International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Adaptive Image Segmentation with YOLO for Visual Positioning System for Rhinestone Machine

Linling Jia [1], Anton Louise De Ocampo*[2]

Submitted:10/03/2024 **Revised**: 25/04/2024 **Accepted**: 02/05/2024

Abstract: In order to improve the visual positioning effect of rhinestone for greeting card paste rhinestone machine, shorten positioning time, and reduce positioning errors, this study conducted a YOLO based visual positioning method for greeting card paste rhinestone machine. This method first completes the design of the rhinestone image acquisition structure, performs de-noise processing on the collected image, and then repairs the ROI area of the rhinestone image based on target neighborhood interpolation and region median filling. It uses seed region growth technology to complete image adaptive segmentation processing. Finally, based on the YOLO3 network, a visual positioning model for the greeting card paste rhinestone machine is constructed to achieve the recognition and positioning function of the greeting card paste rhinestone machine. The experimental results show that the proposed method has a positioning time of less than 1.31 ms and a positioning error of less than 0.41mm, which is superior to the comparison method and the positioning effect is better.

Keywords: *YOLO algorithm, greeting card paste rhinestone machine, rhinestone, visual positioning*

1. Introduction

At present, the production of greeting cards has been mechanized, and rhinestone pasting, as a key link to improve the beauty and added value of greeting cards, its accuracy and efficiency directly affect the quality and market competitiveness of the final product [1]. Traditional greeting card drilling machines usually use mechanical positioning or manual teaching to position rhinestone. However, this method has problems such as low positioning accuracy and low efficiency, which cannot meet the needs of modern greeting card production enterprises [2]. In recent years, the development of computer vision technology has provided new solutions for the positioning of greeting card stickers, drilling rigs, and rhinestone. Through visual positioning technology, precise recognition and positioning of rhinestone can be achieved, thereby improving the accuracy and efficiency of drilling. Therefore, Therefore, it is of great practical significance and application value to research and develop a vision-based rhinestone localization method for greeting card paste rhinestone machines.

In recent years, numerous scholars have conducted research on this aspect and have achieved certain research results. Zhou Qingyi et al. [3] plan to study the workpiece positioning algorithm of a five axis dispensing machine based on computer vision. This method utilizes computer vision and non-linear optimization ideas to solve the feature points of the measured part position, reducing positioning accuracy and actual deviation. However, this method is affected by various factors in practical applications, such as lighting conditions, image noise, etc.,which leads to unstable positioning accuracy and makes it difficult to meet the demand for high-precision positioning. For example, Wang Rui et al. [4] studied the rotational posture of a five axis dispensing mechanism based on computer vision. This method first proposes a three-dimensional coordinate measurement system based on a CCD camera. Then, based on this, the geometric errors of the machine tool are modeled to obtain the transformation relationships between various coordinate systems, and then the initial solutions of the machine tool parameters are obtained; Finally, the L-M algorithm is used to optimize the machine parameters and achieve dispensing control of the machine. However, this method requires high accuracy in image processing and may be affected by factors such as calibration board production accuracy and camera calibration errors in practical applications, resulting in insufficient positioning accuracy and stability. XueHaihui et al. [5] studied the external characteristics of graphene modules and achieved their positioning on the dispensing machine by extracting their geometric boundaries and features, as well as calculating the X, Y, and rotation parameters of arbitrarily arranged graphene modules. However, this method may be affected by factors such as the randomness of module placement, image quality, and algorithm processing speed in practical applications, making it difficult to achieve ideal positioning efficiency and accuracy. According to the requirements of the dispensing process, Xu Jianming et al. [6] studied the control system of a five axis dispensing machine using industrial computers as hardware platforms and CODESYS as software development platforms. The research content includes: implementing one master and multiple slaves

[1] *Batangas State University, The National Engineering University，4200，Batangas City,Philippines.*
ORCID ID : 0009-0005-8043-1496
[2*] *Batangas State University, The National Engineering University，4200，Batangas City,Philippines.*
ORCID ID : 0000-0002-6280-6259
** Corresponding Author Email: antonlouise.deocampo@ieee.org*

control mode using EtherCAT bus; Implement the reading and analysis of the G program using the SM3-CNC library; Using DH parameter method for kinematics modeling and designing forward and backward functional blocks; Using methods such as segmented 3rd Hermite interpolation and 3rd spline interpolation to achieve PVT interpolation. Although this control system improves the automation level of dispensing, due to its main focus on the design and implementation of the control system, the accuracy and efficiency improvement of rhinestone positioning are limited.

