IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 November 2023, accepted 17 January 2024, date of publication 23 January 2024, date of current version 1 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3357519

RESEARCH ARTICLE

Multi Camera Localization Handover Based on YOLO Object Detection Algorithm in Complex Environments

WENNAN WU[1,2] **AND JIZHOU LAI**[1]

[1]College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210000, China
[2]Jiangsu Satellite Navigation Testing Center Company Ltd., Nanjing 210000, China

Corresponding author: Wennan Wu (wu.wennan@nuaa.edu.cn)

ABSTRACT With the development of computer vision, image processing, and other technologies, the management of smart cities has been enhanced, and intelligent visual detection and tracking technology has progressed. A single-camera monitoring system presents challenges, including limited observation range, unstable tracking, and difficulties in recognizing complex scene obstructions. To overcome these obstacles, a multi-camera monitoring system must be implemented. To enhance the accuracy of multiple cameras' positioning and recognition, while also increasing their efficiency in recognizing targets, this study employs a novel approach that combines spatial mapping based on position data and feature matching based on target objects. Firstly, in the overlapping area of multiple camera targets, a uniform spatial constraint method is used to map and match the target object. The color features of the target object are used for matching. Secondly, the You only look once (YOLO) object detection algorithm is introduced to recognize targets within the overlapping area of the camera using homologous transformation. In this way, a multi camera positioning technology based on YOLO object detection algorithm is designed. The test results show that the YOLOv5 algorithm has a maximum mAP accuracy of 97.2% on the test set. At a reasoning speed of 10 ms, the YOLOv5 algorithm has a maximum mAP accuracy of 51.6%. The average values of the classification loss function, target loss function, and GIoU loss function of the YOLOv5 algorithm are 0.001, 0.01, and 0.015, respectively. The error probability of YOLO within 10cm in the DukeMTMC re TD dataset remains above 96.5%. The error probability of YOLO within 9.5cm in the OTB dataset remains above 95%. When the target object is blocked, the highest accuracy of the YOLO positioning system is 0.74. The above results indicate that the multi camera localization technology based on YOLO object detection algorithm can improve the accuracy of localization and recognition. It can also solve the problems of object occlusion recognition and continuous object tracking.

INDEX TERMS Object detection, multiple cameras, location tracking, YOLO algorithm, homomorphic transformation.

I. INTRODUCTION

The rapid development of computer technology and image processing has increased the demand for security management in smart cities. Intelligent visual detection-based monitoring technology is becoming more advanced [1]. Nowadays, video surveillance systems are widely used in daily life and society. The advancement of camera technology and declining costs have accelerated the growth of video surveillance systems, yet problems have come to light in the realms of video surveillance and artificial intelligence. Firstly, the chip core technology relied on for video surveillance is still not available in China. Secondly, China has an advantage in data compared to computational power and algorithms, which puts them at a disadvantage among the three elements of artificial intelligence. Thirdly, the

The associate editor coordinating the review of this manuscript and approving it for publication was Venkata Ratnam Devanaboyina.

development speed of domestic video surveillance systems has been slowed down by the lack of feature vector databases and secure and controllable domestic operating systems [2], [3]. However, multiple cameras use multiple angles to monitor video information. It can cover a wider monitoring range and have better positioning accuracy and tracking efficiency [4]. The intelligent surveillance technology utilizing multiple cameras has gained attention from researchers with significant advancements made in current monitoring technologies. Nonetheless, due to complex real-world environments and various interfering factors such as lighting and angle, the technology for multi-camera target positioning and tracking has yet to achieve optimal results. Object detection is one of the key topics in the computer vision and is widely applied in the industrial field [5]. The You only look once (YOLO) object detection algorithm has the advantages of fast detection speed and simple network structure, and can use deep neural networks to classify and extract image features [6]. The YOLOv8 is a relatively improved algorithm in the YOLO series, which solves the problem of low accuracy in large-scale data based on the YOLO algorithm [7]. Therefore, achieving accurate positioning and recognizing multiple cameras in complex environments requires consideration of both the accuracy of target feature extraction and the handover of targets between cameras. This paper concentrates on developing a multi-camera target recognition and positioning system, enhancing current target recognition and tracking algorithms, and introducing a collaborative strategy utilizing both spatial and color features. This paper compares traditional moving object algorithms and performs corner detection and optical flow feature extraction on the extracted target area to obtain the motion features of the moving object. This study establishes a spatial model between multiple cameras and the target by constraining the spatial position of the cameras under multiple cameras. Then, the color features of the target are studied and a target matching method based on color name features is proposed. This method reduces the dimensionality of the color features of the target and completes color feature matching for the same target. Finally, a color name feature matching strategy based on spatial constraints was established. It compares the color name features of the target under occlusion conditions with the color name features under normal conditions to complete target occlusion determination. The research focus is on multi camera target recognition. The focus of re identification research is to obtain features from two different planes. The focus of this study is to solve the precise positioning and recognition of cross camera intersection in complex environments such as small observation range, unstable target tracking, and the presence of occlusion. To achieve this goal, a target occlusion judgment method combining target spatial position constraints and color features is used. The first step is to obtain the homography matrix between multiple cameras. Then, the spatial position of the target is determined from multiple cameras, and the same target is recognized based on its spatial constraints. Finally, the color features of the target are obtained for feature matching. When different target positions overlap and the color features undergo significant changes, it is determined that the target occlusion phenomenon occurs. There are currently two main methods for solving cross camera object handover. One is the spatial mapping based on location information, and the other is matching based on the features of the target object. However, both methods have limitations. The innovation of this paper is to combine the advantages of the two methods. In the area where multi camera targets overlap, the target object is mapped and matched using the Homography space constraint method, and then matched using the color features of the target object. Combining the advantages of the two methods can ensure the recognition efficiency and improve the recognition accuracy at the same time. In this context, this project conducts research on multi camera positioning technology built on YOLO object detection algorithm to achieve higher positioning and recognition accuracy.

YOLOv8 algorithm is a relatively new improved algorithm in the current YOLO series of algorithms. The YOLOv8 algorithm solves the problem of low accuracy in large-scale data based on the YOLO algorithm. Therefore, achieving precise positioning and recognition of multiple cameras in complex environments improves the accuracy of target feature extraction. It also solves the problem of cross camera positioning and switching recognition. The structure and contribution of this study are shown in Figure 1.

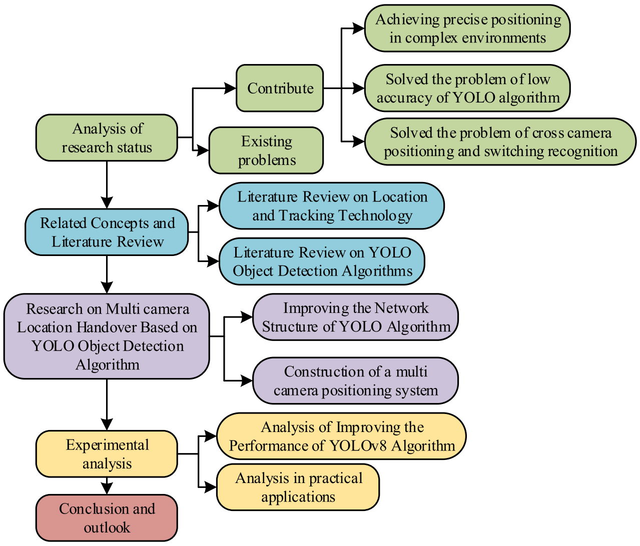

Achieving precise positioning in complex environments
Contribute
Solved the problem of low accuracy of YOLO algorithm
Analysis of research status
Existing problems
Solved the problem of cross camera positioning and switching recognition
Literature Review on Location and Tracking Technology
Related Concepts and Literature Review
Literature Review on YOLO Object Detection Algorithms
Research on Multi camera Location Handover Based on YOLO Object Detection Algorithm
Improving the Network Structure of YOLO Algorithm
Construction of a multi camera positioning system
Experimental analysis
Analysis of Improving the Performance of YOLOv8 Algorithm
Analysis in practical applications
Conclusion and outlook

FIGURE 1. **Structural arrangement and contribution.**

Through comparative analysis and research of existing studies, it can be found that each method has its unique advantages and disadvantages. For example, near eye displays solve the problems of visual quality and natural interaction, but their small field of view limits the user experience; The Markov model increases the confidentiality of user locations, but it can only predict the next state and cannot predict further states; Fuzzy detection strategy technology improves tracking speed and stability in complex environments, but cannot accurately locate vulnerabilities. In addition, the Light Detection And Ranging (LiDAR) ranging sensor system can detect low-light objects, but it cannot operate in extreme weather conditions. Object detection based on deep learning improves object detection efficiency, but each region does not share features, and the training process is relatively cumbersome. The trilateral feature pyramid model based on YOLO algorithm improves the recall and accuracy of object detection systems, but requires multiple changes in image size, which is time-consuming. Although the YOLO algorithm has excellent overall performance, improvements are still needed in the field of multi camera localization and recognition. The contribution of this article is to propose a new object detection algorithm - YOLOv8 algorithm. This algorithm effectively improves the accuracy issue in large-scale data based on the YOLO algorithm. Therefore, in order to achieve precise positioning and target recognition of multiple cameras in complex environments, it not only improves the accuracy of target feature extraction, but also successfully solves the recognition difficulties during cross camera positioning and switching. Compared with existing research, YOLOv8 algorithm has higher accuracy and faster speed, and can be better applied in multi camera systems. It can achieve significant breakthroughs in target detection and positioning technology, promoting the development and innovation of this field. The contribution of this article is to provide reference and inspiration for future research, and to provide better solutions for practical applications.

