Intelligent sorting method for assembly line based on visual positioning and robotic arm model predictive control

Ruining Zhang, Wei Lu*, Xingliang Jian, Hui Luo

(*College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210095, China*)

Abstract: The existing steering device in the fruit and vegetable packaging assembly line cannot adjust the attitude of lettuce to a unified attitude, affecting the input and packaging process of the packaging machine. This study proposes an intelligent assembly line sorting method based on the visual positioning and model predictive control of a robotic arm. First, lightweight improvement based on the YOLOv5 is realized, the lettuce stalk in the background of the conveyor belt is promptly identified, the image of the lettuce stalk in the anchor box area is processed, and the edge contour point set is determined to extract the pixel coordinates of the optimal grasp point and mirror inclination angle of the lettuce. For the intelligent assembly line system, a robot arm kinematics model is constructed and the robot kinematics inverse solutions are calculated. Additionally, the lettuce movement speeds are dynamically measured by the vision system. A combination of the model prediction control, dynamic tracking, and rapid sorting of the lettuce by the robot claw is realized. The results show that the average detection time of a single frame image in the visual positioning part is 0.014 s, which is reduced by 50%; the accuracy and recall are 98% and 95%, respectively. The detection time is significantly reduced by ensuring accuracy. Within the current speed range of the packaging assembly line conveyor belt, the manipulator can grasp lettuce at different speeds stably and fast; the average axial error, average radial error, and adjusted average inclination angle error are 0.71 cm, 1.02 cm, and 3.79°, respectively, verifying the high efficiency and stability of the model. The proposed method of this study enables application in the intelligent sorting operation of industrial assembly lines

Keywords: YOLOv5, deep learning, image recognition, model predictive control, intelligent assembly line

DOI: 10.25165/j.ijabe.20231604.7908

Citation: Zhang R N, Lu W, Jian X L, Luo H. Intelligent sorting method for assembly line based on visual positioning and robotic arm model predictive control. Int J Agric & Biol Eng, 2023; 16(4): 206–214.

1 Introduction

Lettuce is one of the most popular crops in the world. It is cultivated all over the world and has valuable dietary value and medicinal value[1]. With the development of agricultural technology and crop cultivation technology, the total production value of grain production is approximately 500 million t, which creates a huge market for the packaging production industry. Crops need to be packed as soon as possible after harvest to ensure freshness and completeness. In the era of artificial intelligence, the selection, processing, and packaging of lettuce can be completed on an assembly line. However, at the input end of the packaging machine, the random supply of lettuce leads to a random attitude, packaging materials cannot fit lettuce perfectly, and it is difficult to achieve the preset packaging requirements, damaging the packaging machine. Therefore, to change the transportation state of the packaged object, it is necessary to configure the reversing device before entering the packaging machine. The original transportation direction and center of gravity position of the object remain unchanged after reversing. However, the currently available reversing device can only rotate the lettuce at the same angle, creating a risk of lettuce leaves getting stuck in the machine. Moreover, manual sorting has disadvantages such as a small batch, low precision, and an increase in manpower and time consumption. Therefore, it is necessary to design intelligent visual recognition and grasping systems that can automatically adjust the lettuce posture to a unified posture.

Presently, the combination of vision technology and robots has become an important way to improve the level of intelligence in various industries[2]. The real-time, stability, and accuracy of the visual positioning module and the manipulator control grasping module cause challenges in lettuce sorting using the intelligent assembly line system. Therefore, to establish a better combination of vision and control to realize the fast recognition and ranking of pipeline lettuce, a strong real-time lightweight network, and image processing method to realize the real-time recognition, positioning, and speed measurement of lettuce is required, as well as an accurate and efficient feedback control system to drive the fast-tracking and ranking of the robot arm. Ayyad et al.[3] first estimated the initial pose of the workpiece, performed multi-view reconstruction, and then performed accurate localization based on a position and image visual serving method. Ruan et al.[4] reviewed two main methods of fruit location and recognition, including digital image processing techniques and algorithms based on deep learning, and conducted target tracking in dynamic jamming environments. Do et al.[5] generated both target and reference model point clouds based on a depth camera system. Lim et al.[6] extracted the coordinates of feature points of boxes with different shapes and then controlled the robot. However, the grasping method lacks real-time state feedback and adjustment of the tracking path, and the grasping accuracy is

Received date: 2022-09-11 **Accepted date:** 2023-01-27

Biographies: Ruining Zhang, Undergraduate college, research interest: deep learning, machine learning. Email: 32219227@njau.edu.cn; **Xingliang Jian**, master, research interest: electromagnetic non-destructive testing, non-destructive testing. Email: jianxingliang@njau.edu.cn; **Hui Luo**, PhD, research interest: fault testing and diagnosis of analog electronic system, detection of agricultural products based on spectral analysis technology. Email: huiluo@njau.edu.cn.

***Corresponding author: Wei Lu**, PhD, research interest: intelligent robot and Artificial intelligence technology, intelligent sensing and non-destructive testing technology. Mailing address: College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210031, China. Email: njaurobot@njau.edu.cn.

relatively low. It is only suitable for situations where the shape of the target object is simple and the path changes slightly. Therefore, to establish a better combination of vision and control to realize the fast recognition and ranking of pipeline lettuce, a strong real-time lightweight network, and image processing method to realize the real-time recognition, positioning, and speed measurement of lettuce is required, as well as an accurate and efficient feedback control system to drive the fast-tracking and ranking of the robot arm.

With the continuous development of deep learning represented by convolutional neural networks, object detection has been widely applied in various fields[7,8]. Object detection algorithms can be divided into two categories, one is the two-stage algorithm, including the R-CNN series and FPN algorithms. However, there occurs a double calculation of images, which consumes time, and causes low efficiency. The other is the one-stage algorithm, including the YOLO series and SSD algorithms. The YOLO algorithm conducts object classification and candidate box prediction directly, and the detection speed is significantly improved. The SSD algorithm combines the advantages of the YOLO[9] algorithm; the faster R-CNN algorithm proposes feature maps of different scales to detect objects; and CNN directly makes predictions[10]. This kind of algorithm has a robust real-time performance, which promotes the application of object detection algorithms based on deep learning in industrial applications. The YOLOv5 algorithm is suitable when the detection object is a big target and the speed is pursued.

