

Research Article

An Intelligent Vision-Based Method of Worker Identification for Industrial Internet of Things (IoT)

Shuai Wang,^{1,2} Xiaoyu Li,¹ Wei Chen ,^{1,3} Weiqiang Fan ,¹ and Zijian Tian¹

¹School of Mechanical Electronic & Information Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China

²Department of Inner Mongolia, Administration of Coal Mine Safety Ordos Division, Ordos 017000, China

³School of Computer Science & Technology, China University of Mining and Technology, Xuzhou 221116, China

Correspondence should be addressed to Wei Chen; chenwdavior@163.com

Received 22 October 2021; Revised 20 December 2021; Accepted 3 January 2022; Published 27 January 2022

Academic Editor: Chi-Hua Chen

Copyright © 2022 Shuai Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of Internet of things (IoT) and computer vision (CV), the application of combining the IoT platform and CV technology to monitor the worker safety has attracted more and more attention in the field of industrial information. Worker identification is a prerequisite for safety management in industrial production, and safety helmet can not only protect worker's head from accidental injuries but also help to identify the work types of workers through different colors. Therefore, this study proposes an intelligent method for worker identification based on moving personnel detection and helmet color characteristics. First, the motion objects that contain personnel and nonpersonnel are detected by the Gaussian mixture model (GMM) and extracted to generate the region of interest (RoI) images. Then, the multiple-scale histogram of oriented gradient (MHOG) features of the RoI images are extracted, and the personnel images are identified by the support vector machine (SVM). Third, the workers' head images are obtained by the OpenPose model and personnel mask, and the GoogLeNet-based transfer learning network is established to extract the head images features and realize worker identification. This method is tested on our dataset, and the average accuracy of worker identification for multiple helmet color combinations reaches 99.43%, which is robust to workers' angle, scale, and occlusion.

1. Introduction

Safe production is the top priority of modern industrial development. With the advancement of Industry 4.0 era, it is an inevitable trend to use intelligent vision-based technology to identify worker information, and then feed it back to the data center through Internet of things (IoT) for analysis and processing. This not only automatically improves the efficiency of human work but also strengthens the guarantee of workers' life safety. In recent years, the vision-based technology of worker safety monitoring has shown significant advantages in industrial production and has played an important role in promoting the development of industrial IoT. Despite this, the industrial sectors report thousands of casualties every year, and a large proportion of the accidents are caused by falling objects [1]. In order to protect workers

from being struck by falling objects, wearing helmet is the most convenient and effective measure of safety protection [2]. Therefore, to strictly require workers to wear helmets, a series of rules and regulations have been formulated by relevant security departments, especially for construction sites, coal mines, oil fields, factories, and other places.

From the above introduction, we know that personnel of any identity must wear safety helmets as long as they enter the production site. So, the helmet not only represents the existence of worker in a specific area but also reflects the different work types and worker identities in a certain industry. For example, in the power industry, white helmet represents the leader, blue helmet represents the management personnel, yellow helmet represents the construction worker, and red helmet represents the outsider. In the construction industry, red helmet represents the leader,

yellow helmet represents the ordinary worker, blue helmet represents the technician, and white helmet represents the manager or safety supervisor. In the coal mine industry, red helmet represents the technician or leader, yellow helmet represents the security worker, and black helmet represents the production worker. It can be seen that for each industry, the worker identities can be directly identified through their helmet colors, which is of great significance to the safety monitoring management and personnel scheduling of the enterprise.

To enhance workers' self-protection consciousness and ensure their personal safety, intelligent identification of safety helmet wearing is crucial for improving the safety management quantity of factories and supervisors [3]. At the same time, compared with traditional labor-intensive and time-intensive management methods, intelligent vision-based identification is helpful to promote the automation of industrial production. In the past years, some researchers have developed related technologies to identify safety helmets. Since the human face is the most prominent part near the helmet and the application of face recognition has become more and more mature recently, some scholars studied the discrimination algorithm about helmet wearing through face positioning. For instance, Chen et al. [4] introduced a safety monitoring system based on multiattribute face recognition to meet the safety detection of intelligent factory. Zhang et al. [5] determined whether the head wears a helmet by locating the upper area of the face and extracting its features of color and shape. Although the face-based methods have strong robustness and high accuracy for helmet detection, they cannot identify the helmet when the worker's back facing the camera or his face is obscured. Currently, with the rapid development of the deep learning (DL) and computer vision (CV), algorithms that directly rely on the convolutional neural network (CNN) for helmet recognition and detection are emerging. Huang et al. [6] used the YOLOv3 model to replace manual monitoring of site safety regulations to detect safety helmet. Zhou et al. [7] proposed a helmet detection algorithm based on the attention mechanism (AT-YOLO) for objects with small scales and obstructions. Wang et al. [8] presented a novel SHW detection model on the basis of improved YOLOv3 to heighten the capability of target detection on the construction site. Li et al. [9] developed a method based on the SSD-MobileNet algorithm for real-time detection of a safety helmet at the construction site. Wang et al. [10] introduced an approach to train and evaluate 8 DL detectors based on YOLO architectures, and samples include 4 colors of helmets, person, and vest. Chen and Demachi [11] provided a novel solution to identify improper use of personal protective equipment (PPE) by combining the DL detector and individual detector that uses geometric relationships analysis. Xiong and Tang [12] presented an extensible pose-guided anchoring framework for detecting proper use of PPE. Wu et al. [13] designed a one-stage system based on CNN to automatically monitor whether construction personnel are wearing hardhats and identify the corresponding colors. Fang et al. [14] used the high

precision, high speed, and widely applicable Faster R-CNN method to detect construction workers' NHU in different construction site conditions. Long et al. [15] presented a DL approach for accurate safety helmets wearing detection by employing a single shot multibox detector. It can be seen from the above research that the integration of CV and DL technology has achieved remarkable results in the industrial field in recent years and has been maturely used in the application of helmet detection and recognition. However, for places with poor lighting and environmental conditions, such as underground coal mines, workers are not clear and sample collection is difficult. Especially for small sample problems, the poor fitting effect of the CNN-based training model leads to unsatisfactory detection results. In addition, for the image played by the video frame, the CNN method has high requirements on the detection time and equipment performance, so it has a higher application cost.

