



A high-performance framework for personal protective equipment detection on the offshore drilling platform

Xiaofeng Ji¹ · Faming Gong¹ · Xiangbing Yuan² · Nuanlai Wang¹

Received: 4 August 2022 / Accepted: 25 February 2023
© The Author(s) 2023

Abstract

In order for the offshore drilling platform to operate properly, workers need to perform regular maintenance on the platform equipment, but the complex working environment exposes workers to hazards. During inspection and maintenance, the use of personal protective equipment (PPE) such as helmets and workwear can effectively reduce the probability of worker injuries. Existing PPE detection methods are mostly for construction sites and only detect whether helmets are worn or not. This paper proposes a high-precision and high-speed PPE detection method for the offshore drilling platform based on object detection and classification. As a first step, we develop a modified YOLOv4 (named RFA-YOLO)-based object detection model for improving localization and recognition for people, helmets, and workwear. On the basis of the class and coordinates of the object detection output, this paper proposes a method for constructing position features based on the object bounding box to obtain feature vectors characterizing the relative offsets between objects. Then, the classifier is obtained by training a dataset consisting of position features through a random forest algorithm, with parameter optimization. As a final step, the PPE detection is achieved by analyzing the information output from the classifier through an inference mechanism. To evaluate the proposed method, we construct the offshore drilling platform dataset (ODPD) and conduct comparative experiments with other methods. The experimental results show that the method in this paper achieves 13 FPS as well as 93.1% accuracy. Compared to other state-of-the-art models, the proposed PPE detection method performs better on ODPD. The method in this paper can rapidly and accurately identify workers who are not wearing helmets or workwear on the offshore drilling platform, and an intelligent video surveillance system based on this model has been implemented.

Keywords Personal protective equipment · Object detection · Random forest · Offshore drilling platform

Introduction

With the development of the oil industry [1], the development of offshore oil and gas resources has gradually become a hot

spot and focus. The offshore drilling platform has a complex environment, with a large number of equipment and a dense distribution of pipelines. In the course of equipment maintenance, workers are at risk of drowning, injuries caused by falling equipment, falls from heights, and other risks. Worker safety can be improved by the proper use of personal protective equipment (PPE), such as helmets and work clothing. The uniform color and logo of the workwear can not only prevent the invasion of outsiders, but also timely detection of drowning staff; helmets can not only prevent workers from being injured by falling objects, but also prevent head injuries when falling from heights.

In spite of the fact that PPE can effectively prevent worker injuries, some workers do not wear helmets or workwear during operations, which exposes them to potential hazards and results in significant losses to corporate productivity. Consequently, it is necessary to identify workers who are not wearing PPE in a timely manner to ensure the platform's safe

✉ Faming Gong
19930033@upc.edu.cn

Xiaofeng Ji
xiaofengji1996@163.com

Xiangbing Yuan
yuanxiangbing.slyt@sinopec.com

Nuanlai Wang
nuanlaiwang@163.com

¹ College of Computer Science and Technology, China University of Petroleum (East China), Changjiang Road Street, Qingdao 266580, Shandong, China

² Offshore Oil Production Plant Information Service Center, Sinopec Grp Offshore Oil Prod Plant, Heilongjiang Road, Dongying 257000, Shandong, China

production. Currently, the offshore drilling platform is primarily used to detect unsafe human behavior through manual review of surveillance videos. However, manual supervision is not without challenges: (1) fatigued supervisors can make mistakes, overlook omissions, and make mistakes in judgment; (2) subjective influences from the monitor's emotions, state of mind, etc., can distort judgment; and (3) the platform has a large number of surveillance cameras, making it difficult to view all of the surveillance images manually. Therefore, it is urgent to eliminate manual intervention and to achieve automation and intelligence in the detection of PPE on the offshore drilling platform.

Initially, researchers focused on sensor-based methods to achieve PPE detection on construction sites [2,3]. However, the method required expensive equipment, which not only increased the cost of industrial production but also posed a potential risk to workers' health. With the rise of the Industry 4.0 era [4], the use of artificial intelligence technology to achieve the intelligence of industrial production is gradually becoming a trend. Before the emergence of deep learning, image processing combined with machine learning was mainly used to detect whether workers' helmets and workwear were worn correctly [5–7]. However, this method is only applicable to scenes with few distractors and performs poorly in complex scenes. After the emergence of deep learning, the use of techniques such as object detection [8] to achieve PPE detection has become the mainstream approach [9–12]. The method has greatly improved the detection speed and accuracy and is gradually being applied to industrial production. However, most of them were developed for construction sites and indoor industrial production scenarios, and there was no PPE detection method developed for offshore drilling platform.

This paper proposes a multi-algorithm fusion framework that enables fast and accurate PPE detection of the offshore drilling platform, with a simple deployment process that does not require the procurement and installation of complex sensors; moreover, the modified object detection algorithm improves the accuracy of object localization and identification, and is more suitable for object detection in complex scenarios on offshore drilling platforms. In addition, our further processing of the object detection results can satisfy the PPE detection under complex postures and can accurately identify the behavior of abnormal helmet wearing such as holding helmets. Therefore, the proposed framework in this paper is of great significance for PPE detection on offshore platforms. The main contributions of this paper are as follows.

- (1) We propose an RFA-YOLO object detection algorithm, which is based on a residual feature augmentation network with YOLOv4 as the baseline for reducing the loss of high-level feature information.
- (2) A method is proposed to construct position features based on the object bounding box, which characterizes the relative offsets of the person and the helmet, as well as the person and the workwear.
- (3) In this paper, the detection of PPE on the offshore drilling platform based on object detection and classification is an important breakthrough in the realization of monitoring intelligence in the offshore oil field.

Literature review

The current mainstream PPE detection methods fall into two categories: sensor-based methods, and computer vision-based methods. The sensor-based approach involves installing sensors and analyzing their signals to determine whether workers are wearing workwear or safety helmets based on the signals. Zhang et al. [13] developed a visualized smart helmet identification system based on Radio Frequency Identification (RFID) using an Internet of Things (IoT) architecture. Agnes Kelm [3] used a mobile RFID to check the PPE compliance of personnel. Another example is using RFID technology to determine when a worker makes use of the PPE, which performs mesh network communications using Zigbee [14]. However, sensor-based methods require significant investments in procurement, installation, troubleshooting [15] and maintenance, making them difficult to scale up in practice.

With the continuous progress of information technology and industrial technology, modern industrial production is developing in the direction of high speed, precision and intelligence, and data-driven [16,17] abnormal event diagnosis methods are gaining more and more attention and development, including the application of visual data for PPE detection. The vision-based methods can be divided into two categories: one is the traditional method of image processing combined with machine learning [5–7,18–20]; the other is using deep learning technology, e.g., object detection [9–12,21–28]. In the traditional approach, the region of interest (e.g., the head region or torso region) is usually first located using image processing techniques, and then image features [29–31] are extracted and machine learning methods are used to train a classifier to determine whether the region is a helmet or workwear. For example, Li et al. [19] used the ViBe background modeling algorithm and the pedestrian classification framework to identify workers. Then, they localized the head region and used color space transformations and color feature recognition to detect helmet. Cai et al. [7] constructed edge images of safety helmets at different angles, extracted four directional features, and designed a safety helmet-non-safety helmet classifier by modeling the feature distribution with a Gaussian function.

With the development of deep learning, many researchers apply object detection and other technologies to PPE detection. The proposal of R-CNN [32] makes the object detection algorithm get out of the bottleneck period and greatly improve the detection speed and accuracy. Subsequently, a series of object detection algorithms based on the candidate boxes have been proposed and are being used in intelligent video surveillance. For example, the method based on Faster R-CNN [33] solves the problem of detecting construction workers wearing helmets under remote monitoring [9]. Fan et al. [10] compared the principles and performance of several object detection algorithms for helmet detection, where Faster R-CNN has high accuracy for large-scale target detection.

Although the object detection algorithm based on candidate frames achieves high accuracy, it has poor real-time performance. Therefore, many researchers have applied single-stage object detectors to the detection of safety helmets and workwear. Han et al. [12] added an attention mechanism to the feature fusion phase of the SSD [34] algorithm to refine the feature information in the target region to improve the safety helmet detection accuracy. Iannizzotto et al. [26] proposed a PPE detection framework combining object detection and manual judgment, and embedded it into end devices to achieve real-time detection. Gallo et al. [27] completed a real-time PPE detection system based on deep neural networks (DNNs) by employing an edge computing model. In addition to this, the YOLO series of algorithms [35–37] are also widely used for helmet detection. Fan et al. [10] and Wang et al. [11] used different methods to improve the YOLOv3 [37] algorithm for industrial helmet detection. Wang et al. [24] constructed the high-quality dataset and used different versions of YOLO to achieve the detection of six classes of objects (four colors of helmets, person, and vest), which validated the excellent performance of YOLOv5x for PPE detection.

With the complexity and diversity of industrial environments, target detection techniques alone are no longer able to meet the accuracy requirements that are required for PPE detection. The multi-algorithm fusion model combining object detection with other algorithms has gradually become a trend for solving the problem of PPE detection in complex scenes and complex poses in recent years. Nath et al. [22] proposed three PPE detection frameworks based on the YOLO algorithm combined with convolutional neural networks and decision trees, respectively. Xiong et al. [23] broke the traditional approach by obtaining the human key point areas through pose estimation, localizing partial attention areas (head and upper body) using the intrinsic relationships of body parts, and then using a CNN-based classifier to identify the classes of partial attention areas.