Over the past few decades, the control and design of robotic arms have become a major area of interest in robotics. In the 21st century, researchers began to pay attention to the application of agricultural robots. PID controller has been widely used in the early stage of the robot industry because of its simple structure and acceptable performance. However, because the manipulator is a complex system with nonlinear, strong coupling and time-varying characteristics, how to obtain the optimal PID parameters is a big challenge. In order to improve the control accuracy of robots, people have been focusing on the implementation of robust optimal control, such as fuzzy logic controllers[11], sliding mode controllers[12], controllers based on reinforcement learning[13], Model Predictive Control (MPC)[14], A and other optimal control strategies. Among them, Model Predictive Control (MPC) is an effective way to optimize future behavior over a limited range to track desired end-effector locations[15]. Tracking control, online obstacle avoidance, and path tracking of end-effector attitude have been realized, and the calculation time is short.

With the intelligent assembly line system as the research object, object detection is realized by the lightweight improvement of the YOLOv5 algorithm, the lettuce pose is extracted by real-time image processing, and the lettuce speed is detected by coordinate transformation from the lettuce clipping pixel coordinate system to the physical coordinate system of the manipulator. The manipulator conducts dynamic tracking, grasping, and sorting based on model predictive control to provide a technical reference for the target object sorting process of the modern crop packaging line.

2 Experimental setup and method flow chart

In this study, to solve the problems of the input and packaging process caused by the randomness of the delivery posture of lettuce in the packaging line, an intelligent assembly line sorting method based on visual positioning and robot arm model predictive control was proposed. The method consisted of vision and control parts. The main components of the system included a camera, a robot arm, an industrial computer, and an assembly line conveyor belt. The vision part was responsible for the recognition, acquisition, and image processing of the lettuce. The positioning and speed information were input into the IPC control system to realize the real-time signal input in the control part. The IPC and manipulator transmitted the target information and optimized the feedback in real-time based on the model's predictive control. The specific experimental setup diagram, flow chart, and closed-loop control diagrams are shown in Figures 1-3.

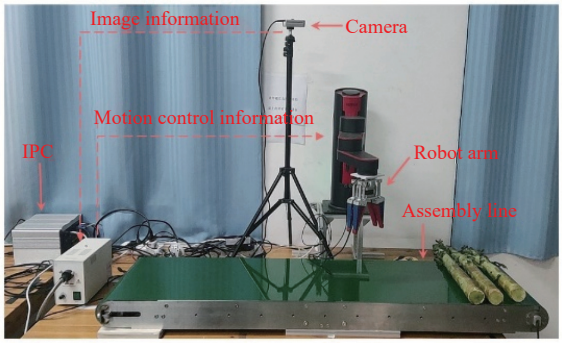


Figure 1 Specific experimental setup of this study

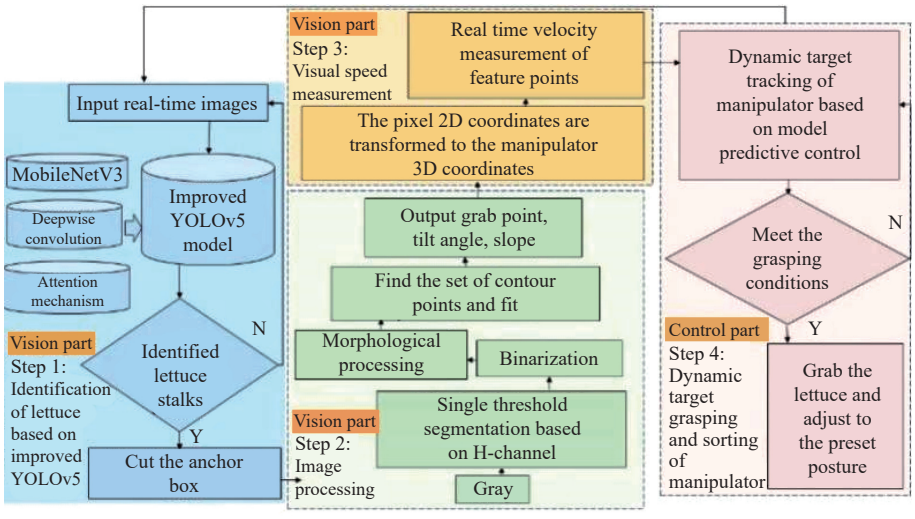


Figure 2 Overall model flow chart

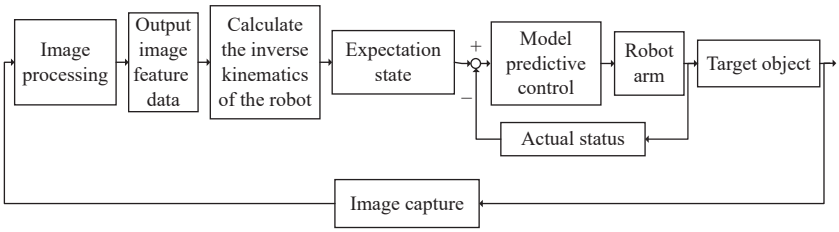


Figure 3 Closed loop control chart

3 Rapid identification and dynamic positioning method of lettuce stalk based on vision system

3.1 Lettuce image collection and data set production

The image acquisition device was the Shengyue USB HD wide-angle camera (60 frames, 6 mm focal length, USB interface, resolution of 1920×1080 pixels). The images were orthophotos of a single intact asparagus lettuce. To ensure the diversity of the samples, 794 RGB images were collected under different background and light conditions. To enrich the dataset, three methods of image flipping, brightness adjustment, and blur were used. Labeling software was used for manual annotation, and the data set was created according to the VOC format. The minimum outer rectangular box of the lettuce was regarded as the real box.

3.2 YOLOv5 network model

The YOLOv5 model uses the Pytorch framework to directly predict input images. Compared with the earlier YOLO algorithm, its detection speed and accuracy are significantly improved. The intelligent packaging pipeline studied in this study took lettuce as the research object. In order to achieve better real-time and lightweight detection, YOLOv5 was improved.

3.2.1 Backbone network

The backbone network of YOLOv5 was replaced by the backbone network of MobileNetV3[16,17]. MobileNetV3 used H-SWICH to replace the time-consuming SWICH activation function, shown in Equation (1), which makes the operation speed faster while maintaining accuracy and is more convenient for application in embedded devices.

$$h-swich(x)=x\cdot\frac{\mathrm{ReLU6}(x+3)}{6} \tag{1}$$

where, x represents the input size of the neuron.

The light attention module SE[18] (squeeze and excitation SE) was used for the color difference between the lettuce and assembly line; the structure is shown in Figure 4. The purpose of SE is to create different weights of different regions of the image, to obtain significant feature information[19,20]. This improves the attitude of the lettuce in the image and improves the recognition accuracy.

