The International Journal of Engineering and Science (IJES)
|| Volume || 10 || Issue || 5 || Series I || Pages || PP 59-66|| 2021 ||
ISSN (e): 2319-1813 ISSN (p): 20-24-1805

Object Detection under Similar Interference Based on EfficientDet

Zhenwen Sheng [1], Linzhong Wang [2*], Jing He [2]

[1] Shandong Xiehe University, *jinan, China*

[2] *College of Electrical and Information Engineering, Hunan University of Technology, Zhuzhou, China*

Corresponding Author: Linzhong Wang

---***ABSTRACT***---

In object detection, the most difficult interference is similar interference. The interference is relatively similar to the characteristics of the target object. When the interference feature extraction is less difficult than the target object, the use of the interference feature can better identify the target object. Therefore, a twin strategy, which uses interference with low detection difficulty, is proposed to provide the target object a non-single detection scale. Therefore, the interference and target objects are labeled at the same time, and the single classification problem is transformed into the multiclassification problem of detecting interference and target objects. Thus, interference is detected initially and does not interfere with the target object detection. Then, the EfficientDet model structure based on the twin strategy is proposed to study the object detection algorithm under similar interference. Finally, an experimental study is carried out. Experimental results show that the detection accuracy of the target object can be improved by more than 4%.

KEYWORDS: *Object detection, EfficientDet, Similar interference*

Date of Submission: 01-05-2021 Date of Acceptance: 15-05-2021

I. INTRODUCTION

The purpose of object detection is to find the target object in the measured image for precise positioning and then classify and even track. However, image processing is computationally intensive and complex. Deep learning has a strong ability to learn features from data automatically; thus, object detection technology based on deep learning is widely used in recent years[1, 2].

In the object detection process, many interference problems occur, such as background interference, missing interference, and laser interference. Background interference refers to the interference of light changes, camera shake, dynamic background, and shadows[3]. In some occasions, the information cannot be obtained comprehensively, resulting in incomplete image information called missing interference [4]. Laser negatively influences the detection characteristics of infrared photoelectric sensors and other targets; this condition is called laser interference[5].

The commonly used method to deal with background interference is background subtraction. Wang et al.[6] proposed a background-driven salient object detection (BD-SOD) method. When the image contains cluttered background and different target parts, the proposed model can greatly reduce the negative interference of cluttered background and different object parts. The image information are limited because they are difficult to collect completely; thus, signal model reconstruction and image restoration are commonly used in processing [7, 8]. To deal with laser interference, the location information of the interference point and the target should be determined in advance. However, with the development of deep learning, the information may be known, and the deep features of the image and the sensitivity to occlusion are used to eliminate the interference[9, 10].

Some interferences are similar, indicating that the interference is very similar to the target object, thereby causing great interference to the target object to be detected. For example, when the image has a frequency feature similar to the defect frequency signal, the similar frequency feature has greater interference than the experimental results, and the defect feature it is usually difficult to extract, thereby affecting the recognition result[11]. For this type of similar interference problem, when the detection difficulty of interference is lower than the target object, the proposed twin strategy is combined with EfficientDet to achieve target detection under similar interference.

The structure of this paper is as follows. The second section introduces the structure, experimental configuration, and experimental content of EfficientDet based on the twin strategy. The third section verifies the effectiveness of EfficientDet based on the twin strategy through comparative experiments. The fourth section presents the conclusion.

DOI:10.9790/1813-1005015966

II. EFFICIENTDET MODEL BASED ON TWIN STRATEGY

A twin strategy is proposed to address the problem of similar interference. To verify the effect of the twin strategy, it is combined with EfficientDet, a object detection network with speed and accuracy. Then, the twin strategy, the EfficientDet network, and the EfficientDet network based on the twin strategy are introduced.

2.1 Twin strategy

A twin strategy is proposed to deal with similar interference problems in object detection. Our detection target is A, interference B is relatively similar to A, and the detection difficulty of interference B is much lower than A; A and B are detected at the same time to improve the accuracy of A. Detection of A is unnecessary to avoid interference characteristics in addition to learning its own characteristics. Therefore, different from the traditional method of removing interference, the proposed approach continuously improves the accuracy of the network to eliminate the impact of interference. Informally, A and B are similar to twins. Interference A is shown in Figure 1 with interference B. Compared with A, B has fringe and wears a watch; thus, it is more distinctive. Therefore, in object detection, training the network to mine the gap between the target and the interference is unnecessary. We only need to identify the interference at the same time, transform a single classification into a multiclassification, and use the interference detection result to identify the target and avoid few boundary conditions that can be used for single classification[12].