In summary, many scholars have conducted in-depth research on PPE detection and achieved remarkable results,

but the above methods still have certain limitations when applied to offshore drilling platform scenarios. For example, the wet environment at sea can cause damage to the sensor, making it fail to work properly, and replacing it would be costly. The pipelines of the offshore drilling platform are crisscrossed and similar to the morphology of the workers and the color of their clothes, and the complex scenes make the error rate of the PPE detection methods based on image processing high. Platform workers have complex postures and often hold helmets in their hands, etc. Existing PPE detection methods based on object detection only recognize the presence of helmets in the image, but it is difficult to accurately determine whether the helmet is located on the worker's head, and there is a lack of relevant research on workwear detection. Therefore, the purpose of this paper is to overcome the above problems and propose a high performance method that enables fast and accurate PPE detection on the offshore drilling platform.

Methodology

In this section, we introduce the proposed framework for PPE detection based on multi-algorithm fusion, detailing the components of the detection framework sequentially.

The proposed PPE detection framework

As shown in Fig. 1, the detection framework proposed in this paper consists of three parts: object detection, feature construction and feature classification. Of which, the object detection part is to identify and locate the person, helmet and workwear targets by the proposed RFA-YOLO algorithm, then outputs the class and the bounding box coordinate information; the feature construction part is to use the coordinate information of the object bounding box to construct the position features; the feature classification part is to use the feature classifier to determine whether the helmet is being worn on the head position and whether the workwear is being worn on the worker. The input image undergoes object detection, feature construction and feature classification, and then outputs the helmet/workwear wearing situation, and finally outputs the detection result of worker's PPE wearing through inference mechanism. The detection results contained four categories: the worker is both wearing a helmet and workwear (PHW); the worker is only wearing a helmet (PH); the worker is only wearing workwear (PW); and the worker is neither wearing a helmet nor wearing workwear (P).

RFA-YOLO-based object detection

Detecting objects is the first step to achieving PPE detection, and the accuracy of the object detection directly affects the

results of the final detection. To improve the localization and recognition accuracy of the person, helmet, and workwear, we propose the RFA-YOLO algorithm based on the residual feature augmentation (RFA) module [38] with YOLOv4 [39] as the baseline.

Figure 2 illustrates the structure of the YOLOv4 network. The input image is passed through the CSPDarknet53 network, which contains five CSP modules for feature extraction and outputs feature maps of various scales. Assume that the feature maps C_3 , C_4 , and C_5 are generated after the third, fourth, and fifth CSP modules, respectively. The feature map C_5 undergoes the Spatial Pyramid Pooling (SPP) [40] module to realize the fusion of local features and global features, which can enhance the semantic information of the feature map. Next, the fusion of feature maps is realized by PANet

[41]. The convolution and sampling operations are performed on three scales of feature maps, and the number of channels and resolution size of feature maps are adjusted to obtain three feature maps of M_3 , M_4 and M_5 . Then, the fusion of feature maps is realized through concat. The final prediction results for the three scales are obtained.

However, the generation of M_5 from C'_5 needs to undergo a 1×1 convolution to reduce the number of channels, resulting in partial loss of feature information. The loss of such information may contain important semantic information, which would prevent the subsequent feature fusion from passing on that semantic information, thereby affecting the accuracy of the final detection. Therefore, this paper proposes the RFA-YOLO algorithm, which adds a RFA module after the SPP module to reduce the loss of feature information. The major

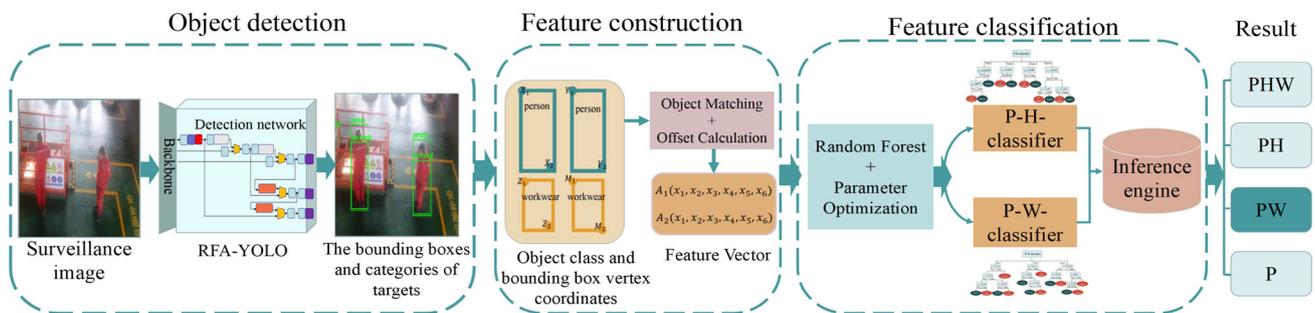


Fig. 1 The proposed framework for PPE detection

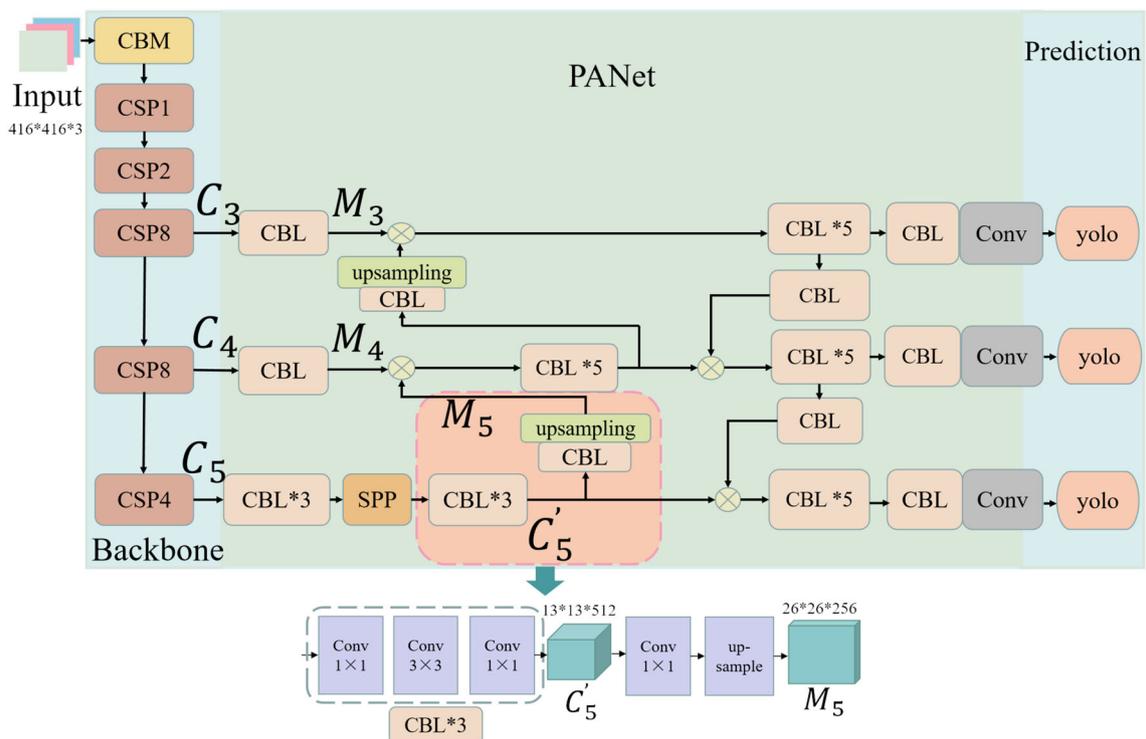


Fig. 2 The network structure of YOLOv4

concept is to use residual branches to integrate different spatial contextual information into the original branches, which can enrich the information contained in the feature map M_5 . Figure 3 shows the processing of the RFA.