II. RELATED WORKS

Currently, multi camera localization is one of the key research areas in computer vision. With the development of Internet of Things technology, intelligent monitoring systems have broad application prospects. The massive monitoring data is difficult to efficiently utilize solely by manpower. Multi camera positioning is not limited by field of view and is of great significance in fields such as transportation and security. Compared to a single camera, multi camera joint tracking can better meet people's actual needs and is more in line with the current trend of intelligence and networking. Intelligent multi camera joint monitoring systems are gradually being sought after by educational institutions, security departments, and government departments, and have great development prospects. Therefore, it has also attracted the attention of many domestic and foreign experts and scholars, and has conducted many studies on it. Among them, research on positioning and tracking technology has achieved many results. Koulieris G A et al. believed that near eye displays occur an essential position in virtual reality. They proposed a near-eye display that combines tracking technology and near-eye display hardware, citing a structured overview of visual perception principles and tracking technology. This method solved the challenges of visual quality and natural interaction in virtual reality [8]. Guo X's team proposed a fusion positioning system based on multiple technologies to enhance the application value of positioning technology in military and commercial fields. The system integrated three features: algorithm, weight space, and source to form a framework. The results indicated that the system has strong robustness in complex electromagnetic environments [9]. Sangaiah A K and other scholars raised an intelligent device localization program that combines machine learning technology to increase the location confidentiality of users using roaming location services. This method used Markov models to locate user position sequences and utilized a merged decision tree to identify user positions. The confidentiality of this program in location services had reached 90% [10]. Liu S et al. constructed a fuzzy detection strategy technology that combines correlation filters to improve tracking efficiency in complex environments. To avoid template contamination, this method utilized target templates in memory for tracking and conducted testing experiments on the OTB100 dataset. Experiments have shown that this technology improves tracking speed and stability in complex environments [11]. Scholars such as He W proposed adaptive control technology based on a layered framework to achieve autonomous tracking of flapping wing micro air vehicles on the plane. This scheme utilized FWMAV dynamics to calculate the flapping frequency of the wing and the aerodynamic force generated by the tail, and designed a position controller using a hyperbolic tangent function. This technology proved the applicability of the control scheme [12]. To explore and study intelligent building sites, Edirisinghe R conceived the concept of digital skin based on building information visualization and mobile device tracking progress monitoring. His concept has built the future safety management and building procurement management system, and carried out user acceptance testing on the application site, verifying the feasibility of the future intelligent building site [13].

The YOLO algorithm plays a crucial role in localization and tracking techniques. Wang G and other scholars proposed an object detection model built on the TRC-YOLO to achieve the application of object detection methods on embedded devices. This method introduced spatial attention into the convolutional attention module, and also added a Receptive field that simulates human vision. This method achieved real-time performance of 31.8frames/s and was highly efficient in applications on embedded devices [14]. Liang S et al. put forward a deep learning based object detection system to address the low efficiency and high energy consumption issues of object detection technology. The system constructed a lightweight edge cloud collaborative object detection framework and combined compressed feature fusion networks with feature pruning extraction networks. The results showed that the accuracy of this scheme reaches 48.9%, effectively improving the system efficiency [9]. To realize the automatic navigation technology of autonomous vehicle, the research team of Dazlee NMAA proposed a laser radar ranging sensor system based on YOLO. Through the form of pulsed laser, low-light objects are detected, and the various properties of the system can be tested under the same parameter conditions. Tests have proven that the system has achieved ideal levels of accuracy and precision [15]. Lee J et al. raised a high-performance embedded system based on YOLO to improve the high accuracy and ease of use of multi-agent video applications. The system introduced an adaptive control model in the new YOLO architecture and performs object detection performance verification on AI embedded systems. The system that retained the YOLO algorithm had high accuracy and convenience, and was applicable in multi intelligent video applications [16]. Zhang S and other scholars proposed a YOLO network system based on deep learning to improve the efficiency of object detection in vehicle terminals. The system utilized deep separable convolution method to optimize the YOLO network, and in order to improve operational efficiency. The convolution operation was decomposed into point by point convolution. The experiment showed that the detection speed of the system reaches four times that of the original system while maintaining the same detection accuracy [17]. To improve the recall and accuracy of object detection systems, researchers such as Wang G proposed a trigonal feature pyramid model that combines the YOLO algorithm. It improved the network structure based on the YOLOv4 algorithm and constructs a spatial pyramid layer in the model. This method improved the recall and accuracy of the object detection system by 1.9% and 4.5%, respectively, proving the effectiveness of this method [18]. Scholars such as Diwan T explored the difference between single target object detection and multi target object detection. This study suggested that single target object detection mainly focuses on proposing strategies through selective regions with complex structures. Single-target target detection, on the other hand, detects all spatial regions of a target in a single shot by means of a relatively simple structure. The performance of any object detector was evaluated by detection accuracy and inference time. Usually, the detection accuracy of multi target object detection was better than that of single target object detection. However, the inference time for single target object detection was better than its corresponding object. In addition, YOLOs are mainly adopted in various applications due to their faster reasoning rather than considering detection accuracy [19]. Guo Z's team improved the primary detector by introducing the Microsoft-You Only Look Once (MSFT-YOLO) model to achieve industrial object detection while also considering accuracy and real-time performance. This model was suitable for industrial scenarios with high image background interference, easy confusion of defect categories, large change in defect scale, and poor detection performance for small defects. Designing a multi-scale feature fusion structure to fuse features of different scales enhanced the dynamic adjustment of the detector to objects of different scales. The results indicated that the average detection accuracy of MSFT-YOLO on the NEU Detections (NEU-DET) dataset is 75.2% [20]. Scholars such as Diwan T summarized the advantages of multi-objective object detection algorithms by comparing different versions of YOLOs. Guo Z and others improved the primary detector to achieve the accuracy and real-time performance of industrial object detection. The final model studied had an average detection accuracy of 75.2%. The above researchers have achieved some results in the field of target detection. However, the accuracy of the model studied in this paper is 81%, which is 5.8% higher than the accuracy of the model studied by Guo Zheng and other scholars, which is 75.2%. Therefore, the recognition accuracy of the studied model is more advantageous than other research models in complex environments.

In summary, YOLO object detection algorithm plays an important role in computer vision. Moreover, positioning and tracking technology has attracted the attention of many scholars due to its wide range of applications and powerful functions. It leads to the emergence of many positioning and tracking technologies. Multi-camera can expand the surveillance field of view, take advantage of cross-information to accomplish target switching, and discover occluded targets. Therefore, this study explores this technique in depth to achieve higher localization and recognition accuracy. YOLOv8 algorithm is a relatively new improved algorithm in the current YOLO series of algorithms. The YOLOv8 algorithm solves the problem of low accuracy in large-scale data based on the YOLO algorithm. Therefore, the realization of multi-camera accurate positioning and recognition in complex environments not only improves the accuracy of target feature extraction, but also solves the problems of cross-camera localization and switching recognition. The YOLOv8 algorithm effectively improves the accuracy problem in large-scale data based on the YOLO algorithm. Therefore, in order to achieve precise positioning and target recognition of multiple cameras in complex environments, it not only improves the accuracy of target feature extraction, but also successfully solves the recognition problem during cross camera positioning and switching.

III. MULTI CAMERA LOCALIZATION HANDOVER BASED ON YOLO OBJECT DETECTION ALGORITHM IN COMPLEX ENVIRONMENTS

A. IMPROVEMENT OF NETWORK STRUCTURE BASED ON YOLO OBJECT DETECTION ALGORITHM

YOLO algorithm is an object detection algorithm that utilizes deep neural network models for image processing. Figure 2 shows the network model structure of YOLO algorithm.

In Figure 2, the YOLO algorithm consists of an input layer, a fully connected layer, a pooling layer, and a convolutional layer. Figure 3 shows the detection model flow of YOLO algorithm.

In Figure 3, the input image will be first divided into an $S \times S$ -grid format, with each grid detecting objects that fall within it. The N target frames are predicted for the detected object, containing 5 prediction parameters, i.e. the height h, width w, center coordinate (x, y), and confidence score v_i. The mathematical Equation for confidence

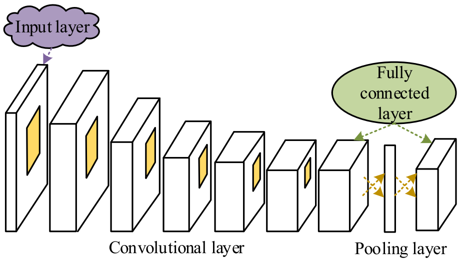


FIGURE 2. Network model structure diagram of YOLO algorithm.

