

## Article

# Research on Identification and Location of Charging Ports of Multiple Electric Vehicles Based on SFLDLC-CBAM-YOLOV7-Tinp-CTMA

Pengkun Quan <sup>1</sup>, Ya'nan Lou <sup>2</sup>, Haoyu Lin <sup>1</sup>, Zhuo Liang <sup>1</sup>, Dongbo Wei <sup>1</sup> and Shichun Di <sup>1,\*</sup>

<sup>1</sup> School of Mechatronics Engineering, Harbin Institute of Technology, Harbin 150001, China; quanpengkun@hit.edu.cn (P.Q.); linhaoyu@hit.edu.cn (H.L.); liangzhuo@hit.edu.cn (Z.L.); weidb@hit.edu.cn (D.W.)

<sup>2</sup> Beijing Institute of Astronautical Systems Engineering, Beijing 100076, China; louyn\_01\_calt@126.com

\* Correspondence: dishichun@hit.edu.cn; Tel.: +86-1390-460-5946

**Abstract:** With the gradual maturity of autonomous driving and automatic parking technology, electric vehicle charging is moving towards automation. The charging port (CP) location is an important basis for realizing automatic charging. Existing CP identification algorithms are only suitable for a single vehicle model with poor universality. Therefore, this paper proposes a set of methods that can identify the CPs of various vehicle types. The recognition process is divided into a rough positioning stage (RPS) and a precise positioning stage (PPS). In this study, the data sets corresponding to four types of vehicle CPs under different environments are established. In the RPS, the characteristic information of the CP is obtained based on the combination of convolutional block attention module (CBAM) and YOLOV7-tinp, and its position information is calculated using the similar projection relationship. For the PPS, this paper proposes a data enhancement method based on similar feature location to determine the label category (SFLDLC). The CBAM-YOLOV7-tinp is used to identify the feature location information, and the cluster template matching algorithm (CTMA) is used to obtain the accurate feature location and tag type, and the EPnP algorithm is used to calculate the location and posture (LP) information. The results of the LP solution are used to provide the position coordinates of the CP relative to the robot base. Finally, the AUBO-i10 robot is used to complete the experimental test. The corresponding results show that the average positioning errors ( $x$ ,  $y$ ,  $z$ ,  $rx$ ,  $ry$ , and  $rz$ ) of the CP are 0.64 mm, 0.88 mm, 1.24 mm, 1.19 degrees, 1.00 degrees, and 0.57 degrees, respectively, and the integrated insertion success rate is 94.25%. Therefore, the algorithm proposed in this paper can efficiently and accurately identify and locate various types of CP and meet the actual plugging requirements.

**Keywords:** deep learning; multi-type vehicle charging ports; YOLOv7; location and posture (LP) estimation; automatic charging



**Citation:** Quan, P.; Lou, Y.; Lin, H.; Liang, Z.; Wei, D.; Di, S. Research on Identification and Location of Charging Ports of Multiple Electric Vehicles Based on SFLDLC-CBAM-YOLOV7-Tinp-CTMA. *Electronics* **2023**, *12*, 1855. <https://doi.org/10.3390/electronics12081855>

Academic Editor: Silvia Liberata Ullo

Received: 11 March 2023

Revised: 10 April 2023

Accepted: 11 April 2023

Published: 14 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the worldwide continuous reduction in the availability of fossil energy, the advantages of new energy vehicles have been gradually highlighted [1–3]. Electric vehicles rely on clean and pollution-free features to get strong support from the government [4–6]. In recent years, the shortage of urban land resources has become increasingly prominent, and the application of stereo charging garages has promoted the development of tram charging towards unmanned direction [7]. Automatic parking and driverless technology are gradually becoming mature. With this technology, a vehicle will arrive at the parking lot by itself and should be charged automatically. For publicly used electric vehicles or those on time-sharing lease, when the user returns a vehicle, the charging is often delayed that affects the user experience and utilization. The charging pile is damaged by weather and human factors, and manual charging will have significant safety risks. At the same

time, the DC charging gun line is heavy, which is not conducive to manual plugging [8]. Based on the aforementioned problems, automatic charging of electric vehicles is an urgent problem that needs to be solved.

At present, some companies and research institutions have proposed their own solutions [9–19]. These solutions show that the core of automatic charging of electric vehicles mainly consists of two parts: the identification and positioning of charging port (CP) and the plug-in mechanism. The identification and positioning of CP is the premise of plugging. Furthermore, the accuracy and universality of CP identification are important guarantees for the successful plugging of robots. Therefore, the high-precision identification and positioning of the CP is of great significance towards realization of automatic charging technology.

At present, the main CP recognition method uses visual positioning, which is divided into two categories: (1) with feature recognition and (2) without feature recognition. In terms of feature recognition, Lv [20] added white labels around the CP, used feature matching for rough positioning, and inserted the CP according to the six-axis force sensor compensation. The author did not provide recognition and positioning accuracy. Pan et al. [21] added five black and white labels around the CP. Based on the contour of the open operation, the geometric solution method was used to calculate the location and posture (LP) of the CP. The LP errors were 1.4 mm and 1.6 degrees, respectively, and the insertion success rate was 98.9%. The CP recognition methods without feature recognition include Li et al. [22] that proposed a CP identification and location method based on the Scale-invariant feature transform and semi-global block matching. The method achieved an average error of 1.51 mm. Zhang et al. [16] improved the canny edge detection and the CP image correlation algorithm of combined morphology. The authors did not specify any recognition accuracy, and the overall insertion success rate was 95.55%. Yao et al. [23], based on the template matching algorithm in Halcon commercial vision software, tested the CP LP error in a room, achieving average errors of 2.5 mm in position and 0.8 degrees in angle. Quan et al. [24] tested the CP identification accuracy in multiple environments using the cluster template matching algorithm (CTMA). The LP errors were 0.91 mm and 0.87 degrees, respectively, and the plugging success rate was 95%.

In recent years, deep learning has achieved rapid development in the field of target recognition, with the emergence of a series of target detection models such as Faster-RCNN, YOLO and SSD [25–32]. These models have improved the universality of target recognition, especially for specific targets, significantly improving the recognition accuracy in complex scenes and light. The YOLO algorithm is highly favored for its relatively high accuracy while ensuring high speed [33,34].

Based on the above research, CP recognition can only adapt to a single type of CP. Although the size of the CP has a unified standard, the CPs from different manufacturers and even different batches of CP from the same manufacturer will result in the inconsistency of the detailed texture and surface roughness of the CP. Due to the limitations of traditional algorithms, different CPs require adjustment of different characteristic parameters, and the universality is poor. At present, among the target detection algorithms, there is no recognition optimization algorithm with structural features. In view of the specificity of CP features, this paper proposes a data enhancement method based on the YOLOV7-tiny algorithm. The locations of similar features determine the label category (SFLDLC), which improved the target classification accuracy of similar feature location-determining categories. At the same time, using the convolutional block attention module (CBAM) attention mechanism combined with the CTMA, the universality and accuracy of the algorithm are improved, and a guarantee for the LP calculation algorithm of CP is provided. The rough positioning stage (RPS) and precise positioning stage (PPS) use the similar projection relationship and EPnP algorithm, respectively, to solve the LP of the recognition results. Subsequently, the robot is guided to complete the insertion work, which realizes automatic charging of various vehicle CPs. Our contributions in this paper are as follows:

- (1) We propose a solution that combines deep learning methods to identify charging port pose information.
- (2) We propose an SFLDLC and CTMA for CP recognition and positioning, which improves the accuracy of recognition.
- (3) We have integrated CBAM into YOLOV7-tinp for CP recognition and positioning, improving recognition accuracy.

This paper is organized as follows: Section 2 introduces the data collection process and the identification and location methods. Section 3 conducts experimental verification in different scenarios, providing positioning accuracy and insertion success rate. Section 4 discusses the sources of positioning errors. Section 5 summarizes the experimental results and further research directions.

## 2. Materials and Methods

### 2.1. Construction of Experimental Test Platform

The experimental test platform mainly includes three components: visual module, control module and plug-in actuator, as shown in Figure 1. To meet the requirements of the experimental insertion workspace, the actuator uses the AUBO-i10 articulated robot with six degrees of freedom. This paper uses the camera manufactured by Daheng Image Vision Co., Ltd. in Beijing, China, where the model of the camera is MER-125-30GM/C-P series industrial camera. The camera lens is the M0814-MP2 lens of Comstar. The light intensity is measured using the Taiwan Taishi TES-1335 digital illuminometer. The specific information is shown in Table 1. This paper adopts the camera calibration method of Zhang [35] and the hand-eye calibration method of Zhu et al. [36].

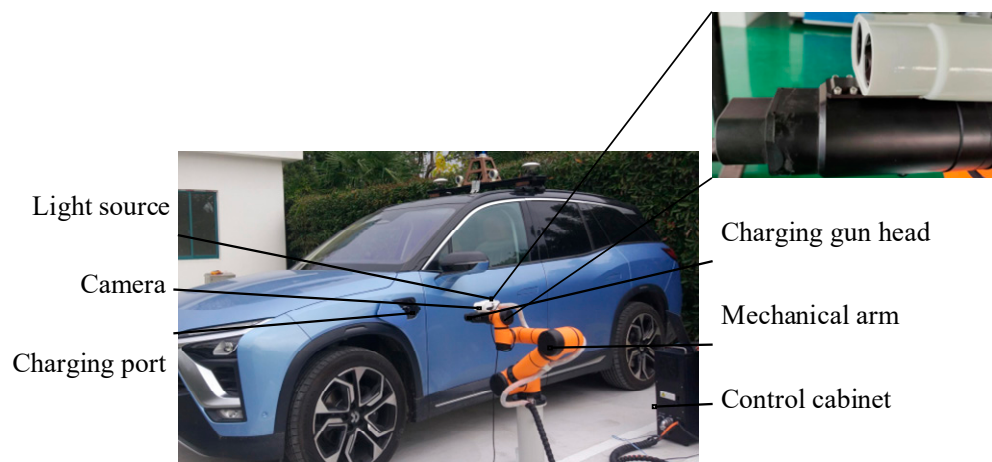


Figure 1. Automatic charging system experimental platform.

