

## Article

# Deep CNN-Based Materials Location and Recognition for Industrial Multi-Crane Visual Sorting System in 5G Network

Meixia Fu <sup>1</sup>, Qu Wang <sup>1</sup>, Jianquan Wang <sup>1,\*</sup>, Lei Sun <sup>1</sup>, Zhangchao Ma <sup>1</sup>, Chaoyi Zhang <sup>1</sup>, Wanqing Guan <sup>2</sup>, Qiang Liu <sup>3</sup>, Danshi Wang <sup>4</sup> and Wei Li <sup>2</sup>

- <sup>1</sup> School of Automation and Electrical Engineering, Institute of Industrial Internet, University of Science and Technology Beijing, Beijing 100083, China; meixiafu1205@ustb.edu.cn (M.F.); wangqu@ustb.edu.cn (Q.W.); sun\_lei@ustb.edu.cn (L.S.); mazhang-chao@ustb.edu.cn (Z.M.); zhangchaoyi@ustb.edu.cn (C.Z.)
- <sup>2</sup> School of Science and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China; wanqingguan@ustb.edu.cn (W.G.); li\_wei@ustb.edu.cn (W.L.)
- <sup>3</sup> College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China; qiangliu@bupt.edu.cn
- <sup>4</sup> State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China; danshi\_wang@bupt.edu.cn
- \* Correspondence: wangjianquan@ustb.edu.cn; Tel.: +86-186-01102992

**Abstract:** Intelligent manufacturing is a challenging and compelling topic in Industry 4.0. Many computer vision (CV)-based applications have attracted widespread interest from researchers and industries around the world. However, it is difficult to integrate visual recognition algorithms with industrial control systems. The low-level devices are controlled by traditional programmable logic controllers (PLCs) that cannot realize data communication due to different industrial control protocols. In this article, we develop a multi-crane visual sorting system with cloud PLCs in a 5G environment, in which deep convolutional neural network (CNN)-based character recognition and dynamic scheduling are designed for materials in intelligent manufacturing. First, an YOLOv5-based algorithm is applied to locate the position of objects on the conveyor belt. Then, we propose a Chinese character recognition network (CCRNNet) to significantly recognize each object from the original image. The position, type, and timestamp of each object are sent to cloud PLCs that are virtualized in the cloud to replace the function of traditional PLCs in the terminal. After that, we propose a dynamic scheduling method to sort the materials in minimum time. Finally, we establish a real experimental platform of a multi-crane visual sorting system to verify the performance of the proposed methods.

**Keywords:** intelligent manufacturing; computer vision; cloud programmable logic controller; deep convolutional neural network; character recognition; dynamic scheduling



**Citation:** Fu, M.; Wang, Q.; Wang, J.; Sun, L.; Ma, Z.; Zhang, C.; Guan, W.; Liu, Q.; Wang, D.; Li, W. Deep CNN-Based Materials Location and Recognition for Industrial Multi-Crane Visual Sorting System in 5G Network. *Appl. Sci.* **2023**, *13*, 1066. <https://doi.org/10.3390/app13021066>

Academic Editors: Junchi Yan, Minghao Guo and Antonio Fernández-Caballero

Received: 29 November 2022  
Revised: 30 December 2022  
Accepted: 9 January 2023  
Published: 12 January 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

Intelligent manufacturing methods [1–3] have presently attracted increasing attention from both academic and industry domains, in which many state-of-the-art technologies improve dramatic industrial revolution, including the Internet of Things [4], artificial intelligence (AI) [5], computer vision (CV) [6], and other technologies. The visual sorting system [7,8] has an enormous application range with the development of CV and industrial cameras, especially, multi-crane visual sorting systems in iron mining, steel metallurgy, coal mining, and other fields. There are three critical modules that we focus on in multi-crane visual sorting systems, which are the location module, the recognition module, and the control module. The location module aims to locate the position and get the world coordinates of the materials. The recognition module recognizes the type of material on the conveyor belt. The control module combines cloud PLC to sort and put the materials in the specified position within the acceptable time. Many visual sorting technologies were researched in the industry.

The first task for the visual recognition module is to detect the materials that are captured by the camera. In CV, many deep CNN-based algorithms were widely used in object detection [9]. There are two main categories, including one-stage detection frameworks [10,11] that is a single CNN network to directly predict class probabilities and bounding box offsets from original images and two-stage detection frameworks [12,13] that first generate object proposals from original images and then extract features of them to predict object label and bounding box offsets for proposal refinement. Recently, these algorithms are also used in the visual sorting system. Shen et al. [14] applied YOLOv4 which was one algorithm of one-stage detection frameworks to locate and place the garbage in the correct recyclable trash. Song et al. [15] proposed a single-stage grasp detection framework based on region proposal network architecture in the robotic grasp system, the complexity of which was lower than two-stage architecture. Wang et al. [7] used Faster R-CNN to automatically detect and classify the wheel hubs, and then send them to the production lines for high efficiency. However, deep learning-based methods are utilized in robotic grasp systems or wheel hubs detection but are not suitable for the multi-crane sorting system. Compared with two-stage algorithms, one-stage performs better in model complexity and computing resources. Considering the requirement of high accuracy and timeliness, a fast detection algorithm based-YOLOv5 [16] is applied in a multi-crane visual sorting system for further enhancement. We get pixel coordinates from the detection algorithm. The control module needs to know the world coordinates of materials for the sorting task. Then, the camera calibration method [17] is used to transform the materials from the pixel coordinate system to the world coordinate system.