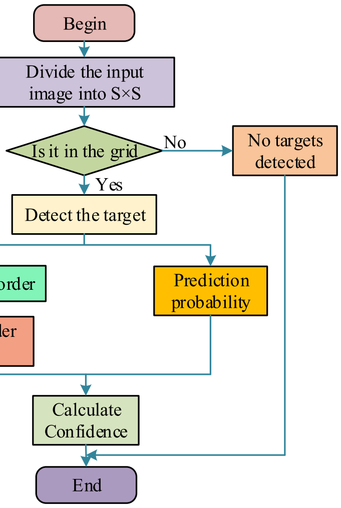


FIGURE 3. Flow chart of YOLO algorithm's detection model.

score is equation (1).

$$v_i = pr\left(S\right) * IoU \tag{1}$$

In equation (1), v_i represents the confidence score. $pr\left(S\right)$ is the probability of the target object existence in the grid. S is the target object. IoU is the intersection and union ratio, indicating the accuracy of the current predicted target border position. The mathematical Equation of IoU is shown in equation (2).

$$IoU_{\beta}^{\alpha} = \frac{box_{\beta} \cap box_{\alpha}}{box_{\beta} \cup box_{\alpha}} \tag{2}$$

α in equation (2) represents the actual target border that exists. β represents the predicted target border. IoU_{β}^{α} represents the intersection and union ratio between the actual target border and the predicted target border. box_{α} represents the actual situation of the target border in existence. box_{β} represents the predicted target border situation. The confidence level of the existence of objects in the target border is equation (3).

$$pr\left(A_i\left|S\right.\right) * pr\left(S\right) * IoU_{\beta}^{\alpha} = pr\left(A_i\right) * IoU_{\beta}^{\alpha} \tag{3}$$

In equation (3), $pr\left(A_i\right)$ represents the probability of the presence of a target object within the target border. $pr\left(A_i\left|S\right.\right)$ represents the probability that the target object within the target border belongs to class-i object. The value range of i is $[1, 2, \ldots, M]$, where M represents the total number of target object types [21]. Although the YOIO algorithm has the benefits of simple network structure and fast algorithm detection speed, there are still some limitations, such as large parameter volume and low detection accuracy. To deal with it, the YOLOv8 is proposed to upgrade the network structure in line with YOLO. Figure 4 lists the structure of the YOLOv8 algorithm.

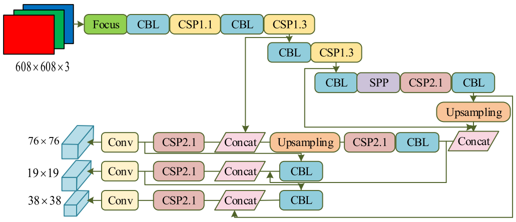


FIGURE 4. Structure diagram of YOLOv8 algorithm.

From Figure 4, the improved YOLOv8 network structure mainly includes Focus module, CSP module, CBL module, and SPP module. The Focus module refers to slicing the image and downsampling the image. This method concentrates the width and height information of the image on the channel dimension, thus expanding the channel space and reducing the loss of information. The CSP module reduces the computational volume in structural design while improving detection accuracy. The forward propagation Equation of CSP structure is shown in equation (4).

$$\begin{cases} x_t = w_t \cdot \left(x''_0, x_1, \ldots, x_k\right) \\ x_k = w_k \cdot \left(x''_0, x_1, \ldots, x_{k-1}\right) \\ x_e = w_e \cdot \left(x'_0, x_t\right) \end{cases} \tag{4}$$

x'_0 in equation (4) represents that the underlying channel is directly connected to the end. x''_0 represents that the underlying channel needs to pass through dense blocks and transition layers. x_t represents the output data obtained through dense blocks and transition layers, while x_e represents the output result of x_t after passing through the transition layer. The backpropagation update Equation is equation (5).

$$\begin{cases} w'_t = f\left(w_t, g''_0, g_1, \ldots, g_k\right) \\ w'_k = f\left(w_k{\cdot}g''_0, g_1, \ldots, g_{k-1}\right) \\ w'_e = f\left(w_e, g'_0, g_t\right) \end{cases} \tag{5}$$

In equation (5), w'_t represents the output data obtained by reverse passing through dense blocks and transition layers. w'_g represents the output result after reverse passing through transition layers. Dense blocks are output within the $\left[x''_0, x_1, \ldots, x_k\right]$ range. The CSP structure prevents duplicate computation of information, thereby reducing computational complexity and improving operational efficiency. The loss function is made of classified loss and regression loss, and the intersection and combination ratio IoU represents the distance between the real box and the prediction box. Thus, the regression loss is often calculated by the intersection and combination ratio. The Equation (6) is the loss function of regression loss.

$$\begin{cases} GIoU = IoU - \dfrac{\left|H/(A \cup B)\right|}{|H|} \\ GIoUloss = 1 - GIoU \end{cases} \tag{6}$$

In equation (6), A is the predicted box. B represents the true box. H represents the minimum closed box of A and B. Equation (7) is the loss function of classified loss.

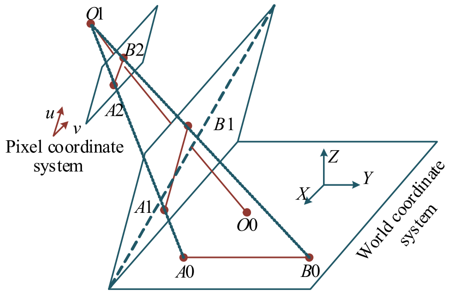
$$\begin{aligned} Loss\,(obj) &= GIoUloss + \sum_{i=0}^{S\times S}\sum_{j=0}^{B} 1_{ij}^{obj}\left[H_i \log(H_i)\right. \\ &\quad \left. + (1-H_i)\log(1-H_i)\right] \\ &\quad - \sum_{i=0}^{S\times S}\sum_{j=0}^{B} 1_{ij}^{noobj}\left[H_i \log(H_i) + (1-H_i)\log(1-H_i)\right] \\ &\quad + \sum_{i=0}^{S\times S}\sum_{j=0}^{B} 1_{ij}^{obj} \sum_{h\in classes} \\ &\quad \left[P_i(h)\log(P_i(h)) + (1-P_i(h))\log(1-P_i(h))\right] \end{aligned} \tag{7}$$

In equation (7), $Loss\,(obj)$ represent classification Loss function. The complete classification loss function includes 3 parts: confidence prediction loss, regression loss and category budget loss. This new network structure can improve the detection quality of medium and small target objects.

B. CONSTRUCTION OF A MULTI CAMERAPOSITIONING SYSTEM

Multiple cameras often have shooting coverage areas in the spatial distribution, which ensures that there are no blind spots in monitoring and can continuously track target objects [22]. To automatically determine the target in the next camera's field of vision, it is necessary to match the target in the overlapping area of the camera. The plane Homography transformation belongs to a plane mapping relationship. Two pictures can be regarded as two planes. By associating the points where two planes overlap, the same part in two planes can be found. The Homography matrix can realize the perspective transformation between views. So the Homography matrix is often used to associate the points in the overlapping area of two planes. The difference between multi camera tracking and single camera tracking is that there is an additional issue of cross camera tracking. The main method of cross camera tracking is to correlate the information in the photographed Overlap zone, and use Homography transformation to match the target objects in the overlapping area. The diagram of homography matrix mapping is displayed in Figure 5.

From Figure 5, to locate the target object within the camera area, it is vital to describe the position of the target object from its feet. Therefore, there is no need to convert the 3D world coordinates to the 2D pixel coordinates, only the 3D world coordinates need to be mapped to the 2D pixel coordinates. The world coordinate system is the absolute coordinate system of the system. Before establishing a user coordinate system, the coordinates of all points on the screen are determined based on the origin of the coordinate system. The world coordinate system is a fixed coordinate system. The X-axis is the horizontal axis and the Y axis belongs to the vertical axis. The Z axis is perpendicular to the XY plane, and the origin is the intersection point (0,0,0) of the X, Y, and Z axes in the lower left corner of the graphic boundary. The user coordinate system is a movable coordinate system, which is defined by the user referring to the world coordinate system. Using the projection and Pinhole camera imaging principle, the real 3D world coordinates can be converted into 2D pixel coordinates. Translating the pixel coordinates in the same plane can obtain the image coordinates. The isomorphic matrix can be obtained by using the Homography matrix to describe the mapping relationship between two planes. First, a homography matrix between the camera and the world coordinate can be calculated. It describes the positional relationship between the world coordinate and the pixel coordinate, thus achieving the positioning of the target object. Then, homography matrix transformation is performed between the two cameras to achieve pixel correlation in the shooting coverage area of the two cameras. This method can simplify the collaboration of position information among multiple cameras, thereby achieving precise positioning of the same target object at different shooting angles [23]. The relationship between points is calculated in equation (8).



错误！未指定书签。

FIGURE 5. Schematic diagram of homography matrix mapping.

$$\begin{cases} p = QP \\ \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \end{cases} \tag{8}$$

In equation (8), Q represents the projection matrix of the camera. p represents the projection point in the camera image plane. $P\,(X, Y, Z)$ represents the coordinates of a 3D space. P represents a spatial point. $p\,(x, y)$ is the coordinates in the plane. p represents a planar point. The spatial point with the center point of the camera in a straight line is connected. The point where the line intersects the plane is the plane point. On the contrary, points on the straight line in three-dimensional space can be represented as spatial points corresponding to planar points [24]. The plane of coordinate $Z = 0$ is selected. The mathematical expression for dimensionality reduction of the projection matrix is equation (9).

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{9}$$

In equation (9), at this point, the image plane points correspond one-to-one to the spatial plane points. This corresponding change relationship is called homography transformation [25]. The difference between multi camera tracking and single camera tracking is that there is an additional issue of cross camera tracking. The main method of cross camera tracking is to associate information in the overlapping area of the shot, and use homography transformation to match target objects within the overlapping area. Figure 6 shows the multi camera localization and tracking framework.

