Nuclear Science and Techniques (2024) 35:74
https://doi.org/10.1007/s41365-024-01435-z

Research on intelligent search-and-secure technology in accelerator hazardous areas based on machine vision

Ying-Lin Ma[1] · Yao Wang[2] · Hong-Mei Shi[2] · Hui-Jie Zhang[3]

Received: 27 June 2023 / Revised: 7 December 2023 / Accepted: 13 December 2023 / Published online: 24 May 2024

Abstract
Prompt radiation emitted during accelerator operation poses a significant health risk, necessitating a thorough search and securing of hazardous areas prior to initiation. Currently, manual sweep methods are employed. However, the limitations of manual sweeps have become increasingly evident with the implementation of large-scale accelerators. By leveraging advancements in machine vision technology, the automatic identification of stranded personnel in controlled areas through camera imagery presents a viable solution for efficient search and security. Given the criticality of personal safety for stranded individuals, search and security processes must be sufficiently reliable. To ensure comprehensive coverage, 180° camera groups were strategically positioned on both sides of the accelerator tunnel to eliminate blind spots within the monitoring range. The YOLOV8 network model was modified to enable the detection of small targets, such as hands and feet, as well as larger targets formed by individuals near the cameras. Furthermore, the system incorporates a pedestrian recognition model that detects human body parts, and an information fusion strategy is used to integrate the detected head, hands, and feet with the identified pedestrians as a cohesive unit. This strategy enhanced the capability of the model to identify pedestrians obstructed by equipment, resulting in a notable improvement in the recall rate. Specifically, recall rates of 0.915 and 0.82 were obtained for Datasets 1 and 2, respectively. Although there was a slight decrease in accuracy, it aligned with the intended purpose of the search-and-secure software design. Experimental tests conducted within an accelerator tunnel demonstrated the effectiveness of this approach in achieving reliable recognition outcomes.

Keywords Search and secure · Machine vision · Camera · Human body parts recognition · Particle accelerator · Hazardous area

1 Introduction

Prompt radiation is generated during the operation of particle accelerators. The prompt radiation area has a high energy level and exhibits intense radiation characteristics. Consequently, individuals within the controlled area are subjected to significant doses of radiation from the generated neutrons and gamma rays [1]. Thus, all personnel must be evacuated within the controlled area before initiating accelerator operations.

Conventionally, skilled personnel are deployed to evacuate personnel from a controlled area. These proficient individuals enter a controlled area and meticulously inspect and evacuate the other personnel according to a pre-established sequence and location [2, 3]. This approach presents several issues: (1) With the continuous advancement of scientific technology, there is a noticeable trend towards the scaling up of accelerator installations. Accelerators measuring several kilometers or even tens of kilometers in length have already emerged, and plans are underway for accelerators approaching a length of 100 km [4, 5]. These large-scale accelerators correspond to a substantial expansion of the controlled area. Consequently, the traditional approach faces increasingly prominent drawbacks such as prolonged time consumption and low efficiency. (2) Large accelerators encompass

✉ Yao Wang
wangyao@bjtu.edu.cn

[1] School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

[2] School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, Beijing 100044, China

[3] Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China

a multitude of components and intricate structures, resulting in numerous blind spots within hazardous areas. These obstructions give rise to significant safety concerns, as they may conceal individuals who have not been found by proficient personnel.

In light of these circumstances, we present a machine-vision-based intelligent search-and-secure technology as a solution. This technology leverages a camera group deployed in a hazardous area and a server with an identification program specifically designed to perform intelligent and rapid identification of stranded individuals within a tunnel.

Owing to the multitude of equipment within the hazardous area of the accelerator, some of which have large dimensions, the line of sight of personnel responsible for searching and security may be obscured. Obstruction also remains a challenge in pedestrian target detection [6]. To address this challenge, Pang et al. introduced a strategy that utilizes masks to guide attention networks, enhancing the detection of obstructed pedestrians by emphasizing the visible parts of the human body and suppressing obscured areas [7]. Zhang et al. proposed an OR-CNN (Occlusion Region Convolutional Neural Network) focusing on both loss and core ROI (Region of Interest) pooling operations in a two-stage detection process [8]. To address the complexities of pedestrian pose variability and mutual occlusion, Khan et al. proposed a novel perspective, asserting that human heads, which are less susceptible to obstructions, could serve as robust focal points for detection across diverse scales in intricate scenarios. Their innovative head detection system demonstrated highly promising results, encouraging the exploration of local detection techniques to identify obstructed pedestrians [9]. Moreover, Chen et al. presented a comprehensive pedestrian detection methodology that integrated both head and full-body information through multi-feature fusion [10]. We drew inspiration from these methodologies by discerning the head, hands, and feet as subsets of a pedestrian's body. Subsequently, we seamlessly integrated these subsets into the overarching pedestrian structure. This integration addresses the concern regarding the shielding of individuals stranded in a tunnel due to equipment obstruction.

This study aimed to develop an intelligent monitoring system tailored for the sweep of an accelerator tunnel, encompassing considerations in both the hardware and software realms. On the hardware front, our primary emphasis was on devising and implementing a camera group that boasts expansive 180° horizontal and vertical field angles. Strategically positioned on both sides of the tunnel, these cameras adeptly alleviate the challenge of pedestrians encountering complete obstruction. On the software facet, we designed the Parts of the Human Body (PHB) model for pedestrian recognition. This model employs a comprehensive approach; covered pedestrians are identified by analyzing their heads, hands, and feet, and intelligent search and security software was designed. By seamlessly integrating the camera group with the PHB model, our system achieves a one-key intelligent clearing of the accelerator tunnel.

The contributions of this study are as follows: (1) A novel machine-vision-based search-and-secure system is introduced, marking a pioneering approach ensuring the evacuation of all individuals from the tunnel before the accelerator activates. The core focus of this study is to tailor the system to suit the specifics of an accelerator tunnel environment. (2) To address the issue of accelerator occlusion, we introduced a novel design featuring a camera array consisting of six units that ensures comprehensive visual coverage, a dimension that has not been previously explored. (3) The YOLOv8 model is enhanced by leveraging body part recognition to detect stranded personnel. This innovative approach significantly increased the recall rate.

2 System architecture

The hazardous area of a large accelerator typically has a width of no more than 10 m, a height of no more than 6 m, and a length ranging from several hundred meters to tens of kilometers. The primary accelerator equipment spans the length of the hazardous area, as illustrated in Fig. 1. In our proposed intelligent search-and-secure system, we organized detection units at 15 m intervals, each equipped with a set of camera groups situated on both sides of the hazardous area, which were connected to both the regular video surveillance server and the intelligent video surveillance server. The regular video surveillance server is responsible for standard functionalities, such as real-time monitoring and video playback, whereas the intelligent video surveillance server incorporates a PHB recognition program specifically designed for intelligent sweep purposes.

