Hindawi
Advances in Multimedia
Volume 2022, Article ID 6049013, 8 pages
https://doi.org/10.1155/2022/6049013

Research Article

Pipeline Multitype Artifact Recognition Method Based on Inception_Resnet _V2 Structure Improving SSD Network

Yi Zheng

Chongqing Industry Polytechnic College, Yubei, Chongqing 401120, China

Correspondence should be addressed to Yi Zheng; zhengyi@cqipc.edu.cn

Received 30 December 2021; Revised 12 January 2022; Accepted 17 January 2022; Published 31 January 2022

Academic Editor: Mohammad R. Khosravi

A fast recognition method for assembly line workpieces based on an improved SSD model is proposed to address the problems of low detection accuracy and lack of real-time performance when existing target detection models face small-scale targets and stacked targets. Based on the SSD network, the optimized Inception_Resnet _V2 structure is used to improve its feature extraction layer and enhance the extraction capability of the network for small-scale targets. The repulsion loss (Reploss) is used to optimize the loss function of the SSD network to solve the problem of stacked workpieces. The issue of difficult detection is improved. The robustness of the algorithm is enhanced. The experimental results show that the improved SSD target detection method improves the detection accuracy by 9.69% over the traditional SSD map. The detection speed meets the real-time requirements, which is a better balance of detection real time and accuracy requirements. The algorithm can recognize small-scale and stacked targets with higher category confidence, better algorithm robustness, and better recognition performance compared to the same type of target detection algorithms.

1. Introduction

With the rapid development of the modern manufacturing industry, the demand for miniaturized, diversified, and personalized workpieces is gradually increasing. The need for production automation and intellectualization by processing enterprises is also increasing [1]. Sorting operations are an essential part of industrial production processes. Two main methods are currently used: the manual sorting method and the conventional target detection combined with the mechanical arm sorting method [2, 3]. The manual sortition process is heavily impacted by subjective factors and has a relatively high requirement for the operating environment. It is easy to cause more sorting errors due to visual fatigue, reduced operational efficiencies, and high production costs, which is not conducive to improving business competitiveness [4]. Traditional target detection and sorting methods generally use the manual, characteristic extraction, SVM, and the logistic regression classifier to classify the characteristics [3, 5, 6]. Compared to manual processes, sorting accuracy and error detection rates have significantly improved, but there are still shortcomings. For example, the generalization capacity of the model is weak, large-scale training samples are difficult to implement, and multiclassification issues are challenging to solve.

In recent years, with the continuous improvement of the performance of the convolution neural network, some researchers have begun to try to use deep learning technology to supplement target detection. Several types of target detection algorithms based on neural networks exist. Compared with Faster R-CNN, Mask-RCNN, and other similar algorithm models, the regression-based SSD algorithm has the characteristics of solid target feature extraction ability and fast training speed because of the advantages of a variety of algorithms [7, 8]. It is suitable for the real-time detector system. However, the SSD algorithm also has some flaws, facing small workpieces with insignificant features, stacked workpieces, workpieces of different types but similar shapes, weak robustness, high missed detection rate, and false detection rate, which cannot meet the requirements for accurate detection [9–12]. To improve the extraction capacity of the network features, Yang et al. [13] proposed a DSSD

algorithm based on the VGG backbone network. By introducing a deconvolution module in the model to improve the algorithm's reconnaissance accuracy, small-scale target detection is slightly improved. However, the implementation process is complex, and the detection rate is slow, which cannot meet the demands of large-scale and real-time detection. Dai et al. [14] designed a gear part recognition model based on Faster RCNN, using ResNet101 as a feature extraction network, with irregular cross-convolution and differential convolution kernels to accomplish the feature fusion of different convolutional groups, which can better accomplish the recognition and detection of small-scale parts with a fast detection speed but unsatisfactory detection accuracy and high mass detection rate. Zhai et al. [15] Using the DenseNet network to improve the SSD model. A new DSOD algorithm is proposed, but it still cannot solve the low accuracy of small-scale target detection. Jeong et al. [16] proposed an R-SSD algorithm. The effect of the SSD algorithm is improved by improving the functionality merging method. Still, the detection speed is slow, and the detection capacity of stacked targets is low, unable to meet industry target detection requirements.

