Based on the improvement of YOLO v3 aluminum alloy door and window profile identification method

LiShuiping¹,a, TongHao²b*,
¹Wuhan University of Technology, Wuhan, Hubei, China
²Wuhan University of Technology, Wuhan, Hubei, China
aemail: lshpmail@163.com
b*email: th1316229666@whut.edu.cn
*Corresponding author’s e-mail: th1316229666@whut.edu.cn

Abstract: In the traditional aluminum alloy door and window profile production process, the identification of profiles mainly rely on RFID electronic labels and staff eye identification, but aluminum alloy doors and windows profiles are a wide variety, while some profile models are similar, staff difficult to judge, RFID electronic labels also have recovery difficulties, higher cost problems. Based on this, this paper proposes an aluminum alloy door and window profile identification method based on machine vision and deep learning, with YOLO v3 as the detection network, adding a separable convex network to the Darknet neural network, improving the detection accuracy and speed, and improving the K-mean clustering method to obtain a more accurate prediction box. The experimental results show that compared with SSD network and YOLO v3 network, the accuracy of improved detection network recognition is up to 0.15s, which can meet the needs of industrial production.

1. Introduction
The initial profile detection model structure based on the DakNet-53 network architecture and YOLO v3 algorithm idea is shown in Figure 3-8. It can be found that the model uses the DarkNet-53 framework as the trunk network, DarkNet-53 is a simple classification network, the whole set pooling layer and the segmentation layer in the network are removed, the remaining network layer is used to extract image features, and then the YOLO v3 decision area is attached to the trunk network as the target detection network, which connects the last three feature layers of the DarkNet-53 network[1], with the volume fusion and upper sampling and other operations divided into different sizes. To achieve multi-scale target detection. The dimensions of the output feature map correspond to the characteristics of its scale× the scale 13×13 feature map size is small, feel wild, detect large-scale objects, and the upper sampling channel fusion scale is 26×26 feature map, detection of medium-scale objects;Further, the upper sampling channel is fused into a feature map × 52 and 52 to detect small-scale objects[2].

2. Method
In order to further improve the real-time performance of the profile detection model, the profile detection model is improved by adding a deep, detachable convolution network structure to the Darknet network structure[3]. Deep detachable convolution consists of two convolutions:
channel-by-channel convolution and point-by-point convolution.

![Figure 1 Network structure improvement program](image)

In Figure 1, the input and output feature graphs are \( D_f \), the coil size is \( D_k \), the number of input channels is \( M \), and the number of output channels is \( N \). As can be seen from Figure 3-5a, the complexity of standard resuming operations is \( d_k \times d_k \times M \times N \times D_f \times D_f \), because the Darknet network is a pyramid structure, and the number of input and output channels increases with the deepening of the number of network layers, which in turn increases the amount of network computing. The calculated amount of the improved tortograms in Figure 3-5b is the number of channel-by-channel and point-by-point reels (1\( \times \)1 reels), i.e., the complexity of the standard reflects \( d_k \times M \times D_f \times D_f \times N \times D_f \times D_f \).

The candidate box generation mechanism (anchor box) of the YOLO v3 model requires clustering of the label boxes in order to obtain the initial dimensions of the candidate boxes. K-mean clustering is an unsecured algorithm that measures the similarity of data objects by distance, primarily for grouping similar samples into the same category, \( K \) is the number of clustering categories, and \( \text{means} \) refers to the mean of data objects in each category. The greater the similarity between the data, the shorter the distance between the data objects, the more likely it is to be divided into the same category, the commonly used similarity calculation method is the European distance method[4].

In the actual clustering process, different scales of label boxes are more sensitive to errors, large size and smaller size are more sensitive, so in order to more accurately extract the information of label boxes, improve detection accuracy, in this paper using IoU (Intersection over Union) instead of K-means clustering algorithm Euclide distance[5], as the distance measurement standard of clustering effect, its calculation formula is shown in the formula (1):

\[
\text{IoU} = \frac{S(R_{A \cap B})}{S(R_{A \cup B})} .
\]

(1)

A and B are the area of the target candidate box and label box, IoU is the ratio of the overlapping area of the two boxes to the merged area, so that the clustering effect will not be affected by the dimensions of the label box. Measurement distance[6], such as formula (2), is the exact dimension of the label box when the adjacent label box clustering heart value no longer changes.

\[
d(x, B) = 1 - \text{IoU}(x, B)
\]

(2)

\( d \) is the distance between the label box and the similar class heart, \( B \) is the similar class heart, and \( x \) is the sample label box.
3. Experiment
In this section of the training set, the aluminum alloy door and window profile detection algorithm is compared with the profile detection algorithm based on the original YOLO network, the profile detection algorithm based on the original YOLO v3 and the SSD-based profile detection algorithm. A total of 10,000 sample sets containing aluminum alloy door and window profile face images were generated by data set enhancement, and 9000 training set images were extracted for network training;

Four networks in the training process, epoch set 100 times, after each training round to verify the network's training performance, four network training models in the first 30 epochs tend to converge, four networks in the verification accuracy and los curve changes as shown in the figure.

From Figure 2, it can be obtained that the SSD network in the four detection networks has played the inherent advantages of the Anchor mechanism, and has the best performance in terms of loss value and accuracy. The improved YOLO v3 network in this paper is similar to the original YOLO v3 network in terms of recognition rate, but lower than the original network in terms of loss value, which indicates that the sethable converge network improvement method of this paper is effective, and the improved loss function curve converges faster.