3 Design of 180° camera group

The main equipment of the accelerator comprises magnets, vacuum beam tubes, high-frequency cavities, beam detection equipment, and various equipment supports [11–13], as illustrated in Fig. 2. Smaller equipment, such as vacuum pipes, allows personnel to perform maintenance tasks in their proximity; however, the majority of the maintainer's body remains uncovered, enabling identification by cameras positioned on either side of the equipment. Conversely, larger equipment, such as magnets and high-frequency cavities, may require personnel to work on the upper and lateral sides. Personnel situated on the magnet side experienced significant body obstruction, rendering the camera's recognition effect ineffective or making them non-identifiable. Therefore, camera groups must be

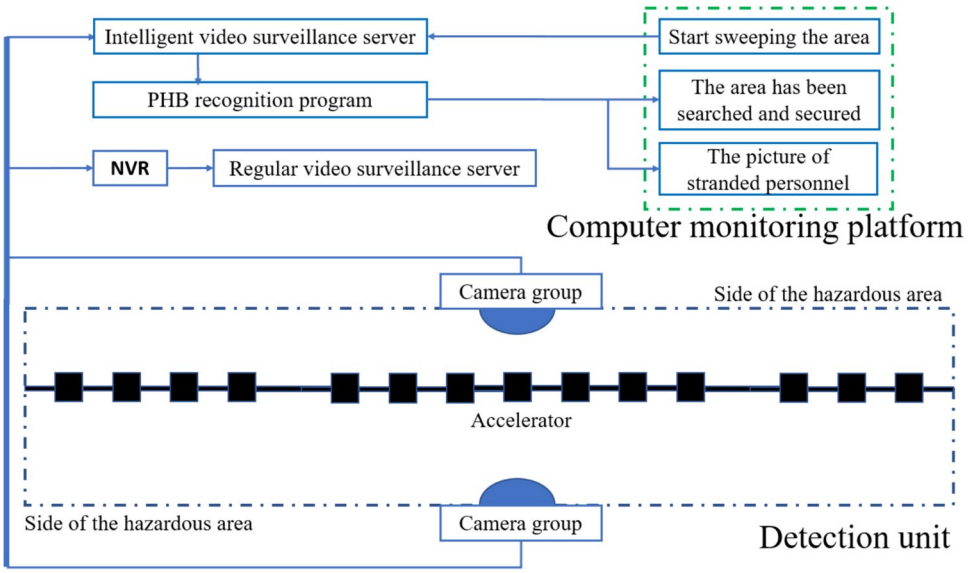


Fig. 1 Architecture diagram of intelligent search-and-secure system

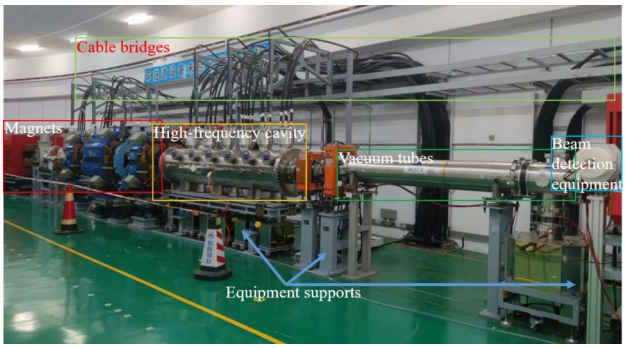


Fig. 2 (Color online) Primary components of the particle accelerator

arranged on both sides of the magnet. The limited space beneath the magnet restricts the full entry of personnel bodies; however, body parts such as the head, hands, and feet are consistently identifiable. In the vertical range of the 3–5 m controlled area, cable bridges and ventilation pipes are typically installed along the walls, with cranes positioned at the top. Personnel may approach these areas for maintenance, which makes it crucial for the camera to have visibility of these individuals. In the longitudinal direction of the controlled area, the monitoring distance of the camera must be maximized. Simultaneously, it is crucial to ensure that the camera can effectively monitor the body parts of individuals in close proximity.

Based on the aforementioned analysis, it can be deduced that to prevent personnel from being overlooked because of the camera's blind spot, the vertical and horizontal field angles of the intelligent search-and-secure system cameras must be close to 180°. The field-of-view angle of a single camera fails to satisfy this criterion, necessitating a combination of multiple cameras to form a camera group [14].

The camera group was affixed to a wall situated below the cable bridge and the ventilation duct. Simultaneously, larger equipment within the accelerator tunnel, such as magnets, can extend up to a height of approximately 2 m. The installation of a camera group at an approximate height of 2 m is recommended to reduce potential obstructions. The imaging size of a single camera with a 1/2.7" CMOS sensor was 5.27 mm × 3.96 mm ($w \times h$). A smaller lens was used to achieve the widest field of view possible. By selecting a 2.8 mm lens, the following formula yields a horizontal field angle of 86.5° and a vertical field angle of 70.5°.

$$\text{Horizontal field angle: } \alpha = 2\arctan(w/2f) \tag{1}$$

$$\text{Vertical field angle: } \beta = 2\arctan(h/2f) \tag{2}$$

where w represents the width of the field of view, h represents the height of the field of view, and f denotes the focal length of the lens [15].

As illustrated in Fig. 3, a camera group consisting of six cameras covers a vertical viewing field of 180° and a horizontal viewing field of 173°. The angular separation between cameras 1 and 4 was measured to be 86.5°. Similarly, the angular separation between cameras 2 and 3 is 109.5°, with the latter pair positioned above and below camera 1. Correspondingly, cameras 5 and 6 were positioned above and below camera 4, respectively, with an angular separation of 109.5°.

The optimal viewing range of the 2.8 mm lens was limited to a maximum of 7.5 m. Meanwhile, owing to the camera group's horizontal viewing angle of 173°, a blind field of view measuring 0.46 m will be present near the wall when the distance exceeds 7.5 m. Taking into account the fact that the actual field of view might slightly exceed the calculated

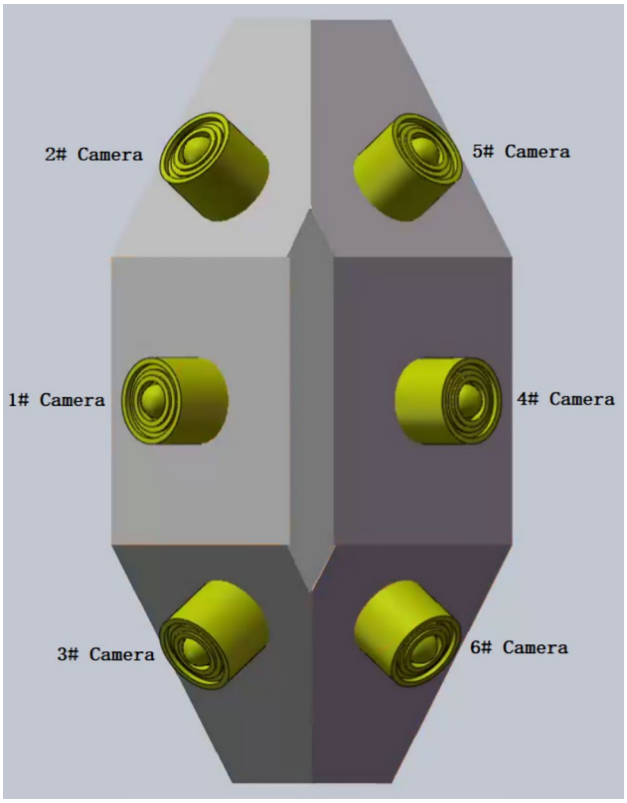


Fig. 3 (Color online) Structure diagram of the 180° camera group

field of view, we conducted a verification test and determined that a distance of 15 m between camera groups would be appropriate. This distance ensured adequate coverage and minimized the occurrence of blind spots in the monitoring area.