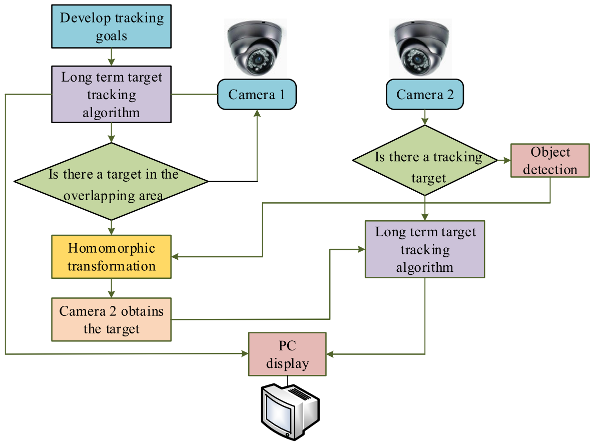

Develop tracking goals
Long term target tracking algorithm
Camera 1
Camera 2
Is there a target in the overlapping area
Is there a tracking target
Object detection
Homomorphic transformation
Long term target tracking algorithm
Camera 2 obtains the target
PC display

FIGURE 6. **Multi camera positioning and tracking framework diagram.**

From Figure 6, camera 1 specifies that the tracking target is in the tracking state, while camera 2 is in the target detection state. When the target enters the overlapping area of the camera, the unio transformation algorithm is used to identify and match the target within the overlapping area. When the distance between the camera 2 and the target detection is less than a threshold value, the matching of the target can be realized. Then the camera 2 can determine the identity of the tracking target and track it until the target leaves the field of view of the camera 2. The main goal of multi-camera positioning and tracking is to track the target object in a large range, but the difficulty of multi-camera lies in the fusion and collaboration between information. How to effectively match the same target under different cameras, and how to identify the target object in the case of occlusion, determine the accuracy of multi-camera recognition [26]. To solve these problems, multiple cameras can be used to identify and match the color and texture of the target object, so as to achieve the accurate recognition and matching of the target object. The color features of the target object can reflect the information of the target object, and the target object can be accurately identified by extracting the color features of the target object. When the target object is occluded by other objects, extract the color features of both objects and match the object close to the camera. Then use an external rectangular box to subtract the object far from the camera from the overlap between the two objects. Finally, the color features of the target object are extracted and matched. The occlusion determination diagram is shown in Figure 7.

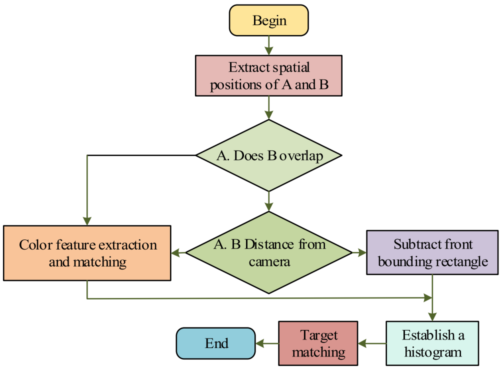

Begin
Extract spatial positions of A and B
A. Does B overlap
Color feature extraction and matching
A. B Distance from camera
Subtract front bounding rectangle
End
Target matching
Establish a histogram

FIGURE 7. **Occlusion detection diagram.**

From Figure 7, after the spatial position of the target is extracted, the color features of the target are first used to determine whether there is occlusion of the target object. To determine whether to subtract the front outer rectangular box based on the distance between the target object and the camera. Finally, to establish a color histogram for target matching. The collected color map is called a histogram. After normalizing the histogram, Euclidean distance method is used to calculate similarity. The smaller the value of Euclidean distance, the greater the similarity [27]. The mathematical Equation for Euclidean distance is calculated in equation (10).

$$M_E\left(R, D\right) = \sqrt{\sum_{K=0}^{N} \left[H_R\left(K\right) - H_D\left(K\right)\right]^2} \tag{10}$$

In Equation (10), N represents the amount of colors in the histogram. R and D represent the original image and the contrast image. $H_R\left(K\right)$ represents the histogram of the original image. $H_D\left(K\right)$ is the histogram of the contrast image. When the value of the Euclidean distance is greater than the threshold, the target object is obstructed. At this time, other cameras can be used for spatial supplementation to mark the relative position information of the current target object, thereby completing the localization and recognition of the obstructed object. In addition to matching based on feature information, image matching can also be based on grayscale, and the common grayscale matching method is also known as normalized matching. When using color features for matching, feature extraction is often affected by lighting. So a large number of mathematical methods are often required for pre processing before extraction. Compared to feature matching, grayscale matching extracts fewer feature points, which can improve the efficiency of matching. Different cameras may be affected by parameters, lighting, angle and other factors when shooting the same target. This leads to differences in the displayed images, making target matching difficult [28]. To reduce the influence of external factors on feature extraction, Scale Invariant Feature Transformation (SIFT) can be used. SIFT has high robustness in scale transformation and rotation of local features. Figure 8 is the flowchart of SIFT feature extraction.

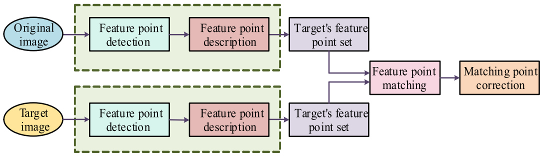

Original image
Feature point detection
Feature point description
Target's feature point set
Feature point matching
Matching point correction
Target image
Feature point detection
Feature point description
Target's feature point set

FIGURE 8. Flowchart of SIFT feature extraction.

Figure 8 shows the steps of SIFT feature extraction. The first step is to generate a Gaussian scale space to ensure constant resolution, and then preliminarily explore key points for spatial extreme point detection. The second step is to accurately locate key points and allocate directional information [29]. Finally, feature point matching is performed through the obtained key point descriptions. The function expression of Gaussian scale space is listed in equation (11).

$$\begin{cases} L(x, y, \alpha) = G(X, Y, \alpha) \cdot R(x, y) \\ G(x, y, \alpha) = \dfrac{1}{2\pi\alpha^2} e^{\frac{x^2+y^2}{2\alpha^2}} \end{cases} \tag{11}$$

In Equation (11), $L(x, y, \alpha)$ is the scale space of the Gaussian kernel. $G(x, y, \alpha)$ represents the Gaussian function. $R(x, y)$ is the original image. α represents the scale space factor, which is used to describe the scale of the fuzzy kernel. The scale space factor is the standard deviation of the Gaussian function. The larger the value of α, the more blurred the image will be. To improve computational efficiency, the Differential of Gaussian (DOG) operator can be used to extract feature points. The mathematical Equation of DOG is equation (12).

$$\begin{aligned} D(x, y, \alpha) &= [G(x, y, k\alpha) - G(x, y, \alpha)] \cdot R(x, y) \\ &= L(x, y, k\alpha) - L(x, y, \alpha) \end{aligned} \tag{12}$$

In equation (12), $L(x, y, \alpha)$ represents the Gaussian scale space, and k is the scaling factor of adjacent Gaussian scale spaces [30]. To avoid edge effects and increase its robustness and anti noise ability of the DOG operator, fitting curves can be used to locate key point positions. The mathematical Equation for fitting curve denoising is shown in equation (13).

$$\begin{cases} M = \begin{bmatrix} D_{xx} D_{xy} \\ \\ D_{xy} D_{yy} \end{bmatrix} \\ \dfrac{Tr(M)^2}{Det(M)} = \dfrac{\left(D_{xx} + D_{yy}\right)^2}{D_{xx} D_{yy} - D_{xy}^2} = \dfrac{(\varepsilon + \delta)^2}{\varepsilon\delta} = \dfrac{(1 + r)^2}{r} \\ \dfrac{Tr(M)^2}{Det(M)} < \dfrac{(1 + r)^2}{r} \end{cases} \tag{13}$$

In equation (13), ε and δ are the gradients of the eigenvalues M in the horizontal and vertical directions, respectively. After removing the noise around the object, more accurate feature points can be obtained. To ensure that the key points do not deform after rotation, a reference direction needs to be set. The gradient expression is as in equation (14), shown at the bottom of the next page.

Equation (14) calculates the gradient value of each key point, and considers the direction with the most statistics as the main direction. Then key points can be described based on vector information such as their position, scale, and direction. When constructing a descriptor, it not only includes information about the target object, but also includes neighborhood information [31]. To ensure the accuracy of neighborhood information, it is necessary to first rotate the coordinate axis of the neighborhood to the main direction of the key points, generate descriptors, and calculate the gradient of each position in the neighborhood. After calculating the gradient information, the feature vectors of each key point are obtained. To reduce the impact of lighting, normalization is used to process the feature vectors, and the initial feature vector is set to $E = (e_1, e_2, L, e_m)$. The normalized feature vector is equation (15).

$$L_i = \frac{e_i}{\sqrt{\sum_{j=1}^{m} e_j^2}} i = 1, 2, 3Lm \tag{15}$$

With the development of computer vision and the popularization of video surveillance systems, video surveillance systems based on multiple cameras have been increasingly valued and developed. Compared with single cameras, multiple cameras have better monitoring effects in both target occlusion and monitoring areas. However, how to utilize the collaborative information between multiple cameras to obtain more target features and higher recognition accuracy is still the research direction and difficulty of multi camera systems. In this paper, we improve the existing target recognition algorithms on the basis of combining spatial feature recognition methods with color feature recognition methods. A color feature matching strategy based on spatial constraints is established [9]. By comparing the color features of the target under occlusion with normal conditions, the occlusion of the target object is determined.