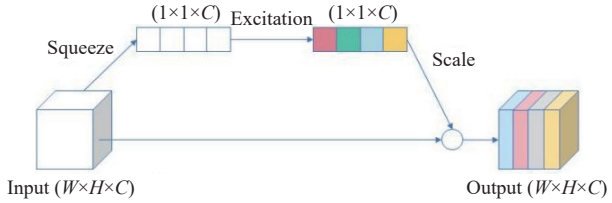


Figure 4 SE attention module

3.2.2 Depthwise separable convolution

Depthwise separable convolution can replace conventional convolution by combining depthwise and pointwise convolution for feature extraction[21,22]. The process of conventional and depth-separable convolution is shown in Figure 5.

Let the input size of the feature map be $D_K\times D_K$, M be the number of input channels, the size of the convolution kernel be $D_F\times D_F$, the number of convolution kernels be N, compared with conventional convolution, the computational burden of the depthwise separable convolution is reduced as Equation (2).

$$\frac{D_K\cdot D_K\cdot D_F\cdot D_F\cdot M+D_K\cdot D_K\cdot M\cdot N}{D_K\cdot D_K\cdot D_F\cdot D_F\cdot M\cdot N}=\frac{1}{N}+\frac{1}{D_F^2} \tag{2}$$

The computational efficiency of the depthwise separable convolution is considerably better than that of conventional convolution, thus, this study uses depthwise separable convolution instead of the convolution in YOLOv5s.

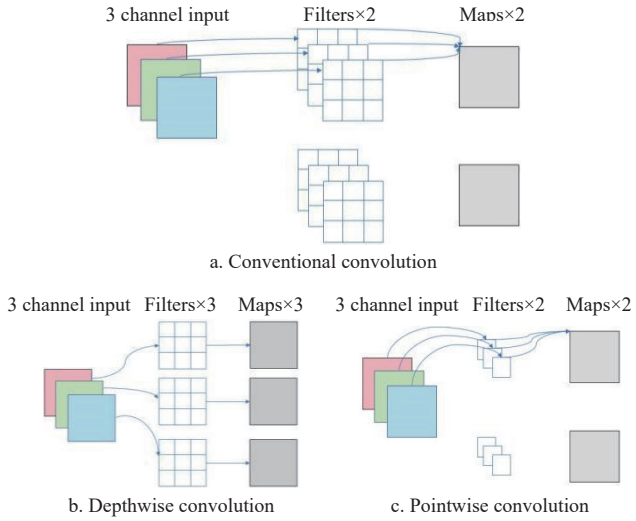


Figure 5 Process of conventional and depth-separable convolution

3.2.3 Loss function

YOLOv5 uses GIOU_LOSS as the loss function, which is composed of bounding box confidence loss L_{conf}, category loss L_{cla}, and coordinate loss L_{GIOU}[23,24]. The calculation of the loss function is shown in Equations (3)-(5).

$$L=L_{conf}+L_{cla}+L_{\mathrm{GIOU}} \tag{3}$$

$$\mathrm{IOU}=\frac{A\cap B}{A\cup B} \tag{4}$$

$$\mathrm{GIOU}=\mathrm{IOU}-\frac{|C-(A\cup B)|}{|C|} \tag{5}$$

where, A is the area of the real box; B is the area of the predicted box; C is the area of the minimum surrounding rectangle of A and B. When the real and prediction boxes exhibit inclusion relation or width and height alignment, the difference set is 0. Therefore, the loss function of CIOU is selected as the coordinate loss, as shown in Figure 6 and Equation (6).

$$\mathrm{CIOU}=\mathrm{IOU}-\frac{\rho^2(b,b^{gt})}{c^2}-\alpha v \tag{6}$$

where, c is the diagonal distance of the smallest enclosing rectangle containing both the prediction and real boxes; d is the distance of the center point of the real and prediction boxes; $\rho^2(b, b^{gt})$ is d

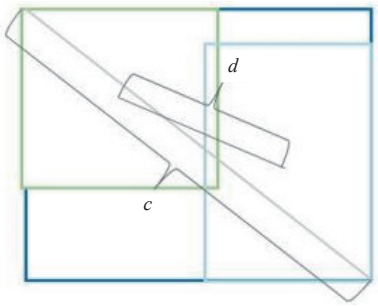


Figure 6 Schematic of CIOU loss function.

squared.

The CIOU[25] loss function can continuously approach the prediction box to the real box through iteration, ensuring that the aspect ratio of the prediction box and the real box is closer, and accelerating the regression convergence speed of the prediction box.

3.3 Lettuce characteristic point localization method

In the assembly line, to use the mechanical arm to grab lettuce, the key is to determine the optimal grab point and tilt angle of the stalk. Therefore, after identifying the lettuce, it is necessary to further determine the specific position and coordinates of the stalk. In this study, the lettuce is identified first, the specific pose of the lettuce stalk is recognized in the image, coordinate transformation is carried out, then the moving speed of feature points on the assembly line is dynamically estimated. The flow chart of the lettuce feature point location method is shown in Figure 7.

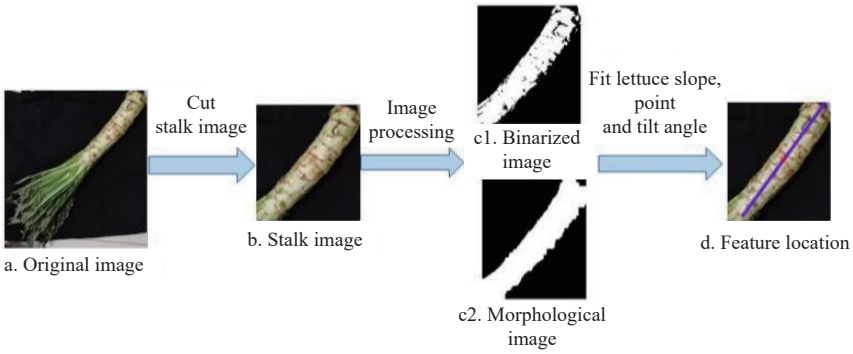


Figure 7 Flow chart of lettuce characteristic point localization method

3.3.1 Lettuce stalk cutting based on anchor box

After the lettuce image is input into the improved YOLOv5 model, the model identifies and selects the correct lettuce target box (anchor box), outputs the coordinates, length, and width of the anchor box, and cuts the lettuce stalk part on the lettuce image. The lettuce leaves are discarded in this step to avoid their influence on subsequent image processing, as shown in Figure 7b.

3.3.2 HSV color space

The color feature is a global feature, which describes the surface properties of the corresponding objects in the image or image region. It is less dependent on the size, direction, and perspective of the image itself, and has high robustness, thus, it is often used for image recognition.