After detecting materials, the classes are necessary to be recognized preparing for sorting different materials, such as different levels of product quality. Due to the limited experiments, we use Chinese character objects instead of industrial materials. In CV, several techniques for recognizing characters have been reported in many fields [18,19]. Albahli et al. [20] presented an effective and efficient hand-written digit recognition system that designed a customized faster regional convolutional neural network to localize and classify the digits into 10 classes. Cao et al. [21] designed a zero-shot handwritten Chinese character recognition framework based on CNN with a hierarchical decomposition embedding method to achieve competitive performance. Xie et al. [22] presented a vehicle license plate recognition system, in which a novel combined feature extraction model was designed for license plate detection, and a backpropagation neural network was used to recognize the characters of plates. Caldeira et al. [23] proposed an optical character recognition system to recognize characters printed on steel coils in the industry, in which the alignment and segmentation methods were used for filtering non-character components, and a CNN-based classification network was used for recognizing the characters. Gang et al. [24] applied EfficientNet based on CNN to recognize characters of components mounted on the printed circuit board for decreasing defect detection. Most of methods are used to recognize English letters and digits in many areas. However, it is more difficult to recognize Chinese characters due to their complex and diverse features. The ResNet-based network [25] performs significantly in many applications and has extensive research in computer vision. In this article, we plan to explore a ResNet-based network to recognize Chinese characters for high recognition accuracy. The position, types, and timestamp of Chinese chesses will be transmitted to cloud PLC that controls local PLC to conduct the multi-crane for completing material sorting.

The control module plays a significant role in the industry, which aims to use PLCs [26] that sense the inputs, execute the program, and write the outputs. Traditional PLCs are connected to the equipment in the terminal, which cannot realize the decoupling of a traditional PLC and equipment. In addition, the data cannot be interconnected between different equipment due to different industrial control protocols. Many approaches were developed to break the traditional pyramid structure of the industrial internet. Park et al. (Ref. [27]) presented the architecture of a PLC programming environment that employed a virtual plant model consisting of virtual devices to support the specification of

discrete event models. A hardware-in-the-loop simulation approach was proposed in the production system, which was directly used for the real plant with minor adaptations [28]. Goldschmidt et al. [29] introduced cloud-based software PLCs to improve the performance of scalability and multi-tenancy, in which the equipment was controlled by the soft PLCs in the cloud that dynamically scaled and assigned workloads. Kalle et al. [30] showed a novel virtual-PLC approach to avoid significant perturbation of the remote attack in industrial control systems. Zhu et al. [31] presented a cloud PLC platform in human-machine interaction for real-time remote monitoring. However, the mentioned PLCs cannot be flexibly deployed in the cloud-terminal-edge. Many emerging technologies cannot be integrated into the industrial structure, such as artificial intelligence, big data, and 5G. In order to meet the requirement of flexibility and scalability in intelligent manufacturing, it is necessary to explore an integrated method to break the data island and improve the coordination between devices. The function of PLCs can be virtualized in the cloud, which not only decreases the cost of PLCs but also makes full use of computing resources to realize data intercommunication.

Motivated by the above methods, we developed a novel deep CNN-based multi-crane visual sorting system in intelligent manufacturing, which enables the accurate sorting of materials in real time. The system uses cloud PLC to control the collaborative work of a multi-crane in a 5G network for high reliability and low latency. The main contributions can be summarized as follows:

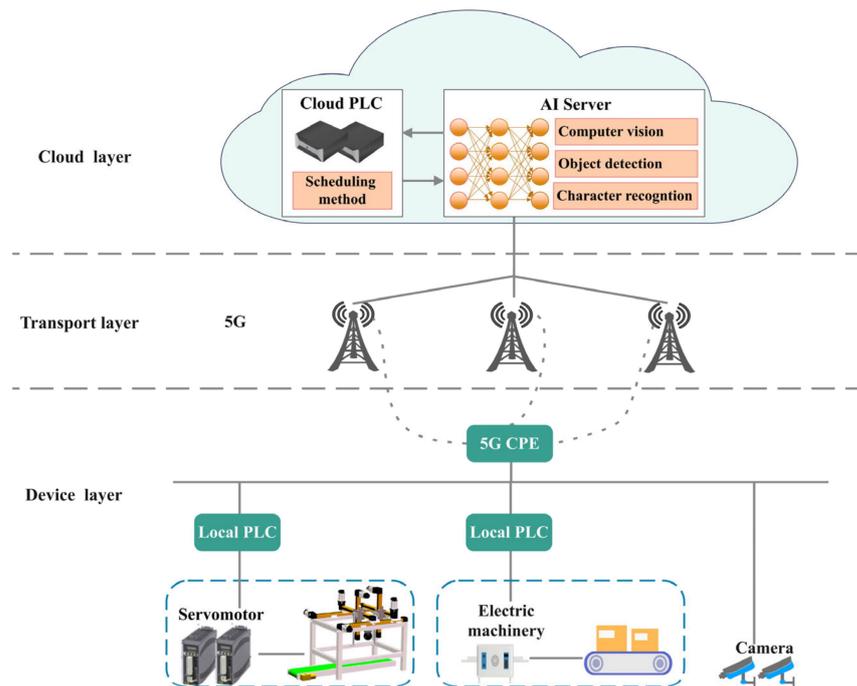
1. We design a novel multi-crane visual sorting system that applies deep CNN algorithm for the material location and recognition, and propose a dynamic scheduling method of materials with cloud PLC in 5G network;
2. We apply the YOLOv5 algorithm to locate the position of materials in the industry and use a camera calibration method to realize coordinate conversion. Additionally, we explore a Chinese character recognition network (CCRNet) to significantly recognize the class of each object from the original image;
3. We propose a dynamic scheduling method of materials, in which the multi-crane is controlled by local PLCs that receive the commands from cloud PLC and sorts the chesses in specified position with minimum time;
4. We establish an experimental platform of the multi-crane visual sorting system using cloud PLC in a 5G network for centralized control and low-latency. Furthermore, we collect the dataset and train the deep CNN-based model. The whole performance of the multi-crane visual sorting system is demonstrated by abundant experiments.

