Automatic detection of safety helmet wearing based on head region location

Yuwan Gu | Yusheng Wang | Lin Shi | Ning Li | Lihua Zhuang | Shoukun Xu

School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, Jiangsu, P.R. China

Correspondence
Shoukun Xu, School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, Jiangsu, 213164, P.R. China
Email: shoukxu@126.com

Funding information
National Natural Science Foundation of China, Grant/Award Number: 61906021

Abstract
In order to solve the problem of difficult and low precision in the detection of safety helmet wearing in the complex pose of construction worker, a detection method of safety helmet wearing based on pose estimation is proposed. In the pose estimation model of OpenPose, the residual network optimized feature extraction is introduced to obtain the skeletal point information of the construction worker, and then the pose of the construction worker is estimated based on the skeletal point information, three-point localization method is proposed for the front and back pose, and skin colour detection method is proposed for the side pose, and then to determine the head region. The YOLO v4 is used to detect the safety helmet region, and then the construction worker’s safety helmet wearing is judged according to whether the head region intersects the safety helmet region or not. Experimental results show that the detection accuracy of the method is higher than other methods, and the adaptability to the environment is stronger.

INTRODUCTION
Safety helmet recognition is an important means of management transformation and upgrade in intelligent construction site, which provides strong technical support for safety personnel on-site supervision. In the daily production, workers injured due to inadequate safety measures accounted for the majority of production safety accidents. As a result of special national conditions, the quality of workers is uneven, although the units often carry out safety education, there is always a fluke for a variety of reasons cannot be guaranteed to wear at all times.

Due to the difficulty of safety helmet recognition in the case of complex pose, the method of safety helmet wearing detection based on pose estimation is proposed. The contributions of the paper are: (1) The pose of human body is determined by skeletal point information of construction worker. (2) According to the pose of human body, the method of three-point localization or skin colour detection is chosen to determine the head region, it solves the problem that the relative position of safety helmet and human body is difficult to be determined under the complex pose of construction worker. (3) Recognition of safety helmet wearing is realized by using target detection network of YOLO v4.

RELATED WORK
Many related studies have been conducted for safety helmet wearing detection. Kalantari et al. [1] divided the entire detection process into two parts. First, combining the frequency domain information and the Histogram of Oriented Gradient (HOG) of the images to detect workers. Then, the colour and Circle Hough Transform (CHT) feature were combined for safety helmet detection. The method achieved a certain detection effect, however, the overall accuracy of the method is low and only a specific colour safety helmet can be detected. Wu et al. [2] used the Local Binary Patterns (LBP), Hu Moment Invariant (HMI) and Colour Histogram (CH) of the image to extract the feature of different colour helmets, then the hierarchical support vector machine (SVM) is used to classify safety helmets. Above methods are based on traditional machine learning method for object detection. These methods are mostly based on subjective feature selection, which required...
solid professional foundation and rich experience. Moreover, feature selection is time consuming and its generalization ability is poor, hard to adapt to changes in conditions, such as lighting.

With the development of safety helmet wearing detection, the traditional machine learning method has been changed to the deep learning method. Some scholars used the popular target detection network to detect safety helmet wearing. Felix Wilhelm Siebert et al. [3] used the RetinaNet network model to detect safety helmet wearing, and used multi-scale feature pyramid and focal loss to solve the limit of one-stage detector in accuracy. The method improves the accuracy of the object localization, but the accuracy of safety helmet wearing detection is not high enough. Qi Fang et al. [4] used the Faster R-CNN network model to detect safety helmet wearing. The Faster R-CNN introduces a Region Proposal Network (RPN) that can generate high-quality region proposals and was used to detect and classify objects based on the region proposals and sharing full-image convolutional features with RPN. Faster R-CNN has a frame rate of 5 fps (including all steps) on a GPU, so it is a practical object detection system in terms of both speed and precision. HYOJOO SON et al. [5] introduced a residual network (ResNet-152) on the basis of Faster R-CNN, the detection process is divided into two stages: First, extracting feature maps via very deep residual networks (ResNet-152), second, bounding box regression and labelling from the original image via Faster R-CNN. Mingliang Zhong et al. [6] used YOLO v3 model to perform real-time detection without safety helmet, and used image pyramid structure to obtain feature maps of different scales for position and category prediction, in the course of training iteration, the size of the input image is changed to increase the adaptability of the model to the scale, so as to improve the precision of safety helmet wearing detection. The method can improve the detection speed with the loss of precision, and cannot adapt to the complex scene of the actual construction site because of the single scene of the data set. Although the accuracy and real-time performance of the method are improved, it is not enough to solve the relative position between safety helmet and human body.