The color parameters in HSV are hue (H), saturation (S), and lightness (V)[26], which are closer to human visual perception than the RGB color space, hence they are widely used to segment objects with specified colors. The histogram of the HSV color space component of the lettuce image was drawn, and it was found that the fluctuation of the H component was the most obvious compared with other components, as shown in Figure 8. Therefore, single threshold segmentation was conducted based on the H channel.

3.3.3 Threshold split processing

First, the color image acquired by the camera is transformed to the gray level. The threshold is then processed, changing the pixel value within the threshold to 0 and the pixel value outside the specified range to 255[27]. Single threshold segmentation has the characteristic of simple operation. In the lettuce assembly line, the lettuce and pipeline background color difference is obvious, which is suitable for the application of single-threshold segmentation. By analyzing the histogram of the lettuce HSV three-channel component and continuous testing, it was deduced that the best effect of the background removal was achieved when $105<H<125$.

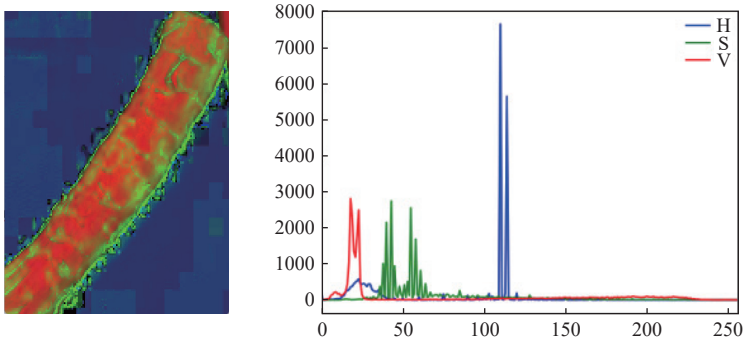


Figure 8 Lettuce stalk HSV transform map and HSV three-channel histogram

Owing to the difference in the H component between the stem and non-stem nodes of lettuce, the binary image of the lettuce stalk is disconnected, as shown in Figure 7c1. Therefore, it is necessary to obtain the ideal lettuce image through morphological processing.

3.3.4 Morphological processing

To connect the stem elements in the figure above, the noise can be eliminated by the morphological operation of corrosion and

expansion, the independent stem elements in the image can be segmented, connecting the stem. The image becomes smaller after the corrosion operation and can be used to remove isolated noisy pixels.

After the expansion operation, the image becomes larger, which can be used to connect the isolated pixels with close locations.

After several operation tests, it was found that the optimal operation of first expansion and then corrosion is carried out using structural elements, as shown in Figure 7c2.

3.3.5 Lettuce stalk identification and labeling

After morphological treatment, the minimum outer rectangular box is determined and the slope of the lettuce is calculated and marked with purple. The pixel coordinate system of the image is different from that of the manipulator, and coordinate transformation is needed to calculate the mirror tilt angle of the lettuce and the optimal grasping point, which are marked in red, as shown in Figure 7d.

3.4 Robot arm control system based on machine vision

3.4.1 Calculation of lettuce moving speed

The lettuce speed can be obtained by the ratio of the displacement (ΔS) and time generated within a certain time (ΔT), and the calculation is as follows:

$$v = \frac{\Delta S}{\Delta T} = \frac{s_1 - s_0}{t_1 - t_0} \tag{7}$$

In this study, the labeled optimal grab point of lettuce is regarded as the feature point, and the difference between the positions of the feature points of the five adjacent frames of lettuce is shown in Figure 9. By calculating the video frame rate, the frame rate obtained in this study is 28.13 fps.

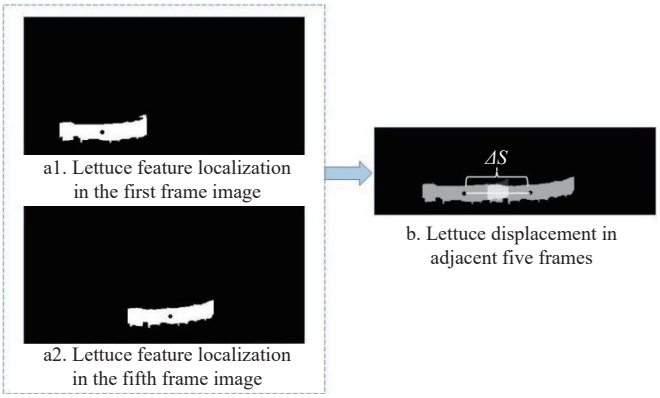


Figure 9 Lettuce characteristic point displacement

3.4.2 Lettuce visual velocimetry model

The camera used in this study is based on the pinhole camera model. The overall coordinate model is shown in Figure 10, and the relationship between the image's physical and pixel coordinate systems is shown in Figure 11. X_W, Y_W, and Z_W represent the world coordinate system, that is, the coordinates of the object in the real world; X_L, Y_L, and Z_L represent the coordinate system of the manipulator; X_C, Y_C, and Z_C represent the camera coordinate system; xOy represents the physical coordinate system of the image; and uOv represents the image pixel coordinate system. To achieve cutting image pixel coordinates to the mechanical arm coordinate system, the coordinate transformation is required to go through five steps: cutting image pixel coordinates into pixel coordinates; image pixel coordinates transformation to image physical coordinates; image physical coordinates transformation to camera coordinates; camera coordinates transformation to world coordinates; and world coordinate conversion to the mechanical arm coordinate.

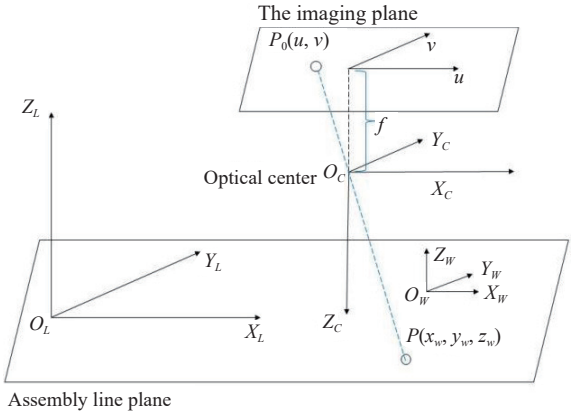


Figure 10 Overall coordinate model

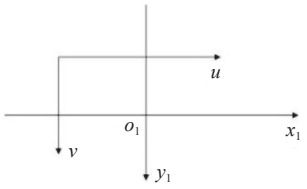


Figure 11 Relationship between image physical coordinate and image pixel coordinate

We assume that there exists a point $P_C(x_c, y_c, z_c)$ in the camera coordinate system, whose coordinates in the picture physical coordinate system are denoted as $P_1(x, y)$, and the coordinates in the picture pixel coordinate system are denoted as $P_0(u, v)$. First, the cutting image pixel coordinates are converted to the image pixel coordinates through the coordinate origin translation. Subsequently, the image pixel coordinates are converted to the camera coordinates. The relationship between the three coordinate points can be expressed as Equations (8) and (9).