In order to avoid the shortcomings of the discussed methods, Mneymneh et al. [16] created an integrated framework that can automatically and efficiently detect any noncompliance with safety rules, which is achieved by isolating construction workers from the captured scene and then detecting the hardhat in the identified region of interest (RoI). Shine and Jiji [17] designed an automated system to identify motorcyclists without a helmet from traffic surveillance videos, which uses a two-stage classifier to extract motorcycles from surveillance videos, and the detected motorcycles are further fed to a helmet identification stage. Yogameena et al. [18] also presented a system for automatically detecting the motorcyclists with and without safety helmets, which segment the foreground object and detect the motorcycle, and then identify the motorcycle with and without helmet. Wu and Zhao [19] put forward an intelligent vision-based approach that used motion detection and pedestrian detection as the foundation technique to identify safety helmet. This kind of stage-based methods are not easily affected by the factors such as scale changes, sample size, and the posture of helmet carrier, so it has better practicability and higher identification efficiency compared with the methods previously discussed.

It is worth noting that although the approach in [19] is a cost-effective helmet identification algorithm, the usage of conventional histogram of the oriented gradient (HOG) descriptor will not only increase the number of redundant calculations but also cannot express the overall features of the personnel comprehensive. In addition, the layer-by-layer classification efficiency of hierarchical support vector machines (H-SVM) is relatively low. In view of the advantages of stage-based methods and the deficiencies of [19], this study proposes an algorithm for worker identification based on moving personnel detection and helmet characteristics. This can not only detect whether a worker wears a safety helmet on his head but also identify different worker identities through the helmet colors, and so, it has important research significance for the worker safety management in industrial production sites.

The major contributions of this study can be described as follows:

- (1) Compared with popular pure CNN-based methods, this research provides a simple and easy to implement scheme for worker identification, which can be applied to a variety of application scenarios and small sample datasets. In addition, it is more effective and robust for multiscale workers, angle changes, and occlusions.
- (2) We have designed a multiscale histogram of oriented gradient (MHOG) descriptor to extract features of RoI images, which can reduce the calculation complexity of single-scale HOG descriptor while extracting comprehensive information of RoI images
- (3) We have provided a method of locating the worker's head by the OpenPose model, which is simple and effective, and is not affected by worker's posture
- (4) We have proposed a method of head images identification based on the transfer learning network, which can identify different head samples at one time with a small number of training samples.

The rest of this study is organized as follows: Section 2 presents the overall architecture of the worker identification algorithm. Section 3 describes the methodology of personnel identification. Section 4 describes the methodology of helmet identification. Section 5 shows and discusses the experimental results. Section 6 concludes the current work and direction for future work.

2. Framework of the Worker Identification

This section presents the overall architecture of the proposed algorithm, as shown in Figure 1. The architecture is mainly composed of three parts: motion extraction, personnel identification, and helmet identification.

2.1. Motion Extraction. Extraction of foreground objects is necessary for many advanced image analysis tasks.

The application scenario of this study aims at the static background video image, and the motion object is the foreground object we are interested in this section. The motion detection is to extract objects that have changed spatial positions in a video. Over the years, scholars have proposed many different motion detection methods, mainly including the frame difference method, optical flow method, background subtraction, feature matching method, KNN, and variants of these methods. Among them, the background difference method is easy to implement, has fast detection speed and high accuracy, and can provide more comprehensive features of the motion area. The Gaussian mixture model (GMM) is the representative of the background subtraction method, which can get relatively ideal detection result. The process of motion extraction is shown in Figure 2; the GMM is first used to detect the motion objects in the video frames. Then, the morphological operations are applied to the resulting motion mask to eliminate noise. Finally, to further optimize the

detection result of motion objects, blob analysis is adopted to detect groups of connected pixels that are likely to correspond to moving objects. So far, the motion mask has been extracted. In order to convenient the subsequent processing of the motion image, the detected motion pixel coordinates are used as the benchmark to draw rectangular region on the original frames, and the RoI images of motion extraction are obtained with the size of $256 \text{ pixels} \times 128 \text{ pixels}$.

2.2. Personnel Identification. Since not all the motion objects extracted by GMM are personnel, the task in this section aims to separate personnel objects from nonpersonnel objects, which can be treated as the classic issue of pedestrian detection. Considering that there may exist problems of poor lighting conditions and uneven illumination in the industrial environment, to more accurately identify personnel, the preprocessing operation of brightness [20] is performed on the RoI images before personnel identification.

HOG feature [21] is the most widely used pedestrian feature descriptor; it has achieved a detection success rate of nearly 100% on the MIT pedestrian database and has reached a detection rate of approximately 90% on the INRIA pedestrian database that includes changes in perspective, illumination, and background. Detailed and comprehensive features are more helpful to distinguish personnel from nonpersonnel; this study designs multiple block scales [22] to extract MHOG vectors that contain global features and local features. After that, the SVM classifier [23] is trained through MHOG vectors to classify the RoI images, and only the positive sample images are reserved.

2.3. Helmet Identification. The helmet identification is the core content of worker identification in this article, and the wearing of helmets directly reflects the workers' identities and their safety status. The focus in this section is to extract and identify the workers' head images, so as to realize the worker identification by helmet features. In order to facilitate the extraction and analysis of helmet features, the first step is to extract the workers' head images using the OpenPose model [24] and personnel mask and then resize the head images with the size of $64 \text{ pixels} \times 64 \text{ pixels}$. Furthermore, a transfer learning network based on GoogLeNet [25] is established to extract the head image features and identify whether workers wear helmets and the helmet colors.

3. Methodology of Personnel Identification

3.1. Multiscale Histogram of Oriented Gradient (MHOG). For the original HOG, when a block with the step size of one cell scans the image, it will cause any adjacent blocks to overlap each other, which means that redundant information will be brought into the final HOG feature vectors. However, if there is no overlap between the scan windows, the gradient histogram information on the spatial structure between adjacent blocks cannot be obtained, so the HOG features of the RoI image will be incomplete. For this reason, this section will introduce the MHOG descriptors to solve the problem. The core idea of MHOG is to decompose the