We arrange the remainder of this paper as follows. In Section 2, we introduce the architecture of the multi-crane visual sorting system. Section 3 concretely describes YOLOv5 algorithm for object location, CCRNet for Chinese character recognition, and a novel dynamic scheduling method for the controlling process of the materials. Section 4 shows the established experimental platform and the results. Section 5 obtains the conclusion and future work of this article.

## 2. Materials and System

The architecture of the multi-crane visual sorting system with cloud PLC in 5G network is shown in Figure 1, which consists of three layers, including the device layer, the transport layer, and the computing layer. The data of all devices are transmitted by a 5G network, and the commands from the cloud PLC also are sent to control the devices by the 5G network.

There are three types of devices in the device layer, including a multi-crane, a conveyor belt, and the cameras. The movement of the multi-crane and the conveyor belt are controlled by local PLCs that receive the commands from the cloud PLC. In order to locate and recognize the materials on the conveyor belt in real-time, the images or the video are transmitted by a 5G network to the cloud, in which an AI server is deployed that integrates deep learning methods to process data.



**Figure 1.** The architecture of multi-crane visual sorting system.

In the transport layer, a 5G system consists of an access network, a bearer network, and a core network. We deploy two 5G base stations as the access network. The bearer network is as a bridge to transmit the signal between the access network and the core network. In order to meet the requirement of high reliability and low latency about reliable low latency communication (URLLC) [32] as one typical application of 5G, some function of the core network sink to the edge. We can deploy cloud PLC in the edge or in the cloud.

In the computing layer, we design the cloud PLC to coordinatively control remote devices and cooperate with the high-performance server to support many applications that need a large amount of computing, like CV applications and precise location.

For the multi-crane visual sorting system, the first step is to locate the position and obtain the world coordinates of the materials. Second, the class of all materials are recognized from the original image. The process results are transmitted to the cloud PLC by the 5G network. Third, the cloud PLC sends commands to the local PLC that controls the multi-crane according to the dynamic scheduling method. In this article, we focus on three critical technologies that are the location module, the recognition module, and the control module, the details of which are introduced as follows.

### 3. Methodology

#### 3.1. The Introduction of CNN

CNN has obtained achievements in many applications, such as person re-identification, action recognition, object detection, and image segmentation. CNN is one of the most important neural networks in deep learning, which performs dramatically in feature representation [33]. It is always the backbone of deep CNN. An example of CNN structure is presented in Figure 2, which consists of convolution layers, pooling layers, and fully connection layers. In convolution layers, small convolutional kernels are utilized to convolve the original image and intermediate feature maps to learn edge features, color features, texture features, and other features. These kernels are updated adaptively for many epochs training by back propagation. The pooling layers still use a small kernel following convolution layers to reduce the dimension of feature maps, which also decreases the computing, avoiding overfitting. There are two common types: max pooling and average pooling. After convolution and pooling layers, the feature maps are flattened to one-dimensional vectors through the global average pooling with (1, 1) kernel. In full connection layers, all



difference between area C and the union area is obtained. GIoU is the ratio of the difference and area C. The regression loss is given by:

$$L_{reg} = 1 - GIoU = 1 + \frac{A^C - A^U}{A^C} - IoU \tag{1}$$

where  $A^C$  is the area of C that is the smallest box that includes A and B.  $A^U$  is the area of the union between the truth box A and the prediction box B.

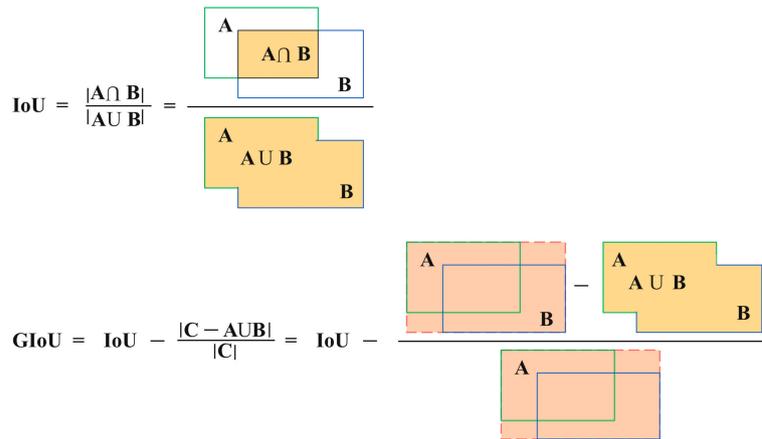


Figure 4. Diagrams of IoU and GIoU.

**Classification Loss.** Binary Cross-Entropy with Logits Loss is used to recognize and optimize the classification of the materials. The classification loss is summarized as:

$$L_{cls} = \frac{1}{N} \sum_{i=1}^N [\hat{C}_i \ln(\text{sigmoid}(C_i)) + (1 - \hat{C}_i) \ln(1 - \text{sigmoid}(C_i))] \tag{2}$$

where  $N$  is the value of the min-batch,  $C_i$  is the predicted class, and  $\hat{C}_i$  is the truth class.  $\text{sigmoid}$  is the sigmoid function that is  $[0, 1]$ .

**Confidence Loss.** Binary Cross-Entropy with Logits Loss is used to optimize the confidence of the bounding box. The confidence loss is summarized as:

$$L_{con} = \frac{1}{N} \sum_{i=1}^N [\hat{P}_i \ln(\text{sigmoid}(P_i)) + (1 - \hat{P}_i) \ln(1 - \text{sigmoid}(P_i))] \tag{3}$$

where  $N$  is the value of the min-batch,  $P_i$  is the prediction of the confidence, and  $\hat{P}_i$  is the truth of the confidence. The value of  $\hat{P}_i$  is  $[0, 1]$ .

The total loss of YOLOv5 is:

$$L_{total} = \lambda_1 L_{reg} + \lambda_2 L_{cls} + \lambda_3 L_{con} \tag{4}$$

where  $\lambda_1, \lambda_2,$  and  $\lambda_3$  are weights of the three losses.