IV. VERIFICATION OF MULTI-CAMERA POSITIONING SYSTEM BASED ON YOLO OBJECT DETECTION ALGORITHM

V. VERIFICATION OF PERFORMANCE BASED ON IMPROVED YOLOv8 ALGORITHM

To verify the performance of a multi-camera positioning system based on the YOLO object detection algorithm, the improved YOLOv8 algorithm was first compared and verified with the YOLO series of algorithms. The experimental configuration is exhibited in Table 1. The deep learning training framework used in this experiment is PyTorch, and the optimization method is Momentun with a coefficient of 0.9. Batch_size is set to 64 and the learning efficiency is 0.01. YOLOv8 algorithm was compared with YOLO series algorithms and experimental results are obtained.

TABLE 1. Experimental environment configuration table.

Operating system	Ubuntu16.04.7LTS
Internal storage	32GB
Processor	Intel Core i7-9700
Video storage	11GB
GPU	NVIDIA GeForce GTX 1080Ti
Order	Python3.7
Deep learning framework	PyTorch1.5

The YOLOv8 algorithm and a series of algorithms such as YOLO, YOLOv2, YOLOv3, and YOLOv4 were tested in the above experimental environment. The algorithms were compared using the evaluation metric mAP and runtime of deep learning models. The accuracy comparison of each algorithm is Figure 9. In Figure 9, the horizontal axis refers to the running time and the vertical axis is mAP. YOLOv8 achieved a maximum mAP of 97.2% on the test set, which is 8.5% higher than YOLOv4's 88.7%: YOLOv8 algorithm has high accuracy. In terms of runtime, YOLOv8 has a maximum runtime of 148ms, which is still within an acceptable range, although it runs slower than other algorithms. Overall, the improved YOLOv8has higher accuracy than the YOLO series algorithm.

To verify the performance of the target detection algorithms in terms of inference speed, recall, and loss function, the experiments were conducted using the MS COCO dataset to test the YOLOv8 and YOLO series algorithms in the same experimental environment. The inference speed of the target detection algorithm is shown in Figure 10. In Figure 10, the abscissa is the inference speed of the model on the GPU. The vertical axis represents the accuracy achieved by the model on the MS COCO dataset. Within the inference speed of 10ms, YOLOv8 has the highest accuracy of 51.6, which is 3.5 times higher than YOLOv4's highest accuracy of 48.1.

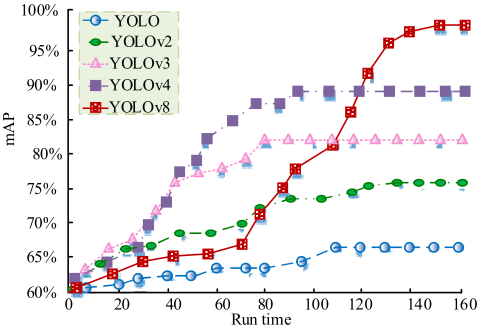


FIGURE 9. Accuracy comparison chart of various algorithms.

Although YOLO has a maximum accuracy of 52.1, slightly higher than YOLOv8, its inference speed reaches its highest value at 54ms, while YOLOv8 reaches its highest value at 6ms. Overall, it proves that the YOLOv8 has higher accuracy and faster inference speed, indicating good practicality.

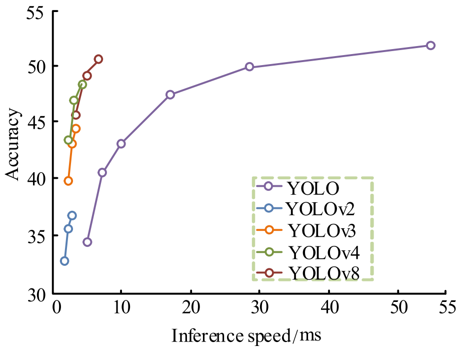
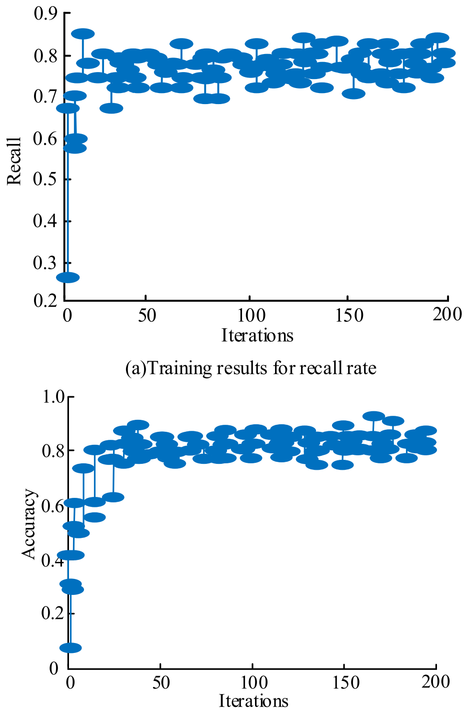


FIGURE 10. Inference speed graph of target detection algorithm.

To test the recall rate, another important performance indicator of target detection, the confidence level was set to 0.001 to verify the training results of YOLOv8 accuracy and recall rate. The training results for recall and accuracy are shown in Figure 11. In Figure 11 (a), the horizontal axis represents the iterations and the vertical axis represents the recall rate. The recall rate almost shows a vertical upward trend within the range of 0 to 10 iterations. After 10 iterations, the recall rate fluctuates around 0.79, with a fluctuation range of no more than 0.1, indicating a good training effect. In Figure 11 (b), the horizontal axis is the iterations and the vertical axis represents accuracy. Within the range of 0-10 iterations, accuracy also shows a vertical upward trend; After 10 iterations, it shows a fluctuation state of around 0.81, with a relatively concentrated fluctuation range indicating that accuracy training is in the optimal stage.

$$\begin{cases} \theta(x,y) = \tan^{-1}\dfrac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)} \\ m(x,y) = \sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2} \end{cases} \tag{14}$$



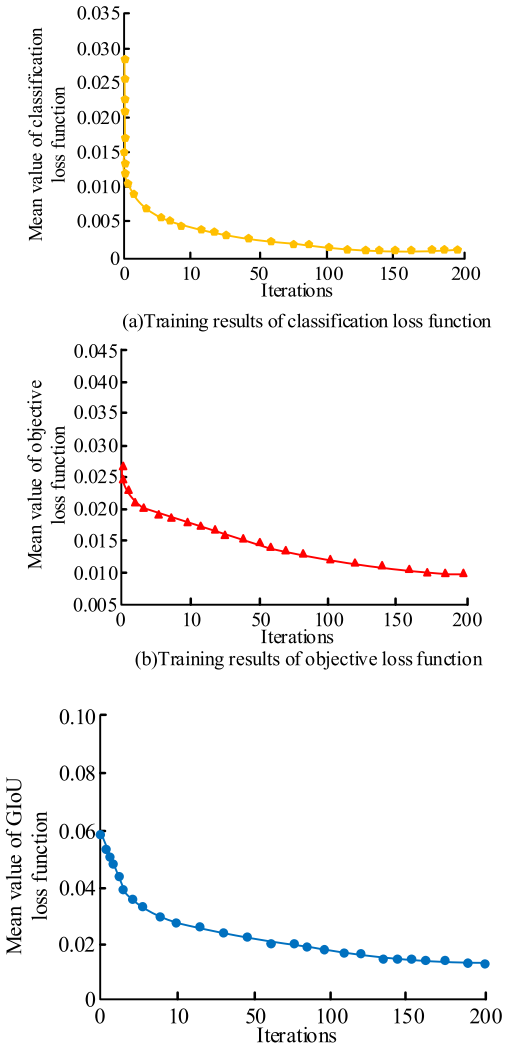
(a)Training results for recall rate

(b)Accuracy training results

FIGURE 11. Training result chart for recall and accuracy.

To verify the performance of the loss function of YOLOv8, the classification, target and GIoU loss function are intensively trained. The training results of the loss function of YOLOv8 algorithm are listed in Figure 12. In Figure 12, the abscissa of the three graphs is the iterations, and the ordinate is the mean value of the classification, target and GIoU loss function. The three types of loss function show a straight downward trend at the initial stage of the iteration, and the downward trend gradually slows down with the increase of the iterations. In Figure 12 (a), the mean value of the classification loss function is at least 0.001, which indicates that YOLOv8 algorithm has high classification accuracy. In Figure 12 (b), the mean value of the target loss function is at least 0.01, which indicates that YOLOv8 algorithm has high target detection accuracy. In Figure 12 (c), the mean value of GIoU loss function is at least 0.015, indicating that the minimum circumscribed matrix of YOLOv8 algorithm fits well. Overall, the loss function of YOLOv8 presents an ideal state.

The centralized verification results of the classification loss function, target loss function and GIoU loss function of YOLOv8 are displayed in Figure 13. In Figure 13 (a), the classification loss function drops vertically to about 0.008 at the initial stage of iteration, and then the decline speed slows down. The mean value of the classification loss function fluctuates in the range of 0.005∼0.010 within the range of 5∼200 iterations, indicating that the classification accuracy of YOLOv8 algorithm is high. In Figure 13 (b), the target loss function drops vertically to around 0.009 at the initial stage of the iteration, and then rises significantly in the range of 5∼200 iterations. The mean value of the target loss function rises to about 0.021 at the highest, proving that the target detection accuracy of YOLOv8 is high. In Figure 13 (c), the GIoU loss function drops vertically to around 0.028 at the initial stage of the iteration, and then shows a turbulent and slow downward trend in the range of 5∼200 iterations. The mean value of GIoU loss function drops to about 0.016 at the lowest, which indicates that the fit verification result of the minimum circumscribed matrix is relatively ideal. Overall, the improved YOLOv8 algorithm exhibits high accuracy and stability in overall performance.