4. Test results and analysis.
Take advantage of the improved YOLO based on the completion of the training v3's aluminum alloy door and window profile detection network with the isolated consumables improved DarkNet-53 feature extraction network test sample set test results as shown in Table 1. Leak detection problem, in the double hook frame and solid slide and other profile detection problems, window profile detection is all correct; The comprehensive accuracy rate of the algorithm in this paper is 97.3%, the
The comprehensive error detection rate is 1.9%, the comprehensive leakage rate is 0.8%, and all three indicators meet the design requirements of the detection system.

| The type of profile | The number of samples | correct sample | error sample | Missing samples | Accuracy | Error rate | Missed rate |
|--------------------|-----------------------|----------------|--------------|----------------|----------|------------|-------------|
| 1                  | 100                   | 99             | 1            | 0              | 99%      | 1%         | 0%          |
| 2                  | 100                   | 97             | 2            | 1              | 97%      | 2%         | 1%          |
| 3                  | 100                   | 95             | 4            | 1              | 95%      | 4%         | 1%          |
| 4                  | 100                   | 98             | 2            | 0              | 98%      | 2%         | 0%          |
| 5                  | 100                   | 96             | 3            | 1              | 96%      | 3%         | 1%          |
| 6                  | 100                   | 100            | 0            | 0              | 100%     | 0%         | 0%          |
| 7                  | 100                   | 94             | 4            | 2              | 94%      | 4%         | 2%          |
| 8                  | 100                   | 99             | 0            | 1              | 99%      | 0%         | 1%          |
| 9                  | 100                   | 98             | 1            | 1              | 98%      | 1%         | 1%          |
| Total              | 900                   | 876            | 17           | 7              | 97.3%    | 1.9%       | 0.8%        |

Figure 3 shows the test set image of some aluminum alloy door and window profiles based on the detection results of the profile detection algorithm.

![Figure 3 Aluminium alloy door and window profile detection results](image)

The above test sample set is detected by the template matching algorithm of image processing, and the SSD network and YOLO v3 network are combined with the aluminum alloy door and window profile end face detection algorithm to detect the above test sample set. The detection time comparison table of 4 algorithms in network training ginseng, profile detection accuracy and single image is shown in Table 4-5. The accuracy of the four algorithms and the comparison of single detection time are shown in table 2.

| Serial Number | Profile Detection Algorithm | Number of network ginseng (millions) | Accuracy | Accuracy Check time (s) |
|---------------|-----------------------------|--------------------------------------|----------|-------------------------|
| 1             | Template matching           | 0                                    | 95.4%    | 0.09                    |
| 2             | This article algorithm      | 6.21                                 | 97.3%    | 0.15                    |
| 3             | YOLO v3                     | 5.33                                 | 91.1%    | 0.13                    |
| 4             | SSD                         | 12.80                                | 98.2%    | 0.65                    |

From Tables 2, it can be found that the template matching algorithm of image processing can obtain high accuracy by setting strict thresholds, which lags behind only the profile detection algorithm and SSD-based detection algorithm, and because there are no complex volumes. Product calculation, its single detection time is the shortest, but the accuracy of this method does not meet the requirements of the profile detection system, while the production of matching templates on the operator's technical
requirements of a higher level, compared with the depth of learning class method in the practical and intelligent degree is low. The profile detection network algorithm based on YOLO network is only 91.5% accurate due to the lack of multi-scale feature extraction model, which is the lowest of the four detection algorithms and can not meet the requirements of the detection system in this paper. In this paper, the profile detection calculation based on YOLO v3 network improves the DarkNet residual network by deep separable network, provides multi-scale feature prediction, improves the detection accuracy to 98.5%, and its network parameter is 5.33 million, the model is as lightweight as YOLO network, the single image detection time is 0.15s, and the detection speed of real-time detection requirements is obtained with lower lightweight model. The profile detection algorithm based on SSD network can achieve the detection accuracy of the first-order network because of the Anchor mechanism, which makes its overall detection accuracy the highest of the four algorithms, but its network parameter scale is up to 12.8 million, and the training time is longer. And the maximum detection time of a single image is 0.65s, when the detection task is more, the system completes the image acquisition, the aluminum alloy door and window profile detection is not completed, can not give the detection results, will delay the subsequent profile size measurement and hole detection tasks, so can not meet the online real-time detection requirements.

5. Conclusion.
In summary, compared with the other three algorithms, the aluminum alloy door and window detection algorithm proposed in this paper achieves a good balance between detection rate and single image detection time, and the detection accuracy, detection speed, system economy and intelligence of the algorithm meet the design requirements of the detection system in this paper.

References
[1] Li, J.X. (2020) Study on workware identification methods based on improved YOLO v3[J]. Combine machine tools with automated machining techniques,(08):92-96+100.
[2] Ma, F.Q. (2020) Research on surface defect detection techniques for large field-of-view work parts based on deep learning.[D] Harbin University of Technology.
[3] Deng, L (2019) Study on the detection method of PCB surface assembly defect based on machine vision[D]. Wuhan University of Technology
[4] Zeng, S. (2020) Part recognition system design and research based on deep learning. [D] Shanghai Normal University
[5] Zhang, J. (2020) The study of industrial part recognition algorithm based on YOLO v3 in intelligent assembly[J]. Photoelectronics laser.(10):1054-1061
[6] Zhu, X.C. (2020) Hard hat wearing detection with improved YOLO v3 algorithm[J]. Journal of Nanjing Institute of Engineering (Natural Sciences Edition)(04):23-26.