Table 1. Detailed information of robots.

Composition of Robots	Parameter	Detailed Information
Robotic arm	Maximum execution force	10 kg
	Arm span length	1563.2 mm
	Repeated positioning accuracy	0.05 mm
Camera	Pixel size	$3.75 \mu\text{m} \times 3.75 \mu\text{m}$
	Camera resolution	$1292 \times 964$
	Lens size	8 mm

### 2.2. Image Data Acquisition

This study is aimed at the DC CP with the national standard number of GBT 20234.3-2011. During data collection, the robot is fixed on the base in order to obtain the actual LP information of the CP relative to the camera. The base world coordinates are kept

unchanged, and the robot is moved into the CP in the state of teaching, which is the zero LP state. The robot is then moved out of the CP randomly within the recognition range, and the LP information of the camera relative to the CP is obtained based on the robot’s LP information and zero LP information.

Four types of CPs are used in this paper. In order to reduce the interference of the environment, i.e., the collected image is too dark or too bright, this paper designs an automatic exposure algorithm to adjust the average brightness value of the image between 100–160. The data in this paper were collected in the Songjiang District, Shanghai (120.5924 E, 31.3036 N). This article considers indoor, outdoor, morning, afternoon, noon, evening, sunny, and cloudy environments. There are 12 scenarios in total. Because of the similarity of the scenarios, six scenarios are finally determined, as shown in Table 1. In order to improve the actual positioning accuracy, this paper divides positioning into two stages: RPS and PPS, as shown in Figure 2.

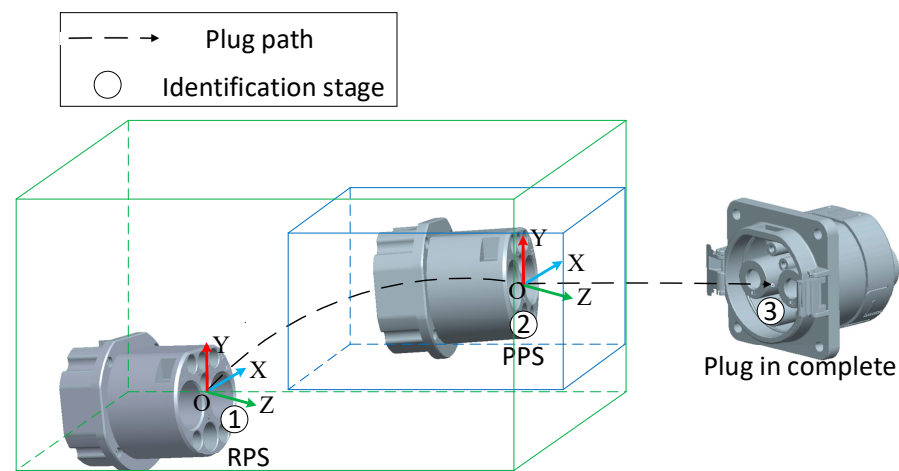


Figure 2. Schematic diagram of identification stage scope.

The purpose of RPS is to find the CP target and achieve rough positioning. According to the actual application scenario, the ranges of the x, y, and z directions are [−150, 150] mm, [−100, 100] mm, and [245, 600] mm, respectively, and the ranges of the angles in the Rx, Ry, and Rz directions are [−15, 15] degrees, [−15, 15] degrees, and [−10, 10] degrees, respectively. The data information is shown in Table 2. The PPS is the secondary positioning near the focal length and is aimed at achieving accurate positioning. Based on the actual application scenario, the ranges of the x, y, and z are [−6, 6] mm, [−6, 6] mm, and [250, 275] mm, respectively, and the ranges of the angles in the Rx, Ry, and Rz directions are [−15, 15] degrees, [−15, 15] degrees, and [−10, 10] degrees, respectively. The data are shown in Table 3.

Table 2. Collection of CP data in RPS.

Scenes	Weather	Time	Minimum Light Intensity (Klux)	Maximum Light Intensity (Klux)	Quantity
Indoor	All	All	3.3	5.6	800
Outdoor	Sunny	A.M./P.M.	7.5	44.9	800
	Sunny	Noon	11.6	54.5	800
	Overcast	A.M./P.M.	6.1	14.2	800
	Overcast	Noon	5.1	21.6	800
All	All	Night	0.6	3.1	800

**Table 3.** Collection of CP data in PPS.

Scenes	Weather	Time	Minimum Light Intensity (Klux)	Maximum Light Intensity (Klux)	Quantity
Indoor	All	All	4.6	5.7	800
Outdoor	Sunny	A.M./P.M.	8.6	45.2	800
	Sunny	Noon	13.4	53.6	800
	Overcast	A.M./P.M.	7.1	18.0	800
	Overcast	Noon	7.8	22.4	800
All	All	Night	3.1	3.6	800

### 2.3. Feature Recognition Method

#### 2.3.1. Feature Selection

The feature recognition is divided into RPS and PPS in order to ensure accurate feature location. The RPS is mainly used to identify and locate the CP target with a long distance and a wide range. The main problem is that when the target is far away, the image is blurred in the non-focal position, the characteristics of the CP vary greatly, and the proportion of the target in the field of vision is small. To deal with the aforementioned problems, we choose a larger target as the feature of the CP. Although the outermost feature of the CP is the most obvious, the outermost dimension of the CP from each manufacturer is different and there is no standard size. Therefore, we choose a relatively large feature and regard the round feature of the CP as a whole. The PPS is mainly aimed at the target near the focal length, which requires a high recognition accuracy. Therefore, we consider the individual circular feature contour as the target feature for recognition, and the feature range of RPS is as follows:

$$\begin{cases} X_{min} = \min(x_1 - w_1, x_5 - w_5, x_7 - w_7) \\ Y_{min} = \min(x_1 - h_1, x_2 - h_2, x_3 - h_3) \\ X_{max} = \max(x_3 - w_3, x_6 - w_6, x_9 - w_9) \\ Y_{max} = \max(x_7 - w_7, x_8 - w_8, x_9 - w_9) \end{cases} \quad (1)$$

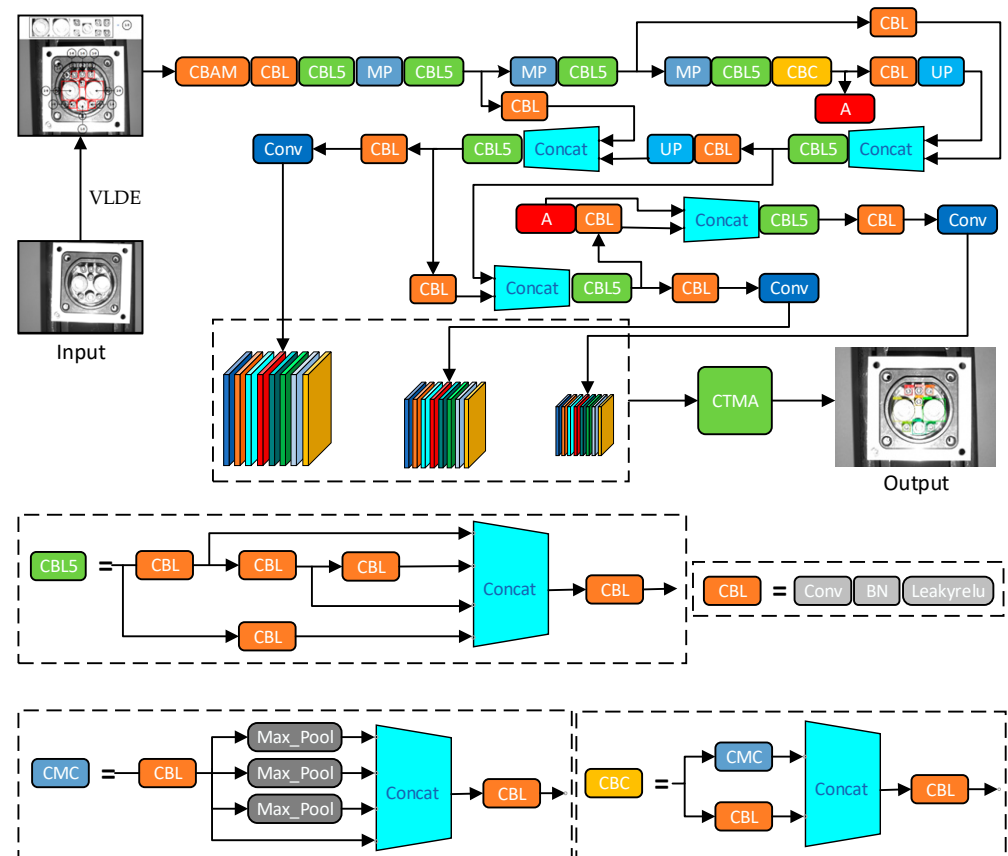
where  $x_n, y_n$  represent the center point coordinates of the  $n$ th feature;  $w_n, h_n$  represent the length and width of the  $n$ th feature, and  $(X_{min}, Y_{min})$  and  $(X_{max}, Y_{max})$  represent the pixel coordinate positions of the upper left and lower right corners of the feature box, respectively. The center position of the CP is calculated according to the obtained characteristic information of the CP as

$$\begin{cases} X_m = x_m \\ Y_m = (y_m - a_m h) \end{cases} \quad (2)$$

where  $(x_m, y_m)$  and  $(X_m, Y_m)$  represent the recognized feature center point coordinates and the actual feature center point coordinates, respectively, and  $a_m$  represents the conversion coefficient between pixel features and physical features. Using the calculated  $(X_m, Y_m)$ , the coordinate value of the CP in the physical coordinate system can be obtained according to the similar projection relationship.