In response to the shortcomings in the above aspects, this study conducted a visual positioning method for greeting card stickers drilling rig rhinestone based on YOLO. I hope that through this research, the automation level of greeting card production can be improved, labor costs can be reduced, product quality and efficiency can be improved, and useful reference and inspiration can be provided for the automation production of other similar products, promoting the transformation and upgrading of the entire manufacturing industry.

2. Automatic image acquisition and preprocessing design

2.1. The rhinestone image acquisition design

Image acquisition is a crucial first step in the visual positioning method of greeting card stick rig. Effective image acquisition design ensures high quality, representative diamond image data to provide a solid foundation for subsequent processing and analysis [7,8]. The rhinestone image acquisition architecture of this design is shown in Figure 1.

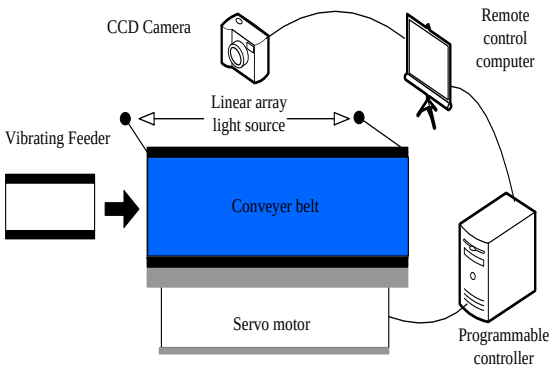


Fig. 1. Architecture of rhinestone image acquisition

As shown in Figure 1, the hardware device in the acquisition structure includes a CCD camera, a conveyor belt, a light source device, an image processing system, a control device, and the like. When collected, the camera is placed directly on the conveyor belt to obtain the image in a fixed frame, and then the color of the conveyor belt is set to pure blue to distinguish the background from the rhinestone object. In conclusion, the automatic extraction of greeting card rhinestone images was completed.

2.2. Denoising processing of rhinestone images

After completing the collection of rhinestone images, the initial image may be affected by various factors, such as equipment noise and uneven environmental lighting, which can affect the subsequent accuracy of rhinestone recognition and positioning [9]. Therefore, this experiment will perform de-noise on it to eliminate or reduce the impact of these noises on image quality. De-noise processing is an indispensable part of digital image processing, aimed at improving the signal-to-noise ratio of images. Although common de-noise methods can reduce noise in images to a certain extent, they also have some shortcomings. Median filtering performs well in handling salt and pepper noise, but may have limited effectiveness in removing other types of noise; Gaussian filtering may cause blurred edge information in the processed image; Bilateral filtering preserves edges while denoising, but may encounter texture replication artifacts when dealing with complex textures [10-12]. These issues may all affect the accuracy of subsequent rhinestone identification and positioning. To overcome these shortcomings, an improved guided image filtering algorithm was proposed in this experiment.

Select a local filter window w_k with radius r and center pixel value k. The index parameter that can measure the structural difference between guide image G and target image I in this window is called the normalized cross-relation η. The calculation formula is as shown in the formula (1):

$$\eta = \left| \frac{\operatorname{cov}(G_k, I_k)}{\sqrt{\operatorname{var}(G_k)\operatorname{var}(I_k)}} \right| \tag{1}$$

Where $\operatorname{var}(G_k)$, $\operatorname{var}(I_k)$ are the variance of G_k and I_k in the local window w_k, respectively; $\operatorname{cov}(G_k, I_k)$ is their covariance in the filter window. The values of the normalized correlation number η generally range from -1 to 1, in the absolute measure to measure the difference of the image structure, the normalized cross relation absolute value closer to 0, the difference between G and I in the image structure of the larger, if the absolute value is close to 1, then that G and I in the image structure similarity is greater, has an approximate linear relationship [13].

The determination of image structural differences by the normalized cross-correlation coefficient η is easily influenced by the local filtering window radius r. If the window half setting is relatively small, the accuracy of detecting structural relationships at the edges of the image

is higher. However, there will be errors in the detection results in flat and textured areas. The reason is that noise will affect the calculation of covariance and variance in the region, resulting in incorrect calculation results. In the case of setting a larger window radius, the situation is the opposite, with errors in the detection of the edge part and higher accuracy in the detection of other parts [14,15].