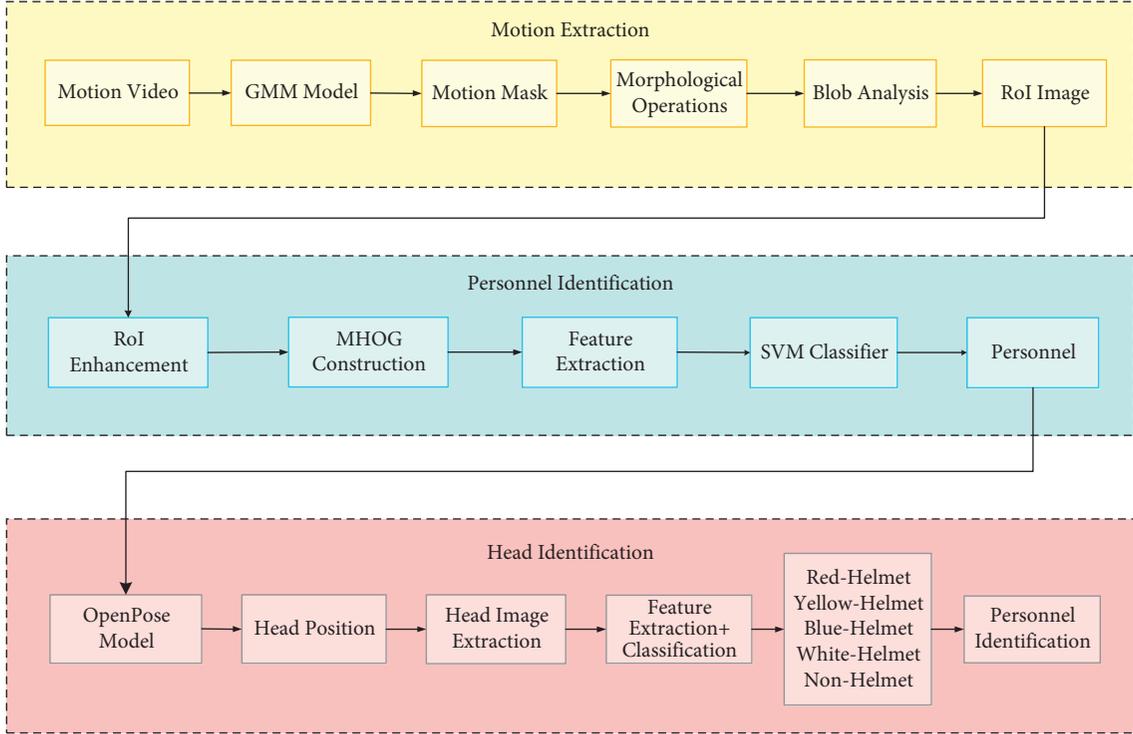


FIGURE 1: The architecture of the worker identification algorithm.

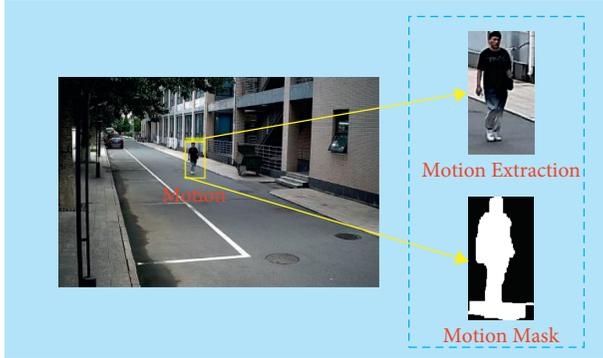


FIGURE 2: Motion extraction using the GMM model.

image in the form of a variety of different block scales, the large-scale HOG represents the global overall features of the object, and the small-scale HOG represents the local detail features of the object; then, the HOG feature vectors at each scale are connected to generate the MHOG feature descriptor.

The theoretical idea of MHOG is shown in Figure 3. The image is divided into blocks of different scales, and the cell size is a quarter of the block, that is, its length and width are all halved on the block. It can be seen that for an image with the size of 128 pixels \times 64 pixels, if the block size of scale 1 is set as its own size of 128 pixels \times 64 pixels, the cell number is 4 and the cell size is 64 pixels \times 32 pixels. We specify that the proportional relationship between two adjacent block scales is that the length and width of the next block are half of the previous block, so the block size of scale 2 is 64 pixels \times 32 pixels, whose cell number is 16, cell size is 32 pixels \times 16 pixels; the block size of scale 3 is 32 pixels \times 16 pixels, cell

number is 64, cell size is 16 pixels \times 8 pixels, and so on. Since each cell of HOG produces a 9-bin histogram, the image at scale 1 generates 36 feature vectors, which is defined as $F_i^{\text{Scale}1}$ ($i = 1, 2, \dots, 36$); the image at scale 2 generates 144 feature vectors, which is defined as $F_j^{\text{Scale}2}$ ($j = 1, 2, \dots, 144$); the image at scale 3 generates 576 feature vectors, which is defined as $F_k^{\text{Scale}3}$ ($k = 1, 2, \dots, 576$); and the image at scale 4 generates 2304 feature vectors, which is defined as $F_l^{\text{Scale}4}$ ($l = 1, 2, \dots, 2304$).

We know that the cell size of scale 3 and scale 4 is 16 pixels \times 8 pixels and 8 pixels \times 4 pixels, respectively, and 8 pixels \times 8 pixels cell size of HOG is between scale 3 and scale 4. Obviously, MHOG feature dimensions at scale 3 and even scale 4 are all smaller than 3780 of HOG; moreover, the total dimensions number (3060) of MHOG at the four scales is also less than 3780 of HOG. Therefore, we take the connection result of the feature vector at four scales as the feature vector F_m^{MHOG} of MHOG, which can be expressed as

$$F_m^{\text{MHOG}} = [F_i^{\text{Scale}1}, F_j^{\text{Scale}2}, F_k^{\text{Scale}3}, F_l^{\text{Scale}4}, \dots], \quad (1)$$

where m is the highest scale level of 4, and i, j, k , and l are the feature dimensions at each scale, respectively.

3.2. Support Vector Machine (SVM). As a representative algorithm in the machine learning field, SVM has developed rapidly since it was proposed and has achieved excellent results in many application scenarios. SVM is a two-class classifier model, whose task in worker identification is to separate F_m^{MHOG} of positive samples (personnel images) from the negative samples (nonpersonnel images). The basic idea of SVM learning is to solve the separation hyperplane that

can correctly divide the training dataset and has the largest geometric interval. As shown in Figure 4, w is the normal to the hyperplane, $\|w\|$ is the Euclidean norm, $|b|/\|w\|$ is the perpendicular distance from the hyperplane to the origin, and $w \cdot x + b = 0$ is the separating hyperplane. For a linearly separable dataset, there are infinitely many such hyperplanes, but the separating hyperplane with the largest geometric interval is the only one. The solution method of SVM is as follows:

Given, a training dataset on the feature space $T = \{(v_n, l_n)\}$, where v_n denotes the number of MHOG features, and $l_n \in \{1, -1\}$ denotes the two class labels of personnel and nonpersonnel. First step aims to construct and solve the convex quadratic programming problem:

$$\min_{\alpha} \frac{1}{2} \sum_{a=1}^n \sum_{b=1}^n \alpha_a \alpha_b l_a l_b (v_a \cdot v_b) - \sum_{a=1}^n \alpha_a, \quad (2)$$

$$\text{s.t. } \sum_{a=1}^n \alpha_a l_a = 0. \quad (3)$$

From this, the optimal solution $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*)^T$ can be obtained. Second step aims to calculate w^* and b^* from the following formula:

$$b^* = l_b - \sum_{a=1}^n \alpha_a^* l_a (v_a - v_b). \quad (4)$$

The third step is to achieve the solution of the separating hyperplane as follows:

$$w^* \cdot x + b^* = 0. \quad (5)$$

The classification decision function is

$$f(x) = \text{sign}(w^* \cdot x + b^*). \quad (6)$$

4. Methodology of Helmet Identification

4.1. Head Image Extraction. The OpenPose model [24] is a human posture recognition model developed by Carnegie Mellon University (CMU) based on CNNs and supervised learning and can realize posture estimation of human movements, facial expressions, and limb movements. It is suitable for single and multiple people and has excellent robustness. In order to accurately locate the worker's head, OpenPose is used to estimate the joint points of the worker body, and the joint point related to the head is extracted. In general, each worker has a set of joint points, but when two workers are severely occluded, sometimes only the joint points of the worker who is not occluded can be obtained. Therefore, it is necessary to perform detailed analysis on the worker images with one set of joint points. The schematic diagram of the human body with 18 joint points by OpenPose is shown in Figure 5, the right side of the figure lists the name of the human body organ corresponding to the number of each joint point.

For the worker image with one set of joint points, consider that different angles and postures of the worker

may cause symmetrical joints (such as the shoulders, wrists, and knees) to be hidden to varying degrees, and the absence of front face makes head joint points undetectable. Hence, as shown in Figure 6(a), only the position of worker neck (no.1 joint point) is used as a reference, mark this point as p_1 and make a horizontal straight line through it, locate the highest point p_2 of the worker mask in the vertical direction passing through the no. 1 joint point, and make a horizontal straight line through point p_2 ; these two horizontal straight lines form a horizontal band-shaped area. Then, search for the two outermost points p_3 and p_4 on the left and right sides of the worker mask in the upper half of the horizontal strip area and draw two straight lines in the vertical direction through these two points to form a vertical area. Next, analyze the changing characteristics of the worker mask boundary in the upper half of the common area where the two bands intersect. If there is only one highest point on the mask boundary and the highest point in the horizontal direction only falls but does not rise along both sides of the point p_2 , the image is regarded as a true single-worker image. As shown in Figure 6(b), the worker's head image is extracted from the blue dotted box formed by points of p_1 , p_2 , p_3 , and p_4 shown in Figure 6(a). If the highest point in the horizontal direction first falls and then rises along one side of point p_2 , the image is regarded as a double-worker occlusion image. As shown in Figure 7(a), the lowest point that falls in the middle is marked as p_0 , and the highest point that rises later is marked as p_2' . Draw a virtual straight line through point p_2' in the vertical direction and take the line segment $p_1'p_2'$ on this straight line so that the length of $p_1'p_2'$ is equal to p_1p_2 ; thereby, the position of p_1' point is obtained. In addition, take the midpoint p_5 of points p_2 and p_2' in the horizontal direction and move p_5 to the horizontal line where point p_4 is located. As shown in Figure 7(b), the head image of the worker who is not blocked is extracted from the blue dotted box formed by points p_1 , p_2 , p_4 , and p_5 shown in Figure 7(a), and the head image of the worker who is blocked is extracted from the red box frame formed by points p_1' , p_2' , p_3 , and p_0 shown in Figure 7(a).

For the worker image with two sets of joint points, it also contains two situations: one situation is that the bodies of the two workers are blocking each other, but their heads are independent of each other, and the other situation is that the bodies and heads of the two workers are both blocking each other. First, select one of the no. 1 joint points p_1 , and use the method of true single-worker image shown in Figure 6 to locate the other three points p_2 , p_3 , and p_4 . If the change characteristics of the highest point p_2 consist with the true single-worker image, the method shown in Figure 6 is used to extract the two head images, respectively. If the change characteristics of the highest point p_2 consist with the double-worker occlusion image, the method shown in Figure 6 is used to locate the two sets of points p_1 , p_2 , p_4 and p_1' , p_2' , p_3 , and the method shown in Figure 7 is used to determine points of p_0 and p_5 . Finally, the two head images are extracted by the two sets of points p_1 , p_2 , p_4 , p_5 and p_1' , p_2' , p_3 , p_0 separately.

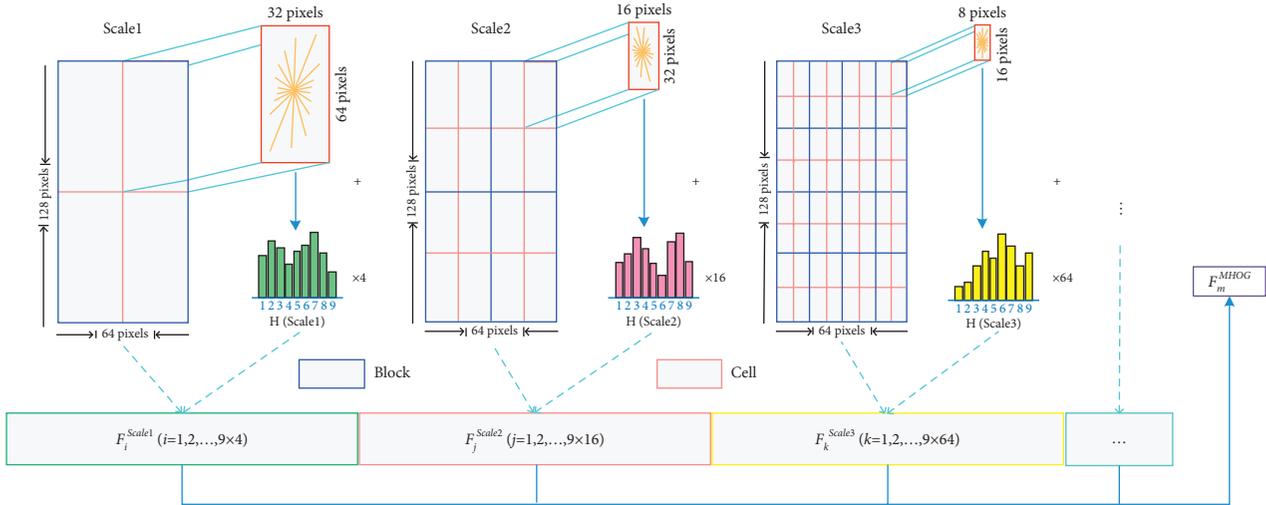


FIGURE 3: The extraction procedure of MHOG features.