To begin, scale-invariant adaptive pooling is used to generate three scale feature maps from the C_5 with scale S : X^1 , X^2 , and X^3 ($X^i = \alpha_i \times S$). Then, the contextual features of each scale are separately input to the 1×1 convolution layer, reducing the feature channel dimension to 256. The reduced dimension feature maps are upsampled separately to produce three scale S feature maps: Y^1 , Y^2 , and Y^3 . Since the interpolation of the up-sampling process can cause confounding effects, the fusion of feature maps is achieved by an adaptive spatial fusion (ASF) module. ASF takes the sampled feature maps Y^1 , Y^2 , and Y^3 as input and obtains the spatial weight map W_i for each feature after the path L_1 . Meanwhile, the features Y^1 , Y^2 , and Y^3 are fused based on the weights after the path L_2 to obtain the feature map M_6 . Suppose m_{ij} denotes the feature vector of feature map M_6 at position (i, j) , which is a weighted fusion of the vectors of the three feature maps at position (i, j) , as shown in Eq. 1, where Y_{ij}^l denotes the ij -th feature vector on feature map Y^l ($l \in \{1, 2, 3\}$) and W_{ij}^l denotes the spatial importance weights of different feature maps, and they are shared among all channels. After the feature fusion to generate M_6 contains multi-scale contextual information. Meanwhile, the M'_5 is obtained after the 1×1 convolution of C'_5 , and the feature map M''_5 is obtained by fusing the information of the corresponding channels of M_6 and M'_5 , as shown in Eq. 2, where M'_{5ij} denotes the feature vector of the feature map M'_5 at position (i, j) , M''_{5ij} implies the (i, j) -th vector of the output feature maps M''_5 among channels. Then, M''_5 is upsampled to obtain M_5 . Through ASF, it is possible to increase the amount of information in M_5 in each dimension while keeping the dimensionality constant. The generated M_5 is finally fused with the low-level feature maps M_3 and M_4 .

$$m_{ij} = Y_{ij}^1 \cdot W_{ij}^1 + Y_{ij}^2 \cdot W_{ij}^2 + Y_{ij}^3 \cdot W_{ij}^3, \tag{1}$$

$$M''_{5ij} = M'_{5ij} + M_{6ij}. \tag{2}$$

Feature construction based on the object bounding box

After the object detection, the person, helmet, and workwear in the image are located and identified, and the category and bounding box coordinates are output. It can only detect if the image contains a person, a helmet, and workwear, but it cannot detect whether the helmet is positioned on the head of the worker or whether the worker is wearing workwear. To determine the position relationship of the objects, this paper constructs position features based on the object bounding box coordinates, which are used to characterize the relative offsets of the person and the helmet, as well as the person and the workwear. First, we propose an object matching mechanism to pair the person, safety helmet, and workwear according to the intersection ratio size of the bounding boxes. Then, the relative differences of the vertices of each group of bounding boxes are calculated and normalized to obtain the position features.

Object matching mechanism

As shown in Fig. 4, the classes and bounding boxes are output after the object detection, and the group of bounding boxes consisting of the person, helmet, and workwear is obtained through the object matching mechanism. We suppose that P_i, H_i, W_i represent the person, helmet, and workwear bounding boxes, respectively.

$$IoU = \frac{area(P_i) \cap area(E_i)}{area(P_i) \cup area(E_i)}. \tag{3}$$

The overall structure of the object matching mechanism is shown in Algorithm 1. First, we select the box P_i in the set of person bounding boxes, then select the helmet/workwear bounding box $E_j \in \{H, W\}$ from the set of PPE, calculate the intersection ratio between the bounding box P_i and the bounding box E_i with Eq. 3, and save the calcu-

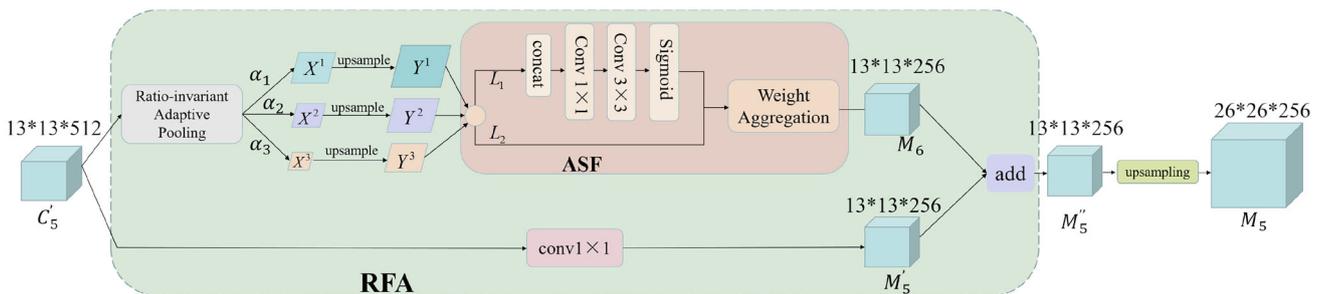


Fig. 3 The network structure of RFA

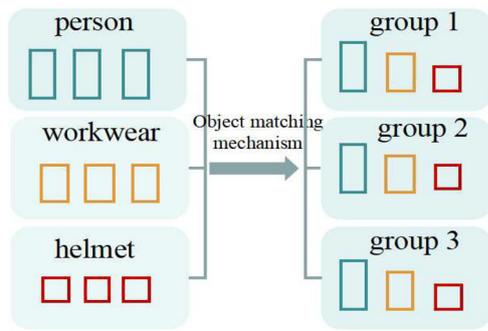


Fig. 4 Match the object detection results

Algorithm 1 The object matching mechanism

- 1: Let S be the set of the intersection of person and the PPE
- 2: Let C be the set of bounding boxes after the matching
- 3: **for** each detected box of person P_i **do**:
- 4: **for** each type of PPE $E \in \{H, W\}$ **do**:
- 5: **for** each detected box E_i of type E **do**:
- 6: Calculate the intersection ratio IoU_i between P_i and E_i (Eq. 3)
- 7: **end for**
- 8: Insert IoU_i in the set S
- 9: Calculate the E_{max} corresponding to the maximum and non-zero value in the set S
- 10: **end for**
- 11: Insert E_{max} in the set C
- 12: **return** C
- 13: **end for**

lation result into the set S . Then, find the maximum and non-zero value in the set S corresponding to the protective equipment E_{max} , which is the bounding box matching the person box P_i . Finally, we obtain the object matching result $C = \langle P_i, E_{max} \rangle$, and $E_{max} \in \{(H, W) \| H \| W \| \emptyset\}$.

The construction of the position feature

After object matching, the set of bounding boxes consisting of the person, helmet, and workwear is obtained, and we use the offsets between the bounding boxes to construct position features that represent the relative positional relationships of the objects. The workers who are incorrectly wearing PPE such as holding a helmet in their hands are further identified based on the position features. Figure 5 shows the bounding box of the object detection output, where the left subfigure is correctly wearing PPE and the worker holding a helmet in the right subfigure. As can be seen from the figure, the relative offset between the helmet and the person in the right subfigure is significantly different from that in the left subfigure.

During the construction of the position features, the top-left vertex, the center point, and the bottom-right vertex of the bounding box are selected as the reference points for calculating the relative offsets. We assume that P_i represents the person’s bounding box and $E_i \in \{H, W\}$ represents the per-

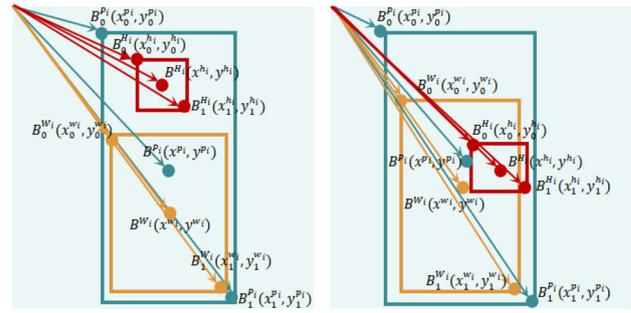


Fig. 5 The object bounding boxes. The left subfigure is the set of bounding boxes that contend with the correct wearing of PPE and the right subfigure is the set of bounding boxes with the incorrect wearing of PPE

sonal protective equipment bounding box. The coordinates of the center point of the bounding box are calculated according to Eq. 4, where $C_i \in \{P, H, W\}$, the point (x_0, y_0) represents the coordinates of the top-left vertex of the bounding box, and (x_1, y_1) represents the coordinates of the bottom-right vertex of the bounding box, which are obtained from the object detection. The point (x, y) represents the coordinates of the center point of the bounding box. Then, the height and width of the person box P_i are calculated using Eq. 5, and the offset between box P_i and box E_i is calculated using Eq. 6. The difference between the three vertices is first calculated and then normalized using the height and width of box P_i to avoid the effect caused by the different sizes of the object scale. Finally, the offset is transformed into a position feature matrix using Eq. 7, which contains six feature elements.

$$x^{C_i} = \frac{x_0^{C_i} + x_1^{C_i}}{2}, y^{C_i} = \frac{y_0^{C_i} + y_1^{C_i}}{2}, \tag{4}$$

$$\begin{bmatrix} w^{P_i} \\ h^{P_i} \end{bmatrix} = \begin{bmatrix} x_1^{P_i} \\ y_1^{P_i} \end{bmatrix} - \begin{bmatrix} x_0^{P_i} \\ y_0^{P_i} \end{bmatrix}, \tag{5}$$

$$\begin{bmatrix} x_0^{E_j, P_i} & x^{E_j, P_i} & x_1^{E_j, P_i} \\ y_0^{E_j, P_i} & y^{E_j, P_i} & y_1^{E_j, P_i} \end{bmatrix} = \begin{bmatrix} \frac{1}{w^{P_i}} & 0 \\ 0 & \frac{1}{h^{P_i}} \end{bmatrix}. \tag{6}$$

$$F = \begin{bmatrix} x_1 & x_3 & x_5 \\ x_2 & x_4 & x_6 \end{bmatrix} = \begin{bmatrix} x_0^{E_j, P_i} & x^{E_j, P_i} & x_1^{E_j, P_i} \\ y_0^{E_j, P_i} & y^{E_j, P_i} & y_1^{E_j, P_i} \end{bmatrix}. \tag{7}$$

Feature classification based on optimized random forest

With the position features, a classifier needs to be designed based on the features. When the position feature is input to

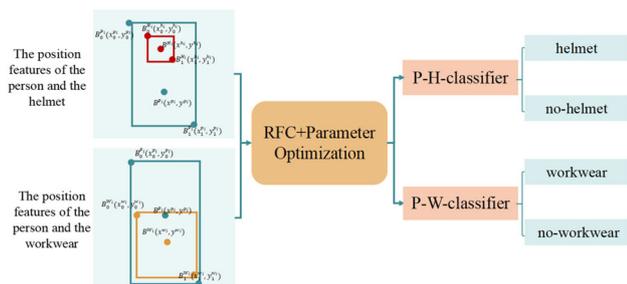


Fig. 6 The construction of feature classifier based on optimized random forest

the classifier, it is able to output the category to which the feature belongs. As shown in Fig. 6, to improve the accuracy of the classifier, this paper adopts parameter optimization to improve the classification model based on the random forest classification (RFC) algorithm to construct the person and helmet feature classifier (P–H-classifier) and the person and workwear feature classifier (P–W-classifier), respectively. According to the output of the classifier, the position relationship between the person and the helmet is evaluated: on the head or not on the head; at the same time, the position relationship between the person and the workwear is also evaluated: on the body or not on the body.