We only obtain the pixel coordinates of objects that cannot be used by cloud PLC. Camera calibration [17] is used to describe the collection between the pixel coordinate system and the world coordinate. We define the pixel coordinate system  $(u, v)$  and the world coordinate system  $(X_w, Y_w, Z_w)$ .  $P_w$  presents a real P point. The relation between the pixel coordinate system and the world coordinate system can be obtained as:

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = MP_w \tag{5}$$

where  $Z_c$  is the scale factor.  $M$  is the projection matrix of the camera.

For performance evaluation, we use precision, recall, and mean Average Precision (mAP) as the standard methods for performance [34]. After the location module, the world coordinates of the materials are achieved, which are transmitted to cloud PLC.

### 3.3. The Recognition Module of Chinese Characters

For the sorting task, we need to recognize the class of each object. In the experiments, we use Chinese characters instead of the real materials in the industry due to the limited condition. A CCRNet is designed to recognize the characters, as shown in Figure 5. It uses ResNet [25] as the backbone network and compares them in the experiments. The backbone is used to extract the features-of-interest region. After backbone, we use the fully connected layer to output the class of Chinese character. For example, the input is the image of a Chinese character and the output is the class of “Jing”. We collect the dataset to train and test these models.

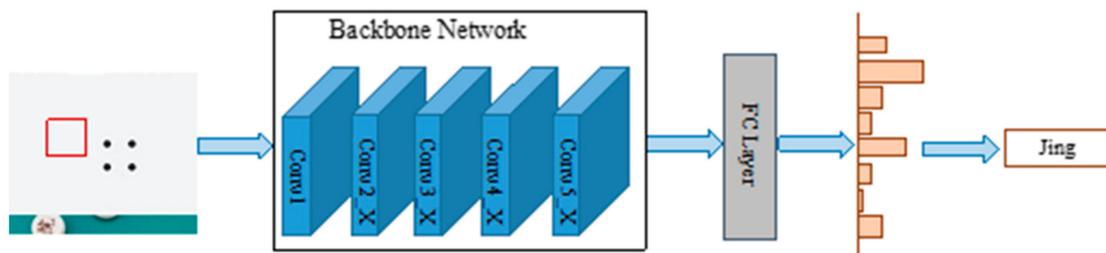


Figure 5. ResNet-based Chinese character recognition network.

The structure of ResNet18, ResNet34, and ResNet50 is introduced in Table 1. All of them mainly consist of Conv1, Conv2\_x, Conv3\_x, Conv4\_x, Conv5\_x, and the fully connected layer. The difference is the number of convolutional layers that are described in Table 1. In this article, the number of the output is 18 classes. We compare the performance of these models about Chinese character recognition in the experiments.

Table 1. The structure of ResNet18, ResNet34, and ResNet50 [25].

Layer Name	ResNet18	ResNet34	ResNet50
Conv1		(7, 7), 64, stride 2	
		(3, 3), max pooling, stride 2	
Conv2_x	$\left\{ \begin{matrix} (3, 3), 64 \\ (3, 3), 64 \end{matrix} \right\} \times 2$	$\left\{ \begin{matrix} (3, 3), 64 \\ (3, 3), 64 \end{matrix} \right\} \times 3$	$\left\{ \begin{matrix} (1, 1), 64 \\ (3, 3), 64 \\ (1, 1), 256 \end{matrix} \right\} \times 3$
Conv3_x	$\left\{ \begin{matrix} (3, 3), 128 \\ (3, 3), 128 \end{matrix} \right\} \times 2$	$\left\{ \begin{matrix} (3, 3), 128 \\ (3, 3), 128 \end{matrix} \right\} \times 4$	$\left\{ \begin{matrix} (1, 1), 128 \\ (3, 3), 128 \\ (1, 1), 512 \end{matrix} \right\} \times 4$
Conv4_x	$\left\{ \begin{matrix} (3, 3), 256 \\ (3, 3), 256 \end{matrix} \right\} \times 2$	$\left\{ \begin{matrix} (3, 3), 256 \\ (3, 3), 256 \end{matrix} \right\} \times 6$	$\left\{ \begin{matrix} (1, 1), 256 \\ (3, 3), 256 \\ (1, 1), 1024 \end{matrix} \right\} \times 6$
Conv5_x	$\left\{ \begin{matrix} (3, 3), 512 \\ (3, 3), 512 \end{matrix} \right\} \times 2$	$\left\{ \begin{matrix} (3, 3), 512 \\ (3, 3), 512 \end{matrix} \right\} \times 3$	$\left\{ \begin{matrix} (1, 1), 512 \\ (3, 3), 512 \\ (1, 1), 2048 \end{matrix} \right\} \times 3$
		Average pooling	
		Fully connected layer	

We use softmax cross-entropy loss [35] to optimize CCRNet. The loss function is summarized as:

$$L_{scls} = -\frac{1}{N} \sum_{i=1}^N -\log\left(\frac{e^{Z_i}}{\sum_{j=1}^C e^{Z_j}}\right) \quad (6)$$

where  $N$  is the batch size, and  $Z_i$  is the prediction output of the CCRNet.