In recent years, scholars have been able to identify safety helmet wearing by identifying the head region of a person, LinuShine et al. [7] and Silva et al. [8] proposed a method based on the segmentation of moving objects to locate pedestrians quickly, and based on the results of pedestrian detection, the detection of safety helmet wearing is realized by head position, colour space transformation and colour feature discrimination, in the method, the head position is judged according to one-fifth of the body proportion of pedestrians. But the method is not useful for the complex pose of the construction worker. Kheireddine Aziz et al. [9] determined the skeleton point by detecting the visible segment of the corresponding body part in the image, then obtained the line segment corresponding to the extreme node, and calculated its inclination with the vertical axis. If tilted within a certain angle, the line segment within the angle is classified as the head. Again, the method is aimed at standing pedestrians and is not suited to the complex pose of the construction worker. In a word, the above methods are aimed at the safety helmet wearing detection of standing pedestrians, but the actual construction site, the pose of construction worker is very complex and diverse.

From the research on the previous methods found that in the case of complex pose (such as bending, squatting, backward etc.), the difficulty of the right to identify the safety helmet wearing is increased. To solve this problem, a method to estimate the pose of the workers before the safety helmet wearing detection is proposed, so as to improve the accuracy of the detection. Ke et al. [10] used a robust multi-scale structure perception neural network. The deep conv-deconv hourglass model is improved to estimate human body pose by the scheme of adding multi-scale supervision, multi-scale features are globally optimized, structure sensing loss and effective location of key points by adjacent matching. Ning et al. [11] used stacked hourglass designs and initial resnet modules to construct a fractal network to regress human pose into heat maps, and used visual features to encode external knowledge, these visual features can represent the constraints of the human model and evaluate the suitability of the intermediate network output. Cao et al. [12] proposed a key point connection method that encodes the position and direction of human limbs, and designed a framework that can learn the position and connection of human body parts simultaneously to produce the result of human pose detection. The above methods are bottom-up pose estimation methods, and have achieved good results in public datasets. The method in ref. [12] meets to the method in the paper, and can recognize the information of human bone points well.

1 | THE DETECTION METHOD OF SAFETY HELMET WEARING

The flow of safety helmet wearing detection method based on the head region is shown in Figure 1. First, the improved OpenPose human pose estimation method is used to estimate the pose of the construction worker. Then, using different head region location method according to different pose (front, back and side), a three-point localization method is proposed for the construction worker with front or back pose, the method determines the head region according to the detected bone points (head centre point, left shoulder and right shoulder) of the construction worker, and a skin colour detection method is proposed for the construction worker with side pose, the method determines the face region according to the skin colour of the face, and determines the head region using geometric transformation. Finally, a YOLO v4 target detection network is used to detect the safety helmet region, and then the construction worker’s safety helmet wearing is judged according to whether the head region intersects the safety helmet region or not.
1.1 The method of pose estimation

1.1.1 OpenPose

OpenPose is a bottom-up method based on supervised learning, which can estimate human motion, facial expression, and finger movement and so on. First, OpenPose extracts features $F$ from images using the first ten layers of the VGG-19 network. Then the features $F$ is processed through a continuous multi-stage network, each stage $(t)$ of the network contains two branches, whose input results are Part Confidence Map $S'$ and Part Affinity Map $L'$, respectively. Finally, the feature representation of Part Attachment Fields PAFs is used to store the position and direction information of the supporting region of the body.