In this paper, the Gini index is used as the criterion for splitting decision branches to train the classification model. The conditions for nodes to stop recursion are set as follows: the Gini index is less than the threshold; the number of samples is less than the threshold or no selected features are available. The dataset consisting of the position features of the person and helmet is used as the sample for training the P–H-classifier model, and the dataset consisting of the position features of the person and workwear is used as the sample for training the P–W-classifier model. To improve the accuracy of the model, the number of decision trees and the maximum number of candidate features are adjusted to find the optimal combination of parameters. Usually, as the number of decision trees increases, the loss of the test dataset decreases gradually, and when it reaches a certain number, the loss of the test dataset is almost unchanged; if the number of decision trees continues to increase, the loss of the test dataset becomes larger instead, and the overfitting phenomenon appears. The maximum number of candidate features is the number of features that can be randomly selected by each node when generating the decision tree. The maximum number of candidate features not only affects the model accuracy, but also affects the classification speed of the model. After the parameter optimization, an optimal classification model consisting of multiple decision trees is obtained. Once the position features are input to the classification model, multiple classification results are output, and the category with the highest number of votes is designated as the final output.

Results and evaluation

Since there is no public dataset of the offshore drilling platform, this paper collected the offshore drilling platform monitoring images to produce an experimental dataset. Also, to verify the effectiveness of the proposed framework, we design a series of comparison experiments to evaluate the performance of the method through the experimental results.

The experimental dataset

The offshore drilling platform dataset (ODPD) consists of three parts: object detection dataset (ODD), feature classification dataset (FCD), and personal protective equipment dataset (PPED). Of these, the ODD is used to train and test the object detection model, the FCD is used to train and test the performance of the classifier, and the PPED is used to test the effectiveness of the framework.

(1) ODD. We collected surveillance images containing workers on the platform and used the graphical image annotation tool LabelImage to annotate the person, helmet, and workwear in the images. The XML file containing the categories and coordinates is obtained and used to train and evaluate the object detection model. The ODD contains a total of 10,000 images, including surveillance images from all cameras.

(2) FCD. We divided the ODD into four groups: wearing helmets correctly, wearing helmets incorrectly, wearing workwear correctly, and wearing workwear incorrectly. Then, the category and coordinate in the XML file are obtained, and the matching mechanism in Section “Object matching mechanism” is used to match the person box P_i and the PPE box E_i , and the coordinate are processed by the proposed position feature construction method to obtain the position features $F_i = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ of P_i and E_i . When E_i is equal to H , the position features of the person and the helmet are obtained, and the correct wearing of the helmet is defined as a positive sample and the incorrect wearing of the helmet is defined as a negative sample to obtain the dataset P–H $= \{F_i, y\}, y \in \{0, 1\}$, where F_i is the feature vector and y is the category, which is used to train and evaluate the P–H-classifier; when E_i is equal to W , the feature dataset of the person and workwear P–W is obtained, which is used to train and evaluate the P–W-classifier. The P–H dataset and the P–W dataset contain 6600 samples, respectively, of which 3300 are positive samples and 3300 are negative samples.

(3) PPED. The images containing workers were extracted from the historical surveillance video of the offshore drilling platform, and the workers in the images were annotated using the graphical image annotation tool LabelImage, where the samples wearing both helmets and workwear were annotated as “PHW”, those wearing only helmets were annotated

Table 1 Comparison results of various object detection algorithms

Model	Backbone	Proposed time	mAP	Person	Helmet	Workwear
YOLOv3	Darknet-53	2018	83.32	85.45	80.63	81.18
YOLOv4	CSPDarknet-53	2020	85.34	86.96	82.32	82.75
YOLOv5s	CSPDarknet-53	2021	86.63	87.91	83.45	83.82
RFA-YOLO	CSPDarknet-53	–	88.41	89.32	84.21	84.72

Bold values indicate the highest values for the same metric in different models. In the above object detection model, the mAP had the highest value of 88.41, the average precision (AP) of Person had the highest value of 89.32, the AP of Helmet had the highest value of 84.21, and the AP of Workwear had the highest value of 84.72

as “PH”, those wearing only workwear were annotated as “PW”, and those wearing neither helmets nor workwear were annotated as “P”. The PPED dataset contains a total of 2,000 images, in which the number of samples for “PHW” was 821, the number of samples for “PH” was 534, the number of samples for “PW” was 682, and the number of samples for “P” has a sample size of 473.

The experimental design and analysis of results

This section evaluates the performance of each part of the proposed framework. The experiments are divided into three parts: the evaluation of object detection algorithms; the evaluation of position features and the classifier; and the comparison and analysis of PPE detection algorithms.

The evaluation of the object detection model

The YOLO series algorithms are widely used in the field of intelligent video surveillance because of the high detection speed and high detection accuracy. The proposed RFA-YOLO in this paper adopts YOLOv4 as the baseline and uses the RFA module to enhance detection accuracy. To evaluate the performance of the proposed RFA-YOLO algorithm, this paper compares the method with YOLO series algorithms, e.g., YOLOv5s, YOLOv4, and YOLOv3. The mAP and AP are used as evaluation metrics to assess the performance of various object detection algorithms. We randomly divide the ODD dataset into the training set, validation set, and test set according to the ratio of 7:2:1. During the training process, the training epoch is set to 120, and the model with the lowest loss in the validation set is used as the model for object detection, and the model performance is tested on the test dataset.

Table 1 shows the comparison results of model performance with the backbone network structure of the algorithm, proposed time, mAP, and the AP of the person, helmet and workwear. As shown in the table, our algorithm performs better on the ODD dataset. Therefore, we can conclude that the residual feature augmentation module can reduce the loss of effective features, and the inclusion of the RFA module

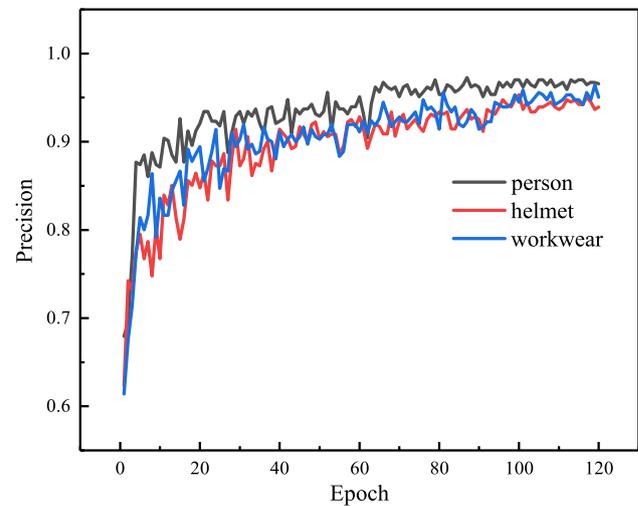


Fig. 7 The precision of each class based on RFA-YOLO

can effectively improve the accuracy of detection compared with the YOLOv4 algorithm.

Figure 7 shows the precision of the person, helmet and workwear based on the RFA-YOLO. With the increase of the training iterations, the precision gradually increase first and then stabilize. Of which, the precision of the helmet and workwear is significantly lower than that of the person, which may be related to the size and characteristics of the object, the helmet is small and cannot be easily detected, besides, the repetition rate of the features of the workwear and the person is larger, which will affect the detection precision of the workwear. Figures 8, 9 and 10 show the visualization of the detection results of different methods. The white box in the figure indicates the case of missed detection or wrong detection. From the figure, the proposed RFA-YOLO algorithm can more accurately locate and identify the person, helmet, and workwear.