4 Design of intelligent search-and-secure software

The workflow of the intelligent search-and-secure software is illustrated in Fig. 4. After a sweep of the hazardous areas is initiated by the computer monitoring platform in the control room, the detained personnel identification program is activated. Simultaneously, all detection units within the corresponding hazardous area begin capturing continuous videos for a duration of 3 min. Subsequently, the captured images are segmented, enlarged, and enhanced. A PHB recognition model is employed to determine whether an individual is present in the captured images. If no person is detected, the intelligent search-and-secure server sends a signal indicating that the hazardous area has been searched and secured. However, if stranded personnel are identified in the images, the monitoring platform displays the corresponding

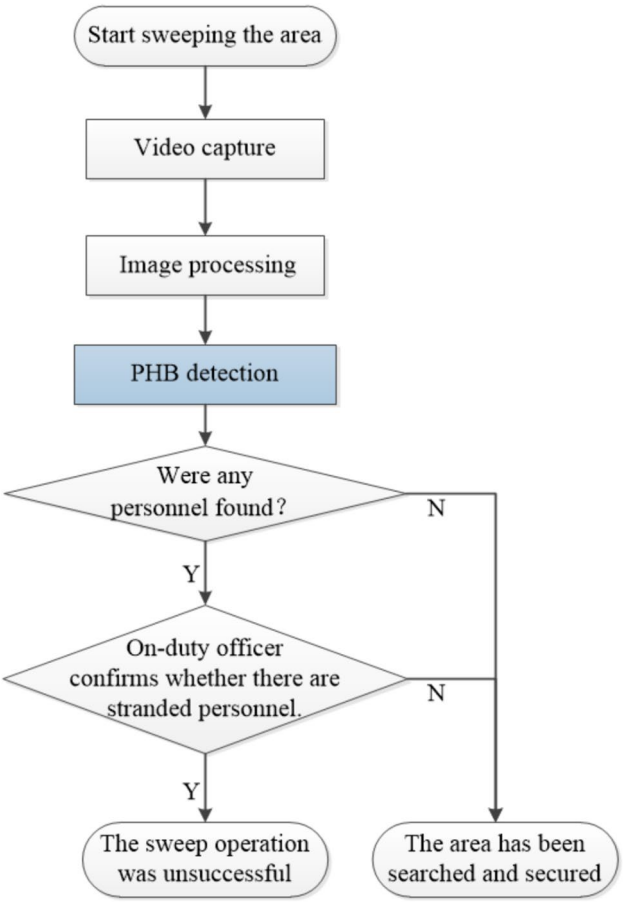


Fig. 4 Working flow chart of intelligent search-and-secure software

images, allowing on-duty personnel to confirm or initiate rescan procedures.

4.1 Design of pedestrian recognition model with fusion of body parts

The primary objective of intelligent search and security software is to identify individuals trapped in hazardous areas through the analysis of images captured by cameras, with pedestrian target detection as its fundamental technology. With the rapid advancement of deep learning, technologies rooted in deep learning, such as image recognition and data processing, have also gained popularity in the nuclear technology domain [16, 17]. Visual inspection technology leveraging deep learning is advancing rapidly. For instance, Tang et al. employed machine vision technology for the precise detection of crack widths [18]. Similarly, they utilized binocular vision methods to accurately measure the deformation of concrete columns [19]. As a crucial subset of visual detection technology, pedestrian target detection has extensive applications in diverse fields such as autonomous

driving, robotics, intelligent monitoring, and human behavior analysis [20, 21].

The search and security process within a hazardous area demands a high level of reliability, posing challenges for existing pedestrian detection technologies in scenarios where pedestrians are obstructed by equipment [22, 23]. Through a comprehensive analysis of the equipment layout and pedestrian occlusion within the hazardous area of the accelerator, we observed that certain body parts, such as the head, hands, and feet, were less likely to be completely occluded when using our developed camera group arrangement on both sides of the hazardous area. Considering these characteristics, we propose a novel pedestrian recognition model that incorporates the distinctive features of different body parts, thereby enhancing the reliability of intelligent searches and secure systems.

The PHB recognition model is an enhanced design based on the YOLOv8 network model, as depicted in Fig. 5. The network model comprises three main components: feature extraction, detection, and header modules.

The feature extraction module follows the YOLOv8 backbone network, which consists of five CBL layers that perform operations such as convolution and normalization on the input feature map. The module includes four C2f modules that facilitate learning of the residual characteristics. To improve the receptive field of the network, the spatial pyramid pooling fusion (SPPF) module performs feature extraction through a parallel input using multiple maximum pooling layers [24, 25]. Building on the three detection layers of YOLOv8, the detection module introduced minimal and maximum target detection layers. The minimal target detection layer focuses on detecting small targets such as hands and feet. It processes the feature map after the 14th layer of the original network and expands it. In the 21st layer, the resulting 160×160 feature map is ConCat fused with the feature map from the second layer of the backbone network, enabling the detection of very small targets [26, 27].

By contrast, the maximum target detection layer addresses cases in which individuals approach the camera too closely, leading to super-large targets. It fuses the 10×10 feature map obtained from the 11th layer of the original network with the 8th layer feature map of the backbone network to obtain the minimum feature map for detecting the maximum targets. After splicing and fusing the features from different layers, namely layers 22, 25, 28, 31, and 34, they are passed to the detection head. The detection head consists of five detector modules that output the prediction information. The final detection results are obtained by further calculations and comparisons.

In the PHB recognition model, the five detection layers corresponded to five sets of initial detection boxes. When the input image size was 640×640 pixels and the distance between the camera and hand target was 8 m, the size of the hand target was approximately 6×6 pixels. The minimal target detection layer has a size of 160×160 pixels and is designed to detect minimal targets larger than 4×4 pixels, thus fulfilling the requirements for hand target detection [28].

The small-target detection layer has a size of 80×80 pixels and is responsible for detecting ordinary small targets larger than 8×8 pixels. The detection layer corresponding to medium-sized targets measures 40×40 pixels and detects targets larger than 16×16 pixels. Similarly, the detection layer corresponding to large targets has a size of 20×20 pixels and can detect targets larger than 32×32 pixels.

Additionally, a super-large target detection layer measuring 10×10 pixels aids in the identification of scenarios in which the large target detection layer encounters challenges in detecting the body occupying the entire image, as depicted in Fig. 6.