To solve the problems, current target detection algorithms have weak detection capability, low robustness, and unsatisfactory localization effect for small-scale targets and stacked targets. In this paper, we introduce the Inception_Resnet _V2 module to reduce the sampling of the low-level function map to expand its perceptual field. In this way, we enhance the ability to extract detailed and localized information from the SSD network and improve the accuracy of small-scale target detection. The reject loss is used to optimize the loss function of its network and fix the problem that stacked artifacts are difficult to detect. The experiments show that the method has more outstanding performance in terms of precision and speed of detection, which is a good reference value for developing intelligent detection in the manufacturing industry.

2. SSD Network Model

SSD (Single Shot MultiBox Detector, SSD) is a target detection algorithm based on the structure of the VGG-16 network [16]. The SSD network uses the first five layers of the VGG-16 network as the backbone network and replaces the fully connected layers (fc6, fc7) of the original network with convolutional layers (Conv6, Conv7). The convolution operation is performed by using 3×3 and 1×1 convolution kernels, and the pooling layer (pool5) is changed from 2×2 with stride = 2 to 3×3 with stride = 1. Conv6 uses null convolution (dilatation = 6) to avoid changing the feature map's resolution while getting an enormous perceptual field. The network results are shown in Figure 1.

A total of 6 convolutional layers (Conv4_3, Conv7, Conv8_2, Conv9_2, Conv10_2, and Conv11_2) are used in the SSD network to extract the features of the images. And the generated six feature maps are used to predict the bounding boxes of different sizes and aspect ratios, respectively. Each convolutional layer has predefined prior boxes of various sizes, of which Conv4_3, Conv10_2, and Conv11_2 have four types, respectively, and Conv7, Conv8_2, and Conv9_2 have six types of prior boxes, respectively. Each pixel on the corresponding feature map generates K (prior box kinds) of prior boxes [17]. Therefore, the SSD network generates 8,732 bounding boxes. Final target location results and confidence in the category are obtained after non-maximum suppression (NMS).

3. Image Detection Method Based on Improved SSD Network

3.1. Detection Process. To ensure the accuracy of workpiece sorting, the image information of workpieces on the assembly line is captured by CCD sensors, and the images are pre-processed using corresponding algorithms. The image detection part uses the SSD network as the detection framework. The improved Inception_ResNet_v2 is used to extract the target features to obtain richer detail and semantic information and improve the accuracy of the network for target detection and recognition. The loss function is optimized using rejection loss to ensure the accuracy of artifact classification and glory in the case of stacking [18]. The workpiece detection process based on the improved SSD network is shown in Figure 2.

The detailed detection steps are as follows:

Step 1: Image capture and preprocessing: the CCD image sensor is calibrated, and the light source is adjusted to obtain the image of the workpiece to be measured on the assembly line. The idea is greyed out, filtered, and less noisy using the weighted average method and bilateral filtering to improve image feature information.

Step 2: Image dataset: preprocessed images of various workpieces are converted to VOC dataset format, and training and test sets are established in a 7 : 1 ratio.

Step 3: Training and testing the network: the enhanced SSD network is trained and tested using the dataset to determine the optimum settings.

Step 4: Multiscale feature prediction: the feature maps of different scales are applied in the SSD network to classify and predict the candidate frames, and obtain prediction frames of different scales.

Step 5: Non-maximum suppression: non-maximum suppression is used to filter out the optimal target bounding boxes, and the identification and localization of the artifacts are completed.

3.2. Image PreProcessing. Considering the simple texture of the selected material, only grayscale information is needed to recognize the material characteristics. Therefore, the original color image is grayed out using the weighted average method to reduce the number of operations in the subsequent nodes and improve image processing efficiency. The original and grayscale processed images are shown in Figures 3(a) and 3(b).

The purpose of filtering the grayed-out image is to reduce background and noise interference. This will result in a clear boundary of the acquired image and improve the