As shown in Figure 2, OpenPose uses an iterative CNN network to detect key parts of the human body. Each CNN network has two branches, CNN_S and CNN_L. The first and subsequent stages of the network are difference in form. The two network branches of each stage are used to compute Part Confidence Maps (the joints) and Part Affinity Fields (the limb trunks), respectively. It can also be seen from the Figure 2 that the first stage of the network received the features $F$, which is processed by the network to get $S^1$ and $L^1$, respectively. Starting with stage 2, the input to the stage $t$ network consists of three parts: $S^{t-1}$, $L^{t-1}$, $F$. The inputs for each stage of the network
are:
\[ S'^t = \rho^t(F, S'^{t-1}, L^{-1}), \forall t \geq 2 \]
\[ L'^t = \phi^t(F, S'^{t-1}, L^{-1}), \forall t \geq 2 \]

It iterates over and over until the network converges. And then, if the two parts \((d_1, d_2)\) are connected or not by the PAFs, the PAFs calculates the linear integral on the line, and if the direction of the line \(\vec{d}_1\) and \(L^*_c(p)\) (join vector) is the same, then the linear integral \(E\) will have a very large value, and there is a very high probability that these two parts are the torso.

\[ E = \int_{u=1}^{u=n} L_c(P(u))g_{\|d_2 - d_1\|_2}du \]

Finally, traversing all the combinations, calculating the integral sum and finding the parts of the torso (upper arm, lower arm, waist etc.), the two adjacent torsos must have a shared joint, through which all the torsos are combined, you get the skeleton of everyone.

### 1.1.2 Improved OpenPose

After the CNN network reaches a certain depth, the classification performance cannot be further improved only by increasing the number of network layers, but the network convergence speed becomes slower and the classification accuracy of the test data becomes worse. Increasing the number of layers after VGG network reaches 19 will lead to the degradation of classification performance. ResNet learns the representation of residuals between inputs and outputs by using multiple parametric layers rather than directly trying to learn the mapping between inputs and outputs, as is the case with CNN networks, thus, the convergence speed and the classification accuracy of the network are improved.

In order to make the OpenPose model suitable for the pose estimation of construction worker, the residual network [13] (ResNet) is used to obtain feature based on the original OpenPose model, in order to enhance the performance. The network structure of the improved OpenPose model is shown in Figure 3.

### 1.2 The method of three-point localization

In order to determine the head region of construction worker with the front and back pose, and thus to achieve the accurate detection of safety helmet wearing, a three-point localization method is proposed to detect the target head region.
OpenPose of real-time multi-person 2D pose estimation is used to detect and identify the key positions of construction worker, and the position information of the key positions is obtained, and detection region (head region) is obtained according to the position information of the centre of the head, the left shoulder, and the right shoulder. First, a triangle is defined by the centre of the head, the left shoulder, and the right shoulder. Then, the triangle region is rotated 180° with the head point as the centre, and the new triangle and the original triangle form a quadrangle. Then take the shorter diagonal of the quadrilateral as the diameter and the intersection of the diagonal as the centre of the circle as the circle. Finally, the circumscribed square of the circle, that is, the detection region (the red rectangle), as shown in Figure 4.

Set the position of the centre of the head, left shoulder, and right shoulder is:

\[ H(x_0, y_0), L(x_1, y_1), R(x_2, y_2) \]  

After rotating 180°, the symmetrical position of the left shoulder and the right shoulder is:

\[ L'(2x_0 - x_1, 2y_0 - y_1), R'(2x_0 - x_2, 2y_0 - y_2) \]  

1.3 The method of skin colour detection

When the left shoulder and right shoulder overlap is detected in the pose estimation, that is, the construction worker is in the side pose, then the three-point localization method is invalid, in order to determine the side pose of the construction worker’s head region and achieve accurate detection of safety helmet wearing, a skin colour detection method is proposed to detect the target head region.

\[ L'(2x_0 - x_1, 2y_0 - y_1), R'(2x_0 - x_2, 2y_0 - y_2) \]  

So the diagonal length of the quadrilateral is:

\[ |Ld| = \sqrt{(x_1 - x_1')^2 + (y_1 - y_1')^2} \]
\[ |Rd| = \sqrt{(x_2 - x_2')^2 + (y_2 - y_2')^2} \]
Under construction, only the skin of the face, neck and hand are exposed. Therefore, the light compensation to the picture is carried on in the YCbCr colour space, YCbCr colour space is easier to analyse than the RGB colour [14], the transformation [15] formula from RGB to YCbCr colour space as shown in Equation (7).