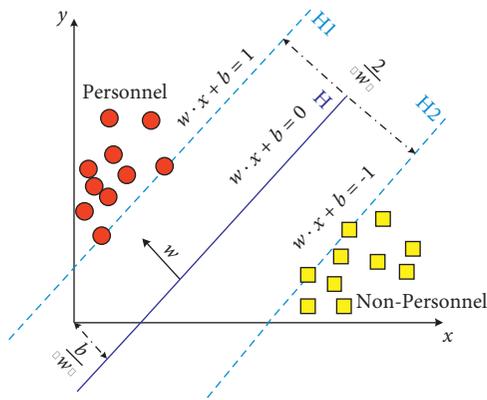


FIGURE 4: Support vector machine.

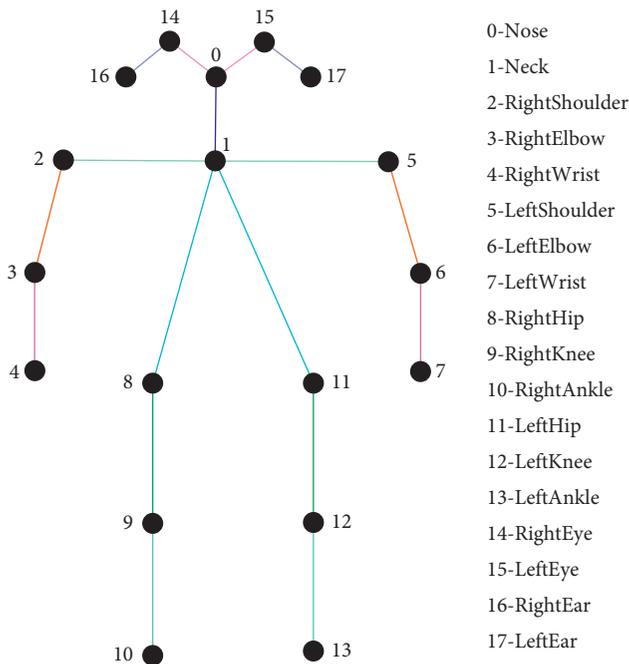


FIGURE 5: Structure diagram of human body joint point.

4.2. Head Image Identification. In this section, the problem of head image identification is achieved by the CNN network, which inputs the head images and outputs their labels together with classification probabilities. Specifically, the strategy of transfer learning is used to classify and identify the workers' head images by fine tuning the pretrained GoogLeNet. GoogLeNet has been trained on over a million images and can classify images into 1000 object categories, so it has learned rich feature representations for a wide range of images. Fine-tuning a network with transfer learning is usually much faster and easier than training a network with randomly initialized weights from scratch, which can help to quickly transfer learned features to the task of head images identification using a smaller number of training images.

The process of transfer learning based on GoogLeNet is shown in Figure 8. First, according to the requirements of the GoogLeNet input layer, the head images are uniformly adjusted to the size of $224 \times 224 \times 3$, and the training dataset and the testing dataset are allocated by a certain ratio. Then, the final two layers, "loss3-classifier" and "output," in GoogLeNet should be replaced with new layers adapted to the head image dataset. The original fully connected layer is replaced with a new fully connected layer that has number of outputs equal to the number of classes. In addition, to make learning faster in the new layers than transferred layers, the parameter values of WeightLearnRateFactor and BiasLearnRateFactor of the fully connected layer are increased. Next, the appropriate values of training option should be set to train the network. For transfer learning, the features from the early layers of the pretrained network (transfer layer weights) are retained. By setting the initial learning rate to a small value, the learning speed in the transfer layers is slowed down. When training the transfer learning network, one epoch is a complete training cycle for the entire training dataset, so there is no need to set a too large epoch and minibatch size. Finally, the trained network is used to predict and classify the testing dataset, the network will

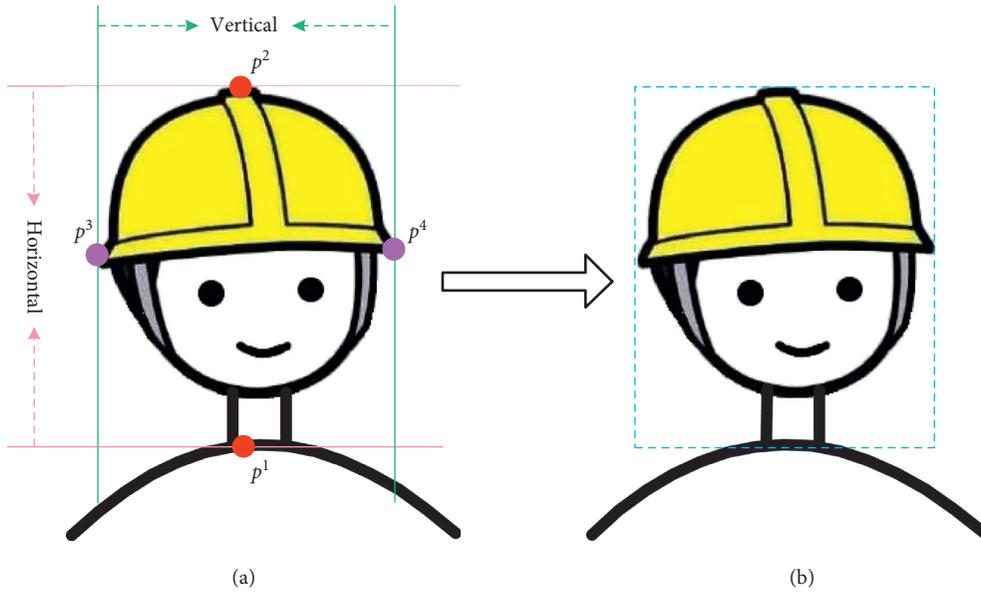


FIGURE 6: Head image extraction of one set of joint points. (a) Positioning of key points on the head. (b) Head extraction template.