(a)Training results of classification loss function

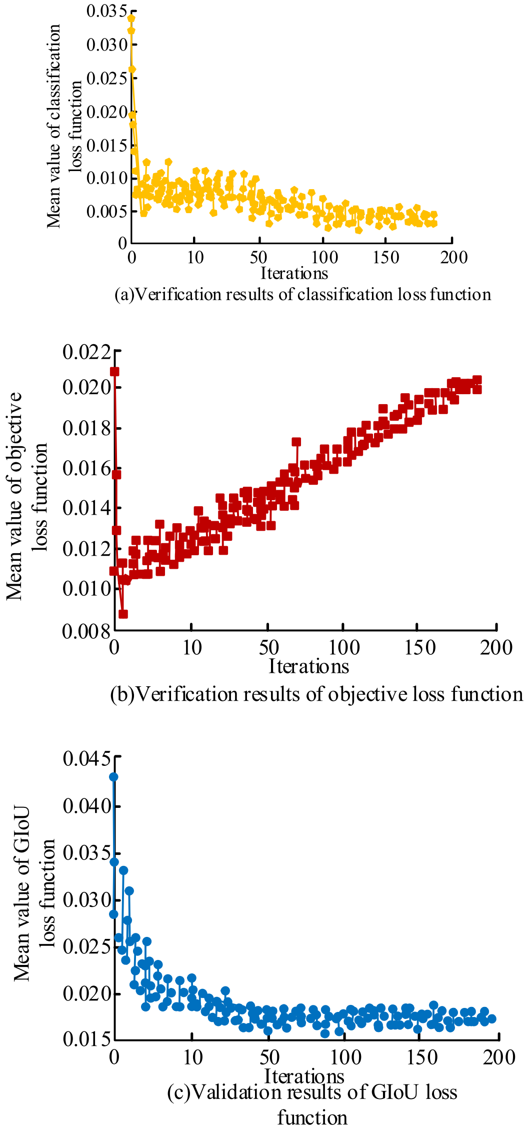
(b)Training results of objective loss function

(c)Training results of GIoU loss function

FIGURE 12. Training result graph of YOLOv8 algorithm loss function.

VI. PERFORMANCE VERIFICATION OF MULTI CAMERA POSITIONING SYSTEM BASED ON YOLO OBJECT DETECTION ALGORITHM IN PRACTICAL APPLICATIONS

To test the performance of the YOLO multi-camera positioning system in practical applications, experiments were conducted to verify the positioning system based on the



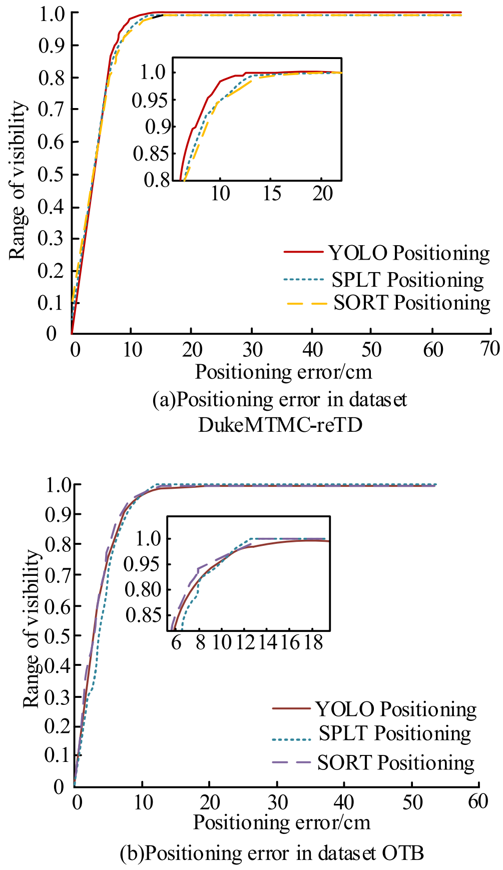
(a)Verification results of classification loss function

(b)Verification results of objective loss function

(c)Validation results of GIoU loss function

FIGURE 13. Verification result of YOLOv8 algorithm loss function.

YOLO algorithm, as well as the positioning system using the SPLT and SORT tracking algorithms. The accuracy of target object positioning and recognition under different sight distances was compared among these three positioning systems. The dataset used in the experiment was the publicly available DukeMTMC-reTD and OTB datasets. DukeMTMC-reTD is a cross camera dataset captured by 8 campus surveillance cameras. The OTB contains a large amount of video data and covers a comprehensive range of influencing factors. Testing different positioning systems on these two datasets, and the positioning errors of different positioning systems are shown in Figure 14. Figure 14 (a) is a comparison of the positioning errors of three positioning systems in the DukeMTMC-reTD dataset. The horizontal axis represents the positioning error and the vertical axis represents the line of sight between the target object and the camera. YOLO's positioning error convergence speed is slightly faster than SPLT and SORT, with the maximum error value exceeding 0.5cm, while YOLO's error probability within 10cm remains above 96.5%. From Figure 14 (b), under the OTB dataset, when YOLO positioning, SPLT positioning, and SORT positioning achieve an error of 95%, the corresponding values are maintained within 9.5cm, 9.6cm, and 8.5cm, respectively. In the performance verification of the multi camera positioning system of YOLO object detection algorithm in practical applications, the accuracy of target object positioning and recognition was compared among three positioning systems under different line of sight conditions. The error probability within 10 centimeters refers to the error probability of target object positioning and recognition at a line of sight of 10 centimeters. Compared to SPLT positioning and SORT positioning, the overall error of YOLO positioning remains within 10cm. This indicates that the accuracy of YOLO positioning has been improved to a certain extent and can satisfy the basic positioning needs of multiple cameras.



(a)Positioning error in dataset DukeMTMC-reTD

(b)Positioning error in dataset OTB

FIGURE 14. Positioning error of the positioning system.

To test the accuracy of the three positioning systems when the target object is occluded, the algorithm position error threshold is set to 10 to 50. The tracking performance of the three positioning systems is tested in real applications to comparatively verify the positioning accuracy of the three positioning systems under the same occluded environment. The positioning accuracy of the three positioning systems are compared and analyzed over a position error threshold range of 10 ∼ 50. As Figure 15, the horizontal axis represents the position error threshold, and the vertical axis represents the accuracy rate. The accuracy of YOLO positioning system is significantly higher than the other two positioning systems in the presence of occlusion. The highest accuracy of the YOLO positioning system is about 0.74. The SPLT positioning system is about 0.63, and the SORT positioning system is about 0.58. The accuracy of the YOLO positioning system has been improved by 0.11 and 0.16 respectively compared to the SPLT and SORT positioning systems. The above data indicates that in the presence of occlusion, the YOLO positioning system has better positioning accuracy performance.

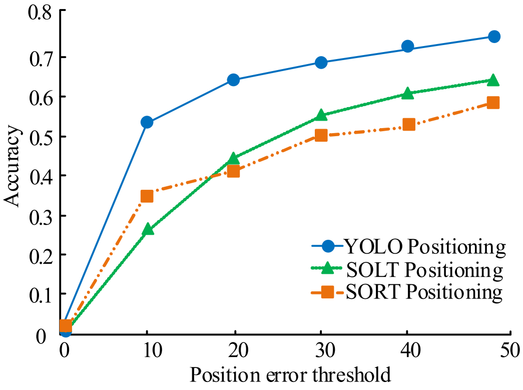


FIGURE 15. Tracking effect of positioning system.

To compare the performance differences between different algorithms more clearly, the performance of YOLOv8 algorithm and YOLO series algorithms in each series is summarized and compared. The performance comparison between YOLOv8 algorithm and YOLO series algorithm is shown in Table 2. As shown in Table 2, YOLOv8 algorithm and YOLO series algorithms have achieved better performance results compared in terms of mAP, runtime, inference speed, and accuracy. Among them, the mAP performance of YOLOv8 algorithm has improved by 30.4%, 21.1%, 14.9%, and 8.5% compared to algorithms such as YOLO, YOLO2, YOLO3, and YOLO4, respectively. Compared with algorithms such as YOLO, YOLO2, YOLO3, and YOLO4, the running time of YOLOv8 algorithm has increased by 40ms, 14ms, 68ms, and 56ms, respectively. The inference speed of YOLOv8 algorithm has been shortened by 48 ms compared to YOLO, and increased by 3.5 ms, 3.2 ms, and 1.9 ms compared to algorithms such as YOLO2, YOLO3, and YOLO4, respectively. The accuracy performance of YOLOv8 algorithm is slightly reduced by 0.96% compared to YOLO, but improved by 42.5%, 15.7%, and 6.4% compared to algorithms such as YOLO2, YOLO3, and YOLO4, respectively. Overall, the overall performance of YOLOv8 algorithm is superior to the YOLO series of algorithms.