4.2 Information fusion strategy

Khan et al. partitioned a broad spectrum of scales into a subscale ensemble encompassing three distinct scales. This segmentation enabled them to effectively process heads aligned with particular subscales. Subsequently, these components were amalgamated into an end-to-end network, yielding highly satisfactory detection outcomes [29]. Inspired by this methodology, our approach extends its concept to address blocked pedestrians. We treated the hands, head, and feet as individual subsets within the overall obstructed pedestrian category. Each subset was detected independently, and a fusion strategy was employed to assemble a comprehensive pedestrian detection framework after detecting these components separately.

Let us consider the overall pedestrian detection box, denoted as box $B^{\text{body}} = \left(x_1^b, y_1^b, x_2^b, y_2^b\right)$, where the coordinates $\left(x_1^b, y_1^b\right)$ and $\left(x_2^b, y_2^b\right)$ represent the upper-left and lower-right points of the detection box, respectively.

In accordance with the observations made in [10], the analysis considered different pedestrian postures, including standing forward and sideways [30]. In this analysis, the upper section of the pedestrian detection frame was designated as the head area, whereas the lower section represented the foot area. Given the flexible nature of hand positioning, the middle and upper regions of a pedestrian's body, along with both sides, are considered potential areas where hands may appear. The head, foot, and hand areas were calculated as follows:

$$\text{Head\_region} = \left(x_1^b, y_1^b, x_2^b, y_1^b + \frac{1}{3}h^b\right), \tag{3}$$

$$\text{Foot\_region} = \left(x_1^b, y_1^b + \frac{2}{3}h^b, x_2^b, y_2^b\right), \tag{4}$$

$$\text{Hand\_region} = \left(x_1^b - w^b, y_1^b, x_2^b + w^b, y_2^b - \frac{1}{3}h^b\right), \tag{5}$$

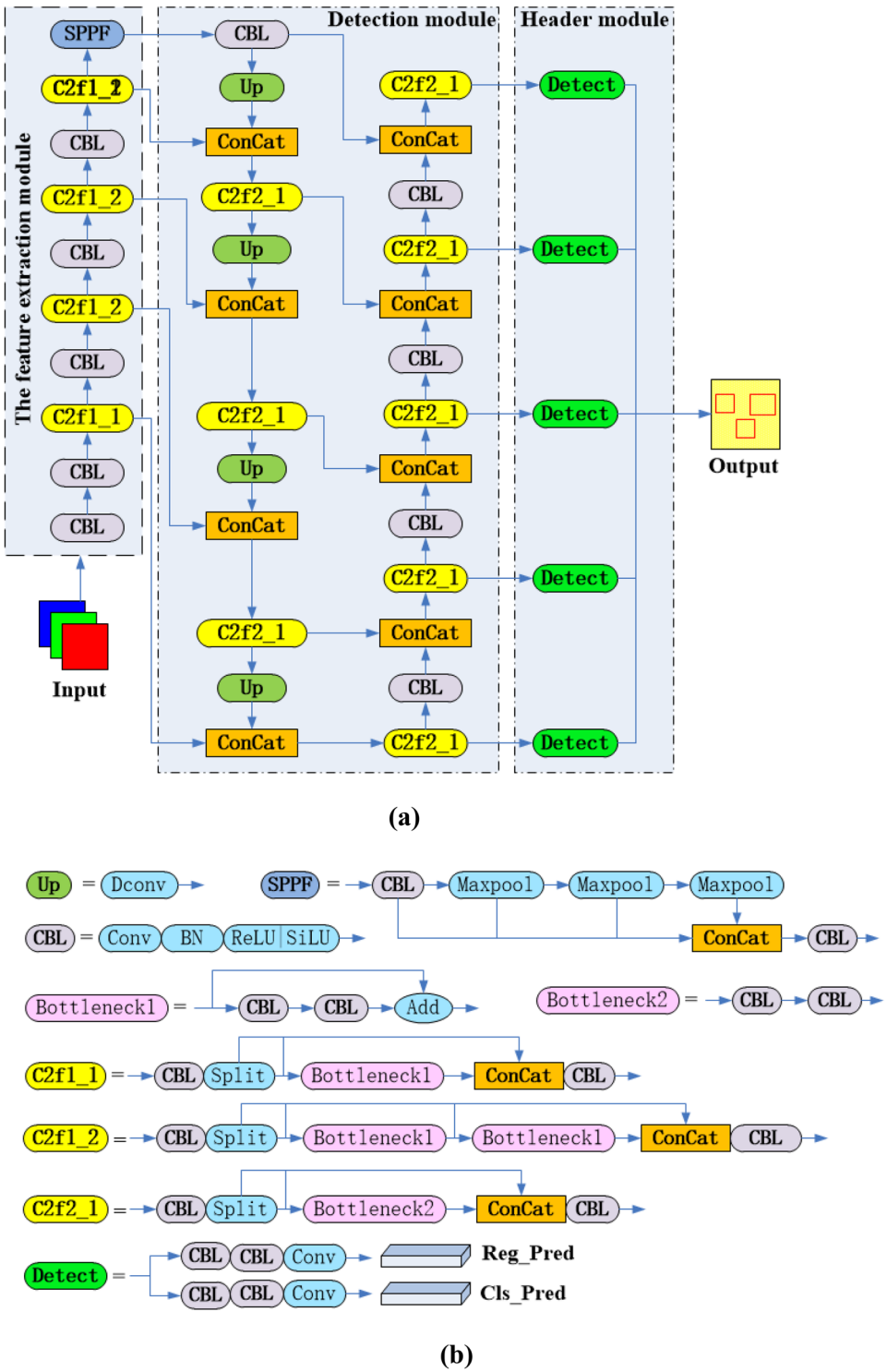


Fig. 5 Network architecture diagram of the PHB model. **a** PHB model main frame. **b** structure of submodules

where w^b represents the width of the overall pedestrian detection frame, and h^b represents the height of the overall pedestrian detection frame.

In crowded scenarios, the body parts of other targets can appear within a pedestrian detection frame. To address this issue, a processing method that involves calculating the distance between a specific type of body part and the center of the target body part was employed. This calculation was performed when the number of body parts within the overall pedestrian detection frame exceeded the expected count. The nearest body part was then matched to the overall pedestrian detection frame.

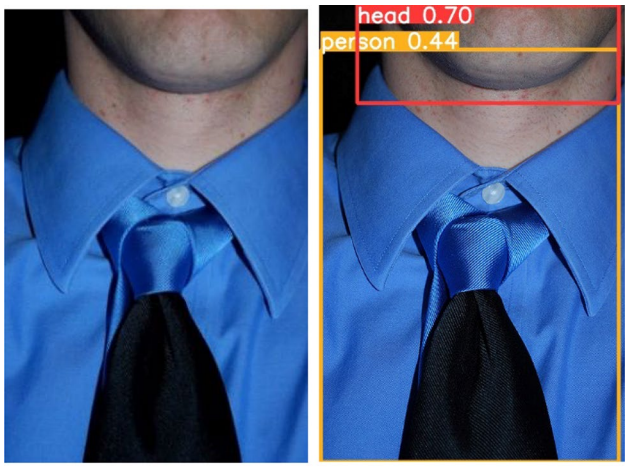


Fig. 6 (Color online) Comparison of object recognition results for oversized targets. **a** unrecognized objects by YOLOv8s. **b** recognized objects by PHB

4.3 Search-and-secure software and interface design

This study used PyQt5 to design the software interface, as depicted in Fig. 7. Upon initiating the search and securing process through the software button, the underlying program proceeds by capturing a screenshot for 3 min. The captured image is then sent to a designated folder for segmentation, followed by the automated execution of the PHB detection program. If a target is detected, the interface displays an image with annotations denoting the entire pedestrian or specific body parts within the scene. On-duty personnel are prompted to confirm or initiate rescanning procedures. In the absence of a detected target, the interface provides a signal indicating a successful sweep.