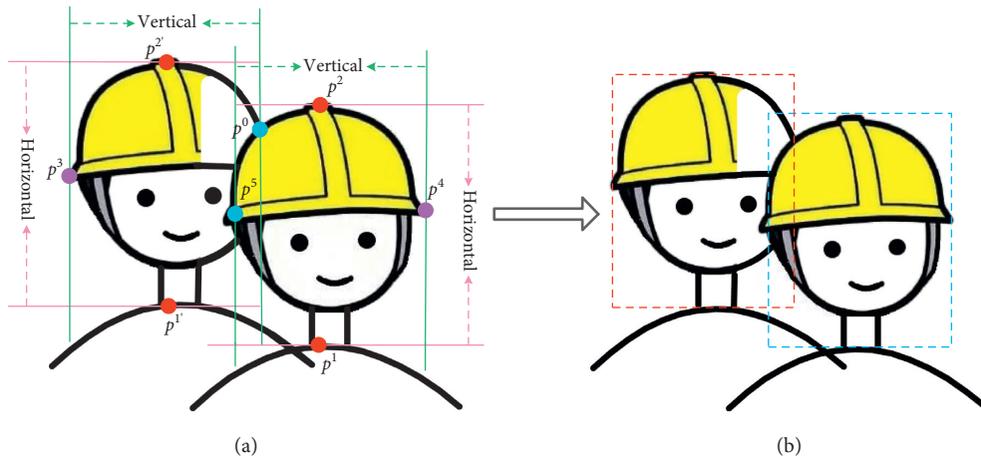


FIGURE 7: Head image extraction of two sets of joint points. (a) Positioning of key points on the head. (b) Head extraction template.

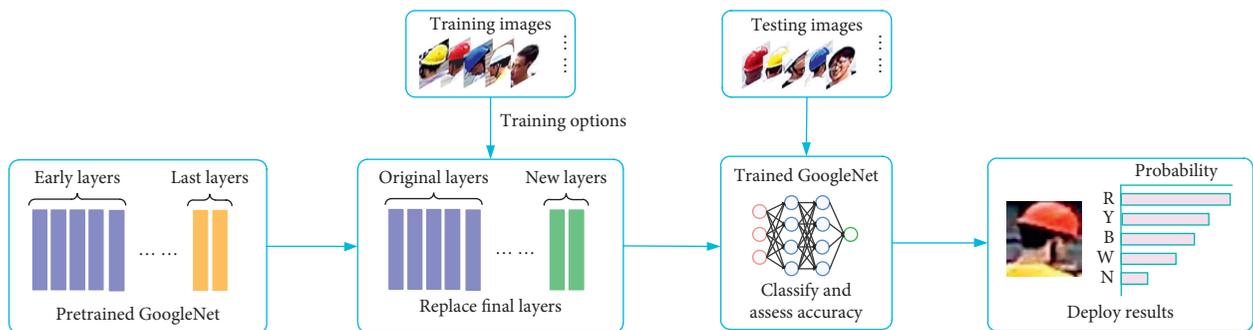


FIGURE 8: Flowchart of transfer learning.

output a classification probability value corresponding to each label for each test sample, and the label corresponding to the maximum probability value is the identification result of the test sample.

5. Experimental Results and Discussion

5.1. Personnel Identification Experiment. The dataset of personnel identification is self-collected motion videos with the size of 1920 pixels \times 1080 pixel, which contains the aspects of different light intensity, distance, angle, and occlusion. Figure 9 shows a small part of the motion objects, in which the personnel images are regarded as the positive samples, and the nonpersonnel images are regarded as the negative samples. We selected a total of 2365 samples from the dataset for experiment, of which there are 1830 positive samples and 535 negative samples.

Table 1 provides the detailed quantity distribution of personnel images in the four aspects of illumination, distance, occlusion, and angle. The illumination conditions include two types of light and dark, the distance includes three types of far, near, and medium, the occlusion includes three types of light, heavy, and no occlude, and the angle includes three types of front, back, and side.

The accuracy of worker identification will directly affect the accuracy of helmet identification. In order to obtain better worker identification results, the data samples are set to three distribution ratios for experimental comparison, namely, 60% training and 40% testing, 70% training and 30% testing, and 80% training and 20% testing. Meanwhile, to improve the performance of the classifier, the automatic hyperparameter optimization method is adopted to train the SVM.

In this study, three conventional indicators [26] of precision, recall, and accuracy are used to evaluate the performance of the proposed algorithm, whose calculation formulas are as follows.

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (8)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}, \quad (9)$$

where TP represents the positive sample that is correctly identified, TN represents the negative sample that is correctly identified, FP represents the positive sample that is incorrectly identified, and FN represents the negative sample that is incorrectly identified.

Table 2 provides the results of personnel identification with the method of MHOG + SVM; it can be seen that the three indicators that corresponding to MHOG (6, 4) perform best, and the recall of 99.73% means that the positive samples can basically be correctly identified. At the same time, in order to better illustrate the performance of MHOG, Table 3 provides the identification results of the HOG + SVM method in the (6, 4) allocation mode for comparison.

Obviously, the precision and accuracy of the MHOG (6, 4) mode and MHOG (7, 3) mode are better than that of HOG + SVM, and recall of the MHOG (6, 4) is equal to that of the HOG (6, 4). This result proves that MHOG has obvious advantages over HOG, and on the basis of HOG, MHOG can effectively reduce the amount of calculation caused by feature dimensions.

In Table 2, the relatively low precision is due to the large FP, and large FP will inevitably reduce the accuracy. But it is worth noting that in Section 4.1, if the OpenPose does not detect any joint point information in the personnel images, the incorrectly classified positive samples can be further corrected into the true negative samples. Therefore, the FP samples of MHOG (6, 4) can be filtered out through the process of worker head extraction. Table 4 provides the identification result of MHOG (6, 4) after the optimization of Section 4.1. It can be seen that the values of precision and accuracy have been effectively improved on the basis of Table 2, and all the indicators prove that the final identification result is excellent.

5.2. Helmet Identification Experiment. All the head samples used in this experiment are obtained from Section 4.1. Figure 10 shows some samples that participated in the experiment, which contains five classes of head images, namely, red helmet, yellow helmet, blue helmet, white helmet, and no helmet.

Since the experiment purpose is to classify and identify any class of the head images, the experimental results should not be affected by the combination types of head images. To verify this view, Table 5 provides all combinations of sample types, which can be summarized as one type of sample, two types of samples, three types of samples, four types of samples, and all five types of samples. In the table, N represents not wearing a helmet, Y represents a yellow helmet, R represents a red helmet, B represents a blue helmet, and W represents a white helmet.

In addition, the experiment selected 65% of the total number of samples as training samples and the remaining 35% as testing samples. The number distribution of various classes of head images in the training dataset and testing dataset is given in Table 6.