The evaluation of position features and the classifier

(1) Parameter optimization. The performance of the random forest classifier is highly dependent on the number of decision trees and the number of candidate features. To obtain

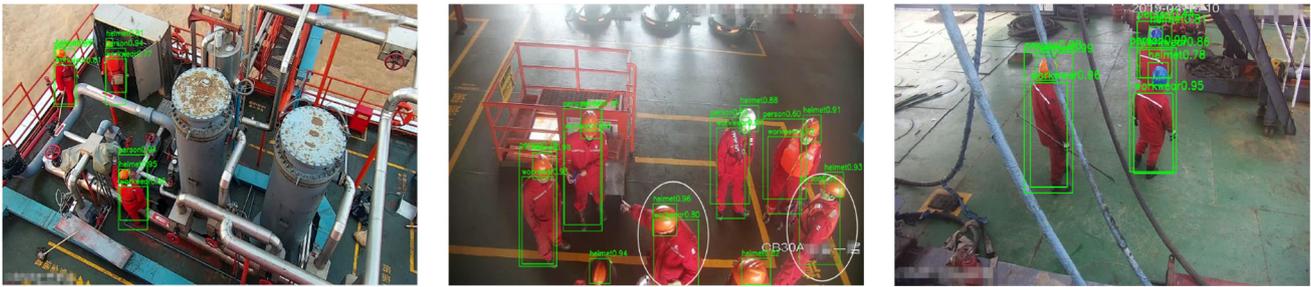


Fig. 8 Detection results of objects based on RFA-YOLO

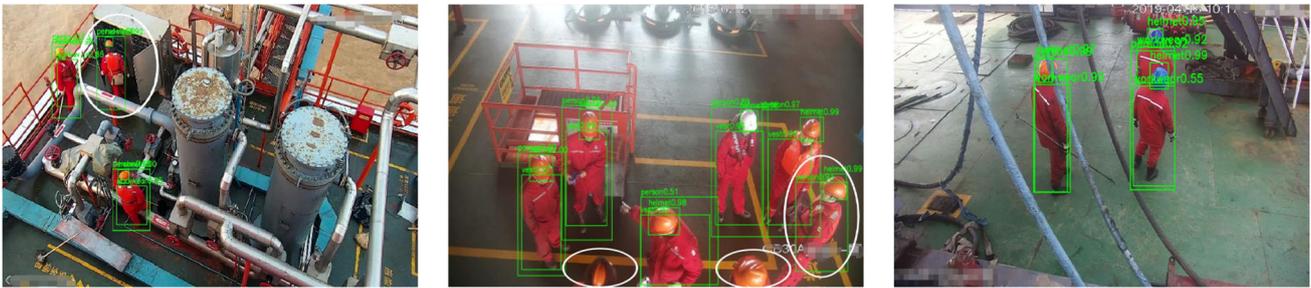


Fig. 9 Detection results of objects based on YOLOv4

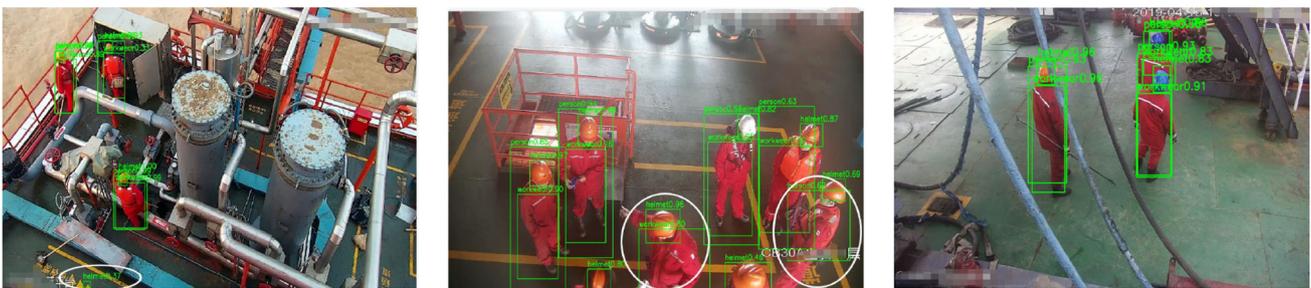


Fig. 10 Detection results of objects based on YOLOv5s

the optimal feature classifier, we explore the combination of parameters that enables the best performance of the classifier. In this paper, each feature vector in the FCD dataset contains 6 features. In the parameter optimization experiments, we set the number of candidate features from 1 to 6, and analyze the change of model classification accuracy as the number of decision trees increases under the condition that the number of features is fixed. Figure 11 shows the accuracy of the P–W-classifier obtained on the P–W dataset with the variation of parameters. As can be seen from the figure, when the number of features is fixed, the classifier performance first tends to increase with the number of decision trees, and then slowly decreases and stabilizes after reaching the peak. Normally, to ensure the diversity and mutual independence of decision trees, the number of candidate features is set to $N/2$ or \sqrt{N} , and in this paper, the corresponding number of features is 3 and 2. As can be seen from Figure, the accuracy of the generated classifier is significantly lower when the number of candidate features is set to 2 than when the number of

candidate features is set to 3. Therefore, we set the candidate feature parameter for training the P–W-classifier to 3 and the number of decision trees parameter to 3.

Figure 12 shows the accuracy of the P–H-classifier obtained by training the P–H dataset with the variation of parameters. As can be seen from the figure, the classifier performance tends to increase with the number of decision trees when the number of candidate features is fixed and stabilizes after reaching a certain value. Based on the diversity and mutual independence of decision trees, we set the number of candidate features for training P–H-classifier to 3 and the number of decision trees to 4 to ensure the accuracy of the classifier.

(2) Feature importance assessment. The location feature vector in this paper contains six features, and to analyze the features that have the greatest impact on the model, we use a Gini index-based approach to evaluate the importance score of each feature in the feature vector. Assuming that the feature importance score is FI and the Gini index of the feature

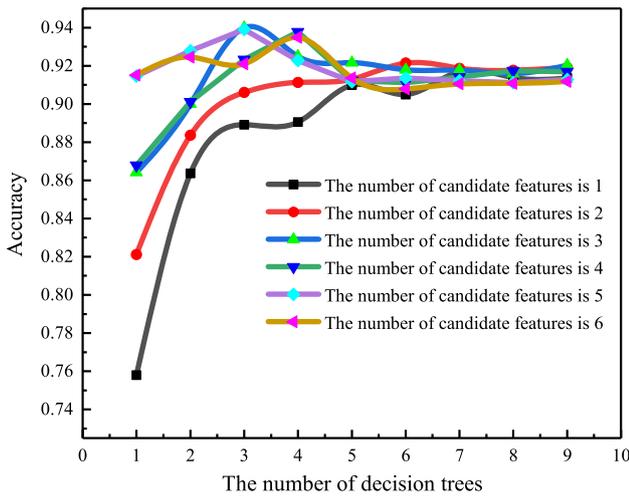


Fig. 11 The performance of P-W-classifier with parameters

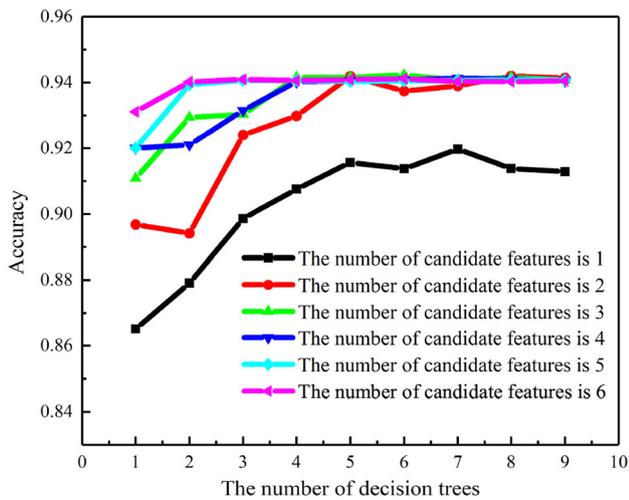


Fig. 12 The performance of P-H-classifier with parameters

is GI , the Gini index of each feature X_j is first calculated, and the Gini index of the i th decision tree node q is calculated according to the Eq. 8, where C denotes the number of categories and indicates the proportion of category c in node q . The importance of feature X_j in the i th decision tree node q is also the amount of change in Gini index before and after the branching of q , which can be calculated by Eq. 9 where GI_l^i and GI_r^i denote the Gini indices of the two new nodes after branching, respectively. If the nodes where the feature X_j appears in the decision tree i are the set Q , then the importance of X_j in the i th tree can be calculated by the Eq. 10. Assuming that there are a total of I decision trees, the importance of feature X_j on I decision tree can be calculated by using Eq. 11. Finally, the importance score of feature X_j is obtained by normalizing it with Eq. 12 where J denotes the number of features contained in the feature vector.