In order to more accurately describe the structure difference between images, this paper calculates smaller window radius and larger window radius under the normalized correlations, and multiply the coefficient of two different radius, with the product ρ as the local filter window to detect the structure difference of the final index parameters, the calculation formula is as follows:

$$\rho = \eta_1 \eta_2 \tag{2}$$

In the formula, η_1, η_2 represent the normalized cross-correlation coefficients of images with smaller and larger window radii, respectively [16]. The indicator proposed in this article has the advantages of both the small radius and large radius normalized cross-correlation coefficients in detecting and measuring image structural relationships compared to the original normalized cross-correlation coefficient. It has strong robustness against noise interference and is more accurate in detecting structural differences between images [17,18]. Introducing the image structure relationship metric proposed in this article into the original guided filtering algorithm, the relationship between the guided image G , the target image I to be processed, and its filtering output q remains linear. The specific formula is as follows:

$$q_i = a_k\left(\rho G_i + (1-\rho) I_i\right) + b_k, \forall i \in w_k \tag{3}$$

Where q_i and G_i represent the pixel values of the filter output image and the guide image, respectively; a_k and b_k are both linear coefficients in the filter window w_k. Minimize the formula (3) as an objective function, such as the formula (4):

$$E(a_k, b_k) = \sum_{i \in w_k}\left(\left(a_k\left(\rho G_i + (1-\rho) I_i\right) + b_k - I_i\right)^2 + \varepsilon a_k^2\right) \tag{4}$$

Where, ε stands for the regularization parameter. Next, evaluate a_k and b_k respectively, and make the derivatives zero. The formula (5) and (6):

$$a_k = \frac{\rho \operatorname{cov}(G_k, I_k) + (1-\rho)\operatorname{var}(I_k)}{\rho^2 \operatorname{var}(G_k) + 2\rho(1-\rho)\operatorname{cov}(G_k, I_k) + (1-\rho)^2 \operatorname{var}(I_k) + \varepsilon} \tag{5}$$

$$b_k = \bar{I}_k - a_k\left(\rho G_k + (1-\rho)\bar{I}_k\right) \tag{6}$$

Where $\bar{I}_k$ represents the average of the pixel values in the filtered window of the treated image I [19] .

The algorithm proposed in this paper will have the problem of filter window superposition, that is, a pixel will appear in multiple filter Windows and will be repeatedly calculated, which makes the calculated filter output have multiple results. In this paper, all the filter outputs in a pixel neighborhood are averaged, and the average value of the pixel is set as the final filter output value. The specific calculation is shown in the formula (7):

$$q_i = u_i\left(G_i - I_i\right) + \bar{a}_i I_i + \bar{b}_i \tag{7}$$

Where u_i is the average of $a_k \rho_i$ in all windows containing pixels i; $\bar{a}_i, \bar{b}_i$ are the average of a_k and b_k in all windows containing pixels i respectively; the three means are calculated as shown in formulas (8) - (10):

$$u_i = \frac{1}{|w|}\sum_{k \in w_i} a_k \rho \tag{8}$$

$$\bar{a}_i = \frac{1}{|w|}\sum_{k \in w_i} a_k \tag{9}$$

$$\bar{b}_i = \frac{1}{|w|}\sum_{k \in w_i} b_k \tag{10}$$

In conclusion, the de-noise treatment of rhinestone image was completed and designed for subsequent rhinestone identification and positioning.

3. Crude segmentation design of ROI area

The rhinestone has similar optical characteristics to diamond, which can replace diamond jewelry. Because the rhinestone has extremely highly transmittance, under the light source, it will produce light refraction in different directions, thus forming a special reflective effect. The refraction of the rhinestone, you can see in the photo, almost every rhinestone target has a piece of white. This region is a singular region in image processing, which will be detrimental to image segmentation and recognition. After the de-noise of the rhinestone image, this area is combined to repair the area of interest in the image. Then, by the coarse segmentation method, quickly locate to the possible rhinestone position.

3.1. Joint repair of ROI area

Image restoration refers to the replacement of singular

regions or objects in an image based on existing image information, such as gray scale, in order to restore the damaged parts. In order to improve the correction results, this article combines target neighborhood interpolation and region median filling the method to complete image processing. The repair steps are as follows:

Firstly, the rhinestone image is transformed into a height field surface, and the area to be repaired for each rhinestone target is identified through geomorphic features. Extract the height values of each point in the repair area in order from the outside to the inside, and fit the height field surface based on the height values of each point in the rhinestone target area [20].

Then, perform surface interpolation repair and fill the refractive area of the rhinestone with the average height. The formula is as follows:

$$h = \frac{\sum_{i=1}^{n} h_i D_i^{-1}}{q_i} \tag{11}$$

Where h is the height information in the neighborhood of the repair point; D is the distance from a point in the neighborhood to the repair point.

Finally, the interpolation of the surface and the mean filling are combined to repair the following formula:

$$H = \frac{\bar{h} \cdot h_0}{\bar{h} + h_0} \tag{12}$$

Where $\bar{h}$ refers to the mean and h_0 is the value after neighborhood interpolation. Moreover, since the height values of all the pixels in the drill repair area are left and right, the height values of all of the original heights should be converted to $h = 255 - h'$ to make the formula more applicable to the drill image.