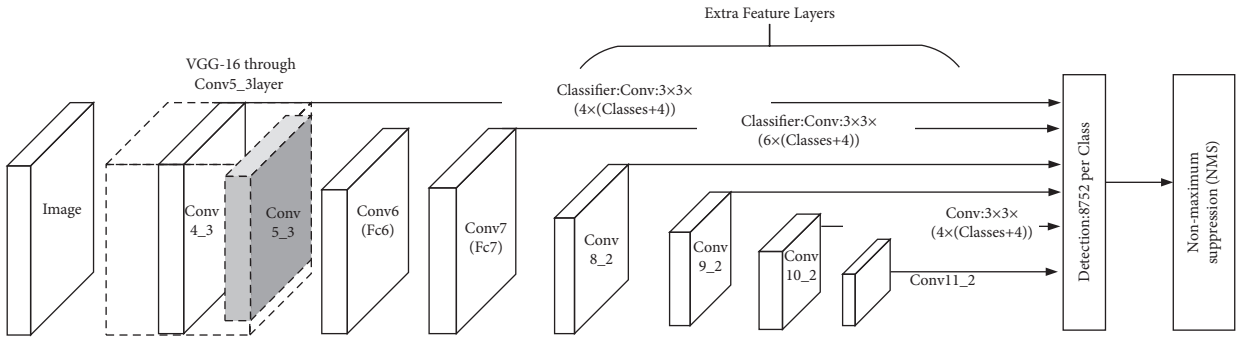

Figure 1: Network structure of SSD.

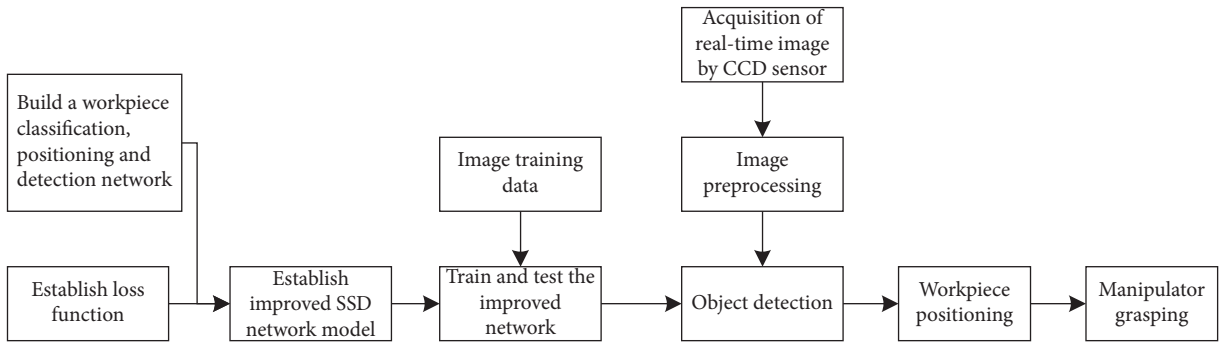

Figure 2: Workpiece inspection process based on improved SSD network.

Figure 3: Comparison between the original image and processed image. (a) The original image. (b) Grayscale processing. (c) Grayscale processing.

feature extraction capability of the network. Bilateral filtering consists of a region and a value domain filter kernel. The nonlinear combination of the two filters allows not only to interpret the correlation in the geometric position of each pixel value in the image but also to derive the similarity in luminance between each pixel value and the centroid pixel value. The edge information of the artifact is preserved to the maximum extent while the image noise is suppressed [19]. The image after bilateral filter processing is shown in Figure 3(c).

3.3. Improved SSD Network Model

3.3.1. Improved Feature Extraction Layer Based on Inception_Resnet _V2 Structure. SSD is weak in detecting small targets. Increasing the network depth is used to improve its performance. It inevitably causes a sharp increase in the number of model parameters and consumes many computational resources. Furthermore, as the number of network layers increases, the model is subject to gradient disappearance and overflow in the back-propagation process, making it

challenging to optimize the model [20]. Considering the practical applications and the requirements to sort and detect parts, the structure of the SSD model and the loss function is adapted and optimized. To improve the extraction capability of the model for target features at different scales and to reduce the leakage and false detection rates of sorting operations in unstructured environments, the Inception_Resnet _V2 module is used to replace the convolutional layer in the SSD network, which allows the network to have sparsity to enhance its feature representation [21]. Moreover, we optimize the Inception_Resnet_v2 module appropriately to enable the network acquire more delicate features. The optimized structure is shown in Figure 4.