#### 2.3.2. Identification Algorithm Model

In the YOLOv7 network, the number of times the model performs feature extraction will increase as the depth of the model increases, which will lead to a high computational complexity. Based on the characteristics of image data set at different recognition stages of the CP, and considering the requirements of image resolution, GPU memory and detection accuracy, the image data sets of the two stages are input into the neural network for training in this paper. The input image resolution is set to  $960 \times 960$  in order to reduce the impact of image compression on the accuracy. Three different sizes of detector heads are used to output the results, including the location information, category information, and confidence of the CP features. Figure 3 shows the network structure.



**Figure 3.** SFLDLC-CBAM-YOLOV7-tinp-CTMA network structure. CONV is the convolution operation, LeakyRelu is the nonlinear activation function, MP is the maximum pooling operation, UP is the upper sampling operation, Concat is the feature fusion function, and Maxpool is the maximum pooling operation. Note: The backbone of YOLOV7-tinp network is divided into four parts: Input, Backbone, Neck, and Prediction.

The feature type and location of the CP have a coupling relationship based on the CBAM-YOLOV7-tinp-CTMA network structure. This paper proposes a data enhancement method for the input image. It is based on the SFLDLC, which enhances the data generalization ability and improves the detection accuracy. The CBAM attention mechanism is fused to identify the feature location information in YOLOV7-tinp. The above methods are combined with CTMA; the accurate feature location and label type are obtained, which provide a guarantee for the LP calculation of the CP relative to the camera.

### 2.3.3. Data Enhancement Method

The RPS feature is unique and there is no mutual substitution relationship between the spatial positions. Therefore, the image is randomly scaled and clipped, and the Mosaic method is used to achieve data enhancement. The characteristic type and position of the CP have a coupling relationship during PPS. The determination of the label type is not based on the characteristics of the label type, but more importantly depends on the position relationship of the label. Therefore, when the data are enhanced, their location characteristics cannot be changed but can be zoomed, cropped, enhanced using the Mosaic data enhancement method, etc. The categories can be changed at the same time, but the original category characteristics of the feature do not change. In addition, this paper proposes a data enhancement method based on the SFLDLC, which enhances the ability of data generalization and improves the target recognition accuracy. In this paper, the round features of the CP from left to right, top to bottom, are defined as features 1 to 9, as shown in Figure 4.



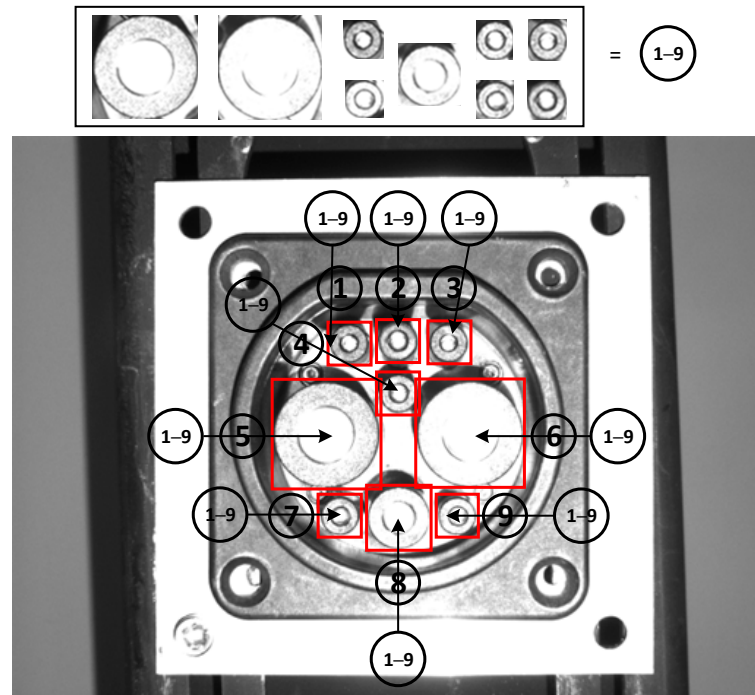


Figure 4. Feature data enhancement method.

During data enhancement, when each feature is enhanced by traditional data methods, the feature is first extracted from the image for use. When the feature is replaced by its position, the label category of the feature is related to its position and has no relation with the feature itself. Different features define the label category according to the replaced position. The constraint conditions of the data during enhancement are as follows:

$$\begin{cases} x_{n+1} > a_1(x_n + w_n + w_{n+1})(n = 1, 2, 7, 8) \\ \max(y_1, y_2, y_3) + \max(h_1, h_2, h_3) < a_2y_4 - h_4 \\ \sqrt{(y_n - y_{n+1})^2 + (x_n - x_{n+1})^2} > \frac{a_3(w_n + w_{n+1} + h_n + h_{n+1})}{2} \\ \max(y_5, y_6) > a_4y_4 \end{cases} \quad (3)$$

where  $x_n, y_n$  represents the center point coordinates of the  $n$ th feature, and  $w_n, h_n$  represents the length and width of the  $n$ th feature. The adjustment coefficients of the distance between the first and third floors, first and second floors, and second and third floors are denoted by  $a_1, a_2$ , and  $a_4$ , respectively, while  $a_3$  represents the degree of adhesion between all features.

### 2.3.4. Attention Mechanism

The CBAM is a convolution attention mechanism module that combines space and channels. As Figure 5 shows, given the intermediate feature graph  $F = R^{C \times H \times W}$  as the input, the CBAM module will judge the attention graph in turn along two independent channels. Subsequently, it will multiply the attention graph with the input feature graph to optimize the features, which not only reduces the size and computation of the feature graph, but also improves the expression ability of the network. In order to extract the effective contour features of the target and obtain the main content of the target detection, the channel attention module is introduced, which is calculated as follows:

$$MC(F) = \sigma\left(W_1\left(W_0\left(F_{avg}^c\right)\right) + W_1\left(W_0\left(F_{max}^c\right)\right)\right) \quad (4)$$

where  $\sigma$  represents the sigmoid function,  $W_0 \in R^{c/r \times c}$  and  $W_1 \in R^{c \times c/r}$ , where  $W_0$  and  $W_1$  represent two inputs shared weights. The ReLU activation function is followed by  $W_0$ ,

and  $F_{avg}^c$  and  $F_{max}^c$  represent the feature map generated in space by using average pooling and maximum pooling, respectively. The height is denoted by  $H$ ,  $W$  is the width,  $C$  is the number of channels, and  $r$  is the reduction rate.

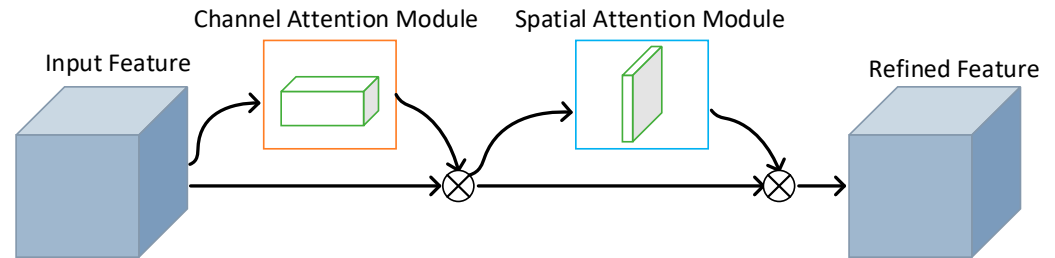


Figure 5. CBAM structure diagram.

In order to accurately locate the detected target and improve the target detection accuracy, the spatial attention module is introduced to focus on key features. It is calculated according to the following expression:

$$Ms(F) = \sigma \left( f^{7 \times 7} \left( \left[ F_{avg}^s; F_{max}^s \right] \right) \right) \tag{5}$$

where  $F_{avg}^s$  and  $F_{max}^s$  represent the characteristics of average pooling and maximum pooling of channels, respectively, and  $f^{7 \times 7}$  represents the convolution operation with a filter size of  $7 \times 7$ .