\[
\begin{bmatrix}
Y \\
C_b \\
C_r 
\end{bmatrix} = \begin{bmatrix}
16 \\
128 \\
128
\end{bmatrix} + \begin{bmatrix}
65.481 & 128.533 & 24.966 \\
-37.797 & -74.203 & 112 \\
112 & -93.786 & -18.214
\end{bmatrix} \begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
\] (7)

Skin colour is identified based on the screening method of \(C_r, C_b\) range of YCbCr colour space. The range of \(C_r\) is \(140 < C_r < 175\), and the range of \(C_b\) is \(100 < C_b < 120\) from experiment. Second, the top hat transformation is mainly used to correct the influence of uneven illumination. The binary image after skin colour detection contains a lot of noise, so the top hat transformation is carried on, in order to reduce the range of non-human face region. The formula of top hat transformation [16] as shown in Equation (8), \(f\) is the grey-level image, \(f \circ b\) is the open operation of the structure of element on an image.

\[
T_{\text{hat}}(f) = f - (f \circ b)
\] (8)

Finally, after the image correlation morphological processing, the larger part of the connected region is located, and the region to be detected as the face region, and the face region is marked by a rectangle frame. The result of skin colour detection based on YCbCr colour space is shown in Figure 5.

### 1.4 The target detection network of YOLO v4

YOLO v4 [17] is the newest detection network of YOLO series, which is innovated by integrating various advanced algorithms based on YOLO v3. Therefore, the target detection network of YOLO v4 is used to detect the safety helmet in order to improve the performance of the detection method.

### FIGURE 7  The part of sample diagram

### TABLE 1  Comparison results of model training under different feature extraction methods

| Method               | head  | shoulder | elbow | wrist | knees | mAP  | speed/s |
|----------------------|-------|----------|-------|-------|-------|------|---------|
| OpenPose (VGG19)     | 92.9  | 90.1     | 81.1  | 72.6  | 72.8  | 81.9 | 0.5     |
| OpenPose (ResNet34)  | 93.4  | 91.4     | 81.6  | 73.2  | 73.7  | 82.7 | 0.53    |
| OpenPose (ResNet50)  | 94.7  | 91.6     | 83.9  | 74.0  | 74.5  | 83.7 | 0.57    |
The innovations of YOLO v4 include the data enhancement of Mosaic, emBN and self-confrontation training of SAT on input terminal, and the innovations of backbone network include CSPDARKNET53, activation function of Mish and Dropblock, and Neck’s innovations include that the target detection network often inserts layers between the BackBone and the final output layer, such as the SPP model, the structure of FPN + PAN, and that the prediction part of the output layer has the same anchor mechanism as the YOLO v3, the main improvements are the loss function of CIOU_Loss on the training, and the nms of prediction box filtering is changed to DIOU_nms.

YOLO v4 uses CSPNet [18] and Darknet-53 as the backbone network of feature extraction. Compared with the ResNet, the target detection accuracy of CSPDarknet53 model is higher, but the classification performance of ResNet is better. However, with Mish and other techniques, the classification accuracy of CSPDarknet53 can be improved.

In order to detect different size targets, a hierarchical structure is used, so that the safety helmet recognition can detect different spatial resolution of the feature map. To make the information of input helmet richer, the bottom-up and top-down data streams are added or connected element before the input helmet. Compared with the FPN [19] network used in YOLO v3, SPP can greatly increase the receptive field, isolate the most significant contextual features, and hardly decrease the network speed. And YOLO v4 selects PANet [20] as the parameter aggregation method from the different backbone layer for the different level detector. Thus, YOLO v4 gradually replaces FPN with modified SPP, PAN and SAM, retaining rich spatial information from bottom-up data streams and rich semantic information from top-down data streams. The network structure of YOLO v4 is shown in Figure 6.

At the same time, YOLO v4 makes reasonable use of bag of freebie and bag of specials for fine tuning. YOLO v4’s AP and FPS are 10% and 12% higher than YOLO v3’s.