Table 2 shows the importance scores of the features obtained through the Gini index-based approach. As can be

seen from the table, the feature with the highest importance score in the P-W-classifier is x_4 , which represents the longitudinal offset between the center point of the person and the center point of the workwear; in the P-H-classifier, the features with the highest importance scores are x_3 and x_6 , which represent the lateral offset between the person and the center point of the helmet and the longitudinal offset of the top point of the lower right corner, respectively.

$$GI_q^i = 1 - \sum_{c=1}^C (p_{qc}^i)^2, \tag{8}$$

$$FI_{jq}^{Gini_i} = GI_q^i - GI_l^i - GI_r^i, \tag{9}$$

$$FI^{Gini_i} = \sum_{q \in Q} FI_{jq}^{Gini_i}, \tag{10}$$

$$FI^{Gini} = \sum_{i=1}^I FI^{Gini_i}, \tag{11}$$

$$FI^{Gini} = \frac{FI^{Gini}}{\sum_{j=1}^J FI_j^{Gini}}. \tag{12}$$

(3) Comparison of classifier models. To evaluate the performance of the proposed classifier, we compare and analyze the classification model in this paper with the position feature classifier proposed by Nath [22]. Nath et al. constructed the feature vector using bounding box coordinates containing four position features and then trained it based on Classification and Regression Trees (CART) to obtain the classification model. Table 3 shows the performance comparison results of the classifiers, and it can be concluded from the table that the classifier obtained by our method performs better. Figures 13 and 14 show the two feature classifiers generated based on the parameter optimization method, and we selected the best performance on the test set as the final classifier.

The evaluation of the PPE detection framework

In this subsection, comparative experiments are designed to compare the performance of the proposed framework in this paper with other PPE detection methods. Since most of the current PPE detection addresses the helmet identification problem, we divide the comparison experiments into two categories, one is the comparison between the method in this paper and other helmet identification methods, in which our framework only identifies helmets and workers in the target detection phase and subsequently uses the P-H-classifier to achieve helmet wear identification; the other is the comparison between the method in this paper and other PPE detection methods (identifying helmets and workwear). In practical applications, it is expected that the model can guarantee high accuracy and real-time performance. From the accuracy point of view, it is expected that the model can identify workers

Table 2 The importance score of each feature

Classifier	x_1	x_2	x_3	x_4	x_5	x_6
P-W-classifier	0.058569	0.052845	0.402222	0.269858	0.164441	0.052065
P-H-classifier	0.005818	0.002023	0.493130	0.001609	0.003884	0.493536

Table 3 The accuracy of classifier based on different method

Feature classifier	Nath [22]	Ours
P-H-classifier	0.903	0.946
P-W-classifier	0.892	0.933

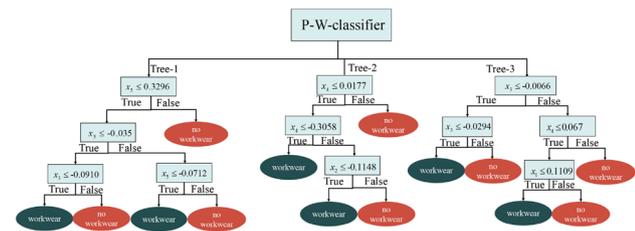


Fig. 13 The generated P-W-classifier

who are not wearing PPE, but also guarantee the reliability of identification, that is, having acceptable false alarm rate and missed alarm rate, and from the real-time point of view, it is expected that the model can complete the identification task quickly. Therefore, we use accuracy, recall, false alarm rate (FAR), missed alarm rate (MAR), and detection time as evaluation metrics. Where, accuracy denotes the probability that the model can correctly identify the dressed PPE and the unworn PPE; Recall₁ denotes the probability that the model correctly identifies the dressed PPE; Recall₂ denotes the probability that the model correctly identifies the unworn PPE; FAR denotes that the model identifies the workers who correctly wear PPE as the unworn PPE, which is the false alarm rate; MAR denotes that the model identifies the workers who do not correctly wear PPE as the correctly dressed PPE, which is the missed alarm rate; and detection time denotes the average time for the model to detect an image. Accuracy and recall can evaluate the model recognition accuracy, FAR and MAR can evaluate the reliability of the model, and the detection time can evaluate the real-time performance of the model.

(1) Comparison experiment of helmet detection.

We compare the performance of this paper’s method with other PPE detection methods (detecting helmets) on the PPED dataset, using wearing helmets as a positive class and not wearing helmets as a negative class. According to the experimental results, the values of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) were counted, respectively, and the statistical results are shown in Table 4. The values of each indicator were

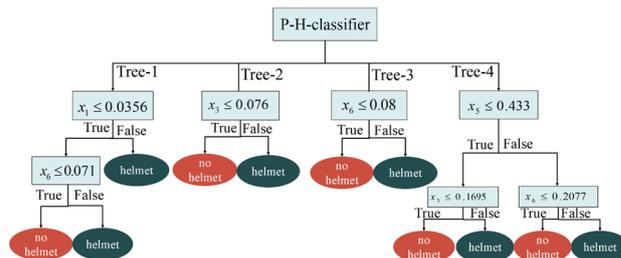


Fig. 14 The generated P-H-classifier

calculated according to Eqs. 13–17, and the performance comparison results of different methods were obtained, as shown in Table 5. As can be seen from table, the proposed method in this paper performs best in terms of model accuracy, reliability and real-time performance. Shen et al. [25] locate the head region based on face recognition results, and then realize helmet recognition using image classification methods. On the offshore drilling platform, many workers in the monitoring screen do not show their faces or do not show their full faces, so the face region cannot be accurately identified, and the head region cannot be located, resulting in model recognition accuracy and reliability is poor. In addition, the method consists of two steps of face recognition and image classification to complete helmet recognition, which leads to poor real-time performance. The proposed method in this paper is based on target detection, which only needs to accurately identify and locate the person and helmet without considering whether the worker shows his face or not, so it performs better on the PPED. Li et al. [19] locate the head position based on the proportional relationship between the head region and the body region, and then use color features to distinguish whether workers wear helmets or not. This method is more applicable to workers walking upright, and the workers on the offshore drilling platform often need to squat or bend down due to work requirements, making it impossible to accurately locate the head position based on the head-to-body ratio, resulting in a low accuracy of final PPE recognition. Since most workers in the PPED do not show their faces or do not show their full faces, Shen et al.’s method performs the worst on this dataset. The method in this paper overcomes these problems and is, therefore, more suitable for helmet detection on the offshore drilling platform.

To better verify our conclusions, we select images containing workers with complex postures from PPED to form dataset CP, select images containing workers with no face or no full face to form dataset NF, and select images con-

Table 4 The statistics of experimental results of different helmet detection methods

Method	TP	FN	TN	FP
Shen [25]	902	453	887	268
Li [19]	978	377	956	199
Ours	1152	203	1053	102

taining workers with full faces and upright postures to form dataset FU. Each dataset contains 300 images each, and conduct comparison experiments on datasets CP, NF and FU, respectively, using accuracy as a metric for evaluation. The experimental results in Table 6 show that the effectiveness of Li et al.'s method is significantly reduced on the CP dataset, and Shen et al.'s method performs the worst on the NF dataset, while when the performance of the above three methods is not much different, it also shows that the proposed method is more applicable to offshore drilling platforms.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \quad (13)$$

$$recall_1 = \frac{TP}{TP + FN}, \quad (14)$$

$$recall_2 = \frac{TN}{TN + FP}, \quad (15)$$

$$FAR = \frac{FN}{TP + FN}, \quad (16)$$

$$MAR = \frac{FP}{FP + TN}. \quad (17)$$

(2) Comparison experiment of PPE detection. To verify the effectiveness of the proposed framework in this paper, we compared the performance with other PPE detection methods on the PPED, with correctly dressed PPE as the positive category and incorrectly dressed PPE as the negative category. The values of TP, FP, TN and FN were counted according to the experimental results, and the statistical results are shown in Table 7. The values of each indicator were calculated according to Eqs. 13 to 17, and the performance comparison results of different methods were obtained, as shown in Table 8. From the table, it can be seen that the proposed framework in this paper can complete PPE detection more accurately and quickly. From the methods used in the literature, Nath [22], Xiong [23] and the method in this paper all use multiple algorithms and multiple stages to achieve PPE detection.

Table 5 The comparison of each metric of different helmet detection methods

Method	Accuracy	Recall ₂	Recall ₁	MAR	FAR	Time(s)
Shen [25]	0.768	0.768	0.666	0.232	0.334	0.12
Li [19]	0.770	0.828	0.722	0.172	0.278	0.09
Ours	0.878	0.912	0.850	0.088	0.150	0.06

Table 6 Accuracy of existing helmet detection methods on different Datasets

Dataset	Shen [25]	Li [19]	Ours
CP	0.764	0.327	0.832
NF	0.223	0.782	0.864
FU	0.923	0.911	0.934

Table 7 The statistics of experimental results of different PPE detection methods

Method	TP	FN	TN	FP
Nath [22]	703	118	1328	361
Xiong [23]	752	69	1421	268
Ours	782	39	1556	133

Among them, Nath et al. [22] also use a combination of object detection and classification, but the object detection model they use is different from this paper, and they use YOLOv3 as the object detection network. As shown by the results in Section “[The evaluation of the object detection model](#)”, their object detection accuracy in offshore drilling platforms is lower than the improved RFA-YOLO model in this paper, and when identifying and locating the human body, helmets and workwear, there are cases of missed detection and false detection, which provides wrong data support for the subsequent classification. In this paper, we improve and optimize the object detection model to make it more suitable for object detection of offshore drilling platform to provide accurate data for subsequent classification, and we optimize the parameters of the classification model to improve the classification accuracy in the subsequent classification stage, as shown by the experimental results in Section “[The evaluation of position features and the classifier](#)”. The classification model in this paper outperforms the classification model proposed by Nath et al. Therefore, based on the above analysis, it can be concluded why our framework can perform better on the PPED.