In practical scenarios, an acceleration tunnel is divided into multiple smaller, controlled areas, each of which is scanned at distinct time intervals. Meanwhile, considering the gradual nature of human movements, we captured images at 30-s intervals for detection purposes. These measures are crucial for reducing the number of captured images and improving overall work efficiency.

The PHB system adopts the image input approach of YOLOv8, which involves resizing the image to a dimension of 640×640 pixels before feeding it into a detection model. However, the camera group outputs images with a size of 1920×1080 pixels. Direct scaling of these images results in a reduction in the number of target pixels, potentially affecting the detection performance for small targets. To mitigate this issue, the search-and-secure program employed in this study divided the original image into 3×3 subgraphs. These subgraphs, along with the original image, were provided as inputs for the PHB program.

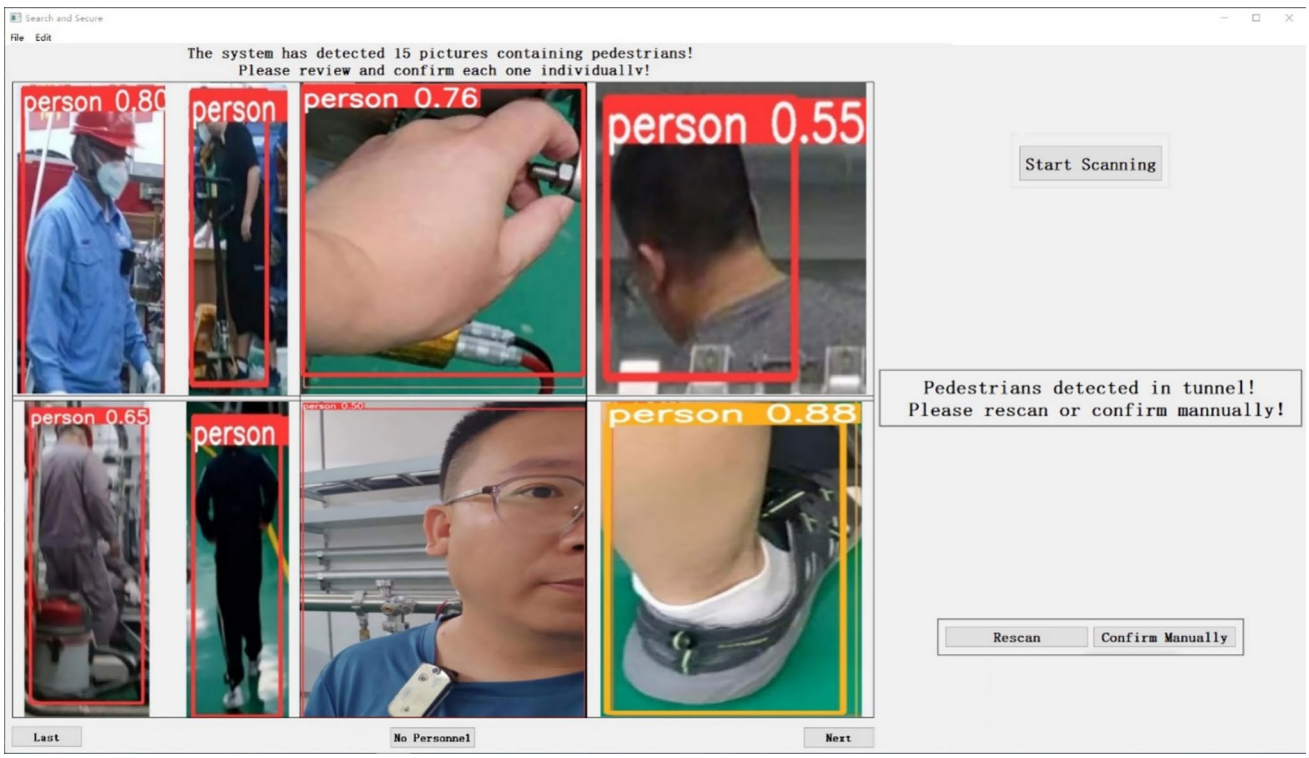


Fig. 7 (Color online) Intelligent search-and-secure software user interface

5 Experimental validation results

5.1 Construction of the validation dataset

Dataset 1: The homemade human body part dataset comprised a collection of 3998 images extracted from scenes within an accelerator tunnel. This dataset encompasses more than 15,000 pedestrian targets. To diversify the dataset, the backgrounds surrounding each pedestrian in the selected images were captured randomly to introduce occlusions. Subsequently, the LabelImg tool was used for precise annotation. The annotations were categorized into four classes: person, head, hand, and foot. The annotated data were then converted from XML to YOLO format and split into training and validation sets, adhering to a ratio of 9:1 for effective model training and evaluation.

Dataset 2: The pedestrian detection fusion dataset comprised a collection of 10,000 images that were randomly sampled from prominent datasets such as COCO, VOC2012, VOC2017, SYSU, and PRW [31]. After a meticulous data cleaning process, the dataset was curated to extract images that specifically contained pedestrians. In total, 9257 images were obtained, encompassing a diverse range of scenarios involving occluded and unoccluded pedestrians, as well as varying distances between the pedestrians and the camera. The dataset was subsequently divided into training and verification sets at a ratio of 1:9. This partitioning scheme ensures an effective evaluation of both the PHB and classical models in terms of their generalization abilities across different pedestrian detection scenarios.

5.2 Evaluation metrics

The detection and evaluation processes used in this study were divided into two main components. The first part focuses on pedestrian-component detection, in which the performance of the detection results is compared with those of YOLOv5s and YOLOv8s. This comparison aimed to validate the impact of the introduced minimal target detection layer (frame) and maximum target detection layer (frame). The evaluation of the detection results was based on conventional metrics such as precision, recall, and mean average precision (mAP). The mAP is computed as the overall average value when the detection threshold ranges from 0.5 to 0.95, denoted as mAP0.5:0.95. The calculation formula is as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{7}$$

$$\text{mAP} = \frac{1}{C} \sum_{c \in C} AP(c), \tag{8}$$

where TP represents cases in which the prediction is positive and aligns with the actual positive instances. FN denotes instances in which the prediction is negative but the actual value is positive. FP indicates cases where the prediction is positive yet the actual value is negative.

The second part of the evaluation focused on the overall pedestrian detection performance. A comparison was made between the detection results obtained using the PHB model and classical models, such as the YOLO series and Faster R-CNN, aiming to assess the generalization ability of the PHB model. The evaluation metrics employed included precision, recall, and average precision (AP) [32].

