 - 0 \rceil = 0.475 < 1$, thus triggering the voice warning "You have entered the regular hazard area, please pay attention to your safety", and the attribute information of this personnel "(T, (0.475, y), 09:23)" is output to the safety personnel data center.

Therefore, a warning of personnel intrusion behavior is completed, and the model saves the result and enters the next monitoring screen for analysis.

5. Discussion

This study combined early warning theory and CNN technology to carry out early safety warning research on intrusion behavior in construction hazard areas, which was mainly divided into two processes: "determination of early warning demand information" and "realization of early warning function".

The personnel attribute information collection system in the study has the advantages of stable identification of personnel and interactive storage of data. The BIM-RFID technology introduced by this system makes it possible to trace the later information and can realize automatic personnel information management. The environment of construction sites is complex and changeable; some special types of construction personnel need to operate in the hazard area at all times. Various factors may lead to deviation in the focus of safety management, and the traditional approach of forming a hazard area by expanding the origin of dangerous sources is difficult to cope with the site conditions. In fact, different types of construction personnel on the construction site most of the time in a different spatial state, in different environments, lead to different sources of danger faced by each type of work, and the different content and process of each type of work so that they are subject to a large difference in safety education. This means that the same construction personnel can have a high safety factor in the professional area, but it may be dangerous to enter the professional construction area of other work types. This situation can be regarded as the construction personnel entering the "hazard area". Therefore, this study can make up for the above problems to a certain extent by dividing the hazard areas of construction sites based on the working areas of each work type.

The key to the realization of the early warning function is to accurately capture the real-time behavior of construction personnel, and efficiently process the data information. The camera group is deployed to acquire images from all perspectives in the construction hazard area, and the acquired real-time videos are transmitted to the terminal in the form of frame images. The complex characteristics of the viewpoint arrays required for temporary characterization of construction personnel and identification of intrusions in hazard areas need to utilize tedious parameters for feature extraction. However, vision-based methods are difficult to identify unsafe behavior data captured using digital images and videos [24]. Therefore, this paper combines the Softmax function and the argmax function to achieve the extraction and classification of image features and uses the unique superiority of the

CNN model in image processing, which can effectively detect the intrusion behavior in the hazard area of the relevant personnel.

To summarize the full text, the main contents are as follows:

(1) A construction site personnel management system based on BIM-RFID was adopted to realize real-time detection of personnel attribute information on the construction site. According to industry norms and operating time characteristics, construction personnel were classified into three categories: civil construction personnel, mechanical operation personnel, and electrical engineering personnel. In combination with the safe operation scope of each professional construction personnel, three types of construction hazard areas, including civil construction hazard areas, machinery moving hazard areas, and power leakage hazard areas, were established.

(2) Based on personnel attributes and hazard area division, the rationale for construction intrusion behavior was elaborated. The center coordinates were determined according to the posture of the construction personnel, and the safety threshold was then compared with the linear distance (the distance between the personnel center coordinate point and the hazard area) and the personnel work area for a two-level differential warning.

(3) The method of multi-scale hierarchical feature extraction was used to obtain the feature map of construction personnel, and then the Softmax classification function and argmax function were employed to establish the recognition model. The model was applied to a specific project for empirical research, which shows that its determination of intrusion behavior of construction site personnel is relatively accurate and can achieve the function of prior warning.

6. Conclusions

On the basis of the characteristics of the early warning scene, this study analyzed the early warning demand information of intrusion behavior. From the perspective of hazard area division, the early safety warning model of personnel intrusion behavior was established by combining CNN to obtain feature mapping of images. Then, the feasibility of the application of the CNN model in engineering construction safety warnings is illustrated. The analysis reveals that the present model shows the advantages of real-time detection, lower data complexity, and efficient extraction and classification in construction early safety warning, and also theoretically provides a new perspective for the early safety warning research of intrusion behavior in the construction site.

Although the established model in this paper shows the prospect of realizing the early warning function through the verification of actual cases, there are some limitations on the determination of the "hazard area division" and "personnel attribute information" required to realize the early warning function because the hazard area and personnel attributes corresponding to different engineering projects and construction conditions may be different. At the same time, the images used in the study have a relatively high definition, and there are not many overlaps and occlusions of the targets, so the errors caused by the dense presence of targets such as occlusions are not considered. Therefore, if this model is adopted by a wide range of field practitioners, further research is needed on the information required for early warning of dangerous construction area intrusion and the accuracy of picture recognition containing complex personnel behavior.