YOLO v4 network is used to detect the safety helmet, and then judge whether the construction workers wear the safety helmet in the head region. In Sections 2.2 and 2.3 the head region has been evaluated, and the position of the safety helmet region and the head region has been determined. If the safety helmet region intersects the head region, the construction worker is judged to be wearing a safety helmet, otherwise, the construction worker is judged not to be wearing a safety helmet.


2 DATASET CREATION

The dataset used in the experiment consists of 8000 images collected by web crawler, public dataset and field dataset. The part of sample images for the dataset are shown in Figure 7. The dataset included images of individual and multiple people in various construction scenarios, and considered the balance between positive and negative sample images (4500 images are positive samples with safety helmet and 3500 images are negative samples without safety helmet), at the same time, in order to improve the accuracy and generalization ability of the network, the dataset is expanded to 40,000 images through the operation of flip, punch, translation and so on.

3 EXPERIMENTAL RESULTS AND ANALYSIS

3.1 Experimental environment and evaluation criteria

Experimental environment configuration: GPU: NVIDIA TITAN Xp×2, CUDA 10.0, Ubuntu 16.04, memory 12 GB. In the paper, precision rate, false positive rate, miss rate, intersection-to-union ratio and mean average precision (mAP) are used to evaluate the validity of the method. The calculation formula is shown as Equations (9)–(12).

\[
\text{precision} = \frac{TP}{TP + FP} \quad (9)
\]

\[
fp = \frac{FP}{FP + TN} \quad (10)
\]

\[
\text{miss} = \frac{FN}{TP + FN} \quad (11)
\]

\[
\text{IoU} = \frac{\text{PartAcreage} \cap \text{OverallAcreage}}{\text{PartAcreage} \cup \text{OverallAcreage}} \quad (12)
\]

Among them: TP (true positive) means a positive sample predicted to be positive by the model, FP (false positive) means a negative sample predicted to be positive by the model, FN (false negative) means a positive sample predicted to be negative by the model, TN (true negative) means a negative sample predicted to be negative by the model. PartAcreage is the safety helmet region from which the rear frame is detected, and OverallAcreage is the marked safety helmet region.

3.2 Comparative experiment of pose estimation model

MPII dataset [21] is used to train pose estimation. There are 16 key points in the MPII dataset: r-ankle, r-knee, r-hip, l-hip, l-knee, l-ankle, pelvis, thorax, upper-neck, head, r-wrist, r-elbow, r-shoulder, l-shoulder, l-elbow, l-wrist. OpenPose removes the pelvis and chest from the training, and adds a centre point. The output is a heat map of 15 key points and a background.

The feature extraction network is improved in the original OpenPose model by replacing the original VGG19 feature extraction network with ResNet34 and ResNet50 respectively. The training precision of the three feature extraction methods is shown in Table 1.

Table 1 shows that the improved OpenPose model improves the average detection accuracy of bone points by 1.8%, and the detection time of bone points has little change. For the whole method, the position information extracted by OpenPose model serves for the selection of head location method (three-point localization method or skin colour detection method), and the improved OpenPose model improves the accuracy of three-point location method, at the same time, there is no significant decrease in efficiency. Figure 8 shows the effect of the three methods on the test image. It can be seen from Figure 8, that the difference between the three methods is not large in the detection of bone points in the conventional pose, whether single or multi-person situation. ResNet50 as the
feature extraction network is the best, and when complex multi-person situation also has good performance. At the same time, the three points of head, left shoulder and right shoulder which are needed in the follow-up head location experiment are also improved.

3.3 The comparison results of head region location

Three-point localization method and skin colour detection method is used to determine the head region. In order to
evaluate the validity, IoU is introduced to evaluate the method. Comparing the methods of head region location in refs. [8] and [9], the IoU value of the three methods is shown in Figure 9, the IoU value of the paper’s method is higher than that of the other two methods, that is, the head region detected by the method in the paper has a higher coincidence degree with the actual head region. Therefore, the method is more effective in locating the head region of the construction workers with complex pose. The experimental results are shown in Figure 9.