The optimized Inception_Resnet_v2 module uses two convolution kernels, 1×3 and 3×1, to replace the second 3×3 convolution kernel in the last layer of the original module (the part in the dotted box in Figure 4). The image features are extracted in parallel to increase the network width and reduce the representation bottleneck of the features and the loss of feature information. The Batch_normalization layer is set after each convolutional layer of the module to speed up the network training, alleviate the vanishing_gradient phenomenon, and prevent the occurrence of the overfitting phenomenon. In the improved model, the 1×1 convolution kernel is used to reduce the number of channels, and the 3×3 convolution kernel is used to extract the features and obtain the corresponding feature map; the 1×3 and 3×1 convolution kernels are used to increase the number of 1×3, and the 3×1 convolution kernel is used to increase the number of channels to extract rich feature information. Finally, the fused information is input to the 1×1 convolution kernel for dimensionality reduction. The dimensionality of in_channels and out_channels can be kept the same, thus improving the nonlinear expression capability of the network.

The optimized Inception_Resnet_v2 module replaces the Conv6, Conv7, and Conv8_2 layers in the SSD network. After completing the operations such as normalization and activation function, we obtain the information of target class, target location, and cell confidence. The NMS operation is performed on the results of different scales. The final detection results are output. The implementation process of the improved SSD model is shown in Figure 5.

3.3.2. Optimization of Loss Function. For the case that stacked workpieces are prone to false detection and missed detection,. the loss function is optimized by applying rejection loss. It is ensured that each prediction box is close to the actual target while staying away from the whole region of other marks and the prediction box. Preventing the prediction box from moving to the adjacent sides obscured the target, making the detection model more robust to targets in unstructured environments. As in Figure 6, an additional penalty will be given when the prediction box of target A is close to position B.

The Reploss expression is as follows:

$$L_{\text{Rep}} = L_{\text{Attr}} + \alpha L_{\text{RepGT}} + \beta L_{\text{RepBox}}, \tag{1}$$

where L_{Attr} denotes the loss of gravitational term so that the prediction box is as close as possible to its target box; L_{RepGT} and L_{RepBox} denote the repulsive term loss so that the prediction frame is as far as possible from the surrounding target frame. The weighting coefficients α and β are used to balance the auxiliary loss, and both coefficients are taken as 0.5 in the experiment.

Let variable $P = (l_P, t_P, w_P, h_P)$ represent the prediction box and variable $G = (l_G, t_G, w_G, h_G)$ represent the real target box; l and t are the coordinates of the upper left corner of the two boxes, w and h are the width and height of the boxes. $P_+ = \{P\}$ is the set of positive samples of the prediction box, and the positive samples are divided by the set Intersection over Union (IoU) threshold. $g = \{G\}$ is the set of real target boxes. The prediction box $P \in \{P_+\}$ is set, and the real target box is treated with the maximum IoU with this prediction box as the target, i.e., $G^P_{\text{Attr}} = \arg\max_{G \in g} IoU(G, P)$. B^P is the prediction box after regression. Then, the gravitation loss function is

$$L_{\text{Attr}} = \frac{\sum_{P \in P_+} \text{Smooth}_{L1}\left(B^P, G^P_{\text{Attr}}\right)}{\left|P_+\right|}. \tag{2}$$

Conversely, if the prediction box P treats the other surrounding targets with its maximum IoU as exclusion targets, then the rejection loss can be expressed as

$$L_{\text{RepGT}} = \frac{\sum_{P \in P_+} \text{Smooth}_{\ln}\left(\text{IoG}\left(B^P, G^P_{\text{Attr}}\right)\right)}{\left|P_+\right|},$$

$$\text{Smooth}_{\ln} = \begin{cases} -\ln(1 - x), & x \le \sigma, \\ \dfrac{x - \sigma}{1 - \sigma} - \ln(1 - \sigma), & x > \sigma, \end{cases} \tag{3}$$

where $IoG(B^P, G^P_{\text{Attr}})$ represents an overlap between B^P and G^P_{Rep}. $\text{Smooth}_{\ln}$ is an LN function in the range (0, 1), and the σ parameter is used to adjust RepLoss sensitivity to outliers. The more the proposal tends to overlap with nontarget ground truth objects, the greater the penalty of RepGT loss on the bounding box regressor, which effectively prevents the bounding box from moving to adjacent nontarget things.

4. Experiment and Analysis

4.1. Experimental Procedure. The workpiece to be measured is sent through the material conveyor to the image acquisition section. The assembly line workpieces are acquired by a CCD camera combined with an illumination light source, and the images are preprocessed [22]. The feature extraction module is used to extract the key features of various types of artifacts and train the model to derive the training results. The improved SSD network generates bounding boxes of different scales in the global feature map, extracts the optimal bounding box using NMS, makes the results compare with the network training results, and then completes the recognition of workpiece types. Finally, the robot is used to grasp the workpieces of different categories. The detection process is shown in Figure 7.