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{x_C}{z_C} f \\ \frac{y_C}{z_C} f \end{bmatrix} \tag{8}$$

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{x}{\mathrm{d}x} + u_0 \\ \frac{y}{\mathrm{d}y} + v_0 \end{bmatrix} \tag{9}$$

where, f represents the focal length of camera imaging, mm; u_o and v_o, represent the image center in the u, v direction pixels; dx and dy denote the physical size of a single pixel in the x and $y$directions of the sensor, mm/pixel.

It was assumed that there is a point $P(X_W, Y_W, Z_W)$ in the world coordinate system, $P(X_C, Y_C, Z_C)$ in the camera coordinate system[27], and $P(X_L, Y_L, Z_L)$ in the manipulator coordinate system. To simplify the calculation, the world coordinate system is set to coincide with the manipulator coordinate system, and the origin is the fixed joint point of the manipulator, thus, the corresponding relationship can be expressed as Equation (10).

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_L \\ Y_L \\ Z_L \\ 1 \end{bmatrix} \tag{10}$$

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}, \text{ and } t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

where, R represents the rotation matrix of the world coordinate

system to the camera coordinate system, which is a 3×3 matrix; t represents the translation vector of the world coordinate system to the camera coordinate system, that is, the position of the camera photocenter in the world coordinate system.

It was assumed that there is a point P in the space whose coordinate in the world coordinate system is $P(x_W, y_W, z_w, 1)$, and the coordinate in the picture pixel plane coordinate system is $(u, v, 1)$ transformed by this model. The corresponding relationship between the two coordinates is as shown in Equation (11).

$$Z_C\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_o & 0 \\ 0 & f_y & v_o & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_W \\ y_W \\ z_W \\ 1 \end{bmatrix} \tag{11}$$

where, f_x and f_y represent the camera internal values, $f_x = \frac{f}{\mathrm{d}x}$, $f_y = \frac{f}{\mathrm{d}y}$.

4 Dynamic tracking and grasping of manipulator based on model predictive control

After obtaining the specific pose of lettuce, the manipulator controller drives the manipulator to complete path tracking, grasping, and pose adjustment. The following is the establishment and explanation of the overall velocity model and predictive control model.

4.1 Model building and MPC controller design

Model predictive control (MPC) is a strategy to achieve control effects through repeated optimization and feedback correction of the predictive model, which has good robustness and control effects [28,29]. Therefore, the predictive control method of the manipulator model is proposed, which regards the dynamic feature points of the assembly line lettuce as the solution and ensures the real-time dynamic tracking of the manipulator. The model predictive control principle used in this study is shown in Figure 12. Considering the fixed height of the manipulator grasping lettuce, the manipulator is simplified into a two-degree[30] of freedom model, and the real and model pictures are shown in Figure 13.

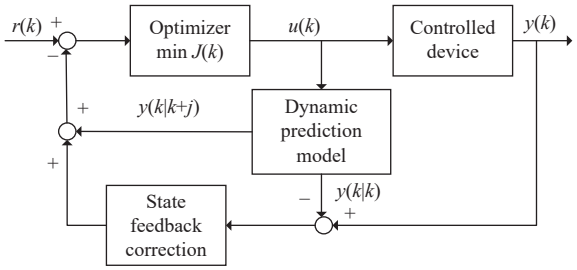
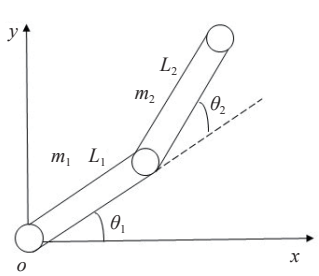


Figure 12 Model predictive control of manipulator

a. Physical picture of the robot arm



b. Two degrees of freedom manipulator model

Figure 13 Robot arm physical object and model.

where, m_1 and m_2 are the masses of the two links, respectively; and L_1, L_2 are the lengths of the two links, respectively. The Lagrange equation is used to establish the dynamic equation of the manipulator[31], as shown in Equation (12).

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} = \tau \tag{12}$$

$$M(q) = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix},\ C(q,\dot{q}) = -m_2L_1r_2\sin(q_2)\begin{bmatrix} \dot{q}_2 & \dot{q}_1+\dot{q}_2 \\ -\dot{q}_1 & 0 \end{bmatrix}$$

where, q is the angular displacement of the joint; $\dot{q}$ is the joint angular velocity; $\ddot{q}$ is the joint angular acceleration; $M(q)$ is the inertia matrix of the manipulator; $C(q,\dot{q})$ is Criolli centrifugation; and τ denotes the control torque vector. The dynamic variable is defined as $x = [q\ \dot{q}]^T$, the control vector as $u = \tau$, and the output vector as $y=q$, At time k, the output vector at time $k+j$ is estimated as $y(k|k+j)$.

After linearization and discretization, the current given input is used for prediction, and the trajectory is tracked by minimizing the cost function while satisfying the constraints. The least square method is used to calculate the distance between the actual and desired states, and the minimum distance is regarded as the cost function and satisfies the constraint conditions[31] to track the desired trajectory. where x_d is the reference trajectory state at time k. At time k and N steps are predicted backward, the control input is U_k, and the error column vector is E_k, as shown in Equation (13).

$$\begin{cases} J_{\min} = \frac{1}{2}U_k^T(G^TQG+R)U_k + (Fx_k + Th - Xd)^TQGU_k \\ E_k = Fx_k + GU_k + Th - X_d \\ F = \begin{bmatrix} A_C^1 \\ \vdots \\ A_C^N \end{bmatrix},\ T = \begin{bmatrix} I \\ A+I \\ \vdots \\ A^{N-1}+\ldots+A+I \end{bmatrix}, \\ G = \begin{bmatrix} B_C^1 & 0 & \ldots & 0 \\ A_C^1B_C & B_C^1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ A_C^{N-1}B_C & A_C^{N-2}B_C & \ldots & B_C^1 \end{bmatrix} \end{cases} \tag{13}$$

where, Q is the control weight matrix; R is the input state weight matrix. The constraint conditions are shown in Equation (14).