To evaluate the superiority of YOLOv8 object recognition algorithm, this algorithm is compared with representative algorithms in the field of object detection such as IENet, TOSO, PIoU, and DRN [35]. IENet is a single stage anchor free rotating target detector. It adopts an Interactive Branch Network structure, aiming to solve the problem of object detection in remote sensing images [36]. OTSU method is an algorithm for determining the threshold of image binarization segmentation [37]. Point Intersection over Union (PIoU) is a viewpoint agnostic monocular detection algorithm used to solve object detection problems [38]. A dynamic relationship network (DRN) for dense multi angle object detection is a deep learning model used for object detection [39]. The performance comparison results of different algorithms are shown in Table 3. According to Table 3, the superior performance of YOLOv8 in terms of mAP and accuracy can be analyzed directly. Compared with algorithms such as IENet, TOSO, PIoU, and DRN, YOLOv8's mAP has improved by 17%, 17.7%, 16.3%, and 7.5%, and its accuracy has improved by 23.44%, 43.73%, 16.48%, and 33.33%. Compared with algorithms such as IENet, TOSO, PIoU, and DRN, the generalization ability of YOLOv8 algorithm has been improved by 13%, 16.9%, 17.9%, and 7.9%, respectively, indicating that YOLOv8 algorithm is more universal and has higher robustness in practical applications.

TABLE 2. Performance comparison table between YOLOv8 algorithm and YOLO series algorithm.

Algorithm	mAP/%	Run time/ms	Inference speed/ms	Accuracy
YOLO	66.8	108	54	52.1
YOLOv2	76.1	134	2.5	36.2
YOLOv3	82.3	80	2.8	44.6
YOLOv4	88.7	92	4.1	48.5
YOLOv8	97.2	148	6	51.6

These data fully demonstrate the excellent performance of YOLOv8 in object detection. Although YOLOv8 algorithm has the longest running time, considering its excellent object detection performance, overall, YOLOv8 algorithm has the most advantages.

TABLE 3. Performance comparison results of different algorithms.

Algorithm	mAP/%	Run time/ms	Accuracy	Generalization ability/%
DRN	89.7	114	38.7	84.6
PIoU	80.9	127	44.3	74.6
TOSO	79.5	77	35.9	75.6
IENet	80.2	92	41.8	79.5
YOLOv8	97.2	148	51.6	92.5

VII. DISCUSSION

To observe the advantages and disadvantages of various methods in related work more intuitively, the following detection methods were compared and analyzed as shown in

Table 4. Each method in Related Work has its advantages but is not perfect. Although near eye displays solve the challenges of visual quality and natural interaction in virtual reality, this method has a small field of view and limits the user experience. The Markov model increases the location confidentiality of users using roaming location services, but it can only predict the next state and cannot predict further states. Fuzzy detection strategy technology improves tracking speed and stability in complex environments, but cannot accurately locate vulnerabilities. The LiDAR ranging sensor system can detect low-light objects, but it cannot operate in extreme weather conditions such as snow and rain. Object detection based on deep learning improves object detection efficiency, but different regions do not share features, making the training process more cumbersome. The trilateral feature pyramid model based on YOLO algorithm improves the check all and accuracy of target detection system. However, it needs to change the image size several times and is time-consuming. The comprehensive performance of YOLO algorithm is more prominent. The detection accuracy study of localization recognition by combining single reactive transform shows excellent performance in the field of multi-camera localization recognition.

TABLE 4. Comparison of advantages and disadvantages of various detection methods.

Method	Advantage	Disadvantage
Near eye display	Solved the challenges of visual quality and natural interaction in virtual reality	The field of view is very small, which limits the user experience
Markov model	Increase location confidentiality for users using roaming location services	The model can only predict the next state and cannot predict further states
Fuzzy detection strategy technology	Improved tracking speed and stability in complex environments	Inability to accurately locate vulnerabilities
LiDAR ranging sensor system	Can detect low light objects	Cannot work in extreme weather such as snow and rain
Object detection based on deep learning	Improved object detection efficiency	Various regions do not share features, and the training process is quite cumbersome
A Trident Feature Pyramid Model Based on YOLO Algorithm	Improved recall and accuracy of object detection systems	Need to change the size of the image multiple times, which is time-consuming

VIII. CONCLUSION

To obtain a multi-camera positioning system with higher positioning accuracy, a multi-camera positioning technology using YOLO object detection algorithm is studied. In line with YOLO, the improved YOLOv8 algorithm is obtained by improving the network structure. Simultaneously the research uses homography transformation for target recognition in overlapping areas of photography. An excellent target following algorithm needs to meet both real-time and accuracy requirements. Compared to a single camera, multi camera localization and tracking also need to consider the problem of target matching in different surveillance videos. In practical situations, the application scenario is very complex, often with issues such as occlusion and large differences in lighting conditions. At the same time, the distance between the target and the camera is constantly changing, and the target is larger in the field of view when it is close to the camera. On the contrary, it is smaller, which will cause scale differences. Therefore, it is difficult to achieve precise cross camera positioning and recognition of multiple cameras. On the basis of previous studies, this paper explores and improves the target tracking method of multiple cameras. When occlusion and overlapping of the field of view occurs, the first step is to find the coordinates of the corresponding matching points to get the single responsiveness constraint relationship. Then it will get the correspondence of all points under the overlap of the field of view of two cameras to realize the target matching. At the same time, the improved matching method with color features is used to match the color features of the target object. Thus, the recognition efficiency and accuracy can be improved at the same time. The work of this article is summarized as follows:

1. Research results:

The performance test results showed that YOLOv8 has a maximum mAP of 97.2% on the test set, which is 8.5% higher than YOLOv4's 88.7%. The accuracy of YOLOv8 reached a maximum of 51.6 within the inference speed of 10ms, which was 3.5 times higher than the highest accuracy of 48.1 of YOLOv4. The recall rate of YOLOv8 showed a vertical upward trend within the range of 0-10 iterations, and fluctuated around 0.79 after 10 iterations. The accuracy of YOLOv8 increased vertically within the range of 0-10 iterations, and fluctuated around 0.81 after 10 iterations.The mean values of the classified loss function, target loss function and GIoU loss function of YOLOv8 reached 0.001, 0.01 and 0.015 respectively. The verification results of the positioning system showed that the error probability of YOLO within 10cm remained above 96.5% in the DukeMTMC re TD dataset. And the convergence speed of YOLO's positioning error was slightly faster than that of SPLT and SORT, with the maximum error value exceeding 0.5cm. When YOLO positioning, SPLT positioning, and SORT positioning achieved an error of 95% in the OTB dataset, the corresponding values were maintained within 9.5cm, 9.6cm, and 8.5cm, respectively. In the case of occlusion of the target object, the accuracy of the YOLO positioning system reached the highest of 0.74, which was 0.11 and 0.16 higher than the accuracy of the SPLT positioning system of 0.63 and the SORT positioning system of 0.58, respectively. Overall, the research on multi-camera positioning technology based on YOLO object detection algorithm has an improvement effect on the accuracy of multi camera positioning. However, due to the limited sample data, the experimental results are not comprehensive enough. Further improvement are needed in this aspect. The under-camera target recognition method is still flawed.

It is unable to achieve 100% target recognition and does not study and utilize the depth information of other objects in the surveillance area. In the future work, deep learning related algorithms can be added to obtain the depth information of the target, recognize other objects in the monitoring area. Recognizing the information characteristics of the target interacting with different objects can obtain the interrelationship between the target and the objects, thus completing the depth information tracking of the target.

Research on multi-camera localization technology still has many problems due to the complexity of multi-target recognition and localization tasks under multi-camera conditions, the complexity of the monitoring area, and the variability of moving targets.

2. Limitations:

Firstly, multi camera localization did not achieve 100% localization recognition because there was no research and utilization of depth information of other objects under the camera. Secondly, the relevant information of the target object is not fully utilized, such as the texture features, category features, and state information of the target object. Finally, without combining sensors to locate the target, such as Bluetooth for position determination, it will more accurately complete the positioning, recognition, and navigation of the target object.

3. Outlook:

In response to the above limitations, further research can be conducted in the future. The first problem can be solved by incorporating deep learning algorithms in future research to obtain the depth information of the target object and to recognize other objects in the monitored area by building a relevant database. Based on the interactive information features between different objects, the mutual relationship between the target and the object can be obtained, thereby completing the depth information tracking of the target object. The second problem can be increased in the future to study the type and attitude of the target object. Feature extraction and obtaining the attitude information of target objects from multiple perspectives will lead to better recognition and matching from the attitude of target objects. For the third issue, combining sensors for target localization and recognition can better combine visual features for target localization, thereby obtaining more accurate positioning information and accuracy.

REFERENCES

[1] A. Budiman, S. Sunariyo, and J. Jupriyadi, "Sistem informasi monitoring dan pemeliharaan penggunaan SCADA (supervisory control and data acquisition)," *Jurnal Tekno Kompak*, vol. 15, no. 2, p. 168, Aug. 2021.

[2] J. Duan, D. Shi, R. Diao, H. Li, Z. Wang, B. Zhang, D. Bian, and Z. Yi, "Deep-reinforcement-learning-based autonomous voltage control for power grid operations," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 814–817, Jan. 2020.

[3] D. Pliatsios, P. Sarigiannidis, T. Lagkas, and A. G. Sarigiannidis, "A survey on SCADA systems: Secure protocols, incidents, threats and tactics," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1942–1976, 3rd Quart., 2020.

[4] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, "Machine learning-based network vulnerability analysis of industrial Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6822–6834, Aug. 2019.

[5] E. Mbunge, B. Akinnuwesi, S. G. Fashoto, A. S. Metfula, and P. Mashwama, "A critical review of emerging technologies for tackling COVID-19 pandemic," *Hum. Behav. Emerg. Technol.*, vol. 3, no. 1, pp. 25–39, Dec. 2021.