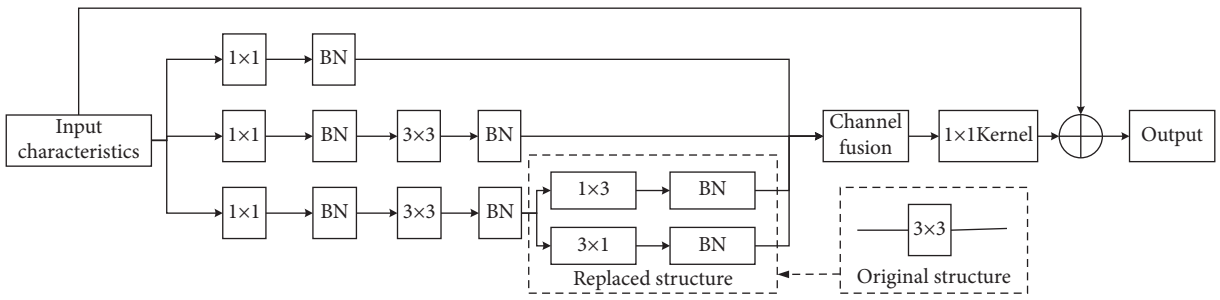

Figure 4: Optimized Inception-v2 module structure.

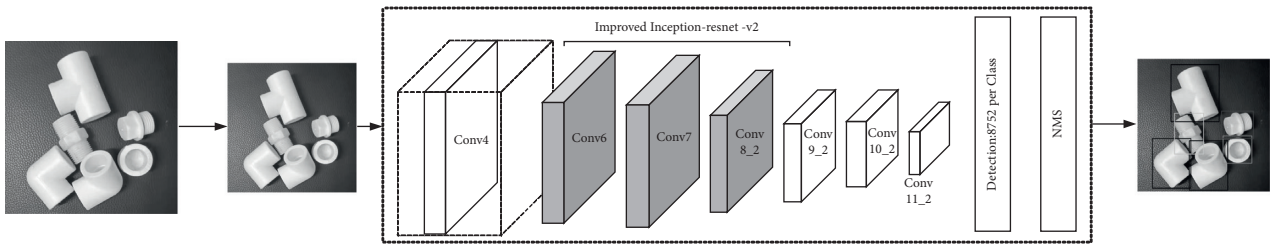

Figure 5: The implementation process of an improved SSD network model.

Figure 6: Schematic diagram of exclusion loss.

The experimental dataset was taken from the images of workpieces collected by the inspection machine of a PPR pipe manufacturer. 1,200 images were taken for each of the five types of pipe such as Equal elbow, Equal tee, plug, union, and cap, totaling 6,000 images, and divided into training and test sets in the ratio of 7 : 1, with 5,250 images for the training set and 750 images for the test set. Firstly, all manifestations of the dataset are preprocessed to obtain images with a resolution of 300×300 pixels, and then the corresponding dataset is annotated and transformed into a standard VOC format. Finally, the training set is used to train the improved SSD network.

The Adam optimizer is used to optimize the network to speed up the training, and an early stopping mechanism is added to the training to monitor the loss value of the model. When the model's loss value is not reduced after 6 epochs, the learning rate decay is halved; when it is not relieved after 10 epochs, training is stopped, and the best loss value weight parameter is retained.

4.2. Experimental Analysis. The method in this paper is tested in the test set with the conventional SSD network and the DSSD network in the literature, and the three types of networks are evaluated using the mAP detection accuracy metric. The experimental results are shown in Table 1.

The comparison results show that the SSD target detection algorithm based on the improved Inception_Resnet_v2 module and optimized loss function proposed in this paper improves 9.69% over SSD in mAP, and the comparison results are shown in Figure 8.