The parameter values of the transfer learning network are given in Table 7.

For the 31 types of the helmet combinations in Table 5, except for the 5 single types, the remaining 26 multiple types are identified, respectively, and the identification results are given in Table 8. In the table, the column data corresponding to Y, R, B, W, and N represent the number of head samples correctly identified, and the last column is the identification accuracy of the proposed method. The identification accuracy is defined as the number ratio of correctly identified samples to the total samples.

It can be seen from Table 8 that 7 of the 10 types for the two colors combination reach 100%, only 2 of the 10 types for the three colors combination reach 100%, and all the 6 types of four colors combination and five colors combination dose not reach 100%. The mean accuracy of the two



FIGURE 9: The data samples for personnel identification. (a) Positive sample. (b) Negative sample.

TABLE 1: Number distribution of the personnel images.

Sample	Illumination		Distance			Occlude			Angle		
	Bright	Dark	Far	Medium	Near	Light	Heavy	No	Front	Back	Side
Number	915	915	785	628	417	190	97	1543	506	587	737

TABLE 2: Identification results of MHOG + SVM.

Method	Precision (%)	Recall (%)	Accuracy (%)
MHOG (6, 4)	87.53	99.73	88.79
MHOG (7, 3)	87.52	99.64	88.72
MHOG (8, 2)	86.87	99.45	87.95

TABLE 3: Identification results of HOG + SVM.

Method	Precision (%)	Recall (%)	Accuracy (%)
HOG (6, 4)	87.43	99.73	88.69

TABLE 4: Identification results of MHOG + SVM after optimization.

Method	Precision (%)	Recall (%)	Accuracy (%)
MHOG (6, 4)	99.45	99.73	99.37

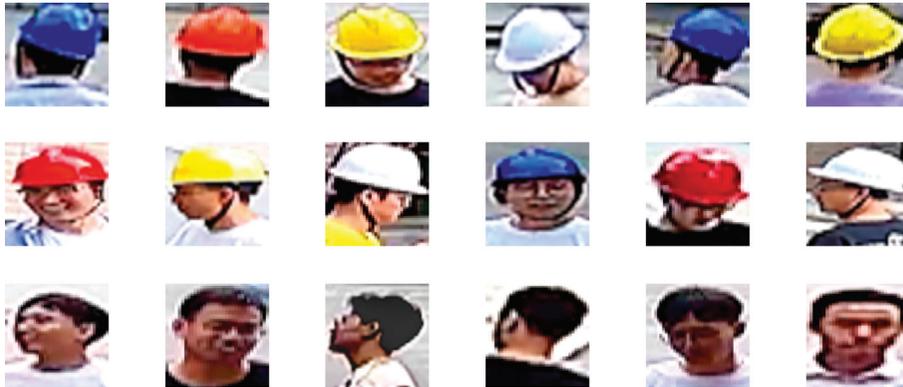


FIGURE 10: Data samples of workers' heads.

TABLE 5: Identification of types of helmets.

One	Two	Three	Four	Five
	RN	YRN		
	YN	YBN		
Y	BN	YWN	YRBN	
R	WN	BWN	YRWB	
B	RY	BRN	YBWN	
W	RW	RWN	RWYN	YRBWN
N	RB	YRB	RBWN	
	YB	YBW		
	YW	RBW		
	BW	YRW		

TABLE 6: Number distribution of workers' head samples.

Sample	Red helmet	Yellow helmet	Blue helmet	White helmet	No helmet
Training	496	205	174	199	157
Testing	267	110	94	107	85

TABLE 7: The parameter settings of transfer learning training.

Parameter	Value
MiniBatchSize	32
MaxEpochs	6
InitialLearnRate	0.0001

TABLE 8: Values of helmet identification accuracy.

Classes	Head sample					Accuracy (%)
	Y	R	B	W	N	
RN	×	267	×	×	85	100
YN	110	×	×	×	85	100
BN	×	×	92	×	84	98.32
WN	×	×	×	105	85	98.96
RY	109	266	×	×	×	99.47
RW	×	267	×	107	×	100
RB	×	267	94	×	×	100
YB	110	×	94	×	×	100
YW	110	×	×	107	×	100
BW	×	×	94	107	×	100
YRN	110	265	×	×	84	99.35
YBN	110	×	93	×	84	99.31
YWN	109	×	×	105	84	98.68
BWN	×	×	92	106	84	98.60
BRN	×	266	92	×	85	99.33
RWN	×	267	×	106	84	99.56
YRB	109	267	93	×	×	99.58
YBW	110	×	93	106	×	99.36
RBW	×	267	94	107	×	100
YRW	110	267	×	107	×	100
YRBN	110	266	93	×	85	99.64
YRWB	108	267	93	107	×	99.48
YBWN	109	×	91	106	84	98.48
RWYN	110	267	×	105	84	99.47
RBWN	×	266	91	106	83	98.73
YRBWN	108	265	90	106	84	98.49

colors is 99.68%, the three colors is 99.38%, and the four colors is 99.16%. Although the identification accuracy of the proposed method shows a downward trend with the increase of helmet types, its overall average accuracy is as high as 99.42%, and the identification accuracy is basically stable with various test schemes. This shows that the algorithm has strong robustness and adaptability and can well identify head images of different colors, angles, and postures.

In addition, the method in [19] and the proposed method are the same type of the helmet identification algorithm, so [19] is used as the comparison algorithm. For the head images identification, the tree structure of H-SVM is used in [19] to classify the samples layer by layer, and each layer classifies one class of samples. The test results of the H-SVM classifier on our dataset show that for the first-level classification, B-N + R + Y + W has the highest classification accuracy of 97.87% for the blue helmets among the five combinations. For

the second-level classification, R-N + Y + W has the highest classification accuracy of 95.88% for the red helmets among the four combinations. For the third-level of classification, N-Y + W has the highest classification accuracy of 97.65% for the no-helmet samples among the three combinations. For the last level of classification, the YW has the classification accuracy of 98.18% and 97.20%, respectively. The average classification accuracy of H-SVM for the five layers is 97.36%.

The comparison shows that the average identification accuracy of the H-SVM method is 2.06% lower than that of the proposed method. More importantly, the H-SVM method requires hierarchical processing, and each level can only identify one type of head image, while the proposed method can identify multiple types of helmet colors at one time. In summary, the proposed method is not only higher than the H-SVM method in the identification accuracy but also far ahead of [19] in identification efficiency.