[6] D. Sadykova, D. Pernebayeva, M. Bagheri, and A. James, "IN-YOLO: Real-time detection of outdoor high voltage insulators using UAV imaging," *IEEE Trans. Power Del.*, vol. 35, no. 3, pp. 1599–1601, Jun. 2020.

[7] G. Li, Z. Ji, X. Qu, R. Zhou, and D. Cao, "Cross-domain object detection for autonomous driving: A stepwise domain adaptative YOLO approach," *IEEE Trans. Intell. Vehicles*, vol. 7, no. 3, pp. 603–615, Sep. 2022.

[8] G. A. Koulieris, K. Akşit, M. Stengel, R. K. Mantiuk, K. Mania, and C. Richardt, "Near-eye display and tracking technologies for virtual and augmented reality," *Comput. Graph. Forum*, vol. 38, no. 2, pp. 493–519, Jun. 2019.

[9] X. Guo, N. Ansari, F. Hu, Y. Shao, N. R. Elikplim, and L. Li, "A survey on fusion-based indoor positioning," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 566–594, 1st Quart., 2020, doi: 10.1109/COMST.2019.2951036.

[10] A. K. Sangaiah, D. V. Medhane, T. Han, M. S. Hossain, and G. Muhammad, "Enforcing position-based confidentiality with machine learning paradigm through mobile edge computing in real-time industrial informatics," *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 4189–4196, Jul. 2019.

[11] S. Liu, S. Wang, X. Liu, C.-T. Lin, and Z. Lv, "Fuzzy detection aided real-time and robust visual tracking under complex environments," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 1, pp. 90–102, Jan. 2021.

[12] W. He, X. Mu, L. Zhang, and Y. Zou, "Modeling and trajectory tracking control for flapping-wing micro aerial vehicles," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 1, pp. 148–156, Jan. 2021.

[13] R. Edirisinghe, "Digital skin of the construction site: Smart sensor technologies towards the future smart construction site," *Eng. Constr. Archit. Manage.*, vol. 26, no. 2, pp. 184–223, Sep. 2019.

[14] G. Wang, H. Ding, Z. Yang, B. Li, Y. Wang, and L. Bao, "TRC-YOLO: A real-time detection method for lightweight targets based on mobile devices," *IET Comput. Vis.*, vol. 16, no. 2, pp. 126–142, Mar. 2022.

[15] N. M. A. A. Dazlee, S. A. Khalil, S. Abdul-Rahman, and S. Mutalib, "Object detection for autonomous vehicles with sensor-based technology using YOLO," *Int. J. Intell. Syst. Appl. Eng.*, vol. 10, no. 1, pp. 129–134, Mar. 2022.

[16] N. Saqib, "Positioning—A literature review," *PSU Res. Rev.*, vol. 5, no. 2, pp. 141–169, Sep. 2021, doi: 10.1108/prr-06-2019-0016.

[17] S. Zhang, Y. Wu, C. Men, and X. Li, "Tiny YOLO optimization oriented bus passenger object detection," *Chin. J. Electron.*, vol. 29, no. 1, pp. 132–138, Jan. 2020.

[18] G. Wang, H. Ding, B. Li, R. Nie, and Y. Zhao, "Trident-YOLO: Improving the precision and speed of mobile device object detection," *IET Image Process.*, vol. 16, no. 1, pp. 145–157, Jan. 2022, doi: 10.1049/ipr2.12340.

[19] Y. Du, N. Pan, Z. Xu, F. Deng, Y. Shen, and H. Kang, "Pavement distress detection and classification based on YOLO network," *Int. J. Pavement Eng.*, vol. 22, no. 13, pp. 1659–1672, Nov. 2021.

[20] Z. Guo, C. Wang, G. Yang, Z. Huang, and G. Li, "MSFT-YOLO: Improved YOLOv5 based on transformer for detecting defects of steel surface," *Sensors*, vol. 22, no. 9, p. 3467, May 2022, doi: 10.3390/s22093467.

[21] S. Oslund, C. Washington, A. So, T. Chen, and H. Ji, "Multiview robust adversarial stickers for arbitrary objects in the physical world," *J. Comput. Cognit. Eng.*, pp. 152–158, Sep. 2022, doi: 10.47852/bonviewjcce2202322.

[22] P. Nguyen, K. G. Quach, C. Nhan Duong, N. Le, X.-B. Nguyen, and K. Luu, "Multi-camera multiple 3D object tracking on the move for autonomous vehicles," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, New Orleans, LA, USA, Jun. 2022, pp. 2568–2577, doi: 10.1109/CVPRW56347.2022.00289.

[23] D. Ristić-Durrant, M. A. Haseeb, M. Banić, D. Stamenković, M. Simonović, and D. Nikolić, "SMART on-board multi-sensor obstacle detection system for improvement of rail transport safety," *Proc. Inst. Mech. Eng., F, J. Rail Rapid Transit*, vol. 236, no. 6, pp. 623–636, Jul. 2022, doi: 10.1177/09544097211032738.

[24] Q. Yu, Y. Yin, Y. Zhang, W. Chen, B. Hu, and X. Liu, "Displacement measurement of large structures using nonoverlapping field of view multi-camera systems under six degrees of freedom ego-motion," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 38, no. 11, pp. 1483–1503, Jul. 2023.

[25] Y. Jiang, L. Zhang, Z. Miao, X. Zhu, J. Gao, W. Hu, and Y.-G. Jiang, "PolarFormer: Multi-camera 3D object detection with polar transformer," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 1, 2023, pp. 1042–1050.

[26] L. Xu, Z. Su, L. Han, T. Yu, Y. Liu, and L. Fang, "UnstructuredFusion: Realtime 4D geometry and texture reconstruction using commercial RGBD cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 10, pp. 2508–2522, Oct. 2020.

[27] M. Maheepala, A. Z. Kouzani, and M. A. Joordens, "Light-based indoor positioning systems: A review," *IEEE Sensors J.*, vol. 20, no. 8, pp. 3971–3995, Apr. 2020.

[28] B. Cao, M. Li, X. Liu, J. Zhao, W. Cao, and Z. Lv, "Many-objective deployment optimization for a drone-assisted camera network," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 4, pp. 2756–2764, Oct. 2021.

[29] S. Liu, Q. Wang, and Y. Luo, "A review of applications of visual inspection technology based on image processing in the railway industry," *Transp. Saf. Environ.*, vol. 1, no. 3, pp. 185–204, Dec. 2019.

[30] X. Wu, S. Aravecchia, P. Lottes, C. Stachniss, and C. Pradalier, "Robotic weed control using automated weed and crop classification," *J. Field Robot.*, vol. 37, no. 2, pp. 322–340, Mar. 2020.

[31] H. Cui and S. Shen, "MMA: Multi-camera based global motion averaging," in *Proc. Conf. Artif. Intell.*, Feb. 2022, vol. 36, no. 1, pp. 490–498.

[32] V. Guizilini, I. Vasiljevic, R. Ambrus, G. Shakhnarovich, and A. Gaidon, "Full surround monodepth from multiple cameras," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 5397–5404, Apr. 2022.

[33] Y. Wu, H. Sheng, Y. Zhang, S. Wang, Z. Xiong, and W. Ke, "Hybrid motion model for multiple object tracking in mobile devices," *IEEE Internet Things J.*, vol. 10, no. 6, pp. 4735–4748, Mar. 2023.

[34] A. R. Sekkat, Y. Dupuis, V. R. Kumar, H. Rashed, S. Yogamani, P. Vasseur, and P. Honeine, "SynWoodScape: Synthetic surround-view fisheye camera dataset for autonomous driving," *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 8502–8509, Jul. 2022.

[35] G. Zhaoxin, L. Han, Z. Zhijiang, and P. Libo, "Design a robot system for tomato picking based on YOLO v5," *IFAC-PapersOnLine*, vol. 55, no. 3, pp. 166–171, 2022.

[36] J. Wang, Y. Chen, Z. Dong, and M. Gao, "Improved YOLOv5 network for real-time multi-scale traffic sign detection," *Neural Comput. Appl.*, vol. 35, no. 10, pp. 7853–7865, Apr. 2023.

[37] W. Jia, S. Xu, Z. Liang, Y. Zhao, H. Min, S. Li, and Y. Yu, "Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector," *IET Image Process.*, vol. 15, no. 14, pp. 3623–3637, Dec. 2021.

[38] P. Hurtik, V. Molek, J. Hula, M. Vajgl, P. Vlasanek, and T. Nejezchleba, "Poly-YOLO: Higher speed, more precise detection and instance segmentation for YOLOv3," *Neural Comput. Appl.*, vol. 34, no. 10, pp. 8275–8290, Feb. 2022.

[39] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *Multimedia Tools Appl.*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023.

WENNAN WU was born in Nanjing, Jiangsu, China, in 1982. He is currently pursuing the Ph.D. degree in satellite and vision navigation with the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing. He is with Jiangsu Satellite Navigation Testing Center Company Ltd., Nanjing.

JIZHOU LAI was born in 1977. He received the Ph.D. degree in navigation, guidance and control from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2005. Since 2013, he has been a Professor with the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics. From 2013 to 2014, he was a Visiting Scholar with the Institute of Aeronautics and Astronautics, University of Toronto, Toronto, ON, Canada. His research interests include multi-sensor fusion intelligent navigation, GNSS-independent autonomous navigation, multi-agent collaboration and fault-tolerant navigation, and intelligent perception and detection technology for unmanned systems.