### 2.3.5. CTMA

The classification accuracy of model recognition in the output layer is improved in this paper by introducing the CTMA. It defines all contour pixel positions and peripheral rectangle contour information that meet the feature points as  $(x_{pn}, y_{pn}, w_{pn}, h_{pn})$ ,  $(n = 1, 2, 3 \dots n)$ , and thus establishes the contour matching function between features 1, 2 and 8. The specific optimization method is as follows:

$$\left\{ \begin{array}{l} D_{nm} = \left[ \sqrt{(y_n - y_m)^2 + (x_n - x_m)^2} - (r_{om} + r_{on}) \right] \cdot \frac{c_{bn}w_{bn} + c_{bn}h_{bn}}{2r_{on}} \\ \sqrt{(x_{bn} - x_{bm})^2 + (y_{bn} - y_{bm})^2} = c_{length_{nm}} [(c_{bn}w_{bn} + c_{bn}h_{bn} + c_{bm}w_{bm} + c_{bm}h_{bm}) / 4 + D_{nm}] \\ \sqrt{(x_{bn} - x_{bj})^2 + (y_{bn} - y_{bj})^2} = c_{length_{nj}} (c_{bn}w_{bn} + c_{bn}h_{bn} + c_{bj}w_{bj} + c_{bj}h_{bj}) / 2 \\ \log_a \left( 1 - \left| 1 - c_{length_{nm}} \right| \right) + \log_a \left( 1 - \left| 1 - c_{length_{nj}} \right| \right) = R(n) \end{array} \right. \tag{6}$$

where  $D_{nm}$  represents the nearest distance between the outer surface of features  $n$  and  $m$ ;  $c_{length_{nm}}$  represents the deviation coefficient of features  $m$  and  $n$ ;  $c_{length_{nj}}$  represents the deviation coefficient of features  $j$  and  $n$ ;  $a$  represents the adjustment coefficient, and  $R$  represents the contour matching degree.

According to all the detected contour information, use Equation (6) to match and locate the contour points, and use the located contour information to calculate the labels of all features based on Equation (7). The specific optimization method is as follows:

$$\left\{ \begin{array}{l} x_{bn} = \sin(\theta_{bn}) \cdot \sqrt{x_n^2 + y_n^2} \\ y_{bn} = \cos(\theta_{bn}) \cdot \sqrt{x_n^2 + y_n^2} \\ \theta_{bn} = \operatorname{arccot} \left( \frac{y_{on}}{x_{on}} \right) - \operatorname{arccot} \left( \frac{2\sqrt{(x_{b5} - x_{b6})^2 + (y_{b5} - y_{b6})^2}}{x_{b6} - x_{b5}} \right) \cdot s + \operatorname{arccot} \left( \frac{2(y_{o8} - y_{o5})}{x_{o6} - x_{o5}} \right) \cdot s \end{array} \right. \tag{7}$$

where  $s$  represents the direction of the feature point:  $s$  in features 1, 3, 4, 6, 7 and 8 is 1, and  $s$  in features 2, 5 and 9 is  $-1$ ;  $\theta_{bn}$  represents deflection angle.



Based on the above matching conditions, the label of the feature is reassigned to ensure the accuracy of the feature label and reduce the situation where the LP cannot be solved due to the abnormal classification label.

### 2.3.6. Loss Function

The loss function used in YOLOV7-tiny is CIoU-Loss. It is calculated as follows:

$$CIoU = 1 - IOU + \frac{\rho^2(A, B)}{c^2} + \alpha v \quad (8)$$

where  $IOU$  represents the overlapping area of the prediction box;  $A$  represents the prediction box;  $B$  represents the real box;  $\alpha$  is the weight function;  $v$  is the consistency of the aspect ratio;  $\rho(A, B)$  is the Euclidean distance between the center point coordinates of the  $A$  box and the  $B$  box, and  $c$  is the diagonal distance of the smallest box wrapping box  $A$  and box  $B$ .

### 2.3.7. Model Evaluation

The main evaluation indicators selected to verify the effectiveness of the proposed model are precision ( $P$ ), recall ( $R$ ), and mean average precision ( $mAP$ ). The formulas for calculating these indicators are as follows:

$$\begin{aligned} P &= \frac{T_p}{T_p + F_p} \times 100\% \\ R &= \frac{T_p}{T_p + F_N} \times 100\% \\ mAP &= \frac{\sum_{i=1}^C AP_i}{C} \end{aligned} \quad (9)$$

where  $T_p$  represents the actual positive case and is judged as a positive case by the classifier;  $F_p$  represents the actual negative case is judged as a positive case by the classifier;  $F_N$  represents the actual positive case but is judged as a negative case by the classifier; and  $C$  represents the number of detection categories. This study only needs to identify the circular features of the CP; therefore,  $C \leq 9$ . The average value of  $AP$  is represented by  $mAP$ , which can measure the overall performance of the target detection algorithm.

## 2.4. Calculation of CP LP

### 2.4.1. Location Solution in RPS

Using the calculated  $(X_m, Y_m)$ , the coordinate value of the CP in the physical coordinate system can be deduced according to the similar projection relationship. It is calculated as follows:

$$\begin{cases} X = \frac{L_w}{L_{iw}} \cdot x \\ Y = \frac{L_h}{L_{ih}} \cdot y \\ D_z = \frac{\sqrt{X^2 + Y^2}}{\sqrt{(s_c \cdot (u - s_w))^2 + (s_c \cdot (v - s_h))^2}} \cdot f \end{cases} \quad (10)$$

where  $(X, Y, D_z)$  is the actual spatial position of the target point relative to the camera;  $(x, y)$  is the position of the target point in the pixel in the image;  $L_w$  and  $L_h$  represent the length and width of the target circular feature, respectively;  $L_{iw}$  and  $L_{ih}$  represent the pixel size of the CP feature in the length and width directions, respectively;  $s_c$  represents the pixel size of the camera, and  $s_w$  and  $s_h$  represent the length and width pixel sizes of the image, respectively.

The location information of the CP can be calculated using (10), which guides the robot end to move to the focal length of the camera and provide guarantee for the PPS of the CP.

### 2.4.2. PPS LP Solution

The pixel position information of effective feature points can be obtained using the above algorithm. Combined with the three-dimensional spatial position of the CP, it can be converted into a PNP problem [37]. Therefore, we use the pixel coordinate  $(x_{apn}, y_{apn})$  and its corresponding spatial position coordinate  $(x_{on}, y_{on}, z_{on})$ . Subsequently, the position information  $(x_{pos}, y_{pos}, z_{pos})$  and attitude information  $(x_{ang}, y_{ang}, z_{ang})$  of the CP coordinate origin relative to the camera center point can be obtained. For the solution to the PNP problem, different methods need different number of effective feature points.

In space, based on the vector set composed of three-dimensional coordinates of at least six feature points, the position of any coordinate point can be represented by setting the weight size as follows:

$$\begin{cases} p_i^w = \sum_{j=1}^n \alpha_{ij} c_j^w \\ \sum_{j=1}^n \alpha_{ij} = 1 \end{cases} \quad (11)$$

where  $p_i^w$  is the point with known three-dimensional coordinates in the world coordinate system;  $c_j^w$  is the  $j$ th control point of  $p_i^w$  in the world coordinate system, and  $\alpha_{ij}$  is the weight coefficient. Figure 6 shows the EPnP algorithm location process.

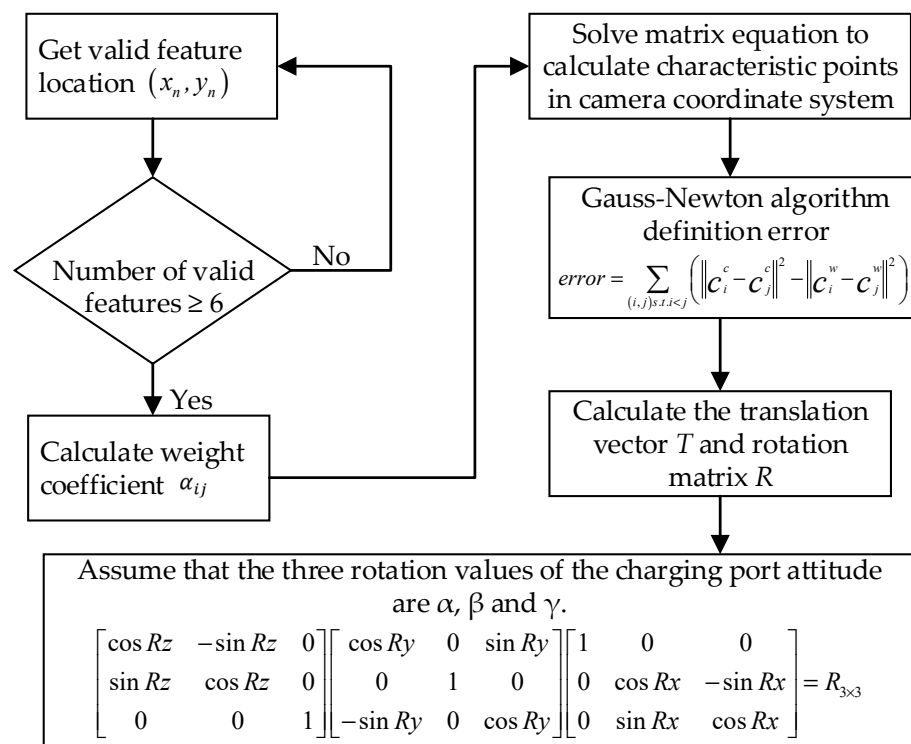


Figure 6. EPnP algorithm location process.

According to the above positioning process, the pixel coordinates of each feature based on the target recognition result are taken as input. In order to ensure the calculation accuracy, the weight coefficient  $\alpha_{ij}$  is calculated only when the number of feature points is not less than six. Subsequently, the feature points are calculated in the camera coordinate system. The error is defined by the Gauss–Newton algorithm, and the translation vector  $t$  and rotation vector  $R$  are calculated. Finally, the position and orientation information of the CP can be obtained.

### 3. Results

The test process is conducted under the Windows 10 operating system. A processor of model Intel (R) Core (TM) i7-10700K CPU @ 3.80 GHz, 3.79 GHz memory, and Nvidia GeForce RTX 3080 graphics card is used. The programming language used is Python 3.9 on the PyCharm programming platform, and PyTorch 1.6 is selected as the deep learning framework. The training is based on the GPU. During the performance test, the CPU is used for comparative testing in order to ensure that it is similar to the actual application scenario.