6. Conclusions

In this study, an intelligent vision-based method for worker identification is proposed. This method is implemented through three parts: motion extraction based on the GMM model, personnel identification based on MHOG and SVM, and helmet identification based on the OpenPose and transfer learning CNN.

The proposed method can be successfully used to identify the worker types in a variety of industrial scenarios, such as construction site, factory workshop, power construction site, and interior decoration. In addition, the method can be applied to small-scale datasets, and the identification accuracy cannot be affected by helmet occlusion, worker size, and different lighting conditions. The testing results on our self-collected dataset show that the accuracy values vary little under different types of identification situations, and the mean value of the accuracy reaches 99.43%.

It is worth considering that the visual features of the self-collected dataset in this study are relatively ideal, so the identification results are relatively excellent. Our next work will capture some worker videos in actual industrial environments and conduct in-depth research on the worker images to promote the application of vision-based worker identification.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (52074305 and 51874300), the National Natural Science Foundation of China and Shanxi Provincial People's Government Jointly Funded Project of China for Coal Base and Low Carbon (U1510115) and the Open Research Fund of Key Laboratory of Wireless Sensor Network and Communication, Shanghai Institute of Microsystem and Information Technology, and Chinese Academy of Sciences (20190902 and 20190913).

References

- [1] W. Fang, L. Ding, H. Luo, and P. E. D. Love, "Falls from heights: a computer vision-based approach for safety harness detection," *Automation in Construction*, vol. 91, no. 7, pp. 53–61, 2018.
- [2] H. Li, X. Li, X. Luo, and J. Siebert, "Investigation of the causality patterns of non-helmet use behavior of construction workers," *Automation in Construction*, vol. 80, no. 8, pp. 95–103, 2017.
- [3] J. Shen, X. Xiong, Y. Li, W. He, P. Li, and X. Zheng, "Detecting safety helmet wearing on construction sites with bounding-box regression and deep transfer learning," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 2, pp. 180–196, 2021.
- [4] X. Chen, W. Chen, and P. Xu, "Attribute-based Face recognition and application in safety detection of intelligent factory," in *Proceedings of the 2018 International Conference on Computing & Artificial Intelligence (ICCAI)*, pp. 76–80, Chengdu, China, March 2018.
- [5] G. Zhang, L. Lv, and L. Dan, "The method for recognizing recognition helmet based on color and shape," in *Proceedings of the 2017 5th International Conference on Machinery, Materials and Computing Technology (ICMMCT)*, pp. 1219–1223, Beijing, China, March 2017.
- [6] L. Huang, Q. Fu, and M. He, "Detection algorithm of safety helmet wearing based on Deep learning," *Concurrency and Computation-Practice & Experience*, vol. 33, no. 13, Article ID e6234, 2021.
- [7] Q. Zhou, J. Qin, X. Xiang, Y. Tan, and N. Xiong, "Algorithm of helmet wearing detection based on AT-YOLO deep mode," *Computers, Materials & Continua*, vol. 69, no. 1, pp. 159–174, 2021.
- [8] H. Wang, Z. Hu, and Y. Guo, "A real-time safety helmet wearing detection approach based on CSYOLOv3," *Applied Sciences-Basel*, vol. 10, no. 19, p. 6732, 2020.
- [9] Y. Li, H. Wei, and Z. Han, "Deep learning-based safety helmet detection in engineering management based on convolutional neural networks," *Advances in Civil Engineering*, vol. 2020, Article ID 9703560, 13 pages, 2020.
- [10] Z. Wang, Y. Wu, and L. Yang, "Fast personal protective equipment detection for real construction sites using deep learning approaches," *Sensors*, vol. 21, no. 10, p. 3478, 2021.
- [11] S. Chen and K. Demachi, "Towards on-site hazards identification of improper use of personal protective equipment using deep learning-based geometric relationships and hierarchical scene graph," *Automation in Construction*, vol. 125, Article ID 103619, 2021.
- [12] R. Xiong and P. Tang, "Pose guided anchoring for detecting proper use of personal protective equipment," *Automation in Construction*, vol. 130, Article ID 103828, 2021.
- [13] J. Wu, N. Cai, and W. Chen, "Automatic detection of hardhats worn by construction personnel: a deep learning approach and benchmark dataset," *Automation in Construction*, vol. 106, Article ID 102894, 2019.
- [14] Q. Fang, H. Li, X. Luo et al., "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos," *Automation in Construction*, vol. 85, pp. 1–9, 2018.
- [15] X. Long, W. Cui, and Z. Zheng, "Safety helmet wearing detection based on deep learning," in *Proceedings of the 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Chengdu, China, March 2019.
- [16] B. Mneymneh, M. Abbas, and H. Khoury, "Vision-based framework for intelligent monitoring of hardhat wearing on construction sites," *Journal of Computing in Civil Engineering*, vol. 33, no. 2, Article ID 04018066, 2019.
- [17] L. Shine and C. Jiji, "Automated detection of helmet on motorcyclists from traffic surveillance videos: a comparative analysis using hand-crafted features and CNN," *Multimedia Tools and Applications*, vol. 79, no. 19–20, pp. 14179–14199, 2020.
- [18] B. Yogameena, K. Menaka, and S. Saravana Perumaal, "Deep learning-based helmet wear analysis of a motorcycle rider for intelligent surveillance system," *IET Intelligent Transport Systems*, vol. 13, no. 7, pp. 1190–1198, 2019.
- [19] H. Wu and J. Zhao, "An intelligent vision-based approach for helmet identification for work safety," *Computers in Industry*, vol. 100, pp. 267–277, 2018.
- [20] D. Xuan, G. Wang, and Y. Pang, "Fast efficient algorithm for enhancement of low lighting video," in *Proceedings of the 2011 IEEE International Conference on Multimedia and Expo (ICME)*, July 2011.
- [21] N. Dalal and B. Triggs, "Histograms of oriented gradients for Human detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 886–893, San Diego, CA, USA, June 2005.
- [22] X. Tian, H. Bao, and C. Xu, "Improved HOG algorithm of pedestrian detection," *Computer Science*, vol. 41, no. 9, pp. 320–324, 2014.
- [23] C. Saunders, M. Stitson, and J. Weston, "Support vector machine," *Computer Science*, vol. 1, no. 4, pp. 1–28, 2002.
- [24] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: realtime multi-person 2D pose estimation using Part Affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2021.
- [25] C. Szegedy, W. Liu, and Y. Jia, "Going deeper with convolutions," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, June 2015.
- [26] P. David Martin Ward, "Evaluation: From precision, Recall and F-factor to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 2229–3981, 2011.