Xiong et al. [23] used pose estimation to obtain human keypoints, localized head and torso regions based on human keypoints, and achieved PPE detection by image classification. From the data in table, it can be seen that the performance of this method on the PPED dataset is inferior to that of the method in this paper. The reason is that the method proposed by Xiong et al. is based on pose esti-

Table 8 The comparison of each metric of different PPE detection methods

Method	Accuracy	Recall ₂	Recall ₁	MAR	FAR	Time (s)
Nath [22]	0.809	0.786	0.856	0.214	0.144	0.08
Xiong [23]	0.865	0.841	0.915	0.159	0.085	0.14
Ours	0.931	0.921	0.952	0.079	0.048	0.08

mation, and the accuracy of the localization and recognition of human key points will directly affect the goodness of the model. On the offshore drilling platform scenario, the complex postures of workers such as squatting and bending over lead to key points obscuring each other, resulting in erroneous detection results, and the dense pipelines also obscure the workers' bodies, affecting the accuracy of posture estimation and thus reducing the PPE recognition accuracy. Unlike the Xiong et al. method, the proposed framework in this paper does not rely on human posture estimation. In addition, the proposed method by Xiong et al. involves a more complex model structure and has a poorer real-time performance compared to the method in this paper. In summary, from the comparison results and the above analysis, it is clear that the proposed framework in this paper has better detection accuracy, reliability and real-time performance on the PPE dataset compared to the existing state-of-the-art PPE detection methods.

Figure 15 shows the results of the PPE detection by our method on the offshore drilling platform. The detection results show that the method proposed in this paper can accurately identify workers wearing PPE (e.g., subfigures (b), (c), (e)) and those not wearing PPE (e.g., subfigures (a), (d), (f)).

The ablation experiment

In this paper, we design ablation experiments to analyze the effect of classifiers P-H-classifier and P-W-classifier on PPE detection results, and also analyze the effect of P-H-classifier and P-W-classifier obtained from different candidate feature parameters and number of decision trees on PPE detection results. Table 9 shows the experimental results on the PPE dataset with Recall, FAR and MAR as evaluation metrics.

From the results in Table 9, it is clear that the classifier has an impact on the final PPE detection results. The results of Recall₁ and FAR are unaffected when there is no classifier or when a high-performance classifier is used; however, when the classifier performance is low, it leads to a decrease in Recall₁ and an increase in FAR value at the same time, which is mainly explained by the fact that the classifier is based on the relative offset of the target to distinguish whether the worker is wearing PPE or not, and for workers who wear PPE correctly, the target detection result only can also be relied on to determine the presence of PPE, thus obtaining the same effect as using a high-performance classifier. Conversely, the

use of a classifier with lower performance leads to incorrect relative position judgments, resulting in the identification of workers wearing PPE as not wearing PPE. We also found that the use of a classifier increases the value of Recall₂ and decreases MAR, because relying on object detection only cannot yield correct results for workers not wearing PPE, especially those holding helmets in their hands. The use of the classifier reduces the probability of identifying unwearing PPE workers as wearing PPE, which is the expected result in practical applications. When we continue to improve the classifier accuracy, there is no improvement for the results, which indicates that the classification parameters chosen in this paper are based on the proposed framework to achieve the optimal results, and if we want to continue to improve the performance of the framework, we need to start from other stages, such as the object detection process, the reliability of the data provided in this stage is crucial, therefore, the follow-up can continue to explore methods to improve the performance of this stage.

The analysis of algorithm applications

We analyze the feasibility of the results from two perspectives of real-time and accuracy, as shown in Fig. 16. In the existing offshore drilling platform, there are 450 monitoring cameras and 9 servers responsible for identification, each server is responsible for 50 cameras, as can be seen from Table 7, the average time for our model to identify each image is 0.08 s, and each server polls and identifies the images collected by 50 cameras in turn. In other words, each camera is polled once every 4 s. Considering that the human behavior in the surveillance screen is continuous, polling once every 4 s is sufficient for practical applications. If there is a higher requirement for real-time, the number of servers can be increased appropriately, thus reducing the number of cameras undertaken by each server. In practical applications, people prefer models with high recognition accuracy and low false alarm rate, that is, the accuracy is as high as possible and the FAR and MAR are as low as possible. As can be seen from Table 7, our model is more in line with people's expectations compared with other PPE detection methods, the value of MAR is lower than 8 and the value of FAR is lower than 5 in 100 images, which is acceptable.

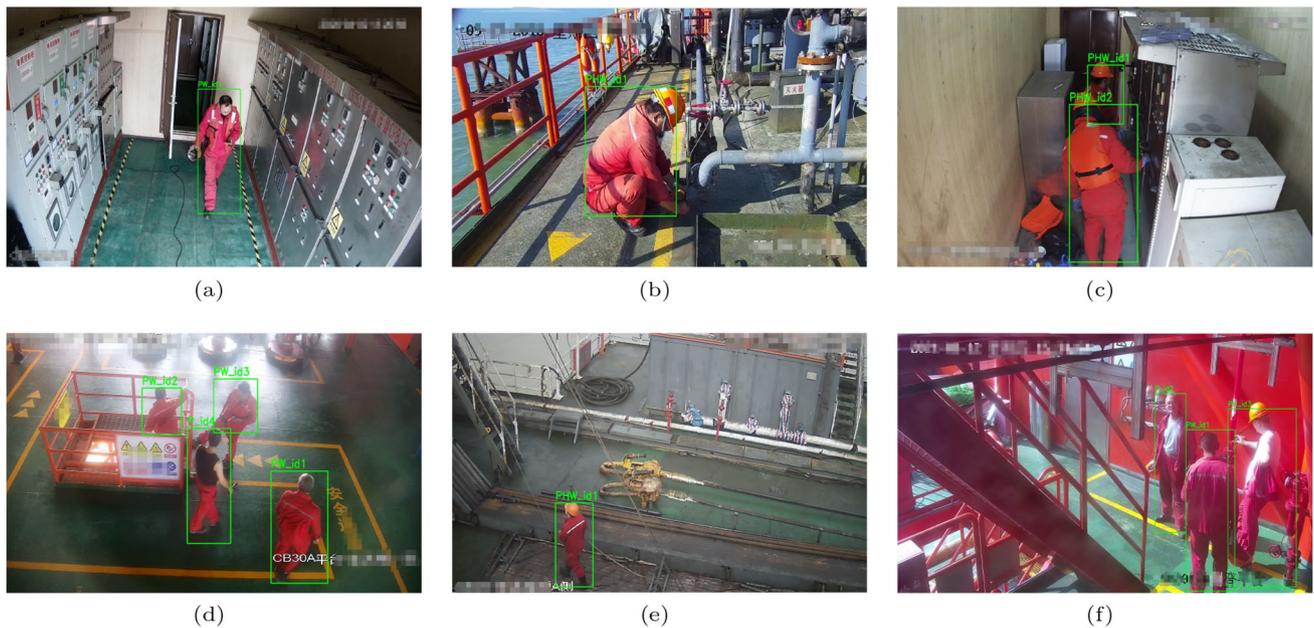


Fig. 15 The detection results of PPE for the offshore drilling platform based on the proposed framework

Table 9 The ablation of classifiers is studied on PPED using the framework proposed in this paper

	Recall ₁	Recall ₂	MAR	FAR
No-classifier	0.952	0.651	0.349	0.048
Classifier (ours)	0.952	0.921	0.079	0.048
Classifier (low)	0.884	0.862	0.138	0.116
Classifier (high)	0.952	0.921	0.079	0.048

No-classifier: no classifier after object detection; Classifier (ours) the classifier used for the experiments in this paper; Classifier (low): classifier with low accuracy composed of other parameters (candidate features of 1 and the number of decision trees of 2); Classifier (high): classifier with high accuracy composed of other parameters classifier (candidate features of 4 and number of decision trees of 4)

Conclusion

In this paper, we propose a high-performance PPE detection framework to rapidly and accurately identify workers on the offshore drilling platform who are not wearing helmets or workwear. This paper transforms PPE detection into object detection and classification and accomplishes detection by means of multi-algorithm fusion. Firstly, we propose a novel object detection algorithm, named RFA-YOLO, to achieve the localization and recognition of the person, helmet, and workwear. RFA-YOLO adds a residual feature augmentation module with YOLOv4 as the baseline, which reduces the loss of effective information in the high-level feature map and improves the object detection accuracy. In order to determine whether workers are wearing helmets and workwear correctly, we propose a method for constructing

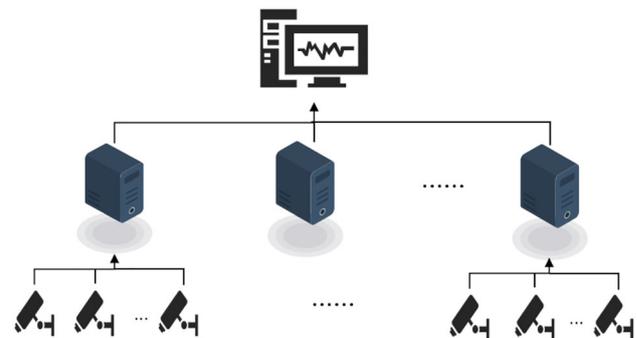


Fig. 16 The implementation diagram of the proposed method on an offshore drilling platform. The bottom layer is the surveillance camera, which is responsible for collecting the site images; the middle layer is the server, where the model is deployed on each server and is responsible for completing the PPE detection; the top layer is the client, where the identification results of each server are aggregated to the client, which is responsible for visualizing the identification results and generating alert messages

position features based on object bounding boxes. The feature vector representing the relative offset of the object is obtained based on the category and coordinate information of the object detection output. With the feature vector, we train the dataset consisting of position features to obtain the classifier by the random forest classification algorithm. In the training process, this paper uses parameter optimization methods to obtain high-performance classifiers. After the surveillance images of the offshore drilling platform are input into the detection framework, the PPE detection is achieved through an inference mechanism after object detection, feature construction, and feature classification. To

verify the performance of the proposed method in this paper, we compare this method with other literature methods, and the experimental results show that the proposed framework performs better for PPE detection on the offshore drilling platform. The framework can detect workers who are not wearing helmets or workwear in time and generate alarm messages. Currently, an operating video monitoring system for the offshore drilling platform has been completed based on this detection framework.