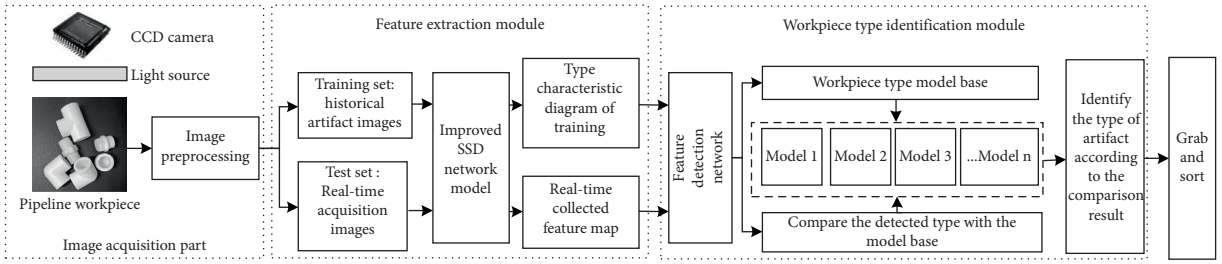

Figure 7: Flowchart of workpiece type detection.

Table 1: Comparison of evaluation parameters of three target detection methods.

Models	Backbone network	mAP (%)	Detection accuracy (%)				
			Equal elbow	Equal tee	Plug	Union	Cap
SSD	VGG-16	73.81	78.15	74.41	71.32	73.59	71.58
RCNN	VGG-16	77.74	72.18	79.68	75.72	78.33	70.21
DSSD	ResNET	78.27	82.08	80.05	73.84	77.14	78.22
DSOD	DenseNet	79.13	80.83	80.32	74.60	74.75	76.86
R-SSD	ResNet	80.04	82.35	81.67	73.62	75.83	80.02
Paper method	VGG-16	83.50	83.68	87.20	84.15	82.14	80.31

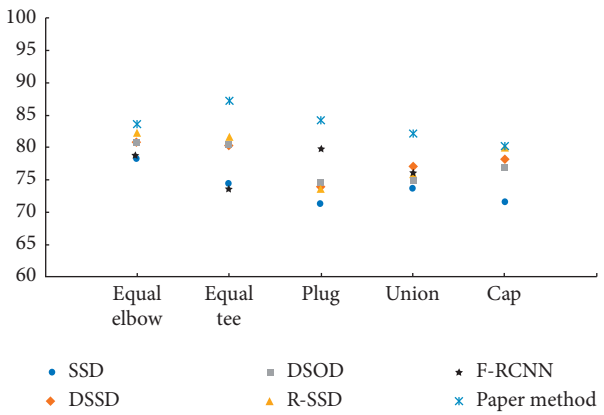

Figure 8: Comparison of recognition accuracy of different target detection algorithms.

Considering the requirement of real-time sorting in practical applications, a part of the target detection algorithm was selected to compare with the algorithm in this paper in terms of mAP index and detection speed, respectively, and the results are shown in Table 2.

From Table 2, we can see that the algorithm in this paper has a more significant improvement in detection accuracy and better performance than similar algorithms; the detection speed has also improved substantially compared to the original SSD algorithm, which can better balance the requirements of real-time detection and accuracy.

To evaluate the algorithm in the paper intuitively, the classical SSD300 model and the model designed in this paper are used for target detection separately. Figure 9 shows the single image detection results of the two models.

Comparison in Figure 9 shows that both models can complete the detection of workpiece images. However, compared with the SSD300 model, the overall classification confidence of the SSD model improved by the feature extraction layer has been significantly enhanced. For example, the category confidence of equal elbow1 (up), equal elbow2 (down), and union in Figure 9(b) is improved by 0.6, 0.5, and 0.3, respectively. In Figure 9(d), not only the overall recognition rate of all types of artifacts is improved but also the obscured targets (which cannot be recognized as a plug) are detected. And the equal elbow on the rightmost side is shown incorrectly in Figure 9(c) (identified as a cap). Thus, it is verified that the improved network model has good feature extraction capability and can further improve the detection accuracy of the workpiece images.

Table 2: Comparison of feature extraction time and pattern recognition time.