5.3 Experimental setup and results analysis

5.3.1 Experimental setup and parameter configuration

The experiments were conducted using a Windows 10 operating system with CUDA 11.1, and the training was performed on a single NVIDIA GeForce RTX 3070 GPU. The input image size was set to 640×640 pixels, and the training process was performed for 300 epochs. Each training batch consisted of 16 images. The gradient descent optimizer utilized a momentum parameter of 0.937 and a weight decay regularization coefficient of 0.0005. The initial learning rate (Lr0) for training was set to 0.01.

5.3.2 Detection results and analysis

The training process for YOLOv5s was completed in approximately 10.4 h, whereas training with YOLOv8s took approximately 8.9 h and PHB took approximately 14 h. Despite the longer training time, PHB outperformed YOLOv5s and YOLOv8s in terms of accuracy, recall rate, and AP [33]. This improvement was particularly notable in the recall rate index of the search and security software, in which the overall recall rate for pedestrians increased by 0.158 (Table 1). The inclusion of the PHB model resulted in an increase in the number of detection layers, which affected the detection speed. However, considering the significance of reliability indicators for intelligent search-and-secure software, the tradeoff of computing time for improved reliability is deemed worthwhile.

In the context of machine vision searches and secure software, the ability to accurately identify all stranded individuals is of paramount importance. However, upon analyzing the results presented in Table 1, while the PHB model shows an improvement in the overall recall rate of pedestrians, the achieved performance falls short of the desired ideal.

Table 1 Performance comparison PHB, YOLOv5s and YOLOv8s

Class	Model	Precision	Recall	Map0.5	Map0.5:0.95
Person	YOLOv5s	0.922	0.737	0.827	0.481
	YOLOv8s	0.938	0.715	0.821	0.488
	PHB	0.941	0.873	0.913	0.706
Head	YOLOv5s	0.97	0.928	0.966	0.714
	YOLOv8s	0.979	0.919	0.962	0.723
	PHB	0.979	0.929	0.969	0.76
Hand	YOLOv5s	0.821	0.71	0.777	0.404
	YOLOv8s	0.862	0.702	0.773	0.442
	PHB	0.873	0.756	0.818	0.475
Foot	YOLOv5s	0.767	0.696	0.734	0.383
	YOLOv8s	0.798	0.683	0.727	0.387
	PHB	0.805	0.715	0.763	0.412

Therefore, this study adopted a two-step approach to the process of information fusion. First, the PHB model was employed to detect the pedestrian body parts within the image. Then, the pedestrian body parts were considered a subset of the overall pedestrian and combined with the overall pedestrian bounding boxes. Specifically, for each overall pedestrian bounding box, the presence of the head, hand, and foot bounding boxes within the region was assessed. If these bounding boxes are identified, the component bounding box with the highest confidence score in that region is selected and paired with the entire pedestrian bounding box. In cases where the pedestrian bounding box has a low score but the body part component bounding box exhibits high confidence, the overall bounding box is retained. Additionally, if a component bounding box demonstrates high confidence but does not match the overall pedestrian bounding box, it is preserved and output as a pedestrian label. This approach aligns with our aim, as depicted in Fig. 8, where the presence of the head, hands, feet, and other body parts indicates the presence of a pedestrian, even if the entire pedestrian is not fully visible.

We conducted a comparative analysis of the YOLOv5s- and YOLOv8s-enhanced PHB models using Dataset 1. The results are presented in Table 2. Notably, the incorporation of information from other body parts led to a significant improvement in the recall rate of the YOLOv5s-PHB model. However, it is essential to acknowledge that the accuracy, as indicated in Table 1, of the overall and head recognition of pedestrians was somewhat diminished. This could be attributed to the influence of the recognition performance associated with other body parts. In contrast, the PHB model based on YOLOv8s exhibited a slightly reduced recall rate compared with its YOLOv5s counterpart. However, this compensates for the improved precision. Consequently, it is crucial to strike a balance between recall and accuracy.

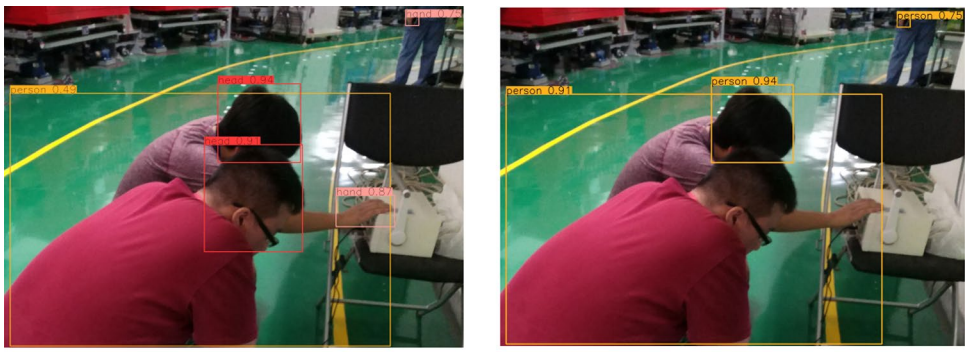
5.3.3 Comparison and analysis of classical algorithms

Upon implementation of the information fusion strategy, the PHB model demonstrated superior pedestrian recognition performance for Dataset 1 compared to YOLOv8s. However, it is important to acknowledge the limitations stemming from the relatively small scale of Dataset 1. Thus, generalization experiments must be conducted on Dataset 2 to validate the generalization capabilities of PHB and assess its effectiveness in diverse scenarios.

Under identical configuration conditions, the PHB-based intelligent search-and-secure algorithm was compared with the classical pedestrian target detection algorithm using Dataset 2. Table 3 presents the results of the study. Notably, despite being designed based on the smaller YOLOv8s

Table 2 PHB person class detection performance

Model		Precision	Recall	AP
YOLOv5s-PHB	Pre-fusion	0.924	0.874	0.916
	post-fusion	0.878	0.921	0.914
YOLOv8s-PHB	Pre-fusion	0.941	0.873	0.913
	post-fusion	0.896	0.915	0.911



(a) (b)

Fig. 8 (Color online) Comparison of effects before and after information fusion. **a** pre-fusion recognition result. **b** post-fusion recognition result

model within the YOLOv8 series, PHB achieves the same precision as the larger YOLOv8l model. The recall rate demonstrated a 13.1% increase, whereas the average detection accuracy improved by 4.4%. Furthermore, when compared to Faster R-CNN, the PHB algorithm outperformed the other algorithms in terms of overall performance. However, the accuracy and recall rates of PHB in Dataset 2 were lower than those in Dataset 1. This discrepancy arises because, in the context of the sweep system, instances in which pedestrians are obstructed by other pedestrians are infrequent. Consequently, Dataset 1, which was used to train the PHB model, prioritizes interclass occlusion and may not effectively address the challenges posed by the severe intraclass occlusion encountered in Dataset 2. In summary, the PHB-based intelligent search-and-secure algorithm guarantees high detection accuracy and a low missed detection rate, specifically in scenarios where pedestrians are obstructed by equipment.