In conclusion, the ROI area repair of the rhinestone image was completed to lay the foundation for further processing of the image.

3.2. Adaptive segmentation design of rhinestone images

On the basis of repairing the ROI area of the rhinestone image, the area is adaptively segmented. Image segmentation is the division of the image in order to ensure the consistency of each area. The purpose of image segmentation is to distinguish the background from the target, so that it is easier to understand the image. Seed region growth technique was chosen for image segmentation. For the rhinestone image, the histogram is fitted with multiple Gaussian functions with the following formula:

$$f(x) = H\sum_{i=1}^{n} a_i e - \frac{(x-b_i)^2}{c_i^2}, a > 0 \tag{13}$$

In the formula, n is the number of fits, a is the peak, b is the mean (center position), and c is the standard deviation. After completing image fitting, combine the gray scale information of the histogram to automatically select the initial seed points for the rhinestone image. Firstly, determine the gray scale range of the rhinestone. Due to the fact that after image restoration, the gray scale value of the rhinestone's light refraction area is similar to the gray scale level of the rhinestone itself, the range of the target gray scale level can be determined based on the average gray scale value of the repaired light refraction area. Then map the range of the histogram to the fitted curve and obtain the maximum points of the fitted curve within that range. And the gray scale of this maximum point is the gray scale of the initial seed of the target. After determining the gray scale level of the initial seed, we can continue to find the reasonable position of the initial seed. The edge information of the rhinestone target can be obtained from the geomorphic features in the height field. So use the center of gravity of the region as the initial point for retrieval. In the process of searching from the inside out, if the gray scale value of a pixel meets the gray scale requirements of the initial seed, and the gray scale levels of each point in the neighborhood of the pixel are within the same gray scale range as the initial search point, then that point is the initial seed point selected. The selection process is shown in Figure 2.

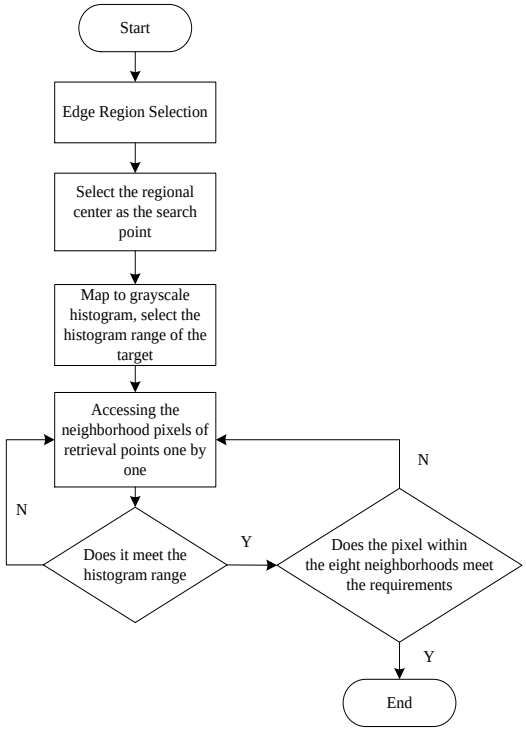


Fig. 2. Flow chart of the initial seed point selection

Based on this result, the growth criterion is constructed to complete the image segmentation. First, the two paths of the maximum and the minimum gradient change were selected as the exploration path according to the principle of gradient descent. There is a critical point exists in each exploration path, that is, only the point in the path within the starting point (the initial seed point) to within this critical point meet the growth criterion requirements.approach:

$$\max\left[\left|P(x)-P(s)\right|\right]\le T\le\left|P(t)-P(s)\right| \tag{14}$$

Where $P(s)$ is the seed point gray value, $P(x)$ is the gray value of each point in the path except the seed and termination points, $P(t)$ is the gray value of the termination point, and T is the growth threshold.

In addition, for these two exploration paths, the regional growth effectively needs to be met:

$$\max\left[\left|P_1(x)-P(s)\right|\right]\le T\le\left|P_2(x)-P(s)\right| \tag{15}$$

Where $P_1(x)$ is the information in the path with maximum gray change and $P_2(x)$ is the information in the path with minimum gray change. If the point in the path does not meet the formula, indicating that there are singularity pixels in the exploration path, then a path needs to be selected according to the gradient difference requirement until the condition is met. To sum up, the adaptive segmentation of rhinestones images is completed, laying a good foundation for the subsequent identification and positioning of rhinestones.