Models	mAP (%)	Detection speed (frame·s^{-1})
SSD	73.81	47.1
DSSD	78.27	12.1
YOLO	71.55	49.4
YOLO-V5	76.49	71.0
Paper method	83.50	24.8

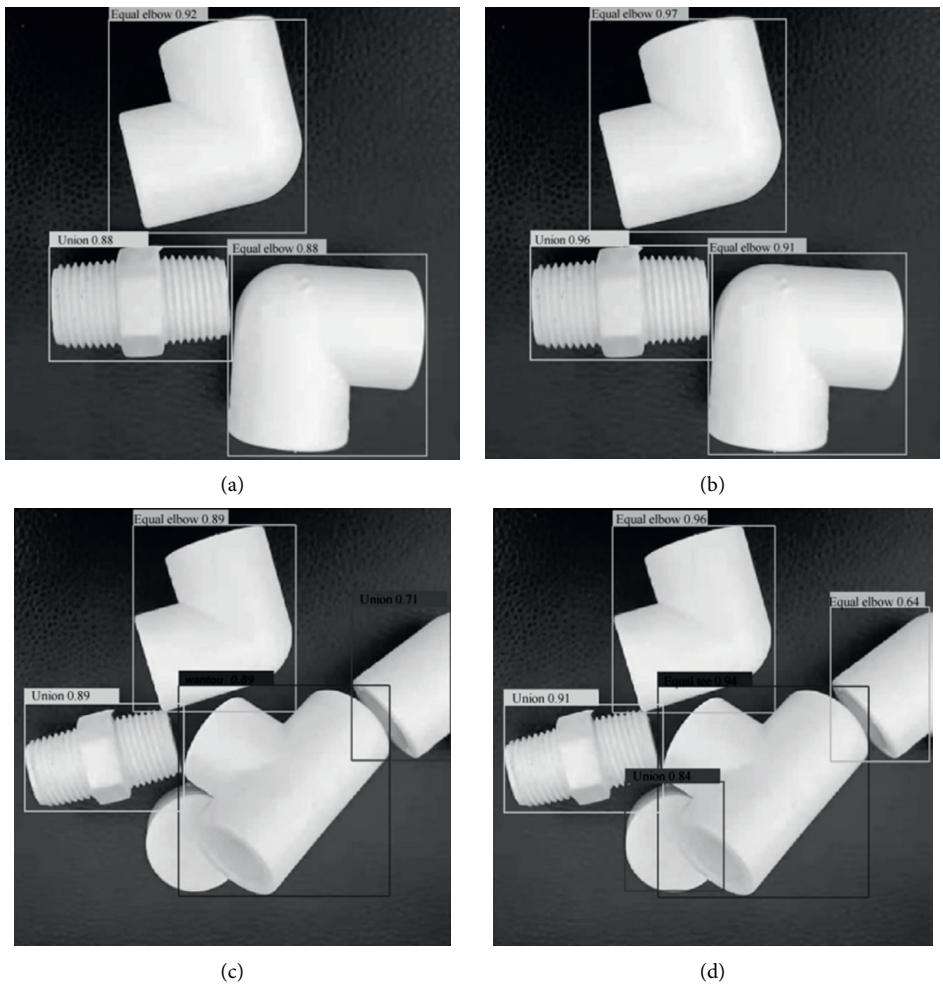

Figure 9: Comparison of single image detection results of two models. (a, c) SSD300 model. (b, d) Improved SSD model.

5. Conclusion

By introducing the Inception_Resnet_v2 structure to replace the feature extraction layer in the SSD model, the small-scale feature extraction capability of the model is improved, and the loss function of the model is optimized by increasing the rejection loss to enhance the detection effect of the network on stacked targets. The experimental analysis and result evaluation show that the improved SSD algorithm has good performance in mAP index and detection speed compared with similar algorithms, which enhances the robustness of SSD algorithm in unstructured scenarios and meets the requirements of real-time and accuracy of workpiece sorting detection, and has good reference value for the development of intelligent detection in the manufacturing industry.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] F. Qian, W. Zhong, and W. Du, "Fundamental theories and key technologies for smart and optimal manufacturing in the process industry," *Engineering*, vol. 3, no. 2, pp. 154–160, 2017.

[2] Y. Li, M. Yan, and X. Liu, "Workpiece intelligent identification and positioning system based on binocular machine vision," in *Proceedings of the 2021 IEEE 9th International Conference on Computer Science and Network Technology (ICCSNT)*, pp. 55–58, IEEE, Dalian, China, 2021, October.

[3] M. H. Ali, K. Aizat, K. Yerkhan, T. Zhantos, and O. Anuar, "Vision-based robot manipulator for industrial applications," *Procedia Computer Science*, vol. 133, pp. 205–212, 2018.