#### 3.1. Judgment Basis of CP LP Error

During data collection, this research fixed the robot on the base in order to obtain the actual position and orientation information of the CP relative to the camera. The world coordinates of the base were kept unchanged, and the robot was inserted into the CP while teaching. This state was considered as the zero LP. The robot was moved randomly out of the CP within the recognition range. Based on the LP information of the robot during data collection and combining it with the zero LP information, the LP information of the CP relative to the end of the manipulator was obtained. Subsequently, the actual LP information of the CP relative to the camera was calculated. The absolute difference between the actual LP information and the theoretical relative LP calculated in this paper was used as the basis for evaluating the accuracy of this algorithm.

#### 3.2. LP Accuracy Test in RPS

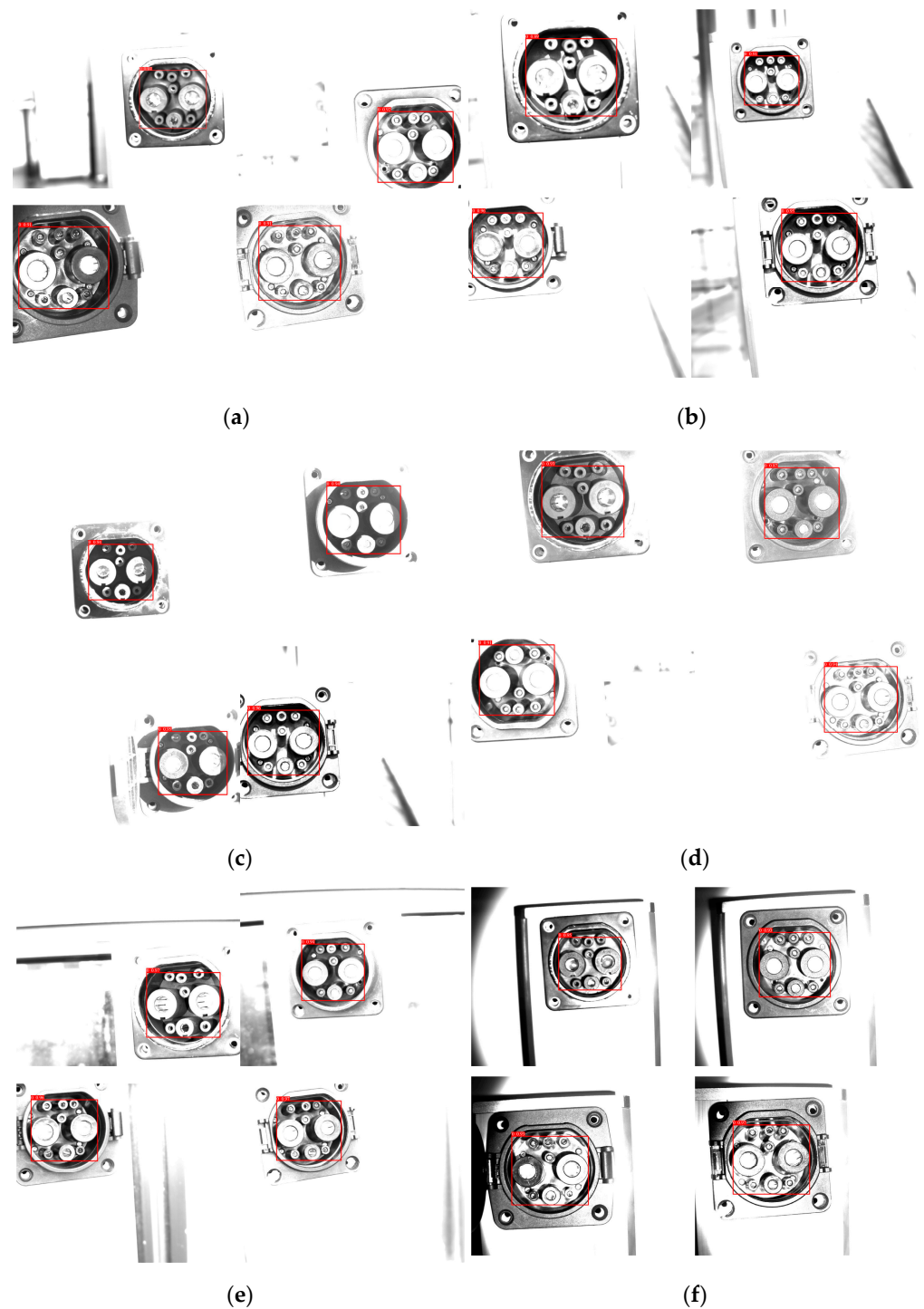
The RPS is mainly divided into feature recognition and LP resolution of the CP. Figure 7 shows the recognition performance of the feature points in different scenarios. The theoretical LP information is obtained based on the LP resolution algorithm proposed in this paper, and subsequently, the actual LP error information is obtained. A comparison of the different recognition methods in the RPS is provided in Table 4.

**Table 4.** Comparison of different recognition methods in RPS.

Stage	Class	Precision	Recall	mAP@ 0.5	mAP@ 0.5:0.95	Time-Consuming (s)
RPS	Faster RCNN	0.997	0.998	0.994	0.984	0.701
	YOLOV3	0.996	0.998	0.993	0.981	0.493
	YOLOV4	0.998	0.998	0.995	0.984	0.439
	YOLOV5s	0.998	0.999	0.995	0.986	0.296
	YOLOV7-tinp	0.998	0.999	0.996	0.990	0.271
	CBAM-YOLOV7-tinp	0.999	1	0.997	0.995	0.305

The experimental results in Table 4 show that the precision of CBAM-YOLOV7-tinp is 0.002 higher than that of Fast RCNN, 0.003 higher than that of yolov3, and 0.001 higher than those of YOLOV4, YOLOV5, and YOLOV7-tinp. The recall value of CBAM-YOLOV7-tinp is 0.02 higher than those of Faster RCNN, YOLOV3, and YOLOV4, and 0.001 higher than those of YOLOV5s and YOLOV7-tinp. In this paper, considering mAP @ 0.5:0.95 as an example, CBAM-YOLOV7-tinp has the highest accuracy, which is 0.005 higher than that without the CBAM. In the actual positioning, we try to improve the detection accuracy by reducing the false recognition in order to avoid damaging the manipulator. Therefore, CBAM-YOLOV7-tinp performs the best in terms of the detection accuracy. Although the detection time increases slightly due to the addition of the attention mechanism, this increased time is acceptable due to the improved accuracy weight in each index.

Based on the comparison of the above results, we use CBAM-YOLOV7-tinp to identify the position of the feature target, substitute the feature position information into the LP solution model, and obtain the LP information in different scenarios, as shown in Table 5.



**Figure 7.** Recognition performance in different scenarios in the RPS: (a) all indoor scenes; (b) outdoor sunny morning/afternoon; (c) outdoor sunny noon; (d) outdoor cloudy morning/afternoon; (e) outdoor cloudy noon; (f) all scenes at night.

The positioning results in Table 4 show that the indoor accuracy is basically the same as that at night, and the relative accuracy is relatively high. The average accuracy values of X, Y, and Z are 2.34 mm, 2.51 mm, and 2.64 mm, respectively. The accuracy in the sunny morning is basically the same as that on the cloudy day. The average accuracy values of X, Y, and Z are 2.72 mm, 2.92 mm, and 2.98 mm, respectively. The accuracy values of X, Y, and Z are 2.81 mm, 2.99 mm, and 3.17 mm, respectively, at noon on the sunny day. The average accuracy values of all cases are 2.61 mm, 2.79 mm, and 2.90 mm, which can meet the needs

of the RPS. The reason for the above accuracy difference is related to the shooting clarity and light difference of the image under different light field conditions.

**Table 5.** LP error in RPS.

Scenes	Weather	Time	X, mm	Y, mm	Z, mm	RX, deg	RY, deg	RZ, deg
Indoor	Sunny	A.M./P.M.	2.36	2.51	2.54	/	/	/
Outdoor	Sunny	A.M./P.M.	2.76	2.95	2.97	/	/	/
		Noon	2.81	2.99	3.17	/	/	/
	Overcast	A.M./P.M.	2.72	2.93	3.02	/	/	/
		Noon	2.68	2.88	2.96	/	/	/
/	/	Night	2.31	2.50	2.74	/	/	/

Note: “/” represents parameters that cannot be calculated or described.

### 3.3. LP Accuracy Test in PPS

The PPS is mainly divided into feature recognition and LP resolution of the CP. Figure 8 shows the effect of feature recognition in different scenarios. The theoretical LP information is obtained based on the LP resolution algorithm in this paper, and subsequently the actual LP error information is obtained. Table 6 shows the LP error of the CP in different scenarios.