4. The visual positioning model of rhinestone was constructed based on YOLOV3

After completing the adaptive segmentation of rhinestone images, in order to further improve the recognition and positioning performance of rhinestone, deep learning theory is introduced in this study, and the YOLOv3 model is selected and improved to complete the recognition and positioning the design of rhinestone. This article optimizes the original output feature map based on YOLOv3. Given the limitations of 8-fold downsampling in YOLOv3 for feature extraction of small targets such as rhinestone, this study specifically improves the original network's three scale feature maps. The specific method is to remove the output branch corresponding to 32 times downsampling of the input image in the original YOLOv3 structure and add a new output feature map with 4 times downsampling. By adopting this improvement method, on the one hand, it can further enhance the network's ability to detect small targets; On the other hand, redundant output feature maps can be removed to optimize the calculation process. The improved network structure is shown in Figure 3.

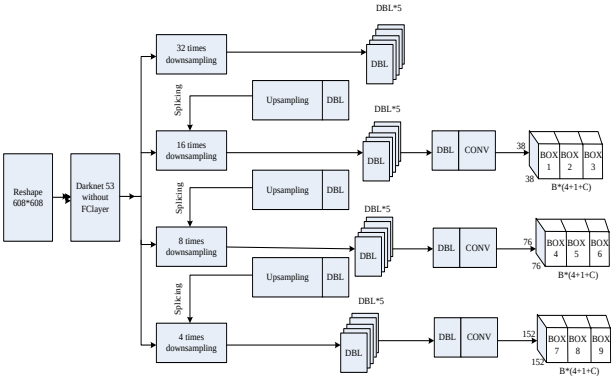


Fig. 3. The improved YOLOv3 network architecture diagram

As shown in Figure 3, the network presupposes three prior boxes for each scale of feature maps. These prior boxes are the sizes and shapes of the most common 9 detection boxes obtained by running the k-means clustering algorithm on all true boxes of the training image used. These 9 prior boxes are (5,5), (6,7), (8,8), (10,11), (12,13), (15,17), (20,22), (27,29), and (39,42), respectively, representing the height and width of the prior boxes. Due to the larger scale of the feature map, the smaller the receptive field, the smaller prior boxes are used to detect small targets. The existence of a prior box enables YOLOv3 to fit the real box based on the offset value of the prior box, which would otherwise require starting over to learn the shapes of the different bounding boxes again. In the output branch of YOLOv3, the dimension of each feature map output is $B\times(C+4+1)$, where B represents the number of feature maps detected at each scale; 4 and 1 respectively represent the size, position, and confidence information contained in each prediction box, denoted as $(t_x,t_y,t_h,confidence)$; C represents the number of categories in the dataset label. Based on the above information, combined with the width and height (p_w,p_h) of the prior box, the bounding box can be predicted, and the prediction result is shown in the formula (16):

$$\begin{cases} b_x=\sigma(t_x)+c_x \\ b_y=\sigma(t_y)+c_y \\ b_w=p_w e^{t_w} \\ b_h=p_h e^{t_h} \end{cases} \tag{16}$$

Among them, b_x,b_y represents the center coordinates of the predicted box, b_w,b_h represents the width and height dimensions of the predicted box, σ is the sigmoid activation function, c_x,c_y is the upper left corner coordinates of the grid,

and formula (16) obtains the absolute position of the predicted box. In general, the final prediction information of YOLOv3 is the offset of the bounding box. Therefore, the sigmoid function is introduced in the formula to normalize the center coordinates, denoted as (t_x, t_y, t_w, t_h). The first two represent the offset of the center coordinates of the prediction box, and the last two represent the scaling of the size of the prediction box. The calculation process is shown in the formula (17).

$$\begin{cases} t_x = G_x - c_x \\ t_y = G_y - c_y \\ t_w = \lg(G_w / p_w) \\ t_h = \lg(G_h / p_h) \end{cases} \tag{17}$$

Among them, G_x, G_y, G_w, G_h represents the center point position parameter and size scale of the real box in the feature map, respectively,By using the first two items, the coordinate offset of the prior box relative to the upper left corner of the grid can be calculated; The latter two items are the ratio between the length and width of the target bounding box and the length and width of the prior box. In order to reduce the unstable gradient caused by training, the scale is scaled to the logarithmic space. When predicting the bounding box, it is then magnified with an exponent based on e. With translation (t_x, t_y) and scaling (t_w, t_h), it is possible to fine tune the prior box to align with the true box. The loss function of YOLOv3 is represented by formula (18):

$$loss = Error_{boxes} - Error_{confidence} - Error_{classes} \tag{18}$$

Where $Error_{boxes}$ is the coordinates and scale error of the bounding box, indicating the position and coordinates of the bounding box, and is expressed as formula (19):

$$Error_{boxes} = \lambda_{coord} \sum_{i=0}^{s^2} \sum_{j=0}^{K} 1_{ij}^{\text{obj}} \left[(t_x - t_x')^2 + (t_y - t_y')^2 + (t_w - t_w')^2 + (t_h - t_h')^2 \right] \tag{19}$$

Where s^2 is the number of cells, K is the number of bounding boxes predicted for each grid cell, and λ_{coord} is the weight of the coordinate error. If there is a target in the i th cell, the t th bounding box of the cell is responsible for predicting the target, if there is a corresponding real box, then $1_{ij}^{\text{obj}} = 1$, otherwise $1_{ij}^{\text{obj}} = 0$.