Author Contributions: Conceptualization, Y.X. and M.L.; Methodology, J.Z. (Jinyu Zhao), W.Z. and Y.X.; Investigation, J.Z. (Jinyu Zhao), Y.X. and M.L.; Software, Y.X. and J.Z. (Jing Zhao); Validation, J.Z. (Jinyu Zhao) and W.Z.; Writing—original draft preparation, J.Z. (Jinyu Zhao), Y.X. and J.Z. (Jing Zhao); Writing—review and editing, J.Z. (Jinyu Zhao), W.Z. and M.L.; Supervision, J.Z. (Jinyu Zhao) All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. National Engineering Quality and Safety Supervision Information Platform. Ministry of Housing and Urban-Rural Development of the People's Republic of China (MOHURD). Available online: https://zlaq.mohurd.gov.cn/fwmh/bjxcjgl/fwmh/pages/default/index.html (accessed on 1 January 2023).
2. Heinrich, H.W. *Industrial Accident Prevention: A Scientific Approach*, 2nd ed.; McGraw-Hill Book Company: New York, NY, USA, 1941; pp. 354–396.
3. Wang, C. Research on the Subway Construction Safety Management Based on Worker's Unsafe Behavior. Master's Thesis, Tianjin University of Technology, Tianjin, China, 2014.
4. Gao, H.; Luo, H.B.; Fang, W.L. Methods of intrusion identification in hazardous areas based on computer vision. *J. Civ. Eng. Manag.* **2019**, *36*, 123–128. [CrossRef]
5. Fouda, M.; Taleb, T.; Guizani, M.; Nemoto, Y.; Kato, N. On supporting P2P-based VoD services over mesh overlay networks. In Proceedings of the GLOBECOM 2009—2009 IEEE Global Telecommunications Conference, Honolulu, HI, USA, 30 November–4 December 2009; pp. 1–6. [CrossRef]
6. Kim, K.; Kim, M. RFID-based location-sensing system for safety management. *Pers. Ubiquitous Comput.* **2012**, *16*, 235–243. [CrossRef]
7. Navon, R.; Kolton, O. Model for automated monitoring of fall hazards in building construction. *J. Constr. Eng. Manag.* **2006**, *132*, 733–740. [CrossRef]
8. Zheng, X.Z.; Wang, X.L.; Liu, H.L.; Sun, Z.G.; Guo, J. Real-time monitoring and early warning system for near-miss incidents of subway station construction. *J. Xi'an Univ. Sci. Technol.* **2019**, *39*, 589–596. [CrossRef]
9. Shuang, Q.; Zhang, Z. Determining critical cause combination of fatality accidents on construction sites with machine learning techniques. *Buildings* **2023**, *13*, 345. [CrossRef]
10. Nnaji, C.; Karakhan, A.A. Technologies for safety and health management in construction: Current use, implementation benefits and limitations, and adoption barriers. *J. Build. Eng.* **2020**, *29*, 101–212. [CrossRef]
11. Altman, E.I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J. Finance* **1968**, *23*, 589–609. [CrossRef]
12. Gu, H.B. Macroeconomic early warning research: Theory, method and history. *Econ. Theor. Bus. Manag.* **1997**, *4*, 1–7.
13. Guo, H.L.; Liu, W.P.; Zhang, W.S. A BIM-PT-integrated warning system for on-site workers' unsafe behavior. *Chin. Saf. Sci. J.* **2014**, *24*, 104–109. [CrossRef]
14. Zhao, T.S.; Xu, K.; Zhou, W. Graded management of hazardous area in construction site. *Ind. Saf. Environ. Prot.* **2018**, *44*, 43–46.
15. Teizer, J.; Cheng, T. Proximity hazard indicator for workers-on-foot near miss interactions with construction equipment and georeferenced hazard areas. *Autom. Constr.* **2015**, *60*, 58–73. [CrossRef]
16. Kim, H.; Lee, H.S.; Park, M.; Chung, B.; Hwang, S. Automated hazardous area identification using laborers' actual and optimal routes. *Autom. Constr.* **2016**, *65*, 21–32. [CrossRef]
17. Maalek, R.; Sadeghpour, F. Accuracy assessment of ultra-wide band technology in locating dynamic resources in indoor scenarios. *Autom. Constr.* **2016**, *63*, 12–26. [CrossRef]
18. Vahdatikhaki, F.; Hammad, A. Risk-based look-ahead workspace generation for earthwork equipment using near real-time simulation. *Autom. Constr.* **2015**, *58*, 207–220. [CrossRef]
19. Li, H.; Dong, S.; Skitmore, M.; He, Q.H.; Yin, Q. Intrusion warning and assessment method for site safety enhancement. *Saf. Sci.* **2016**, *84*, 97–107. [CrossRef]
20. Dong, S.; Li, H.; Skitmore, M.; Yin, Q. An experimental study of intrusion behaviors on construction sites: The role of age and gender. *Saf. Sci.* **2019**, *115*, 425–434. [CrossRef]
21. Guo, H.L.; Yu, Y.T.; Skitmore, M. Visualization technology-based construction safety management: A review. *Autom. Constr.* **2017**, *73*, 135–144. [CrossRef]
22. Xing, Z.X.; Gu, H.L.; Wei, Z.G.; Qian, H.; Zhang, Y.; Wang, L.J. Contrastive Study of the Pedestrian Head Detection Method Based on Convolutional Neural Network. *Saf. Environ. Eng.* **2019**, *26*, 77–82. [CrossRef]
23. Guo, Y.; Su, P.F.; Wu, Y.F.; Guo, J. Object detection and location of robot based on Faster R-CNN. *J. Huazhong Univ. Sci. Technol. (Nat. Sci. Ed.)* **2018**, *46*, 55–59. [CrossRef]
24. Wei, R.; Love, P.; 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*, 21–29. [CrossRef]
25. 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]