[4] T. J. Lukka, T. Tossavainen, J. V. Kujala, and T. Raiko, "Zenrobotics recycler–robotic sorting using machine learning," in *Proceedings of the International Conference on Sensor-Based Sorting (SBS)*, pp. 1–8, Aachen, Germany, 2014, March.

[5] A. Mizushima and R. Lu, "An image segmentation method for apple sorting and grading using support vector machine and Otsu's method," *Computers and Electronics in Agriculture*, vol. 94, pp. 29–37, 2013.

[6] T. Pranckevičius and V. Marcinkevičius, "Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification," *Baltic Journal of Modern Computing*, vol. 5, no. 2, p. 221, 2017.

[7] X. Gao, X. Ding, R. Hou, and Y. Tao, "Research on food recognition of smart refrigerator based on SSD target detection algorithm," in *Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science*, pp. 303–308, Wuhan, China, 2019, July.

[8] M. T. Zeren, S. K. Aytulun, and Y. Kirelli, "İnsansız haugeevshty faster R-CNN ak," *European Journal of Science and Technology*, no. 19, pp. 643–658, 2020.

[9] L. Jin and G. Liu, "An approach on image processing of deep learning based on improved ssd," *Symmetry*, vol. 13, no. 3, p. 495, 2021.

[10] L. Li, M. Fu, T. Zhang, and H. Y. Wu, "Research on workpiece location algorithm based on improved SSD," *Industrial Robot: The International Journal of Robotics Research and Application*, vol. 49, 2021.

[11] G. Shi, Y. Zhang, and M. Zeng, "A fast workpiece detection method based on multi-feature fused SSD," *Engineering Computations*, 2021.

[12] Z. Sun, X. Guo, X. Zhang, J. Han, and J. Hou, "Research on robot target recognition based on deep learning," in *Journal of Physics: Conference Series*vol. 1948, no. 1, IOP Publishing, Article ID 012056, 2021.

[13] D. Yang, C. Bi, L. Mao, and R. Zhang, "Contour feature fusion SSD Algorithm," in *Proceedings of the 2019 Chinese Control Conference (CCC)*, pp. 3423–3426, IEEE, Guangzhou, China, 2019, July.

[14] X. Dai, Y. Zhao, and C. Zhu, "A study of an improved rcnn network model for surface defect detection algorithm of precision workpiece and its realisation," *International Journal of Wireless and Mobile Computing*, vol. 19, no. 1, p. 95, 2020.

[15] M. Zhai, J. Liu, W. Zhang, C. Liu, W. Li, and Y. Cao, "Multi-scale feature fusion single shot object detector based on DenseNet," in *Proceedings of the International Conference on Intelligent Robotics and Applications*, pp. 450–460, Springer, Shenyang, China, 2019 August.

[16] J. Jeong, H. Park, and N. Kwak, "Enhancement of SSD by concatenating feature maps for object detection," 2017, https://arxiv.org/abs/1705.09587.

[17] W. Liu, D. Anguelov, D. Erhan et al., "Ssd: single shot multibox detector," in *Proceedings of the European Conference On Computer Vision*, pp. 21–37, Springer, Amsterdam, The Netherlands, 2016, October.

[18] M. Zhang, K. Pang, C. Gao, and M. Xin, "Multi-scale aerial target detection based on densely connected inception ResNet," *IEEE Access*, vol. 8, pp. 84867–84878, 2020.

[19] P. Lall, D. Iyengar, S. Shantaram, R. Pandher, D. Panchagade, and J. Suhling, "Design envelopes and optical feature extraction techniques for survivability of SnAg leadfree packaging architectures under shock and vibration," in *Proceedings of the 2008 58th Electronic Components and Technology Conference*, pp. 1036–1047, IEEE, Lake Buena Vista, FL, USA, 2008, May.

[20] S. Liu and D. Huang, "Receptive field block net for accurate and fast object detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 385–400, Munich, Germany, September 2018.

[21] L. Bo, T. Hai, and F. Qiang, "A small object detection method based on local maxima and SSD," in *AOPC 2019: AI in Optics and Photonics*vol. 11342, International Society for Optics and Photonics, Article ID 113420Q, 2019, December.

[22] B. Iancu, V. Soloviev, L. Zelioli, and J. Lilius, "ABOships-an inshore and offshore maritime vessel detection dataset with precise annotations," *Remote Sensing*, vol. 13, no. 5, p. 988, 2021.