$Error_{confidence}$ indicates the bounding box of the confidence error containing existing and absent objects, expressed as the formula:

$$Error_{confidence} = \sum_{i=0}^{s^2} \sum_{j=0}^{K} 1_{ij}^{\text{obj}} [c_i' \lg(c_i) + (1 - c_i') \lg(1 - c_i)] + \lambda_{noobj} \sum_{i=0}^{s^2} \sum_{j=0}^{K} 1_{ij}^{\text{obj}} [c_i' \lg(c_i) + (1 - c_i') \lg(1 - c_i)] \tag{20}$$

Where, c represents the number of categories and λ_{noobj} is the weight of the confidence error.

$Error_{classes}$ is the classification loss of the grid cell of the target, expressed as formula (21):

$$Error_{classes} = \sum_{i=0}^{s^2} 1_{ij}^{\text{obj}} \sum_{c \in classes} [p_i'(c) \lg(p_i(c)) + (1 - p_i'(c)) \lg(1 - p_i(c))] \tag{21}$$

Where $p_i(c)$ represents the category probability of the object i.

Based on the above, the visual positioning model construction of rhinestone is completed to realize the identification and positioning function of rhinestone.

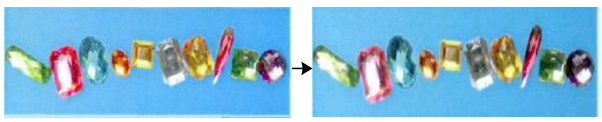
5. Experiments and analysis

5.1. Experimental environment preparation

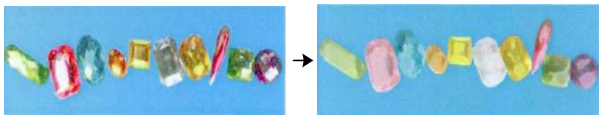
PyTorch was chosen as the deep learning framework and a training environment was built on the Ubuntu 18.04 system. The experimental hardware used an ultra micro 7049GP-TRT computer, equipped with an Intel Xeon Gold 6230 processor, with the powerful computing power of 20 cores. At the same time, NVIDIA RTX 3090 24GB GPU was used, and the server provided 4 GPUs for parallel computing, greatly improving training efficiency. During the training process, an initial learning rate of 0.0004 was set and the Adam optimizer was used for parameter updates. Batchsize is set to 8 to ensure effective utilization of computing resources in each iteration, and the total number of epochs trained is set to 200 to ensure that the model can fully learn the features in the data. To verify the performance of the proposed method, model experiments were conducted on a self-constructed rhinestone image dataset. The dataset was divided into a training set and a testing set in a 4:1 ratio, with the training set used for model training and the testing set used to evaluate the model's generalization ability. Conduct testing through a series of carefully designed experimental settings.

5.2. Functional testing

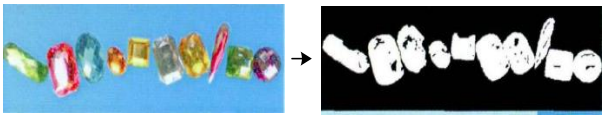
First,the de-noise performance, repair performance, segmentation performance and the positioning performance of the proposed method are verified, and the results are shown in Figure 4.



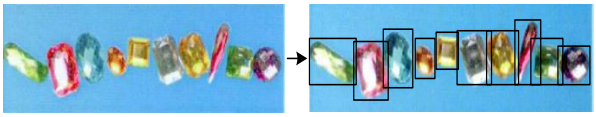
(a) Comparison of denoising effect



(b) Repair effect comparison



(c) Compare the effect of segmentation



(d) Positioning effect comparison

Fig. 4. Functional effect test

As shown in Figure 4, the proposed method can successfully de-noise, repair, segmentation, and positioning of the rhinestone image. As can be seen from the figure, the proposed method performs well in dealing with various image problems, which can provide strong support for the practical application of the visual positioning system of the greeting card paste rhinestone machine.