As can be seen from Figure 10, the effects of the three methods are approximately the same when the workers are in the upright pose. In the case of complex pose, the method of ref. [8] is invalid, and the method of detecting the upper 1/5 region of human body is only suitable for the upright construction worker, the method of ref. [9] is better than ref. [8], however, there is still the phenomenon of missing detection, this is mainly because the position of the head is determined by calculating the inclination of the skeleton point and the vertical axis, it is also suitable for upright construction worker. The three-point

FIGURE 12  The feature visualization of convolution layer

FIGURE 13  The results of the comparative experiment
The results of the four networks for the test dataset localization and skin colour detection method is proposed, which can achieve good results in both upright pose and complex pose.

3.4 The comparison results of safety helmet recognition

The YOLO v4 target detection network is used to detect the region of safety helmet. The dataset is divided into training set, verification set and test set according to 6:2:2. The training set and verification set are annotated.

The learning rate is 0.001, the epoch is 80 times, the iterations is 100 times, the number of iterations is 8000, and the sample number of each iteration is 64. The steps of learning rate are set to (6400, 7200) and the learning rate is adjusted based on the batch_num. Steps and scales correspond to each other. These two parameters set the change of the learning rate. The network training parameters are shown in Table 2.

The final loss value oscillates around 0.5. The training loss values for the dataset are shown in Figure 11 and the convolution layer features are visualized in Figure 12.

The detection effects of four target detection networks, namely, Faster R-CNN [22], RetinaNet [23], YOLO v3 [24], YOLO v4, are evaluated from three indexes: average precision, intersection-to-union ratio and speed, YOLO v3 and YOLO v4 is first-order target detection network, Faster R-CNN and RetinaNet is a second-order target detection network.

The results of the comparative experiment are shown in Figure 13.

As can be seen from the Figure 13, YOLO v4 target detection network is superior to the other three networks in both accuracy and detection time. Furthermore, YOLO v4 is superior to the two second-order target detection networks, namely Faster R-CNN and RetinaNet, in terms of AP and IoU values. YOLO v4 is the newest detection network of YOLO series, which greatly improves the detection precision of safety helmet region. Also the detection time is improved 0.5, 0.4, 0.2 s for Faster R-CNN, RetinaNet and YOLO v3. The results of the four networks for the test dataset is shown in Figure 14.

Figure 14 shows the various indicators of testing images based on four network detection, from the analysis of the data, it can be seen the various indicators of YOLO v4 are better than the other three networks. As shown in Figure 14, the four network tests performed by construction workers in an upright pose were similar, in complex pose, YOLO v4 is superior to Faster CNN, YOLO v3, as a new second-order target detection network, RetinaNet is similar to YOLO v4, but its detection time is relatively weak. So, YOLO v4 is chosen as the target detection network of the safety helmet region.

3.5 Comparison results of safety helmet wearing

In the case of complex pose, the method of ref. [8] is invalid, and the method of detecting the upper 1/5 region of human
According to the data in Table 3, the precision and speed of the proposed method are better than those in refs [8] and [9]. The precision is improved by 13.6% and 9.7%, and the speed is improved 0.2 and 0.6 s. As can be seen from Figure 15, the method of refs. [8] and [9] is invalid for some samples, however, the method in the paper is not limited to worker’s pose.

In the paper, the method judges the wearing situation of the safety helmet by whether the head region intersects the safety helmet region or not. As shown in Figure 16, when the head region intersects the safety helmet region, the worker is judged to be wearing a safety helmet. When the hand-held safety helmet, or no helmet, the head region and the safety helmet region did not intersect, construction worker did not wear a safety helmet.

**FIGURE 15**  Comparison of safety helmet detection methods

**FIGURE 16**  Detecting results of safety helmet wearing
4 | CONCLUSION

In the paper, an improved OpenPose model is used to estimate the pose of the construction worker, and then an appropriate head region location method is selected according to the result of the pose estimation, the head region is determined by three-point localization for workers with the front and back pose, and the head region is determined by skin colour detection for workers with side pose (left and right shoulder points overlap). Finally, the YOLO v4 target detection network is used to identify the safety helmet, and the safety helmet wearing is judged according whether the head region intersects the safety helmet region or not. The experimental results show that the method in the paper has a good result for the detection of safety helmet wearing of construction worker with different pose, and the detection accuracy of safety helmet wearing is also better than the previous method because of the accurate location of the head region, compared with refs [8] and [9], the precision is improved by 13.6% and 9.7%, respectively. However, because the method uses a bottom-up pose estimation strategy for the construction worker, it is less effective when the characters have minimal goals. In order to improve the generalization effect of the model, the next future work is that using 3D space to locate the exact position of the head.