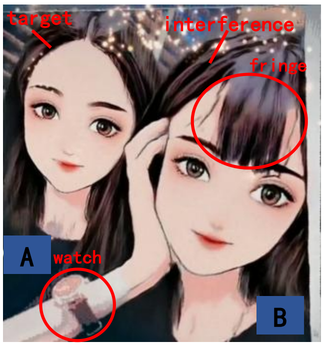

Figure 1: Common explanation of the twin strategy

Figure 2 shows the classification process of separately detecting the target to be tested and using the twin strategy. In the figure, the scattered edges are represented as target objects, and the clear edges are interference because the features of interference are more evident and easier to extract; thus, the detection difficulty is low. Originally, it is a target object and interference with high degree of similarity. The traditional method detects the target object and distinguishes it, as shown on the left side of Figure 2. When a difference in the detection difficulty between the two exists, as shown in the right side of Figure 2, the twin strategy is used to perform detection at the same time, requiring more attention. If it is an interference, then it is easier to detect. The influence of the target object detection result is small.

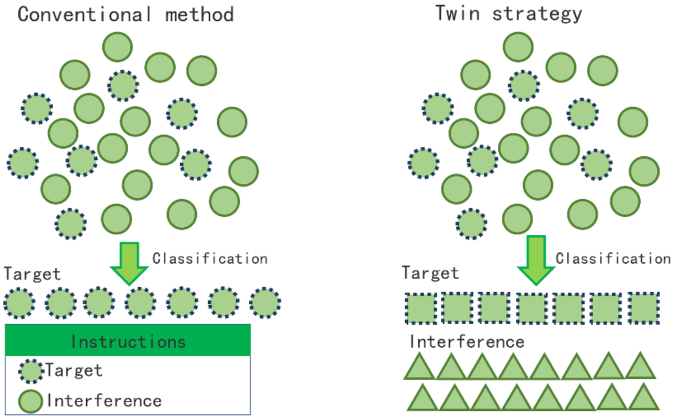

Figure 2: Comparison of single-detection target and twin strategy

2.2 EfficientDet network

EfficientDet[13] is a object detection network proposed by Google; it considers accuracy and speed and is mainly composed of backbone feature extraction network EfficientNet[14], enhanced feature extraction network BiFPN, and efficient head used to obtain prediction results.

The backbone feature extraction network of EfficientDet is EfficientNet. When network accuracy should be improved, network depth, network width, and image resolution are often the starting points. However, whether these three methods can be used together and whether a numerical relationship exists between scaling have not been studied prior to EfficientNet. Therefore, EfficientNet proposes a compound model scaling method, which increases the accuracy of the network from the three dimensions, network depth, network width, and image resolution.

EfficientDet’s enhanced feature extraction network is composed of multiple BiFPNs in series. Its function is to combine the five feature layers P3–P7 output by EfficientNet with weighted features to better balance feature information of different scales, and finally, five effective feature layers are output. Then, efficient head uses these effective feature layers to predict the results.

2.3 EfficientDet network based on the twin strategy

The overall framework of the EfficientDet network based on the twin strategy is shown in Figure 3. When the image containing the target object and the interference enters the EfficientDet network based on the twin strategy, the target object and the interference are extracted separately through EfficientNet, and the enhanced feature extraction of BiFPN is performed. Finally, the prediction result is obtained based on the feature.

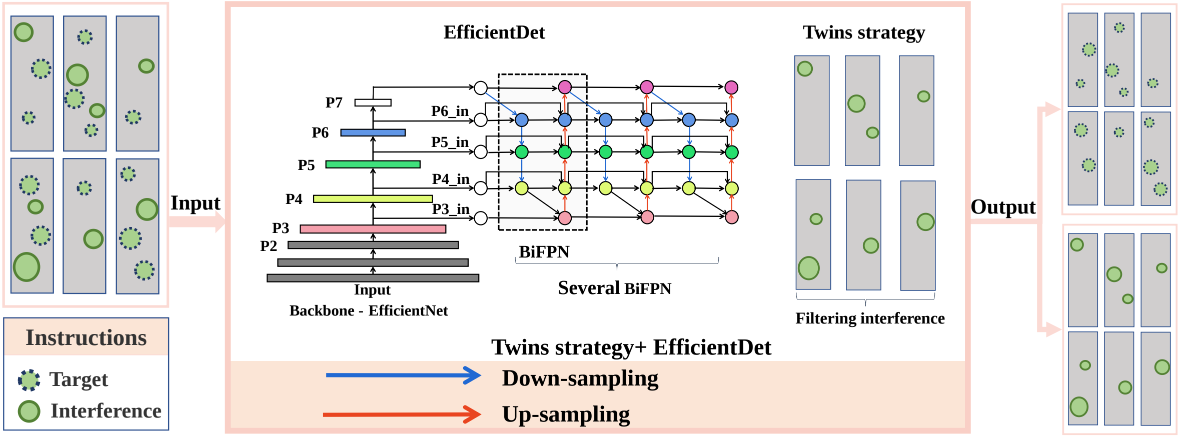

Figure 3: Overall framework of the EfficientDet network based on the twin strategy