5.3. Analysis of the experimental results

To further verify the advancement of the proposed method, the reference [4] method and the reference [5] method were selected as the comparison method, and the localization time and localization error were used as the performance evaluation index. Using three methods to carry out the identification of rhinestone positioning, and using the computer comes with timing software to record the positioning time,to get the results of positioning time as shown in Figure 5.

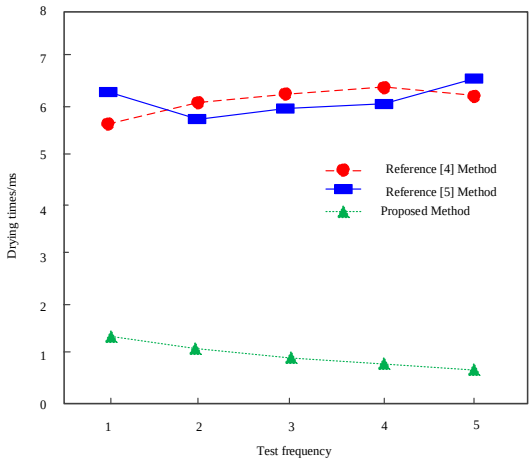


Fig. 5. Comparison of the three methods

As shown in Figure 5, the application of the proposed method, the positioning time is lower than 1.31ms, and the application of two comparison method, the positioning time is higher than 5.7ms, thus, the proposed method positioning efficiency is higher, this is mainly due to the proposed method applied the proposed method by de-noise processing and ROI area repair, effectively improve the image quality, reduce the influence of noise and defects on positioning accuracy. This provides more reliable image data for subsequent localization operations, thus improving the localization efficiency.

Next, the localization error was tested, and the mean value was taken for multiple tests, and the results are shown in Table 1.

Table 1. Comparison of the localization errors of the three methods

Test times	The proposed method has localization error / mm	Reference [4] method localization error / mm	Reference [5] method localization error / mm
1	0.41	1.77	3.43
2	0.37	1.78	3.41
3	0.37	1.72	3.38
4	0.33	1.70	3.33
5	0.32	1.67	3.33
6	0.31	1.65	3.34
7	0.30	1.64	3.26
8	0.30	1.55	3.25
9	0.29	1.58	3.24
10	0.29	1.78	3.21

As shown in Table 1, the application of the proposed method resulted in a positioning error of less than 0.41 mm for the rhinestone. However, the positioning errors of the two comparison methods were both greater than 1.55 mm. Therefore, it can be seen that the proposed method has a relatively small positioning error. This is mainly due to the fact that the proposed method not only performs de-noise and ROI region restoration in image processing, but also adopts more precise image segmentation techniques, making the edges and contours of the rhinestone target clearer and reducing positioning errors caused by poor image quality. In addition, the proposed method is improved based on the YOLOv3 network to make it more suitable for rhinestone positioning tasks. By optimizing the network structure, adjusting parameter settings, and introducing more suitable feature extraction methods for rhinestone recognition, the accuracy and stability of rhinestone recognition have been improved, positioning errors have been reduced, and the application effect is good.

6. Conclusion

To improve the visual positioning effect of rhinestone on greeting card paste rhinestone machine, deep learning theory is introduced in this study, and a YOLO based rhinestone visual positioning method for greeting cardpaste rhinestone machine is proposed. By constructing a YOLO3

rhinestone visual positioning model, automatic recognition and accurate positioning of rhinestone targets are achieved. Through testing, the experimental conclusions are as follows:

1) By applying the proposed method, the positioning time of the rhinestone is all less than 1.31ms, which significantly improves the positioning efficiency compared to the comparison method with a positioning time of over 5.7ms. This data fully demonstrates the real-time advantage of the proposed method, enabling the greeting card paste rhinestone machine to complete the positioning of the rhinestone in a shorter time, improving production efficiency.

2) The proposed method controls the error within 0.41mm, which is much lower than the comparison method's 1.55mm or more. This precise positioning capability ensures that the rhinestone can be accurately attached to the designated position on the greeting card, improving the quality and consistency of the product.

Author contributions

SecondAuthor: Research Direction Guidance, Methodological Support, Experimental Supervision and Guidance, Thesis Review and Revision, Academic Resources and Support.

FirstAuthor: Literature Review and Data Collection, Constructing a dataset, Experimental Design and Implementation, Thesis Writing and Revision.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Chen C.,Lu J.,Zhou M.,Yi J.,Liao M.,Gao Z..A YOLOv3-based computer vision system for identification of tea buds and the picking point[J].Computers and Electronics in Agriculture,2022,198:1-15.