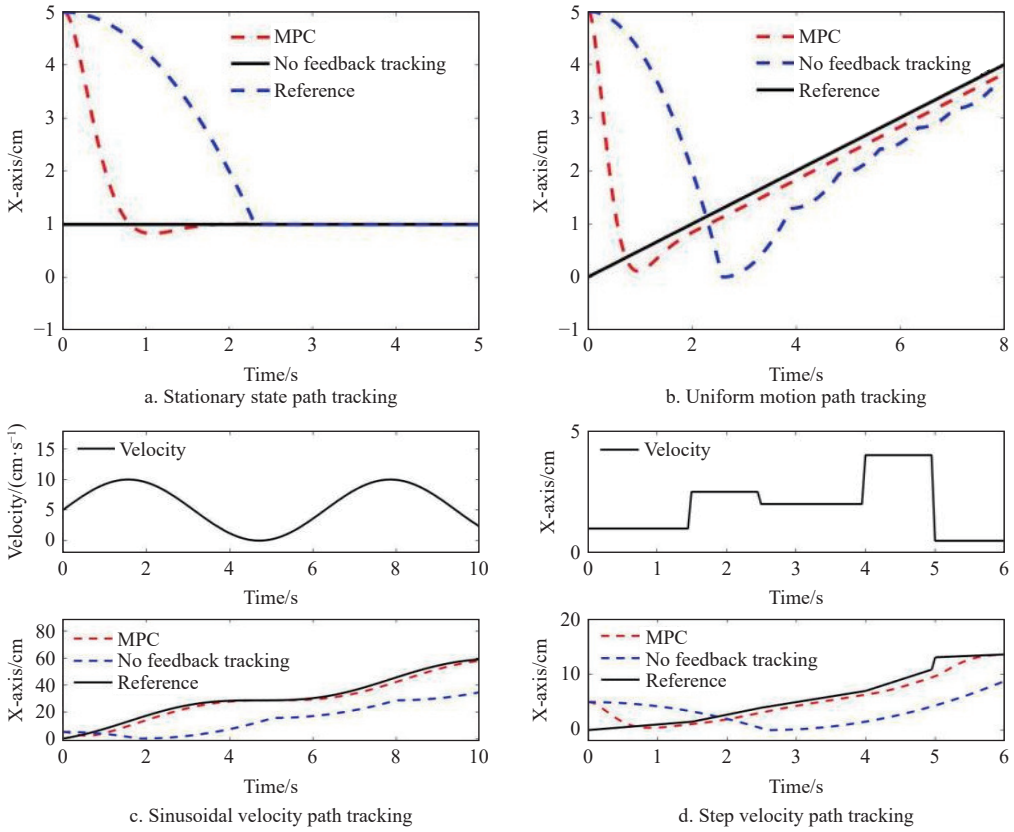
$$\begin{cases} X_{\min} \leqslant Fx_k + GU_k + Th - X_d \leqslant X_{\max} \\ U_{\min} \leqslant U_k \leqslant U_{\max} \end{cases} \tag{14}$$

In the control law of predictive control, the control input of the first step is considered. The current position of the manipulator claw is calculated according to the expression of the manipulator system. The dynamic model is established and the control law is finally obtained until the system reaches the desired state.

The translation vector V_{trans} is the maximum distance δ between the manipulator and lettuce when grasping. When $\|V_{\text{trans}}\|<\delta$ and $V_{\text{rot}}=0$, the controller controls the manipulator to complete the grasping and pose adjustment.

4.2 Simulation and verification

To verify the control effect of this method, a 2-DOF manipulator (Figure 14b) is taken as the control object to conduct data simulation. The parameters of the manipulator are L_1=L_2= 20 cm, m_1=m_2= 0.5 kg, g=9.8 N/kg, and the sampling period T=0.35 s. The state constraint of the manipulator is $\begin{bmatrix} -\frac{\pi}{3} \\ -\frac{\pi}{3} \\ -10\ \text{cm/s} \\ -10\ \text{cm/s} \end{bmatrix} \leqslant x \leqslant \begin{bmatrix} \frac{\pi}{3} \\ \frac{\pi}{3} \\ 10\ \text{cm/s} \\ 10\ \text{cm/s} \end{bmatrix}$



a. Stationary state path tracking
b. Uniform motion path tracking
c. Sinusoidal velocity path tracking
d. Step velocity path tracking

Figure 14 Trajectory tracking of the conveyor belt at different speeds

According to the different running speeds of the pipeline conveyor belt, the static state, constant speed, sinusoidal speed, and step speed path tracking are obtained.

Assuming that the initial position of the manipulator is at x= 5 cm, the characteristic point of the lettuce moves from x=0 in the positive direction of the x-axis, and the coordinates of the manipulator and characteristic point of the lettuce on the y-axis remain unchanged. The simulation results are shown in Figure 14.

The above results can be obtained under conditions satisfying the constraints and for different trajectories, the mechanical arm control under the MPC can meet the requirements in a relatively short tracking time. Consequently, the desired trajectory error is less than δ, and overshoot and oscillation do not occur. This ensures the real-time accuracy of the dynamic trajectory tracking. Error-free tracking can be realized for the desired trajectory with a large velocity variation. The average descending time of the manipulator is 0.5 s, the average rotation time of the manipulator claw is 0.45 s, and the average delay time of the visual part is 0.015 s.

5 Experiment and analysis

5.1 Experimental results and analysis of vision

To prove that all the improved parts of the improved network are better compared with the original network, this study conducts comparative tests based on average precision (mAP), recall (R), and average detection time. Precision (P) refers to the proportion of the retrieved information that is of interest to the user. R refers to the proportion of the information of interest that is detected. AP is the area enclosed by the curve and coordinate axes, and the mAP of each category is the average AP value. The calculation is shown in Equations (15)-(17).

$$P = \frac{TP}{TP + FP} \tag{15}$$

$$R = \frac{TP}{TP + FN} \tag{16}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{17}$$

where, TP is the correct number of positive samples predicted; FN is the wrong number of negative samples predicted; FP is the wrong number of positive samples predicted; TN is the number of negative samples predicted; $p(r)$ is the recall rate under different precision rates r; AP_i is the detection accuracy of class i; and N is the total number of categories. The specific experimental results are listed in Table 1.

Table 1 Comparison of improvement results

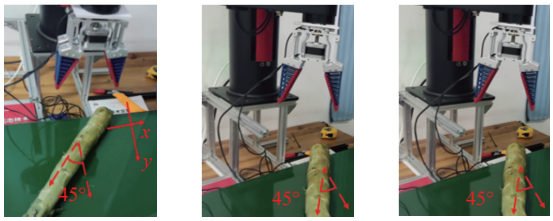
Object detection network	mAP	Recall	Mean detection time
Network before improvement	0.98	0.95	0.021 s
Improved network	0.98	0.95	0.014 s

Compared with YOLOv5 before the improvement, the average detection time of the single frame image of the improved network is 0.14 s, which is 50% faster. The average accuracy rate was 98% and the recall rate was 95%, which remained unchanged. Therefore, the improved network can not only ensure high accuracy and a high recall rate but also significantly reduce the detection time and ensure the real-time performance and high efficiency of the visual part.