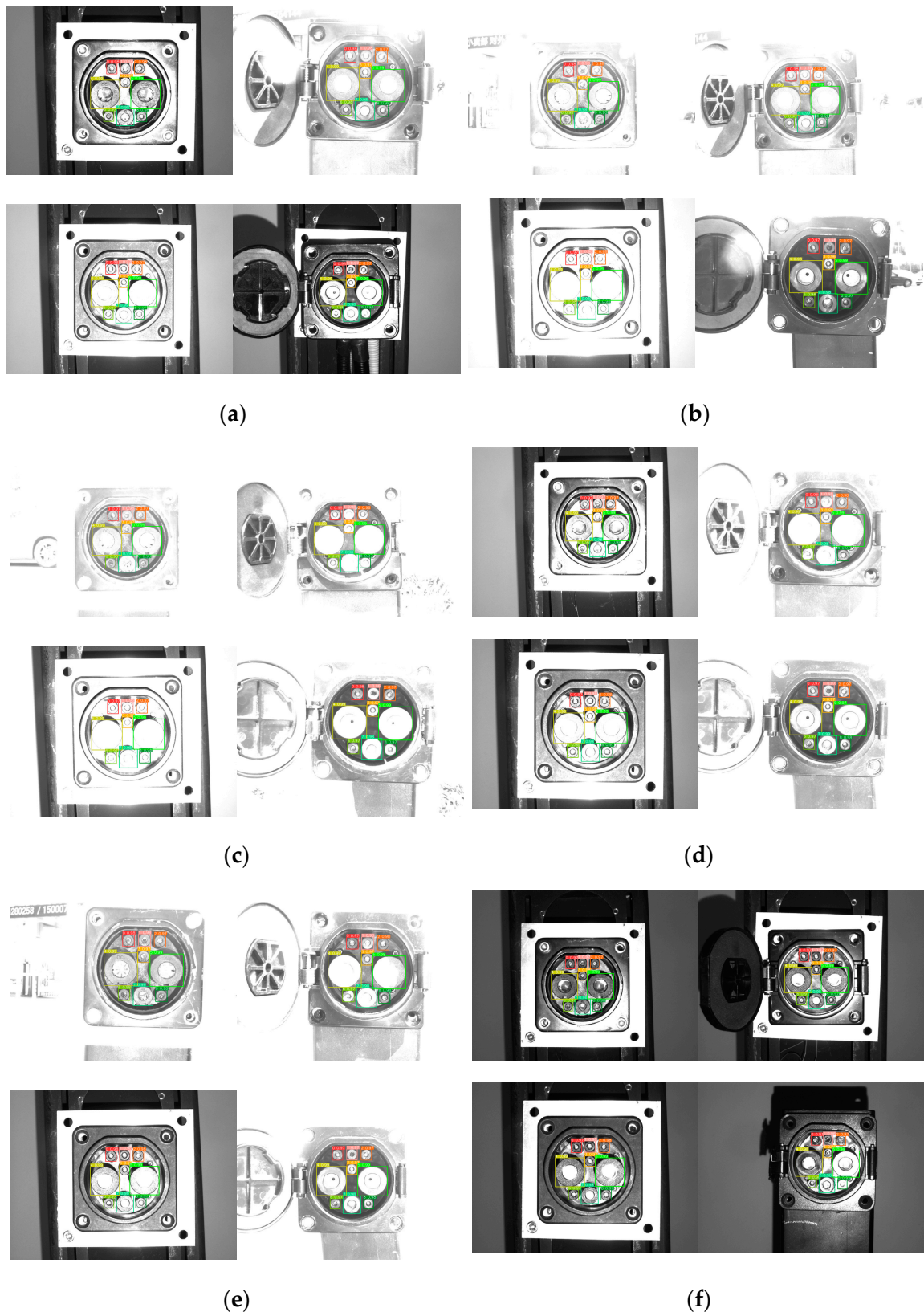
**Table 6.** Comparison of different recognition methods in PPS.

Stage	Class	Precision	Recall	mAP@ 0.5	mAP@ 0.5:0.95	Time-Consuming (s)
PPS	Faster RCNN	0.995	0.996	0.994	0.981	0.704
	YOLOV3	0.995	0.997	0.993	0.980	0.494
	YOLOV4	0.995	0.997	0.994	0.982	0.442
	YOLOV5s	0.997	0.997	0.994	0.983	0.304
	YOLOV7-tinp	0.998	0.998	0.996	0.984	0.272
	CBAM-YOLOV7-tinp	0.999	1	0.997	0.986	0.308
	SFLDLC-CBAM-YOLOV7-tinp-CTMA	1	1	0.998	0.989	0.314

It can be concluded based on the experimental results in Table 5 that out of Faster RCNN, YOLOV3, YOLOV4, YOLOV5s, and YOLOV7-tinp, YOLOV7-tinp outperforms the other models in terms of various indicators. It can further be noted that the results of the PPS directly affect the positioning results. In order to improve the accuracy and meet the insertion accuracy, YOLOV7-tinp is further improved. The precision of SFLDLC-CBAM-YOLOV7-tinp-CTMA algorithm proposed in this paper is 0.002 and 0.001 higher than YOLOV7-tinp and CBAM-YOLOV7-tinp, respectively. Its recall value is 0.002 higher than that of YOLOV7-tinp; mAP @ 0.5 value is 0.002 and 0.001 higher than those of YOLOV7-tinp and CBAM-YOLOV7-tinp, respectively, and the mAP @ 0.5:0.95 value is 0.005 and 0.003 higher than those of YOLOV7-tinp and CBAM-YOLOV7-tinp, respectively. In the actual positioning, damage to the manipulator can be avoided by improving the detection accuracy as much as possible by reducing misidentification. Therefore, SFLDLC-CBAM-YOLOV7-tinp-CTMA performs the best in terms of the detection accuracy. However, its detection time is slightly increased due to the addition of SFLDLC, CBAM, and CTMA. This increased time is acceptable due to the improved accuracy of each index.

Based on the comparison of the above results, we use SFLDLC-CBAM-YOLOV7-tinp-CTMA to identify the position of the feature target, substitute the feature position information into the LP solution model, and obtain the LP information in different scenarios. The corresponding errors are shown in Table 7.





**Figure 8.** Recognition performance in different scenarios in the PPS: (a) all indoor scenes; (b) outdoor sunny morning/afternoon; (c) outdoor sunny noon; (d) outdoor cloudy morning/afternoon; (e) outdoor cloudy noon; (f) all scenes at night.



**Table 7.** LP error in PPS.

Scenes	Weather	Time	X, mm	Y, mm	Z, mm	RX, deg	RY, deg	RZ, deg
Indoor	Sunny	A.M./P.M.	0.63	0.87	1.23	1.15	0.9	0.52
Outdoor	Sunny	A.M./P.M.	0.68	0.93	1.29	1.23	1.13	0.62
		Noon	0.71	0.96	1.32	1.25	1.14	0.65
	Overcast	A.M./P.M.	0.61	0.86	1.19	1.15	0.99	0.57
		Noon	0.59	0.83	1.21	1.17	0.94	0.54
/	/	Night	0.60	0.85	1.20	1.16	0.92	0.51

Note: "/" represents that scenes and weather are not fixed.

The positioning results in Table 6 show that the features of PPS and RPS have a common feature of circular edges. Therefore, the detection and positioning accuracy trends are similar. The positioning accuracy is basically the same in indoor sunny days, outdoor cloudy days, and at night, and the relative positioning accuracy is relatively high. The average accuracy values of x, y, z, Rx, Ry, and Rz are 0.61 mm, 0.85 mm, 1.21 mm, 1.16 degrees, 0.94 degrees, and 0.54 degrees, respectively. The positioning accuracy is low in outdoor sunny days, especially at noon. The average accuracy values of x, y, z, Rx, Ry, and Rz in outdoor sunny days are 0.70 mm, 0.95 mm, 1.30 mm, 1.24 degrees, 1.14 degrees, and 0.64 degrees, respectively. The positioning accuracy can meet the needs of PPS. The reason for the above accuracy difference is related to the shooting clarity and light difference of the image under different light field conditions.

### 3.4. Comparison of Results

In order to evaluate the progressiveness of the algorithm proposed in this paper, this paper compared it with three advanced electric vehicle CP identification and location methods. Table 8 shows the comparison results.

**Table 8.** Comparison of positioning results.

Identification Stage	Method	x	y	z	rx	ry	rz
RPS	our	2.61	2.79	2.90	/	/	/
	quan	/	/	/	/	/	/
	yinkai	/	/	/	/	/	/
	Li	/	/	/	/	/	/
PPS	our	0.64	0.88	1.24	1.19	1.00	0.57
	quan	/	/	/	/	/	/
	yinkai	/	/	/	/	/	/
	Li	/	/	/	/	/	/

Note: "/" represents a high error or inability to locate.

Table 7 shows that when the three advanced methods are used to identify multi-category CPs, the robustness of the algorithm is low, the error is high, and they are unable to identify and locate. Therefore, it is verified that the algorithm proposed in this paper exhibits robustness with respect to the identification of multiple types of CPs and has a significant application value.

### 3.5. Plug Test Verification

As the positioning accuracy in outdoor sunny days is low in the above tests, and the positioning accuracy of other scenes is basically the same, we define two cases as scene 1 and scene 2. We carried out 200 plug-in tests for each of these two situations. In these tests, the algorithm proposed in this paper is used for positioning, combined with the minimum mechanism of three iterations, and the 6-DOF articulated robot of AUBO-i10 is used for plugging. Table 9 shows the test results.

**Table 9.** CP connection test.

Positioning Stage	Scene	Total Experiment (Times)	Success Plug (Times)	Plug Rate (%)
RPS/PPS	scene 1	200	184	92.0
	scene 2	200	193	96.5

Based on the identification and location algorithm proposed in this paper, the average plugging rate of the CP is 96.5% in indoor (sunny/cloudy/night) conditions, and 92.0% in outdoor sunny (morning/noon/afternoon) conditions.

#### 4. Discussion

The LP errors are mainly caused by feature recognition positioning errors and system errors. Next, we discuss these two types of errors.