### 3.4. The Control Module Using Dynamic Scheduling Method

The information of all materials in the visual field is transmitted to the cloud PLC. In the initial state, all cranes are waiting for demands. We design a dynamic scheduling method on cloud PLC to control cranes in minimum time. We make a multi-crane list  $Cra = [i, (x_i, y_i, z_i)]$ ,  $i \in [0, 1, 2 \dots, K]$ .  $i$  is the  $i^{\text{th}}$  crane, and  $(x_i, y_i, z_i)$  is the initial position of the  $i^{\text{th}}$  crane. We define a multi-object list  $Obj = [j, (x_j, y_j, z_j, x'_j, y'_j, z'_j)]$ ,  $j \in [0, 1, 2 \dots, J]$ .  $j$  is the  $j^{\text{th}}$  object.  $(x_j, y_j, z_j)$  is the position captured of the  $j^{\text{th}}$  object by location module.  $(x'_j, y'_j, z'_j)$  is the target position of the  $j^{\text{th}}$  object to be transported. The scheduling strategy is described as Algorithm 1:

---

#### Algorithm 1. Scheduling strategy.

---

**input:** All cranes  $Cra = [i, (x_i, y_i, z_i)]$ ,  $i \in [0, 1 \dots, K]$ . All materials  $Obj = [j, (x_j, y_j, z_j, x'_j, y'_j, z'_j)]$ ,  $j \in [0, 1, 2 \dots, J]$ .

**output:** The relation between the crane and the object for sorting

Initialize multi-crane state

**for** each crane  $i \in Cra$  **do**

Record the initial time  $t_{i0}$  and the initial position  $x_{i0}$  of each crane

**for** each object  $j \in Object$  **do**

        Calculate the arrival time of the  $i^{\text{th}}$  crane  $t_{i1} = t_{i0} + t_{x1}$ ,  $t_{x1}$  is the running time from the initial position to the object position on the x-axis. Record the crane position  $x_{i1}$  at  $t_{i1}$ .

        Calculate the completion time of shipment  $t_{i2} = t_{i1} + t_{x2}$ ,  $t_{x2}$  is the time of shipment. Record the crane position  $x_{i2}$  at  $t_{i2}$ .

        Calculate the time of delivery to the target place  $t_{i3} = t_{i2} + t_{x3}$ ,  $t_{x3}$  is the running time from the object position to the target position.

        Record the crane position  $x_{i3}$  at  $t_{i3}$ .

        Calculate  $(t_{y1}, t_{z1}, t_{y3}, t_{z3})$ . The process is the same with x-axis.

        Calculate the whole time of the  $i^{\text{th}}$  crane that transports the  $j^{\text{th}}$  object:

$$T_j = \max(t_{x1}, t_{y1}, t_{z1}) + t_{x2} + \max(t_{x3}, t_{y3}, t_{z3})$$

    Until all materials are calculated. We get a list  $T = [T_0, T_1, \dots, T_J]$ .

    Sequence the element in  $T$  from small to large. We can get the object's ID that the  $i^{\text{th}}$  crane needs to transport.

Until all cranes are assigned with materials to be transported.

---

## 4. Results

### 4.1. The Experimental Results of Detection Module Based on YOLOv5

We designed a comparative experiment involving YOLOv5 and deployed them on PyTorch using NVIDIA 3090 graphics processing units (GPUs). To improve the robustness of the system, we used data augmentation methods on the original images, including scaling, color space adjustments, and mosaic augmentation. We set the number of epochs for 100 and the batch size to 8. Adam optimizer [36] was utilized for learning the dramatic representations. The initial learning rate was 0.001. We collected 790 images that consisted of chesses and cubes. In more detail, 568 images was selected as the training set, 79 images belong to the testing set, and the other 142 images belong to the validating set.

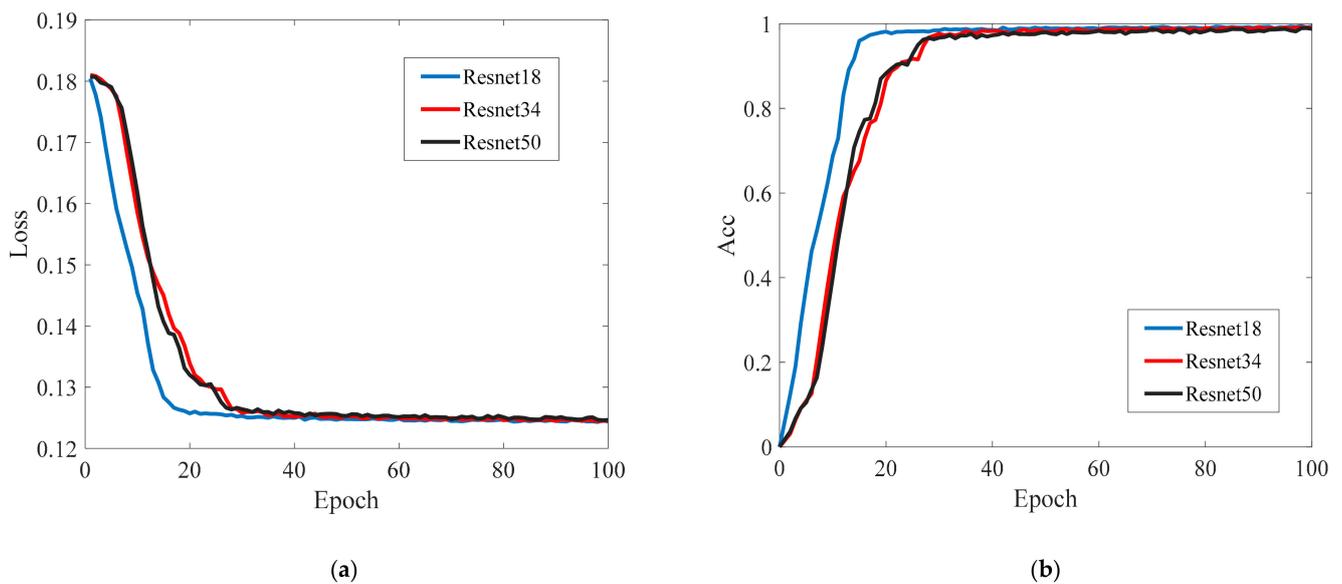
After training and testing, we obtain the precision, recall rate, and mAP on the training set, testing set, and validating set, which are presented based on the YOLOv5 algorithm in Table 2. We achieve the precision, recall rate, and mAP over 99% on the training set. Moreover, the mAP of YOLOv5 on the testing set is up to 98.85%. The precision of the validating set is also up to 96.55%. This further proves the significant performance of YOLOv5 in the multi-crane visual sorting system. From the overall performance, we choose YOLOv5 as the location algorithm and used the process results of YOLOv5 to continue the camera calibration and material sorting experiments.