5.2 Experimental results and analysis of control

To verify the effectiveness of the proposed assembly line intelligent sorting method based on visual positioning and manipulator predictive control, the manipulator's end claw is located in the initial position. The lettuce moves on the conveyor belt, and enters the camera pixel coordinates, and the vision system obtains the speed and position information of the lettuce. Based on the

moving trajectory of the lettuce feature points, the best-expected grasping point of the manipulator claw was obtained and the tracking trajectory was planned. The control system solved the inverse kinematics solution to control the manipulator to track the lettuce feature points and grasp and adjust the posture. The experimental process is shown in Figure 15.



a. Initial position of the manipulator b. Claw rotating c. Manipulator descending

d. Adjusting the lettuce position e. Manipulator descending and closing the claw

Figure 15 Experimental process diagram of the control part

Presently, the transportation speed of the packaging assembly line is between 8 and 50 mm/s, the speeds are 0, 8, 16, 33, and 50 mm/s, respectively. The experiment was divided into five groups, and four expected grasping points were randomly selected for each group to conduct several experiments and record the errors. The results are listed in Tables 2 to 6.

Table 2 Measurement results when the conveyor belt is stationary

Desired grasp point coordinates and inclination angle	Experiment times	Mean axial error/cm	Mean radial error/cm	Adjusted average tilt angle error/(°)
(150, 0) 45°	50	0.86	0.57	5.0
(307, 19) 120°	50	−1.45	−0.56	4.0
(276, 29) 230°	50	1.74	0.43	4.0
(311, 30) 315°	50	1.20	0.46	4.0
Average error		0.59	0.22	4.2

Table 3 Measurement results at a conveyor speed of 8 mm/s

Desired grasp point coordinates and inclination angle	Experiment times	Mean axial error/cm	Mean radial error/cm	Adjusted average tilt angle error/(°)
(150, 0) 45°	50	−0.34	1.10	3.0
(307, 19) 120°	50	−1.56	1.20	4.0
(276, 29) 230°	50	−1.36	0.49	3.0
(311, 30) 315°	50	1.40	0.24	4.0
Average error		−0.46	0.76	3.5

Table 4 Measurement results at a conveyor speed of 16 mm/s

Desired grasp point coordinates and inclination angle	Experiment times	Mean axial error/cm	Mean radial error/cm	Adjusted average tilt angle error/(°)
(150, 0) 45°	50	1.62	0.98	3.00
(307, 19) 120°	50	0.77	2.25	3.00
(276, 29) 230°	50	−1.55	2.48	4.00
(311, 30) 315°	50	−1.30	0.92	5.00
Average error		1.66		3.75

Table 5 Measurement results at a conveyor speed of 33 mm/s

Desired grasp point coordinates and inclination angle	Experiment times	Mean axial error/cm	Mean radial error/cm	Adjusted average tilt angle error/(°)
(150, 0) 45°	50	1.20	1.50	3.0
(307, 19) 120°	50	0.83	−3.20	5.0
(276, 29) 230°	50	2.48	1.36	5.0
(311, 30) 315°	50	1.80	0.57	5.0
Average error			0.06	4.5

Table 6 Measurement results at a conveyor speed of 50 mm/s

Desired grasp point coordinates and inclination angle	Experiment times	Mean axial error (cm)	Mean radial error (cm)	Adjusted average tilt angle error (degree)
(150，0)45°	50	-0.49	3.89	4
(307，19)120°	50	2.60	1.35	2
(276，29)230°	50	1.43	3.10	3
(311，30)315°	50	1.42	1.33	3
Average error			2.42	3

The results show that the average axial error was 0.71 cm, the average radial error was 1.02 cm, and the adjusted average angle error was 3.79° of the different conveyor belt speeds within the scope of the belt speed. The mechanical arm can grab the lettuce steadily and fast at different speeds, which verifies the efficiency of the mechanical arm control method based on model predictive control.

6 Conclusions

1) Because the current reversing device of the packaging machine line was unsuitable for crops, and artificial sorting efficiency was low, the first task in this study was to design a vision system with higher accuracy and real-time. In this study, the assembly line lettuce recognition model was established based on the improved YOLOv5. Using MobileNetV3 as the backbone feature extraction network, the depthwise separable convolution and SE attention mechanism were combined to optimize the extraction effect, and the loss function of the original model was modified to establish an assembly line lettuce recognition model. The optimal grasping point and tilt angle of the lettuce stalk were obtained by image morphology processing. Finally, an assembly line coordinate system was established for coordinate transformation and visual speed measurement. The collected lettuce dataset was used for training and application. The average accuracy rate was 98%, the recall rate was 95%, and the average detection time of a single frame image was 0.014 s, which was improved by 50%. While ensuring the recall rate and detection accuracy, the detection speed was improved to meet the requirements of the assembly line deployment on the embedded side.

2) This study proposed a manipulator control method for a two-link robot system. Guided by the localization of the target feature points and dynamic velocity measurement of the vision system, a kinetics model of the manipulator was constructed. In the time domain, the model predictive control was used for continuous prediction and feedback optimization to track the desired trajectory and complete the dynamic path tracking of the lettuce stalk feature points. When the translation and rotation vectors of the manipulator and lettuce feature points meet the preset requirements, the controller controls the manipulator to complete the grasp and posture adjustment of the lettuce. When lettuce is stacked, multi-

machine coordination can be used for attitude adjustment. At the same time, the more challenging image processing and positioning of lettuce will also be considered in the future. This experiment can realize stable, fast, and accurate lettuce grasping and sorting. This method can significantly improve the production capacity of the assembly line, with high productivity and intelligence, and can liberate workers from complicated manual labor.

Acknowledgements

This work was financially supported in part by the National Natural Science Foundation of China (Grant No. 32071896), Jiangsu Province Science and Technology Plan Special Fund (Key Research and Development Plan of Modern Agriculture) Project (Grant No. BE2022363), Modern Agricultural Machinery Equipment and Technology Demonstration and Promotion Project of Jiangsu Province (Grant No. NJ2021-37), National Foreign Experts Program of China (Grant No. G2021145010L), and the Science and Technology Project of Suzhou City (Grant No. SNG2020039).

[References]

[1] Kim M J, Moon Youyoun, Tou J C, Mou Beiquan, Waterland N L. Nutritional value, bioactive compounds and health benefits of lettuce (*Lactuca sativa* L.). Journal of Food Composition and Analysis, 2016; 49: 19–34.