5.3.4 The impact of fusion strategies in classical models

The PHB model based on YOLOv8s was enhanced, followed by the implementation of an information fusion strategy to enhance model performance. Subsequently, this fusion strategy was directly applied to the classical model, and a comparative evaluation was conducted against the PHB effect. The results are summarized in Table 4. Notably, the SSD model demonstrated significantly inferior performance compared with the PHB model after fusion strategy adoption. Furthermore, the recall rate of the Faster RCNN surpasses that of the PHB effect after incorporating the fusion strategy. However, it is evident that a Faster RCNN also requires nearly twice the processing time of PHB. Considering the high volume of images processed by the search-and-secure system and the emphasis on real-time performance, the PHB model enhanced by the YOLOv8 model proved more suitable [34, 35].

Moreover, our investigation included a comparison with the classical model to evaluate the recognition performance between larger targets simulated by pedestrians approaching the camera and smaller targets, such as hands and feet. Our findings indicate that although the direct application of YOLOv8 exhibited limited effectiveness on smaller targets, our enhancements successfully mitigated this constraint. Consequently, the PHB model demonstrates proficiency analogous to that of the Faster RCNN in recognizing diminutive targets. However, the PHB model excelled at identifying significantly larger targets.

Table 3 Comparison of pedestrian detection performance

Model	Precision	Recall	AP
YOLOv5s	0.822	0.698	0.79
YOLOv5l	0.846	0.74	0.833
YOLOv8s	0.861	0.687	0.793
YOLOv8l	0.867	0.689	0.798
Faster RCNN	0.813	0.796	0.781
PHB	0.869	0.82	0.842

Table 4 Comparison of fusion strategies' impact on classical models

Model		Precision	Recall	AP
SSD	Pre-fusion	0.801	0.681	0.77
	Post-fusion	0.753	0.78	0.759
Faster RCNN	Pre-fusion	0.813	0.796	0.781
	Post-fusion	0.798	0.831	0.776
PHB	Post-fusion	0.869	0.82	0.842

6 Conclusion

Based on the performance evaluation of the model, we installed two sets of 180° camera groups within a section of the China Spallation Neutron Source Accelerator Tunnel [36], as shown in Fig. 9. A relatively enclosed and controlled area was created by strategically introducing partial physical occlusion.

Several field tests were conducted within this controlled area, and the results demonstrated that the intelligent search and security system successfully detected stranded individuals and achieved notable outcomes. However, the tests revealed certain issues that require resolution. For instance, the system incorrectly identified body images within certain promotional photographs in the tunnel as pedestrian targets. These concerns will be addressed in the future as part of ongoing system enhancements.

Machine-vision-based search-and-secure technology has considerable potential for broad applications in diverse settings such as railway yards, chemical plants, museums, and

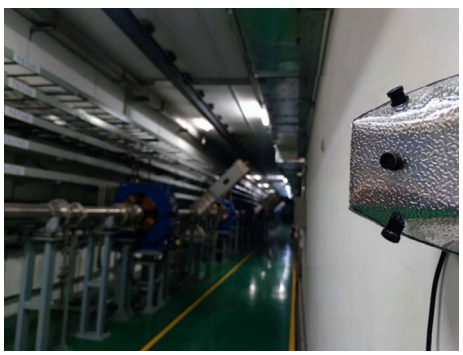


Fig. 9 (Color online) Photograph of the intelligent search-and-secure system deployed in the tunnel

other intermittent hazardous areas [37, 38]. This technology has a significant value and merits further promotion and implementation.

Acknowledgements The authors would like to express sincere gratitude to our research team and the laboratory instructors for their invaluable assistance and insightful discussions throughout the course of this study. Special thanks are extended to the China Spallation Neutron Source for generously providing access to the experimental site and the necessary dataset, which greatly facilitated the progress and outcomes of this research.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ying-Lin Ma, Yao Wang and Hui-Jie Zhang. The first draft of the manuscript was written by Ying-Lin Ma and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability The data that support the findings of this study are openly available in Science Data Bank at https://cstr.cn/31253.11.sciencedb.j00186.00491 and https://doi.org/10.57760/sciencedb.j00186.00491.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