##### 4.1. Feature Recognition Positioning Error

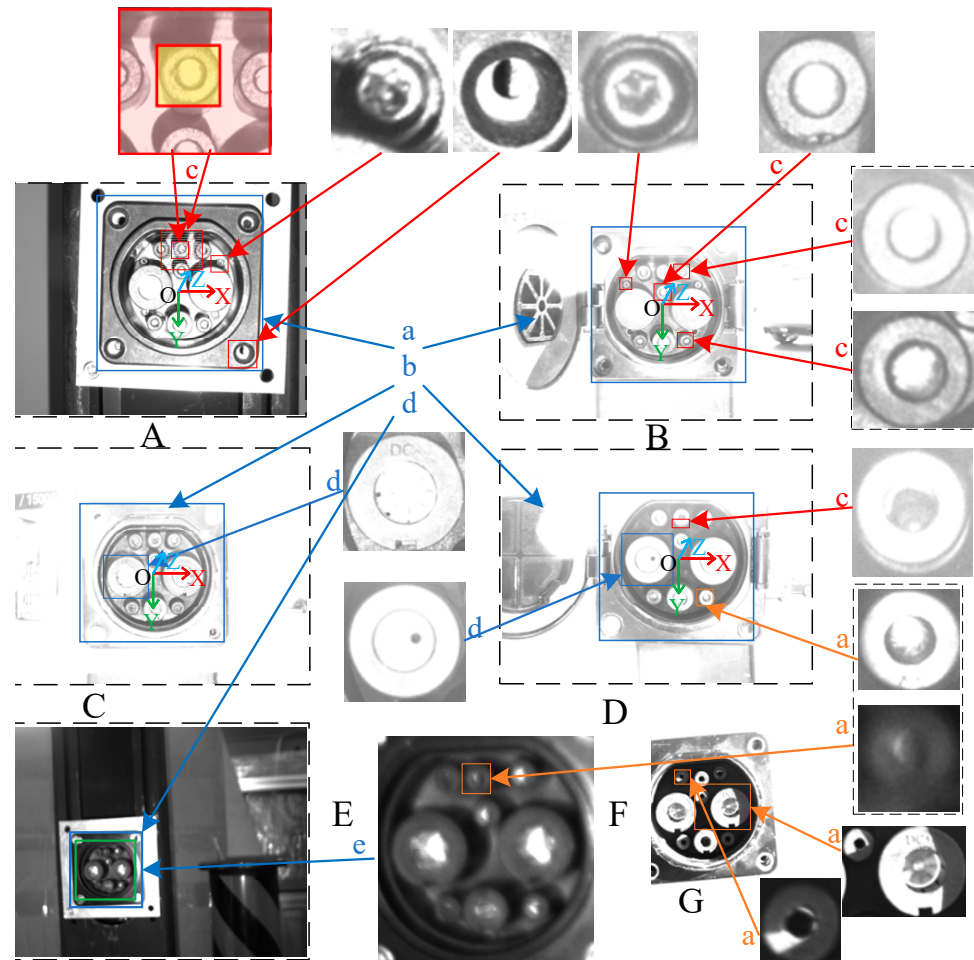
###### 4.1.1. Errors Caused by Complex Scenes and Different Characteristics of CPs

Although the size of CP has a unified standard, its material, smoothness, light angle, light brightness, and the specificity of different manufacturers will affect the recognition of CP characteristics. It can be divided into the following five situations, as shown in Figure 9:

- a. The difference of recognition of the same CP with respect to different times, locations, and scenes: The different time periods include morning and afternoon, noon, and evening. Different scenes mainly include indoor and outdoor. Different orientations mainly include the degree of the camera and the sun, and these differences will increase the recognition difficulty.
- b. The difference of the same CP under different plugging times: As the plugging times increase, there will be bumps on the surface of the CP. The consistency of the surface will be damaged, which degrades the recognition performance of the algorithm.
- c. It is difficult to identify the color and structural characteristics of the same CP. The round feature chamfer of the CP will cause the outline of the CP to deviate under different angles. The round feature and background of the CP are both black in color. The inside of the CP has a circle similar to the feature, which increases the difficulty of identifying the feature.
- d. Difference of different CPs: Chamfer degree of CP feature (d1), as well as different surface materials of CP (d2), the degree of reflection and smoothness of the surface of the CP (d3). The above factors will increase the difficulty of CP identification.
- e. The image at the non-focus position is blurred. To calculate the LP of the CP, the camera used is an industrial camera with an invariable focal length. Therefore, the image will be blurred in the non-focal position.

The aforementioned five conditions are the main reasons for the CP feature recognition difficulty. Therefore, during the recognition process, the occurrence of these conditions should be minimized in order to reduce the interference of complex environment on algorithm recognition. In Tables 5 and 6, the lighting environment of the images taken at night is relatively stable; therefore, it will exhibit a good performance. However, the positioning results are similar to those for indoor daytime, and no higher accuracy is obtained in the stable light field. The main reason is that at the time of data acquisition, the surfaces of some CPs are smooth, which will cause specular reflection. When only fill light is available, the effective surfaces of the CPs and the camera lens will be parallel or form a certain angle, and the effective feature area of the CPs is relatively large. Therefore, when specular reflection occurs at a few feature locations during data acquisition, a large amount of reflected light enters the camera aperture, resulting in overexposure of this part. Due to specular reflection of some features, a large amount of reflected light is reflected away, which will easily cause the loss of feature information. At noon in the outdoor environment, the shadow will appear inside the CP due to the angle between the sun and

the CP. Shadows can express additional interference features. The characteristics of the parts under direct and non-direct sunlight have large differences in the amount of light. This situation is the most complex, causing the largest recognition error.



**Figure 9.** Complex scene analysis of CP. (A) CP category 1; (B) CP category 2; (C) CP category 3; (D) CP category 4; (E) Blurred image of CP at non focal length; (F) Partially enlarged view of (E); (G) Image of CP under sunlight; a, b, c, d, and e represent five difficulty categories, respectively.

#### 4.1.2. LP Calculation Error

The accuracy of EPNP algorithm depends on the pixel coordinate position of feature points and the actual three-dimensional coordinate position. Furthermore, the number of effective features, camera calibration, and hand-eye calibration accuracy affect the solution results. The pixel feature position is directly related to the recognition accuracy of feature points. The actual three-dimensional coordinate position deformation is related to the expression of features on the image. The number of effective feature points is related to the number of features recognized by the algorithm. The calibration accuracy has a relatively small impact on the positioning accuracy.

#### 4.2. System Error

The system error is mainly caused by the robot positioning accuracy, including three aspects. First, the robot's repeated positioning accuracy. Second, the vibration of the base of the robot during the image acquisition and plugging process. Last, the positioning error caused by the gravity interference of the robot at different positions.

## 5. Conclusions

This paper proposed a set of electric vehicle CP identification and location algorithm based on CBAM-YOLOV7-tinp-CTMA, which realized the CP identification and location in multiple categories, multiple scenes, and a wide range. In this paper, the recognition process was divided into two stages, and the recognition and positioning model was established, respectively. The LP was calculated based on the similar projection relationship and EPnP algorithm, and the insertion test was completed by using the mechanical arm.

The two stages were tested in this paper, and the average positioning errors ( $x$ ,  $y$ ,  $z$ ) of RPS CP were 2.61 mm, 2.79 mm, and 2.90 mm, respectively. The average LP errors ( $x$ ,  $y$ ,  $z$ ,  $rx$ ,  $ry$ , and  $rz$ ) of the fine positioning CP were 0.64 mm, 0.88 mm, 1.24 mm, 1.19 degrees, 1.00 degrees, and 0.57 degrees, respectively. In different scenarios, the higher the positioning accuracy, the greater the plugging success rate. The plugging success rate in outdoor sunny days was 92.0%, and in other cases, it was equal to 96.5%. Compared with the existing advanced methods, the algorithm proposed in this paper had a high universality and could identify various types of CPs and complete positioning. It provided a theoretical basis for the positioning of various CPs and could have a high engineering application value.

In the future, more data on CP types and environmental complexity will be added. The improved algorithm will be optimized to improve its adaptability and recognition accuracy, increase the success rate of plugging, and reduce the impact of the plugging process on robots and vehicles. If there are problems with visual positioning, we can use vibration signals to compensate for visual positioning errors in the future, thereby avoiding accidents.