The execution process of the EfficientDet network based on the twin strategy is shown in Figure 4. The image is scaled after entering the network, and then the backbone feature extraction network EfficientNet is entered for gradual downsampling and to obtain feature layers P1–P5. P1 and P2 do not enter the subsequent enhanced feature extraction network because they are shallow downsampling. P5 is down-sampled twice to obtain feature layers P6 and P7. The five feature layers, with the previous feature layers P3, P4, and P5, enter the enhanced feature extraction network connected by multiple BiFPN modules in series, and five effective feature layers are output. Finally, these feature layers are transmitted into ClassNet and BoxNet to obtain the final prediction results.

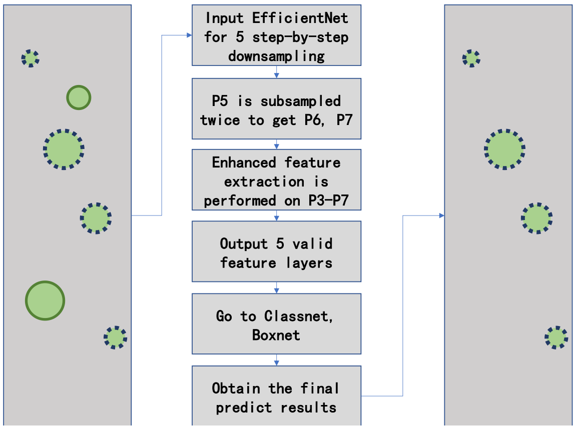
Input EfficientNet for 5 step-by-step downsampling
P5 is subsampled twice to get P6, P7
Enhanced feature extraction is performed on P3-P7
Output 5 valid feature layers
Go to Classnet, Boxnet
Obtain the final predict results

Figure 4: Execution process of the EfficientDet network based on the twin strategy

2.4 Experimental content of EfficientDet network based on the twin strategy

2.4.1 Experimental configuration

This experimental platform is the deep learning framework Pytorch, version 1.6.0, and the torchvision is 0.7.0; based on windows10, the environment configuration is Python 3.7.6 integrated environment under Anaconda 4.8.2; the graphics card is NVIDIA GeForce RTX 2060 SUPER, CuDA 10.1; the Python terminal has been built with OpenCV and Keras environments, and the versions are OpenCV–Python 4.4.0 and Keras 2.3.1.

2.4.2 Experimental parameters

The version used in this article is EfficientDet-D1. EfficientDet uses $Smooth\ L_1$ function to calculate loss. The function $Smooth\ L_1$ obtains the regression loss of the prediction results of all positive label boxes, and the expression of it is as follows[15].

$$Smooth\ L_1 = \begin{cases} 0.5x^2, & |x|<1 \\ |x|-0.5, & x<-1\ or\ x>1 \end{cases} \tag{1}$$

$Focal\ loss$ function is used to obtain the cross-entropy loss of all types of prediction results that have been considered. $Focal\ loss$ is used to control the weight of positive and negative samples, as well as the weights of the samples that are easily and difficultly classified. The specific implementation is as follows [16].

First, the weight of the positive sample is controlled. The cross-entropy loss function of two classifications is considered an example, as follows:

$$CE(p,y) = \begin{cases} -\log(p) & if\ \ y=1 \\ -\log(1-p) & \text{otherwise.} \end{cases} \tag{2}$$

The cross entropy function is simplified using the following formula:

$$p_t = \begin{cases} p & if\ \ y=1 \\ 1-p & otherwise. \end{cases} \tag{3}$$

Equations 2 and 3 can be transformed into the following:

$$CE(p,y) = CE(p_t) = -\log(p_t) \tag{4}$$

To reduce the impact of negative samples on the loss, a coefficient ω_t can be added prior to the loss function,

$$\omega_t = \begin{cases} \omega & if\ \ y=1 \\ 1-\omega & otherwise \end{cases} \tag{5}$$

Then, it is added to equation (4) to obtain the following:

$$CE(p_t) = -\omega_t \log(p_t) \tag{6}$$

All variables are substituted to obtain the following expression:

$$CE(p,y,\alpha)=\begin{cases}-\log(p)*\omega & if \ \ y=1\\ -\log(1-p)*(1-\omega) & if \ \ y=0\end{cases} \tag{7}$$

Second, the weights of the samples, which are easily and difficultly classified, are controlled using the above two-classification example. A modulation factor $(1-p_t)^{\lambda}$ is set, and then focal loss is obtained as follows:

$$FL(p_t)=-(1-p_t)^{\lambda}\log(p_t) \tag{8}$$

Therefore, when the value of p_t of a certain category is very small or even tends to 0 (the classification is very difficult), the value of $FL(p_t)$ is very large. When the value of p_t of a certain category is very large and close to 1 (the difficulty of the classification is small), the value of $FL(p_t)$ is very small. When λ approaches 0, the focal loss is the traditional cross-entropy loss function.