[2] Hao Tian,Wenhai Wu,Huanlong Liu,YaDong Liu,Jincheng Zou,Yifei Zhao.Robotic Grasping of Pillow Spring Based on M-G-YOLOv5s Object Detection Algorithm and Image-Based Visual Serving[J].Journal of Intelligent & Robotic Systems: Theory & Application,2023,109(3):1-17.

[3] Zhou Qingyi, Qu Jiawei, Zhang Chunlei. Research of Workpiece Positioning Algorithm of Five-axis Dispensing Machine Based on Machine Vision [J]. Automation & Instrumentation , 2022,37 (05): 63-69.

[4] Wang Rui, Qu Jiawei, Zhang Chunlei. Solution of Rotation Axis Pose of Five Axis Dispensing Machine Based on Machine Vision [J]. Modular Machine Tool & Automatic Manufacturing Technique, 2022, (05): 36-40.

[5] Xue Haihui, Shi Feng, Liu Jun, et al. Modular Vision Positioning System Based on Intelligent Camera [J]. Automation & Instrumentation, 2022,37 (03): 92-95 + 101.

[6] Xu Jianming, Han Bo. Design of control system for five-axis dispensing machine based on CODESYS [J]. Chinese High Technology Letters , 2022, 32 (01): 13-21.

[7] Zhu Yanfei, Chu Youyang, Li Chuanjiang. Screw Dispensing Valve Model Free Adaptive Control Technology Based on BP Neural Network Optimization Parameters [J]. Chinese Hydraulics & Pneumatics, 2023, 47 (09): 175-181.

[8] Xie Zhenghua, Chen Xiaofei, Wang Changlai. Design and Research of Dispensing Control System for Chip Plastic Package [J]. Machinery Design & Manufacture , 2023, (10): 107-109.

[9] Sun Peng, Wang Xiaodong, He Wenxin, et al. The Integration of Precision Automatic Adhesive Dispensing in Micro-Assembly Equipment [J]. Modular Machine Tool & Automatic Manufacturing Technique, 2023, (01): 56-60.

[10] Tang Xin, Hou Qifeng, Li Xueting, et al. Fast implementation method of data acquisition for FAD2500 dispensing equipment [J]. Qutlook of Electronic Technology, 2023, 30 (01): 52-55.

[11] Linjun Lu,Fei Dai.Accurate road user localization in aerial images captured by unmanned aerial vehicles[J].Automation in construction,2024,158(2):1-15.

[12] Zeynep Turgut,Arzu Gorgulu Kakisim.An explainable hybrid deep learning architecture for WiFi-based indoor localization in Internet of Things environment[J].Future generations computer systems: FGCS,2024,151(2):196-213.

[13] MARIEM GNOUMA,RIDHA EJBALI,MOURAD ZAIED.Deep Hashing and Sparse Representation of Abnormal Events Detection[J].The computer journal,2024,67(1):3-17.

[14] Xu Jianming, Li Xiang. Research on the interpolation algorithm of mobile phone dispensing trajectory based on double B-spline [J]. Journal of Zhejiang University of Technology, 2022, 50 (04): 355-364.

[15] Tang Xin, Zhao Xuefeng, Hou Qifeng, et al. Research on Dispensing Height Measurement Algorithm of Microwave Module Equipment [J]. Qutlook of Electronic Technology, 2022, 29 (07): 90-93.

[16] Liu, Huanlong,Li, Dafa,Jiang, Bin,Zhou, Jianyi,Wei,

Tao,Yao, Xinliang.MGBM-YOLO: a Faster Light-Weight Object Detection Model for Robotic Grasping of Bolster Spring Based on Image-Based Visual Servoing[J].Journal of Intelligent & Robotic Systems: Theory & Application,2022,104(4):77-94.

[17] Li You, Zhang Chen, Wei Wei, Xiang Sen. Cotton Image Segmentation Algorithm Based on Lightweight Symmetric Structure Network[J]. Computer Simulation,2024,41(1):395-400.

[18] Alireza Fadakar,Amir Mansourian,Saeed Akhavan.Deep learning aided multi-source passive 3D AOA wireless positioning using a moving receiver: A low complexity approach[J].Ad hoc networks,2024,154(3):1-15.

[19] Du X.,Cheng H.,Ma Z.,Lu W.,Wang M.,Meng Z.,Jiang C.,Hong F..DSW-YOLO: A detection method for ground-planted strawberry fruits under different occlusion levels[J].Computers and Electronics in Agriculture,2023,214:1-13.

[20] Omar, Karima,Sakr, Rasha H.,Alrahmawy, Mohammed F..An ensemble of CNNs with self-attention mechanism for DeepFake video detection[J].Neural computing & applications,2024,36(6):2749-2765.