**Table 2.** The results of YOLOv5 for object detection on the training set, testing set, and validating set.

Model	Dataset	Number	Precision	Recall Rate	mAP
YOLOv5	Training set	568	99.60%	99.46%	99.44%
	Testing set	79	98.49%	98.81%	98.85%
	Validating set	142	96.55%	95.40%	97.80%

#### 4.2. The Results of CCRNet on Chinese Characters

We collected the dataset for Chinese character recognition. There were 5333 images of 18 classes, in which there were 14 Chinese characters and 4 English letters. 4799 images were selected as the set, and 534 images belonged to the testing set. We added ResNet18, ResNet34, and ResNet50 as the backbone of CCRNet. Epoch was set to 100 and the batch size was 16. Adam optimizer was also used to optimize the network and the initial learning rate was 0.0001. The loss of ResNet-based models is presented in Figure 6a on the training set. We can see that the convergence rate of ResNet18 was faster than ResNet34 and ResNet50. The performance of ResNet50 was better than ResNet34. Meanwhile, the accuracy of ResNet-based models is presented in Figure 6b on the training set. With the increasing of epoch, the accuracy of these three models is becoming better. After 60 epochs, the trend is becoming more and more flat. These three models performed significantly well in character recognition.



**Figure 6.** The loss and accuracy comparison of character recognition on ResNet-based models. (a) The loss comparison. (b) The accuracy comparison.

As seen in Table 3, we trained the CCRNet model based on ResNet18, ResNet34, and ResNet50 for the testing set. The accuracy was over 99%. On the training set, the accuracy of CCRNet\_ResNet18 exceeded 0.01% and 0.08% over CCRNet\_ResNet34 and CCRNet\_ResNet50. On the testing set, the accuracy of CCRNet\_ResNet34 exceeded 0.18% over CCRNet\_ResNet18 and CCRNet\_ResNet50. The average accuracy of three models was over 99.2%, which demonstrates that ResNet-based models performed significantly well on Chinese character recognition.

**Table 3.** The accuracy of ResNet for character recognition on the training set and testing set.

Model	Dataset	Number	Accuracy
CCRNet_ResNet18	Training set	4799	99.50%
	Testing set	534	99.25%
CCRNet_ResNet34	Training set	4799	99.49%
	Testing set	534	99.43%
CCRNet_ResNet50	Training set	4799	99.42%
	Testing set	534	99.25%

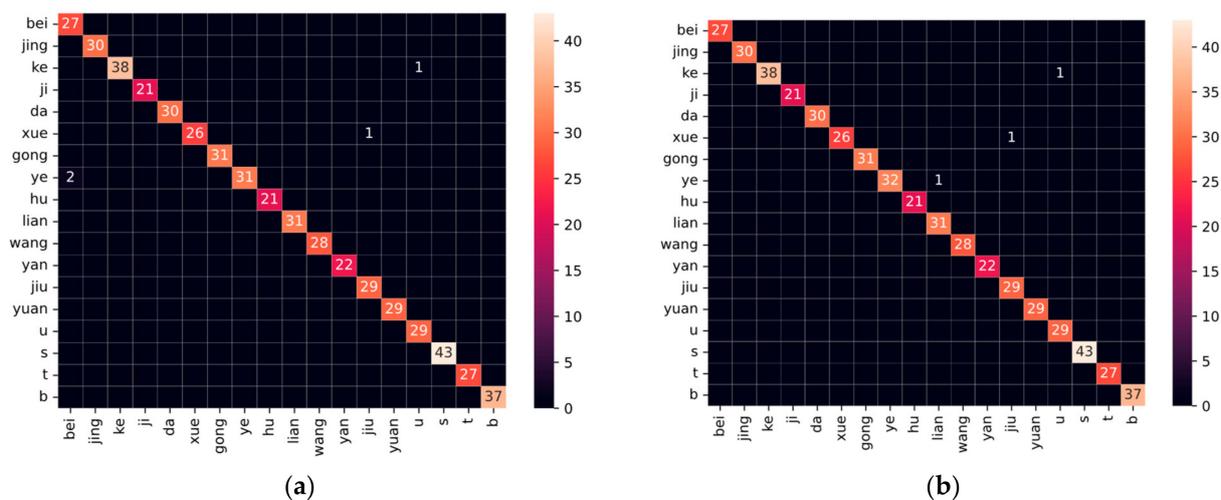
In addition, we compared the accuracy of Chinese characters and English letters, and the results can be seen in Table 4. We can see that the accuracy of English letters is 100% and better than Chinese characters. CCRNet\_ResNet34 achieves 99.25% of accuracy and is 0.26% higher than CCRNet\_ResNet18 and CCRNet\_ResNet50.

**Table 4.** The accuracy of CCRNet on Chinese characters and English letters.

Model	Chinese Character Accuracy	English Letter Accuracy
CCRNet_ResNet18	98.99%	100%
CCRNet_ResNet34	99.25%	100%
CCRNet_ResNet50	98.99%	100%

In Figure 7, the accuracy distribution of each character on CCRNet\_ResNet18, CCRNet\_ResNet34, and CCRNet\_ResNet50 is presented, in which the value on the diagonal represents the number of each character. We can find that four characters are recognized to the error class on CCRNet\_ResNet18 and CCRNet\_ResNet50. Only three characters are recognized to the error class on ResNet34. Considering the performance and complexity, we used CCRNet\_ResNet18 as the final backbone of CCRNet.