**Author Contributions:** Methodology, P.Q. and Y.L.; software, P.Q. and H.L.; validation, P.Q. and Y.L.; data curation, D.W. and Z.L.; writing—original draft preparation, P.Q.; writing—review and editing, S.D. and D.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to thank all of the authors cited and the anonymous referees in this article for their helpful suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

CP	Charging port
RPS	Rough positioning stage
PPS	Precise positioning stage
CBAM	Convolutional block attention module
CTMA	Cluster template matching algorithm
LP	Location and posture
SFLDLC	Similar feature location to determine label category

## References

1. Astiaso Garcia, D.; Dionysis, G.; Raskovic, P.; Duić, N.; Al-Nimr, M.A. Climate Change Mitigation by Means of Sustainable Development of Energy, Water and Environment Systems. *Energy Convers. Manag.* **2023**, *17*, 100335. [[CrossRef](#)]
2. Wang, Y.; Wen, Y.; Zhu, Q.; Luo, J.; Yang, Z.; Su, S.; Wang, X.; Hao, L.; Tan, J.; Yin, H.; et al. Real Driving Energy Consumption and CO<sub>2</sub> & Pollutant Emission Characteristics of a Parallel Plug-in Hybrid Electric Vehicle under Different Propulsion Modes. *Energy* **2022**, *244*, 123076. [[CrossRef](#)]
3. Ouramdane, O.; Elbouchikhi, E.; Amirat, Y.; Le Gall, F.; Sedgh Gooya, E. Home Energy Management Considering Renewable Resources, Energy Storage, and an Electric Vehicle as a Backup. *Energies* **2022**, *15*, 2830. [[CrossRef](#)]

4. Zhang, H.; Cai, G. Subsidy Strategy on New-Energy Vehicle Based on Incomplete Information: A Case in China. *Phys. A Stat. Mech. Its Appl.* **2020**, *541*, 123370. [[CrossRef](#)]
5. Wang, Y.; Fan, R.; Lin, J.; Chen, F.; Qian, R. The Effective Subsidy Policies for New Energy Vehicles Considering Both Supply and Demand Sides and Their Influence Mechanisms: An Analytical Perspective from the Network-Based Evolutionary Game. *J. Environ. Manag.* **2023**, *325*, 116483. [[CrossRef](#)]
6. Liao, D.; Tan, B. An Evolutionary Game Analysis of New Energy Vehicles Promotion Considering Carbon Tax in Post-Subsidy Era. *Energy* **2023**, *264*, 126156. [[CrossRef](#)]
7. Wang, K.; Li, G.; Chen, J.; Long, Y.; Chen, T.; Chen, L.; Xia, Q. The Adaptability and Challenges of Autonomous Vehicles to Pedestrians in Urban China. *Accid. Anal. Prev.* **2020**, *145*, 105692. [[CrossRef](#)]
8. He, C.; Cheng, Z.; Feng, J.; Yin, Q.; Li, X. Safety Analysis and Solution of Electric Vehicle Charging. *Distrib. Util* **2017**, *34*, 12–18. [[CrossRef](#)]
9. Guo, D.; Xie, L.; Yu, H.; Wang, Y.; Xiong, R. Electric Vehicle Automatic Charging System Based on Vision-Force Fusion. In Proceedings of the 2021 IEEE International Conference on Robotics and Biomimetics (ROBIO), Sanya, China, 27 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 405–410. [[CrossRef](#)]
10. Yuan, H.; Wu, Q.; Zhou, L. Concept Design and Load Capacity Analysis of a Novel Serial-Parallel Robot for the Automatic Charging of Electric Vehicles. *Electronics* **2020**, *9*, 956. [[CrossRef](#)]
11. Sachan, S. Automatic Learning-Based Charging of Electric Vehicles in Smart Parking. *Electr. Power Compon. Syst.* **2021**, *49*, 860–866. [[CrossRef](#)]
12. Hirz, M.; Walzel, B.; Brunner, H. Autonomous Charging of Electric Vehicles in Industrial Environment. *Teh. Glas.* **2021**, *15*, 220–225. [[CrossRef](#)]
13. Luo, W.; Shen, L. Design and Research of an Automatic Charging System for Electric Vehicles. In Proceedings of the 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA), Kristiansand, Norway, 9 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1832–1836. [[CrossRef](#)]
14. Miseikis, J.; Ruther, M.; Walzel, B.; Hirz, M.; Brunner, H. 3D Vision Guided Robotic Charging Station for Electric and Plug-in Hybrid Vehicles. *arXiv* **2017**, arXiv:1703.0538.
15. Lou, Y.; Di, S. Design of a Cable-Driven Auto-Charging Robot for Electric Vehicles. *IEEE Access* **2020**, *8*, 15640–15655. [[CrossRef](#)]
16. Zhang, J.; Geng, T.; Xu, J.; Li, Y.; Zhang, C. Electric Vehicle Charging Robot Charging Port Identification Method Based on Multi-Algorithm Fusion. In Proceedings of the Intelligent Robotics and Applications—14th International Conference (ICIRA 2021), Yantai, China, 22–25 October 2021; Springer: Cham, Switzerland, 2021; pp. 680–693. [[CrossRef](#)]
17. Walzel, B.; Sturm, C.; Fabian, J.; Hirz, M. Automated Robot-Based Charging System for Electric Vehicles. In *16. Internationales Stuttgarter Symposium*; Bargende, M., Reuss, H.-C., Wiedemann, J., Eds.; Springer: Wiesbaden, Germany, 2016; pp. 937–949. [[CrossRef](#)]
18. Patel, A.R.; Azadi, S.; Babae, M.H.; Mollaei, N.; Patel, K.L.; Mehta, D.R. Significance of Robotics in Manufacturing, Energy, Goods and Transport Sector in Internet of Things (IoT) Paradigm. In Proceedings of the 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 16–18 August 2018; pp. 1–4. [[CrossRef](#)]
19. Liu, X.-J.; Nie, Z.; Yu, J.; Xie, F.; Song, R. (Eds.) *Intelligent Robotics and Applications: 14th International Conference (ICIRA 2021), Yantai, China, 22–25 October 2021, Proceedings, Part III*; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2021; Volume 13015, ISBN 978-3-030-89133-6.
20. Lu, X. Research on Charging Method of Electric Vehicle Manipulator Based on Monocular Vision and Force Perception. Master's Thesis, Harbin Institute of Technology, Harbin, China, 2020. (In Chinese). [[CrossRef](#)]
21. Pan, M.; Sun, C.; Liu, J.; Wang, Y. Automatic Recognition and Location System for Electric Vehicle Charging Port in Complex Environment. *IET Image Process.* **2020**, *14*, 2263–2272. [[CrossRef](#)]
22. Li, T.; Xia, C.; Yu, M.; Tang, P.; Wei, W.; Zhang, D. Scale-Invariant Localization of Electric Vehicle Charging Port via Semi-Global Matching of Binocular Images. *Appl. Sci.* **2022**, *12*, 5247. [[CrossRef](#)]
23. Yao, A.; Xu, J. Electric Vehicle Charging Hole Recognition and Positioning System Based on Binocular Vision. *Sens. Microsyst.* **2021**, *40*, 81–84. [[CrossRef](#)]
24. Quan, P.; Lou, Y.; Lin, H.; Liang, Z.; Wei, D.; Di, S. Research on Fast Recognition and Localization of an Electric Vehicle Charging Port Based on a Cluster Template Matching Algorithm. *Sensors* **2022**, *22*, 3599. [[CrossRef](#)]
25. Yang, L.; Wang, Z.; Gao, S. Pipeline Magnetic Flux Leakage Image Detection Algorithm Based on Multiscale SSD Network. *IEEE Trans. Ind. Inf.* **2020**, *16*, 501–509. [[CrossRef](#)]
26. Yang, L.; Zhong, J.; Zhang, Y.; Bai, S.; Li, G.; Yang, Y.; Zhang, J. An Improving Faster-RCNN with Multi-Attention ResNet for Small Target Detection in Intelligent Autonomous Transport with 6G. *IEEE Trans. Intell. Transport. Syst.* **2022**, 1–9. [[CrossRef](#)]
27. Sun, X.; Gu, J.; Huang, R.; Zou, R.; Giron Palomares, B. Surface Defects Recognition of Wheel Hub Based on Improved Faster R-CNN. *Electronics* **2019**, *8*, 481. [[CrossRef](#)]
28. Liu, Y.; Sun, P.; Wergeles, N.; Shang, Y. A Survey and Performance Evaluation of Deep Learning Methods for Small Object Detection. *Expert Syst. Appl.* **2021**, *172*, 114602. [[CrossRef](#)]
29. Roszyk, K.; Nowicki, M.R.; Skrzypczyński, P. Adopting the YOLOv4 Architecture for Low-Latency Multispectral Pedestrian Detection in Autonomous Driving. *Sensors* **2022**, *22*, 1082. [[CrossRef](#)]

30. Chen, Z.; Li, X.; Wang, L.; Shi, Y.; Sun, Z.; Sun, W. An Object Detection and Localization Method Based on Improved YOLOv5 for the Teleoperated Robot. *Appl. Sci.* **2022**, *12*, 11441. [[CrossRef](#)]
31. Chen, H.; Guan, J. Teacher–Student Behavior Recognition in Classroom Teaching Based on Improved YOLO-v4 and Internet of Things Technology. *Electronics* **2022**, *11*, 3998. [[CrossRef](#)]
32. Liu, S.; Wang, Y.; Yu, Q.; Liu, H.; Peng, Z. CEAM-YOLOv7: Improved YOLOv7 Based on Channel Expansion and Attention Mechanism for Driver Distraction Behavior Detection. *IEEE Access* **2022**, *10*, 129116–129124. [[CrossRef](#)]
33. Yan, B.; Fan, P.; Wang, M.; Shi, S.; Lei, X.; Yang, F. Real-time Apple Picking Pattern Recognition for Picking Robot Based on Improved YOLOv5m. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 28–38. [[CrossRef](#)]
34. Xie, G.; Zheng, X.; Lin, Z.; Lin, L.; Wen, G. Bird’s Nest Detection of High Voltage Tower Based on Improved YOLOv4 Algorithm. *Electron. Meas. Technol.* **2022**, *45*, 145–152. [[CrossRef](#)]
35. Zhang, Z. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]
36. Hanqi Zhang Hand/Eye Calibration for Electronic Assembly Robots. *IEEE Trans. Robot. Autom.* **1998**, *14*, 612–616. [[CrossRef](#)]
37. He, Z.; Jiang, Z.; Zhao, X.; Zhang, S.; Wu, C. Sparse Template-Based 6-D Pose Estimation of Metal Parts Using a Monocular Camera. *IEEE Trans. Ind. Electron.* **2020**, *67*, 390–401. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.