1. Q.B. Wu, Q.B. Wang, J.M. Wu et al., Study on induced radioactivity of China Spallation Neutron Source. Chin. Phy. C. **35**, 596–602 (2011). https://doi.org/10.1088/1674-1137/35/6/017
2. A. Li, J. Zhang, X. Wang et al., Develop of safety interlock system for linear superconducting accelerator based on PLC. Nucl. Electron. Detect. **41**, 682–686 (2021). https://doi.org/10.3969/j.issn.0258-0934.2021.04.026 (**in Chinese**)
3. J. Cai, J. Xu, J.H. Wang et al., Radiation safety interlock system for neutron physics experiment facility. Nucl. Electron. Detect. Technol. **34**, 1325–1329 (2014). https://doi.org/10.3969/j.issn.0258-0934.2014.11.012 (**in Chinese**)
4. M. Yang, F.S. Chen, Y.F. Wu et al., Development of short prototype of dual aperture quadrupole magnet for CEPC ring. Nucl. Sci. Tech. **34**, 103 (2023). https://doi.org/10.1007/s41365-023-01255-7
5. H.Y. Shi, Q.B. Wang, Z.J. Ma, Preliminary design of beam dump on CEPC-Linac. Nucl. Tech. **42**, 23–28 (2019). https://doi.org/10.11889/j.02533219.2019.hjs.42.100204 (**in Chinese**)
6. H. Xie, W.Q. Zheng, H. Shin, Occluded pedestrian detection techniques by deformable attention-guided network (DAGN). Appl. Sci. **11**, 6025 (2021). https://doi.org/10.3390/app11136025
7. Y.W. Pang, J. Xie, M.H. Khan et al., Mask-guided attention network for occluded pedestrian detection. Paper presented at the IEEE/CVF International Conference on Computer Vision (ICCV), Institute of Electrical and Electronics Engineers, IEEE, 4966–4974 October 2019. https://doi.org/10.48550/arXiv.1910.06160.
8. S.F. Zhang, L.Y. Wen, X. Bian et al., Occlusion-aware R-CNN: detecting pedestrians in a crowd. Paper presented at the European Conference on Computer Vision (ECCV), ECCV Workshop, 8–14 September 2018. https://doi.org/10.48550/arXiv.1807.08407
9. S.D. Khan, Y. Ali, B. Zafar et al., Robust head detection in complex videos using a two-stage deep convolution framework. IEEE Access **8**, 98679–98692 (2020). https://doi.org/10.1109/ACCESS.2020.2995764
10. Y. Chen, W.Y. Xie, H.L. Liu et al., Multi-feature fusion pedestrian detection combining head and overall information. J. Electron. Inf. Technol. **4**, 1453–1460 (2022). https://doi.org/10.1199/JEIT210268. (**in Chinese**)
11. B. Dariusz, M. Piotr, R. Ryszard, Development of particle accelerator technology in Europe: digest of infrastructural and research projects. Paper presented at the Superconductivity and Particle Accelerators, The Henryk Niewodniczański Institute of Nuclear Physics Polish Academy of Sciences (IFJ PAN), 1105403 May 2019. https://doi.org/10.1117/12.2521228
12. J. Tang, L. Zhou, Development status and trend of particle accelerator in China. Atom. Energy Sci. Technol. **56**, 1735–1746 (2022). https://doi.org/10.7538/yzk.2022.youxian.0649 (**in Chinese**)
13. M.Z. Zhang, D.M. Li, L.R. Shen et al., SAPT: a synchrotron-based proton therapy facility in Shanghai. Nucl. Sci. Tech. **34**, 148 (2023). https://doi.org/10.1007/s41365-023-01293-1
14. J. Li, G. Chen, F. Wang, Design and application of a 360° panoramic endoscope for pipe and pipeline. Steel Pipe. **50**, 80–84 (2021). https://doi.org/10.19938/j.steelpipe.1001-2311.2021.6.80.83
15. S.J. Huang, J. Huang, S.J Wang et al., A new method for testing wide range horizontal field angle. Paper presented at the 2021 International Conference on Physics, Computing and Mathematical (ICPCM2021), MATEC Web of Conferences, 355 January 2022. https://doi.org/10.1051/MATECCONF/202235501015
16. Z. He, N. Huang, P. Wang et al., Spatial resolution and image processing for pinhole camera- based X-ray fluorescence imaging: a simulation study. Nucl. Sci. Tech. **33**, 64 (2022). https://doi.org/10.1007/s41365-022-01036-8
17. W.B. He, Y.G. Ma, L.G. Pang et al., High-energy nuclear physics meets machine learning. Nucl. Sci. Tech. **34**, 88 (2023). https://doi.org/10.1007/s41365-023-01233-z
18. Y.C. Tang, Z.F. Huang, Z. Chen et al., Novel visual crack width measurement based on backbone double-scale features for improved detection automation. Eng. Struct. **274**, 115158 (2023). https://doi.org/10.1016/j.engstruct.2022.115158
19. Y.C. Tang, L.J. Li, W.X. Feng et al., Binocular vision measurement and its application in full-field convex deformation of concrete-filled steel tubular columns. Measurement **130**, 372–383 (2018). https://doi.org/10.1016/j.measurement.2018.08.026
20. C. Li, Y.N. Cao, Y.K. Peng, Research on automatic driving target detection based on YOLOv5. J. Phys. Conf. Ser. **2171**, 012047 (2022). https://doi.org/10.1088/1742-6596/2171/1/012047
21. S. Iftikhar, Z.P. Zhang, M. Asim et al., Deep learning-based pedestrian detection in autonomous vehicles: substantial Issues and Challenges. Electronics **11**, 3551 (2022). https://doi.org/10.3390/ELECTRONICS11213551
22. X. Zhu, Research on pedestrian detection method based on YOLOV5. Agric. Equip. Veh. Eng. **60**, 108–111 (2022). https://doi.org/10.3969/J.ISSN.1673-3142.2022.04.024(inChinese)
23. Y.Y. Zeng, Y. Yu, Z.H. Zhou, Research on small target recognition model based on improved YOLOv5. J.Res. Sci. Eng. **4**, 28–33 (2022). https://doi.org/10.53469/JRSE.2022.04(11).06
24. H.W. Chen, G.H. Zhou, H.X. Jiang, Student behavior detection in the classroom based on improved YOLOv8. Sensors **23**, 8385 (2023). https://doi.org/10.3390/S23208385
25. Z. Wang, L.F. Lei, P.B. Shi, Smoking behavior detection algorithm based on YOLOv8-MNC. Front. Comput. Neurosci. **17**, 1243779 (2023). https://doi.org/10.3389/fncom.2023.1243779
26. Y.R. Jin, Z.Y. Shi, X.L. Xu et al., Target localization and grasping of NAO robot based on YOLOv8 network and monocular ranging. Electronics **12**, 3981 (2023). https://doi.org/10.3390/electronics12183981

27. X. Shi, H. Lu, P.J. Qin et al., A long-distance pedestrian small target detection method. Chin. J. Sci. Instrum. **43**, 136–146 (2022). https://doi.org/10.19650/j.cnki.cjsi.J2108848 (**in Chinese**)
28. C.C. Zhu, Y.H. He, M. Savvides, Feature Selective Anchor-Free Module for Single-Shot Object Detection. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition, Computer Vision Foundation and IEEE Computer Society, 21–26 June 2019. https://doi.org/10.48550/arXiv.1903.00621
29. S.D. Khan, S. Basalamah, Scale and density invariant head detection deep model for crowd counting in pedestrian crowds. Vis. Comput. **37**, 2127–2137 (2021). https://doi.org/10.1007/s00371-020-01974-7
30. F. Cheng, L. Mao, D.W. Yang, Multiple body parts distance fusion estimation algorithm based on monocular vision. J. Dalian Minzu Univ. **20**, 412–416 (2018). https://doi.org/10.13744/j.cnki.cn21-1431/g4.2018.05.007 (**in Chinese**)
31. Z. Zou, Z. Shi, Y. Guo, et al., Object Detection in 20 Years: A Survey. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition, Computer Vision Foundation and IEEE Computer Society, 5–11 January 2023. https://doi.org/10.48550/arXiv.1905.05055
32. R. Padilla, S.L. Netto, E.A. Da Silva, A survey on performance metrics for object-detection algorithms. Paper presented at the International Conference on Systems, Signals and Image Processing (IWSSIP), IEEE, 237–242, July 2020.
33. X.F. Qiu, X.R. Sun, Y.C. Chen, et al.,Pedestrian detection and counting method based on YOLOv5+DeepSORT. Paper presented at the 4th International Symposium on Power Electronics and Control Engineering (ISPECE 2021), SPIE, 120800 November 2021. https://doi.org/10.1117/12.2618209
34. G. Li, H. Wei, J. Ai et al., An improved pedestrian detection algorithm based on SSD. J. Guangxi Univ. (Natl. Sci. Ed.) **46**, 1327–1336 (2021). https://doi.org/10.13624/j.cnki.issn.1001-7445.2021.1327 (**in Chinese**)
35. H. Zhang, Y. Du, S. Ning et al., Pedestrian detection method based on Faster R-CNN. Transd. Microsyst. Technol. **38**(2), 147–153 (2019). https://doi.org/10.13873/j.1000-9787(2019)02-0147-03
36. S. Wang, S. Fu, H. Qu et al., Design, development, and commissioning for high-intensity proton accelerator of China Spallation Neutron Source. Atom. Energy Sci. Technol. **56**, 1747–1759 (2022). https://doi.org/10.7538/yzk.2022.youxian.0591 (**in Chinese**)
37. Y. Wang, P.Z. Yu, A fast intrusion detection method for high-speed railway clearance based on low-cost embedded GPUs. Sensors **21**, 7279 (2021). https://doi.org/10.3390/s21217279
38. S. Chen, K. Demachi, Proposal of an insider sabotage detection method for nuclear security using deep learning. J. Nucl. Sci. Technol. **56**, 599–607 (2019). https://doi.org/10.1080/00223131.2019.1611501