The detection and recognition of Chinese characters is shown in Figure 8, in which all objects are precisely detected, and the confidence score is close to 1. In each image, there are different shapes of objects with different classes. All characters are fully recognized and located. From the experimental results, we demonstrated the significance of YOLOv5 and CCRNet on character detection and recognition.



**Figure 7.** Cont.



the performance of the scheduling method and meets the requirement of the sorting task in real-time.

**Table 5.** The consuming time of the crane for the sorting task with scheduling strategy.

Time	Prediction Time (s)	Real Time (s)	$ \Delta t $
1	[(1, 2.972), (2, 3.282), (3, 3.301), (4, 3.616)]	2.756	0.216
2	[(1, 2.804), (2, 3.179), (3, 3.237), (4, 3.648)]	3.016	0.212
3	[(1, 3.146), (2, 3.160), (3, 3.179), (4, 3.398)]	3.056	0.090
4	[(1, 2.958), (2, 3.037), (3, 3.055), (4, 3.370)]	3.064	0.106
5	[(1, 2.838), (2, 3.179), (3, 3.237), (4, 3.648)]	3.036	0.198

The experimental results of the multi-crane visual sorting system are shown in Figure 9. The objects on the conveyor belt are correctly sorted into the designated place in real-time. The proposed method can be applied to many applications in the industry, including separation of different materials and material grading. In the future, we will design a dynamic placement method in a cloud PLC, which can place the objects in the specified position according to certain rules.



**Figure 9.** The experimental diagram of the crane visual sorting system.

## 5. Conclusions and Future Work

The intelligent visual sorting system plays an important role in intelligent manufacturing, in which artificial intelligence and computer vision technologies improve the intelligent and unmanned development of the industry. We establish a deep CNN-based multi-crane visual sorting system with cloud PLCs in a 5G environment. A YOLOv5-based algorithm and a Chinese character recognition network were developed to significantly locate and recognize materials from the original image. mAP of YOLOv5 on validating set is up to 98.85%, and the confidence score is close to 1, which significantly shows the performance of the location module. The accuracy of Chinese character recognition is also up to 99.43% on testing set, which fully met the requirement of different material sorting in the industry. Moreover, cloud PLCs are investigated to be flexibly deployed in the terminal and in the cloud, which allows low-level devices to cooperate with each other. Therefore, the control scheduling method is designed in cloud PLCs that cooperates with AI algorithms to control the multi-crane in minimum time. The average time of sorting is about 3 s, which contains the whole process of the crane from starting to sorting completion. The real experimental platform of the multi-crane visual sorting system verifies the performance of the proposed methods.

Although the experimental results perform dramatically, there is much room to improve the multi-crane sorting system. In the industry, there are massive data streams that are transmitted by a 5G network. We will design 5G and TSN bridge as a transmission layer to highly guarantee deterministic communication. In addition, we will integrate cloud

PLCs and an AI platform on the same server in the cloud for stable performance. We will develop some scheduling method to control the low-level devices for low energy and high efficiency. Furthermore, we will continue to improve the accuracy of AI algorithms that can be used in real industrial applications.

**Author Contributions:** Conceptualization, M.F., J.W., Z.M. and W.L.; methodology, M.F., Q.W., J.W. and L.S.; software, M.F. and Q.W.; validation, M.F., W.G. and C.Z.; formal analysis, M.F. and D.W.; investigation, M.F.; resources, Z.M.; data curation, C.Z. and Q.L.; writing—original draft preparation, M.F.; writing—review and editing, J.W., Q.W. and D.W.; visualization, M.F.; supervision, J.W.; project administration, J.W.; funding acquisition, J.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by National Key Research and Development Program under Grant 2020YFB1708800, Guangdong Key Research and Development Program under Grant 2020B0101130007, Interdisciplinary Research Project for Young Teachers of USTB (Fundamental Research Funds for the Central Universities) FRF-IDRY-21-005, Fundamental Research Funds for Central Universities under Grant FRF-MP-20-37, GuangDong Basic and Applied Basic Research Foundation under Grant 2021A1515110577, and China Postdoctoral Science Foundation under Grant 2021M700385. Central Guidance on Local Science and Technology Development Fund of ShanXi Province under Grant YDZJSX2022B019.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data sharing is not applicable to this article.

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

## References

1. Zhong, R.Y.; Xu, X.; Klotz, E.; Stephen, T.N. Intelligent manufacturing in the context of industry 4.0: A review. *Engineering* **2017**, *3*, 616–630. [[CrossRef](#)]
2. Wang, J.; Xu, C.; Zhang, J.; Zhong, R. Big data analytics for intelligent manufacturing systems: A review. *J. Manuf. Syst.* **2022**, *62*, 738–752. [[CrossRef](#)]
3. Nain, G.; Pattanaik, K.K.; Sharma, G.K. Towards edge computing in intelligent manufacturing: Past, present and future. *J. Manuf. Syst.* **2022**, *62*, 588–611. [[CrossRef](#)]
4. Nguyen, D.C.; Ding, M.; Pathirana, P.N.; Seneviratne, A.; Li, J.; Niyato, D.; Dobre, O.; Vincen, P.H. 6G Internet of Things: A comprehensive survey. *IEEE Internet Things J.* **2022**, *9*, 359–383. [[CrossRef](#)]
5. Zhang, C.; Lu, Y. Study on artificial intelligence: The state of the art and future prospects. *J. Ind. Inf. Integr.* **2021**, *23*, 100224. [[CrossRef](#)]
6. Guo, M.H.; Xu, T.X.; Liu, J.J.; Liu, Z.N.; Jiang, P.T.; Mu, T.J.; Zhang, S.H.; Martin, R.R.; Cheng, M.M.; Hu, S.M. Attention mechanisms in computer vision: A survey. *Comput. Vis. Media* **2022**, *8*, 1–38. [[CrossRef](#)]
7. Wang, Y.; Hong, K.; Zou, J.; Peng, T.; Yang, H.Y. A CNN-based visual sorting system with cloud-edge computing for flexible manufacturing systems. *IEEE Trans. Ind. Inform.* **2019**, *16*, 4726–4735. [[CrossRef](#)]
8. Han, S.; Liu, X.; Han, X.; Wang, G.; Wu, S.B. Visual sorting of express parcels based on multi-task deep learning. *Sensors* **2020**, *20*, 6785. [[CrossRef](#)] [[PubMed](#)]
9. Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Chen, J.; Liu, X.W.; Pietikäinen, M. Deep learning for generic object detection: A survey. *Int. J. Comput. Vis.* **2020**, *128*, 261–318. [[CrossRef](#)]
10. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, R. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
11. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
12. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; p. 28.
13. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
14. Shen, X.; Wu, Y.; Chen, S.; Luo, X. An intelligent garbage sorting system based on edge computing and visual understanding of social internet of vehicles. *Mob. Inf. Syst.* **2021**, *2021*, 5231092. [[CrossRef](#)]
15. Song, Y.; Gao, L.; Li, X.; Shen, W. A novel robotic grasp detection method based on region proposal networks. *Robot. Comput. Integr. Manuf.* **2020**, *65*, 101963. [[CrossRef](#)]