26. Wang, W.; Liu, S.K.; Zhang, Y.G.; Zhao, C.N.; He, H.G. Research on early warning of intrusion into hazardous construction area based on bim and machine vision technology. *Saf. Environ. Eng.* **2020**, *27*, 196–203. [CrossRef]

27. Awolusi, I.; Marks, E.D. Active Work Zone Safety: Preventing accidents using intrusion sensing technologies. *Front. Built Environ.* **2019**, *5*, 21. [CrossRef]

28. Teizer, J.; Allread, B.S.; Fullerton, C.E.; Hinze, J. Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system. *Autom. Constr.* **2010**, *19*, 630–640. [CrossRef]

29. Forsythe, P. Proactive construction safety systems and the human factor. *Proc. Inst. Civ. Eng. Geotech. Eng. Manag. Procure. Law* **2014**, *167*, 242–252. [CrossRef]

30. Arslan, M.; Cruz, C.; Ginhac, D. Visualizing intrusions in dynamic building environments for worker safety. *Saf. Sci.* **2019**, *120*, 428–446. [CrossRef]

31. Huang, Y.; Hammad, A.; Zhu, Z. Providing proximity alerts to workers on construction sites using Bluetooth low energy RTLS. *Autom. Constr.* **2021**, *132*, 103928. [CrossRef]

32. Sakhakarmi, S.; Park, J. Improved intrusion accident management using haptic signals in roadway work zone. *J. Saf. Res.* **2022**, *80*, 320–329. [CrossRef] [PubMed]

33. Hubel, D.H.; Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **1962**, *160*, 106–154. [CrossRef]

34. Li, H.M.; Duan, P.S.; Meng, H.; Guo, H.D. Study on safety early-warning assessment of damaged steel structure reconstruction based on ipso-bp. *J. Saf. Sc. Technol.* **2019**, *15*, 174–180. [CrossRef]

35. Xiao, H.H.; Shi, J.L. Video captioning based on C3D and visual elements. *J. South Chin. Univ. Technol. (Nat. Sci. Ed.)* **2018**, *46*, 88–95. [CrossRef]

36. Yang, L.Q.; Cai, L.Q.; Gu, S. Detection on wearing behavior of safety helmet based on machine learning method. *J. Saf. Sci. Technol.* **2019**, *15*, 152–157. [CrossRef]

37. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; pp. 1–6. [CrossRef]

38. Chauhan, R.; Ghanshala, K.K.; Joshi, R.C. Convolutional neural network (CNN) for image detection and recognition. In Proceedings of the 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 15–17 December 2018; pp. 278–282. [CrossRef]

39. Agarap, A.F. An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification. *arXiv* **2017**, arXiv:1712.03541. [CrossRef]

40. Jahanshahi, M.R.; Masri, S.F.; Padgett, C.W.; Sukhatme, G.S. An innovative methodology for detection and quantification of cracks through incorporation of depth perception. *Mach. Vis. Appl.* **2013**, *24*, 227–241. [CrossRef]

41. Zhu, Z.; Park, M.W.; Koch, C.; Soltani, M.; Hammad, A.; Davari, K. Predicting movements of onsite workers and mobile equipment for enhancing construction site safety. *Autom. Constr.* **2016**, *68*, 95–101. [CrossRef]

42. 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]

43. Lin, P.; Wei, P.C.; Fan, Q.X.; Chen, W.Q. CNN model for mining safety hazard data from a construction site. *J. Tsinghua Univ. (Sci. Technol.)* **2019**, *59*, 628–634. [CrossRef]

44. Xiong, W.; Xu, X.; Chen, L.; Yang, J. Sound-based construction activity monitoring with deep learning. *Buildings* **2022**, *12*, 1947. [CrossRef]

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

46. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]

47. 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]

48. Shahverdy, M.; Fathy, M.; Berangi, R. Driver behavior detection and classification using deep convolutional neural networks. *Expert Syst. Appl.* **2020**, *149*, 122290. [CrossRef]

49. Lung, L.W.; Wang, Y.R. Applying deep learning and single shot detection in construction site image recognition. *Buildings* **2023**, *13*, 1074. [CrossRef]

50. Lai, X.Y.; Zhang, M.G.; Xu, S. Influence of safety attitude and its components on construction workers' safety behaiour. *J. Civ. Eng. Manag.* **2019**, *36*, 74–80. [CrossRef]

51. Liu, W.P. The Schematic Studies of Construction Accident Warning System Based on BIM and Positioning Technology. Master's Thesis, Tsinghua University, Beijing, China, 2015.

52. Zhao, J.; Zhao, J.Y.; Zhang, S.K. Research on construction site personnel management system based on BIM-RFID. *Value Eng.* **2019**, *38*, 12–14. [CrossRef]
53. Paris, S.; Hasinoff, S.W.; Kautz, J. Local laplacian filters: Edge-aware image processing with a laplacian pyramid. *ACM Trans. Graph.* **2011**, *30*, 68. [CrossRef]