ACKNOWLEDGEMENTS

This research is supported by National Natural Science Foundation of China (61906021).

ORCID

Yuan Gu 🌐https://orcid.org/0000-0002-3148-9414

REFERENCES

1. Rubaiyat, A.H.M., et al.: Automatic detection of helmet uses for construction safety. In: International Conference on Web Intelligence Workshops, Omaha, pp. 135–142 (2016)
2. Wu, H., Zhao, J.: An intelligent vision-based approach for helmet identification for work safety. Comput. Ind. 100, 267–277 (2018).
3. Siebert, F., Lin, H.: Detecting motorcycle helmet use with deep learning. Accid. Anal. Prev. 134, 105319(1)–105319(11) (2020).
4. Fang, Q., et al.: Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. Autom. Constr. 85, 1–9 (2018).
5. Son, H., et al.: Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks. Autom. Constr. 99, 27–38 (2019).
6. Zhong, M., Meng, F.: A YOLOv3-based non-helmet-use detection for seafarer safety aboard merchant ships. J. Phys. Conf. Ser. 1325, 012096 (2019).
7. Shine, L., Jiji, C.V.: Automated detection of helmet on motorcyclists from traffic surveillance videos: A comparative analysis using hand-crafted features and CNN. Multimedia Tools Appl. 79, 14179–14199 (2020).
8. Silva R., R., et al.: Detection of helmets on motorcyclists. Multimedia Tools Appl. 77(5), 5659–5683 (2018).
9. Aritz, K., et al.: Head detection based on skeleton graph method for counting people in crowded environments. J. Electron. Imaging. 25(1), 013012(1)–013012(14) (2016).
10. Ke, L., et al.: Multi-scale structure-aware network for human pose estimation. In: The European Conference on Computer Vision (ECCV), Munich, pp. 713–728 (2018).
11. Ning, G., Zhang, Z., He, Z.: Knowledge-guided deep fractal neural networks for human pose estimation. IEEE Trans. Multimedia 20(2), 1246–1259 (2018).
12. Cao, Z., et al.: Realtime multi-person 2D pose estimation using part affinity fields. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, pp. 7291–7299 (2017).
13. He, K., et al.: Deep residual learning for image recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, pp. 770–778 (2016).
14. Qin, X., et al.: DTMNN: Deep transfer multi-ometric network for RGB-D action recognition. Neurocomputing 406, 127–134 (2020).
15. Khan, R., et al.: Color based skin classification. Pattern Recognit. Lett. 33(2), 157–163 (2012).
16. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 3rd ed. Publishing House Of Electronics Industry, Beijing, p. 433 (2010).
17. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, (2020).
18. Wang, C.Y., et al.: CSPNet: A new backbone that can enhance learning capability of CNN. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp. 1–10 (2020).
19. Lin, T.Y., et al.: Feature pyramid networks for object detection. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, pp. 936–944 (2017).
20. Liu, S., et al.: Path aggregation network for instance segmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, Utah (2018).
21. Andriluka, M., et al.: 2D human pose estimation: New benchmark and state of the art analysis. In: Computer Vision and Pattern Recognition (CVPR). Columbus, Ohio, pp. 2999–3007 (2014).
22. Ren, S., et al.: Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Machine Intell. 39(6), 1137–1149 (2015).
23. Lin, T.Y., et al.: Focal Loss for Dense Object Detection. IEEE Trans. Action Analysis & Machine Intelligence. 42(2), 318–327 (2020).
24. Redmon, J., Farhadi, A.: YOLOv3: An incremental improvement. In: The IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, pp. 89–95 (2018).

How to cite this article: Gu, Y., et al.: Automatic detection of safety helmet wearing based on head region location. IET Image Processing, 1–13 (2021). https://doi.org/10.1049/ipt.2.12231.