[2] Qin S. Robot tangram assembly line based on vision. Journal of Physics:Conference Series, 2022; 2229: 012017.

[3] Ayyad A, Halwani M, Swart D, Muthusamy R, Almaskari F, Zweiri Y. Neuromorphic vision based control for the precise positioning of robotic drilling systems. Robotics and Computer-Integrated Manufacuring, 2023; 79: 102419.

[4] Ruan D X, Zhang W T, Qian D. Feature-based autonomous target recognition and grasping of industrial robots. Personal and Ubiquitous Computing, 2023; 27: 1355–1367.

[5] Do Q, Chang W, Chen L. Dynamic workpiece modeling with robotic pick-place based on stereo vision scanning using fast point-feature histogram algorithm. App.Sci. 2021; 11(23): 11522. doi: 10.3390/app112311522.

[6] Lim J, Yang P, Yuanda P, Bakkara R V, Sinaga M, Siagian H, et al. Automated pneumatic vacuum suction robotic arm with computer vision. IOP Conference Series: Materials Science and Engineering. 2020; 801(1): 012134. doi:c 10.1088/1757-899X/801/1/012134.

[7] Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. Electron Markets, 2021; 31(3): 685–695.

[8] Zhao Z Q, Zheng P, Xu S T, Wu X D. Object detection with deep learning: A review. IEEE Transactions on Neural Networks and Learning System, 2019; 30(11): 3212–3232.

[9] Liu C X, Sui H G, Wang J X, Ni Z X, Ge L. Real-time ground-level building damage detection based on lightweight and accurate YOLOv5 using terrestrial images. Remote Sensing, 2022; 14(12): 2763–2763.

[10] Zaidi S S A, Ansari M S, Aslam A, Kanwal N, Asghar M, Lee B. A survey of modern deep learning based object detection models. Digital Signal Processing, 2022; 126: 103514.

[11] Shamshul Haq S, Lenine D, Lalitha S V N L. Performance enhancement of UPQC using Takagi–Sugeno fuzzy logic controller. International Journal of Fuzzy Systems, 2021; 23(6): 1765–1774.

[12] Naik B B, Mehta A J. Sliding mode controller with modified sliding function for DC-DC Buck Converter. ISA Transactions, 2017; 70: 279–287.

[13] Mannucci T, van Kampen E-J, de Visser C, Chu Q P. Safe exploration algorithms for reinforcement learning controllers. IEEE Transactions on Neural Neworks and Learning Systems, 2018; 29(4): 1069–1081.

[14] Jahanshahi H, Sajjadi Samaneh S, Bekiros S, Aly A A. On the development of variable-order fractional hyperchaotic economic system with a nonlinear model predictive controller. Chaos, Solitons and Fractals, 2021; 144: 110698.

[15] Pankert J, Hutter M. Perceptive model predictive control for continuous mobile manipulation. IEEE Robotics and Automation Letters, 2020; 5(4): 6177–6184.

[16] Jia W K, Wei J M, Zhang Q, Pan N N, Niu Y, Yin X, et al. Accurate segmentation of green fruit based on optimized mask RCNN application in complex orchard. Front Plant Sci. 2022; 13: 955256. doi:10.3389/FPLS.2022.955256.

[17] He C C, Li X B, Liu Y S, Yang B Y, Wu Z W, Tan S, et al. Combining multicolor fluorescence imaging with multispectral reflectance imaging for rapid citrus Huanglongbing detection based on lightweight convolutional neural network using a handheld device. Computers and Electronics in Agriculture, 2022; 194: 106808.

[18] de Santana Correia A, Colombini E L. Attention, please! A survey of neural attention models in deep learning. Artificial Intelligence Review, 2022; 55: 6037–6124.

[19] Chen Z Y, Wu R H, Lin Y Y, Li C Y, Chen S Y, Yuan Z N, et al. Plant disease recognition model based on improved YOLOv5. Agronomy, 2022; 12(2): 365.

[20] Yan B, Fan P, Lei X Y, Liu Z J, Yang F Z. A real-time apple targets detection method for picking robot based on improved YOLOv5. Remote Sensing, 2021; 13(9): 1619.

[21] Liu Y T, Zhao J J, Luo Q Y, Shen C T, Wang R, Ding X H. Automated classification of cervical Lymph-Node-Level from ultrasound using depthwise separable convolutional swin transformer. Computers in Biology and Medicine, 2022; 148: 105821.

[22] Asghar M Z, Albogamy F R, Al-Rakhami M S, Asghar J, Rahmat M K, Alam M M, et al. Facial mask detection using depthwise separable convolutional neural network model during COVID-19 pandemic. Front Public Health, 2022; 10: 855254.

[23] Tian Y J, Su D, Lauria S, Liu X. Recent advances on loss functions in deep learning for computer vision. Neurocomputing, 2022; 497: 129–158.

[24] Yao J, Qi J M, Zhang J, Shao H M, Yang J, Li X. A real-time detection algorithm for kiwifruit defects based on YOLOv5. Electronics, 2021; 10(14): 1711.

[25] Dong X D, Yan S, Duan C Q. A lightweight vehicles detection network model based on YOLOv5. Engineering Applications of Artificial Intelligence, 2022; 113: 104914.

[26] Ansari E, Akhtar M N, Abdullah M N, Othman W A F W, Bakar E A, Hawary A F, et al. Image processing of UAV imagery for river feature recognition of Kerian River, Malaysia. Sustainability, 2021; 13(17): 9568.

[27] Song Q S, Li S B, Bai Q, Yang J, Zhang X X, Li Z A, et al. Object detection method for grasping robot based on improved YOLOv5. Micromachines, 2021; 12(11): 1273.

[28] Grandia R, Farshidian F, Dosovitskiy A, Ranftl R, Hutter M. Frequency-aware model predictive control. IEEE Robotics and Automation Letters, 2019; 4(2): 1517–1524.

[29] Pinheiro Tarcisio Carlos F, Silveira Antonio S. Constrained discrete model predictive control of an arm - manipulator using laguerre function. Optimal Control Applications and Methods, 2020; 42(1): 160–179.

[30] Roozbahani H, Alizadeh M, Ahomäki A, Handroos H. Coordinate-based control for a materials handling equipment utilizing real-time simulation. Autom Constr. 2021; 122: 103483. doi:10.1016/j.autcon.2020.103483.

[31] Buizza Avanzini G, Zanchettin A M, Rocco P. Constrained model predictive control for mobile robotic manipulators. Robotica, 2018; 36(1): 19–38.