16. Jocher, G.; Nishimura, K.; Mineeva, T. yolov5. Code Repository. 2020. Available online: <https://github.com/ultralytics/yolov5> (accessed on 30 December 2022).
17. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]
18. Baldominos, A.; Saez, Y.; Isasi, P. A survey of handwritten character recognition with mnist and emnist. *Appl. Sci.* **2019**, *9*, 3169. [[CrossRef](#)]
19. Melnyk, P.; You, Z.; Li, K. A high-performance CNN method for offline handwritten Chinese character recognition and visualization. *Soft Comput.* **2020**, *24*, 7977–7987. [[CrossRef](#)]
20. Albahli, S.; Nawaz, M.; Javed, A.; Irtaza, A. An improved faster-RCNN model for handwritten character recognition. *Arab. J. Sci. Eng.* **2021**, *46*, 8509–8523. [[CrossRef](#)]
21. Cao, Z.; Lu, J.; Cui, S.; Zhang, C. Zero-shot handwritten chinese character recognition with hierarchical decomposition embedding. *Pattern Recognit.* **2020**, *107*, 107488. [[CrossRef](#)]
22. Xie, F.; Zhang, M.; Zhao, J.; Yang, J.; Liu, Y.; Yuan, X. A robust license plate detection and character recognition algorithm based on a combined feature extraction model and BPNN. *J. Adv. Transp.* **2018**, *2018*, 6737314. [[CrossRef](#)]
23. Caldeira, T.; Ciarelli, P.M.; Neto, G.A. Industrial optical character recognition system in printing quality control of hot-rolled coils identification. *J. Control. Autom. Electr. Syst.* **2020**, *31*, 108–118. [[CrossRef](#)]
24. Gang, S.; Fabrice, N.; Chung, D.; Lee, J. Character Recognition of Components Mounted on Printed Circuit Board Using Deep Learning. *Sensors* **2021**, *21*, 2921. [[CrossRef](#)] [[PubMed](#)]
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
26. Biallas, S.; Brauer, J.; Kowalewski, S. Arcade.PLC: A verification platform for programmable logic controllers. In Proceedings of the 2012 27th IEEE/ACM International Conference on Automated Software Engineering, Essen, Germany, 3–7 September 2012; pp. 338–341.
27. Park, S.C.; Park, C.M.; Wang, G.N. A PLC programming environment based on a virtual plant. *Int. J. Adv. Manuf. Technol.* **2008**, *39*, 1262–1270. [[CrossRef](#)]
28. Park, S.C.; Chang, M. Hardware-in-the-loop simulation for a production system. *Int. J. Prod. Res.* **2012**, *50*, 2321–2330. [[CrossRef](#)]
29. Goldschmidt, T.; Murugaiah, M.K.; Sonntag, C.; Schlich, B.; Biallas, S.; Weber, P. Cloud-based control: A multi-tenant, horizontally scalable soft-PLC. In Proceedings of the IEEE 2015 8th International Conference on Cloud Computing, New York City, NY, USA, 27 June–2 July 2015; pp. 909–916.
30. Kalle, S.; Ameen, N.; Yoo, H.; Ahmed, I. CLIK on PLCs! Attacking control logic with decompilation and virtual PLC. In Proceedings of the Binary Analysis Research Workshop, Network and Distributed System Security Symposium, San Diego, CA, USA, 24–27 February 2019.
31. Zhu, Z.Y.; Liu, R.Y. Design of speed reducer testbed based on cloud platform. In Proceedings of the IEEE 2021 5th Advanced Information Technology, Electronic and Automation Control Conference, Chongqing, China, 12–14 March 2021; pp. 53–57.
32. Ren, H.; Wang, K.; Pan, C. Intelligent reflecting surface-aided URLLC in a factory automation scenario. *IEEE Trans. Commun.* **2021**, *70*, 707–723. [[CrossRef](#)]
33. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6999–7019. [[CrossRef](#)] [[PubMed](#)]
34. Henderson, P.; Ferrari, V. End-to-end training of object class detectors for mean average precision. In Proceedings of the Asian Conference on Computer Vision, Cham, Switzerland; 2016; pp. 198–213.
35. Kim, Y.; Lee, Y.; Jeon, M. Imbalanced image classification with complement cross entropy. *Pattern Recognit. Lett.* **2021**, *151*, 33–40. [[CrossRef](#)]
36. Zhang, Z. Improved adam optimizer for deep neural networks. In Proceedings of the IEEE/ACM 2018 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, 4–6 June 2018; pp. 1–2.

**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.