However, there are still limitations of the proposed framework in this paper, for example, in practical application scenarios, a few workers' heads or torsos are partially obscured by pipes during operation, which makes the method in this paper unable to accurately identify the PPE wearing situation; in extreme weather such as fog and heavy rain at sea, the workers in the monitoring screen are blurred, which leads to the inability to accurately locate and identify the workers and also reduces the PPE detection accuracy. Therefore, occlusion and blurred images are the next important problems we need to solve. Therefore, the authors propose that the next piece of work that needs to be improved. The application of knowledge graph in object detection has achieved remarkable performance [42]. Therefore, it is a feasible approach to improve helmet and workwear detection accuracy with the help of inherent relationships between objects (entities). Finally, it is crucial to increase the diversity of the scenes and types of datasets on which more reasonable test and comparison results are obtained.

Acknowledgements This paper is supported by the Natural Science Foundation of Shandong Province (No. ZR2020MF136), and the research project funded by the Central Universities (20CX05016A).

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Canonaco G, Roveri M, Alippi C, Podenzani F, Bennardo A, Conti M, Mancini N (2022) A transfer-learning approach for corrosion prediction in pipeline infrastructures. *Appl Intell* 52(7):7622–7637
2. Dong S, He Q, Li H, Yin Q (2015) Automated ppe misuse identification and assessment for safety performance enhancement. *ICCREM* 2015:204–214
3. Kelm A, Laußat L, Meins-Becker A, Platz D, Khazaee MJ, Costin AM, Helmus M, Teizer J (2013) Mobile passive radio frequency identification (rfid) portal for automated and rapid control of personal protective equipment (ppe) on construction sites. *Autom Constr* 36:38–52
4. Yang F, Gu S (2021) Industry 4.0, a revolution that requires technology and national strategies. *Complex Intell Syst* 7(3):1311–1325
5. Fang W, Ding L, Luo H, Love PE (2018) Falls from heights: a computer vision-based approach for safety harness detection. *Autom Constr* 91:53–61
6. Park M-W, Brilakis I (2012) Construction worker detection in video frames for initializing vision trackers. *Autom Constr* 28:15–25
7. Cai L, Qian J (2011) A method for detecting miners based on helmets detection in underground coal mine videos. *Min Sci Technol (Chin)* 21(4):553–556
8. Li X, He M, Liu Y, Luo H, Ju M (2022) Spcs: a spatial pyramid convolutional shuffle module for yolo to detect occluded object. *Complex Intell Syst* 1–15
9. Fang Q, Li H, Luo X, Ding L, Luo H, Rose TM, An W (2018) Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Autom Constr* 85:1–9
10. Fan Z, Peng C, Dai L, Cao F, Qi J, Hua W (2020) A deep learning-based ensemble method for helmet-wearing detection. *PeerJ Comput Sci* 6:311
11. Wang H, Hu Z, Guo Y, Yang Z, Zhou F, Xu P (2020) A real-time safety helmet wearing detection approach based on csyolov3. *Appl Sci* 10(19):6732
12. Han G, Zhu M, Zhao X, Gao H (2021) Method based on the cross-layer attention mechanism and multiscale perception for safety helmet-wearing detection. *Comput Electr Eng* 95:107458
13. Zhang H, Yan X, Li H, Jin R, Fu H (2019) Real-time alarming, monitoring, and locating for non-hard-hat use in construction. *J Constr Eng Manag* 145(3):04019006
14. Barro-Torres S, Fernández-Caramés TM, Pérez-Iglesias HJ, Escudero CJ (2012) Real-time personal protective equipment monitoring system. *Comput Commun* 36(1):42–50
15. Stojanovic V, He S, Zhang B (2020) State and parameter joint estimation of linear stochastic systems in presence of faults and non-gaussian noises. *Int J Robust Nonlinear Control* 30(16):6683–6700
16. Tao H, Cheng L, Qiu J, Stojanovic V (2022) Few shot cross equipment fault diagnosis method based on parameter optimization and feature mertic. *Meas Sci Technol* 33(11):115005
17. Djordjevic V, Stojanovic V, Tao H, Song X, He S, Gao W (2022) Data-driven control of hydraulic servo actuator based on adaptive dynamic programming. *Discr Contin Dyn Syst S* 15(7):1633
18. Rubaiyat AH, Toma TT, Kalantari-Khandani M, Rahman SA, Chen L, Ye Y, Pan CS (2016) Automatic detection of helmet uses for construction safety. In: 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW), pp 135–142, IEEE
19. Li J, Liu H, Wang T, Jiang M, Wang S, Li X, Zhao X (2017) Safety helmet wearing detection based on image processing and machine learning. In: 2017 Ninth International Conference on Advanced Computational Intelligence (ICACI), pp 201–205, IEEE
20. Bo Y, Huan Q, Huan X, Rong Z, Hongbin L, Kebin M, Weizhong Z, Lei Z (2019) Helmet detection under the power construction scene based on image analysis. In: 2019 IEEE 7th International Confer-

- ence on Computer Science and Network Technology (ICCSNT), pp 67–71, IEEE
21. Huang L, Fu Q, He M, Jiang D, Hao Z (2021) Detection algorithm of safety helmet wearing based on deep learning. *Concurr Comput Pract Exp* 33(13):6234
 22. Nath ND, Behzadan AH, Paal SG (2020) Deep learning for site safety: real-time detection of personal protective equipment. *Autom Constr* 112:103085
 23. Xiong R, Tang P (2021) Pose guided anchoring for detecting proper use of personal protective equipment. *Autom Constr* 130:103828
 24. Wang Z, Wu Y, Yang L, Thirunavukarasu A, Evison C, Zhao Y (2021) Fast personal protective equipment detection for real construction sites using deep learning approaches. *Sensors* 21(10):3478
 25. Shen J, Xiong X, Li Y, He W, Li P, Zheng X (2021) Detecting safety helmet wearing on construction sites with bounding-box regression and deep transfer learning. *Comput-Aid Civ Infrastruct Eng* 36(2):180–196
 26. Iannizzotto G, Bello LL, Patti G (2021) Personal protection equipment detection system for embedded devices based on dnn and fuzzy logic. *Expert Syst Appl* 184:115447
 27. Gallo G, Di Rienzo F, Garzelli F, Ducange P, Vallati C (2022) A smart system for personal protective equipment detection in industrial environments based on deep learning at the edge. *IEEE Access* 10:110862–110878
 28. Ferdous M, Ahsan SMM (2022) Ppe detector: a yolo-based architecture to detect personal protective equipment (ppe) for construction sites. *PeerJ Comput Sci* 8:999
 29. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2):91–110
 30. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol 1, pp 886–893. Ieee
 31. Lienhart R, Maydt J (2002) An extended set of haar-like features for rapid object detection. In: Proceedings, International Conference on Image Processing, vol 1, p IEEE
 32. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969
 33. Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* 28
 34. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC (2016) Ssd: single shot multibox detector. In: European conference on computer vision, pp 21–37, Springer
 35. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 779–788
 36. Redmon J, Farhadi A (2017) Yolo9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7263–7271
 37. Redmon J, Farhadi A (2018) Yolov3: an incremental improvement. [arXiv:1804.02767](https://arxiv.org/abs/1804.02767)
 38. Guo C, Fan B, Zhang Q, Xiang S, Pan C (2020) Augfpn: Improving multi-scale feature learning for object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 12595–12604
 39. Bochkovskiy A, Wang C-Y, Liao H-YM (2020) Yolov4: optimal speed and accuracy of object detection. [arXiv:2004.10934](https://arxiv.org/abs/2004.10934)
 40. Purkait P, Zhao C, Zach C (2017) Spp-net: deep absolute pose regression with synthetic views. [arXiv:1712.03452](https://arxiv.org/abs/1712.03452)
 41. Liu S, Qi L, Qin H, Shi J, Jia J (2018) Path aggregation network for instance segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8759–8768
 42. Xu H, Jiang C, Liang X, Lin L, Li Z (2019) Reasoning-rnn: unifying adaptive global reasoning into large-scale object detection. In: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 6412–6421

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.