When the weights of controlling positive and negative samples are combined with the weights of the samples, which are easily and difficultly classified, the expression of FL is as follows:

$$FL(p_t)=-\omega_t(1-p_t)^{\lambda}\log(p_t) \tag{9}$$

When determining positive or negative samples, intersection over union (IOU)[17] is adopted. If the degree of coincidence between the prior frame and the actual frame is greater than 0.5, then it is a positive sample, and it is ignored in the range of 04–0.5. If the coincidence degree is less than 0.4, then it is a negative sample. If it is a positive sample, then ω_t of focal loss is ω, and p_t is p. If it is a negative sample, then ω_t is $1-\omega$, p_t is p, and the corresponding $FL(p_t)$ is as follows:

$$FL(p_t)=\begin{cases}-\omega(1-p)^{\lambda}\log(p) & If \ the \ anchor \ is \ a \ positive \ sample\\ (\omega-1)p^{\lambda}\log(1-p) & If \ the \ anchor \ is \ a \ negative \ sample\end{cases} \tag{10}$$

2.4.3 Network training process

During the training process, part of the first 50 epochs is frozen for training, and the learning rate is e^{-3}. The next 50 epochs are not frozen, the learning rate is e^{-4}, and the batch size is 4. When training, the hyperparameter ω_t in formula (9) is set to 0.25 and λ to 2.

EfficientDet has nine a priori frames for each region. After the five effective feature layers are obtained, multiple depth separable convolutions are used for feature integration, and one depth separable convolution is used to adjust the number of channels. Finally, the prediction results are obtained.

The training process of the EfficientDet network based on the twin strategy is as follows:

a. The randomly assigned result of the training set, validation set, and test set in the data set containing the target detection object and interference is 3:1:1, and the result is imported into the EfficientDet network;

b. The TXT file that records the classification category and the category in the code are modified to stain and defect, and then the code is operated to save the names of the files in the training set, validation set, and test set to train.txt, val.txt, and test.txt, respectively. The name of an image in each line is recorded;

c. The code is operated to obtain the tag information and generate 2007.train.txt, 2007.val.txt, and 2007.test.txt. This code contains the location information, image name, target category in the image, and the values of the upper, lower, left, and right bounds of the tag in the training set, verification set, and test set. In addition, each line records an image information;

d. The pretrained model EfficientDet-d1.pth of the COCO2017 data set on EfficientDet-D1 is loaded to detect the target object and the interference separately and to compare their detection difficulty. If the detection accuracy of the target object is lower than the interference, then proceed to e; otherwise, proceed to f. The twin strategy cannot be used;

e. Part of the network is frozen and trained for 50 times, and then unfrozen for 50 times;

f. The training is ended.

III. RESULT

Wheelset, as an important support and running component of the train, causes defects in its tread due to the rolling contact of the wheel and rail [18-22]. In addition, the operating environment of the wheelset tread inevitably produces stains on its surface. These stains include mud and other lumps formed on the surface of the wheelset tread; they are very similar to defects. Wheelset tread defects are the target object, and wheelset tread stains are interference. Therefore, the wheelset tread is used as the experimental object to verify the effectiveness of the EfficientDet network based on the proposed twin strategy.

3.1 Verification of the conditions of using the twin strategy

The detection difficulty must be distinguished between the interference and the target object to use the twin strategy. When the detection difficulty of the interference is lower than the target object, the target object can be better detected. The wheelset tread is considered an example, and the EfficientDet network is used to detect the defects and stains of the wheelset tread surface separately. In addition, the self-built wheelset tread data set is adopted. At the same time, due to the large gap between the number of interferences and the number of target objects, the data set is divided into two types with different ratios of interference stains to defects. The ratio of stains to defects in data set 1 is 5:1, and the total number of images in the data set is 909. The ratio of stains to defects in data set 2 is 2.7:1, and the total number of images in the data set is 1061. Under the same experimental conditions, the separate detection results of the wheelset tread defects and stains on the two data sets are shown in Figures 5 and 6.

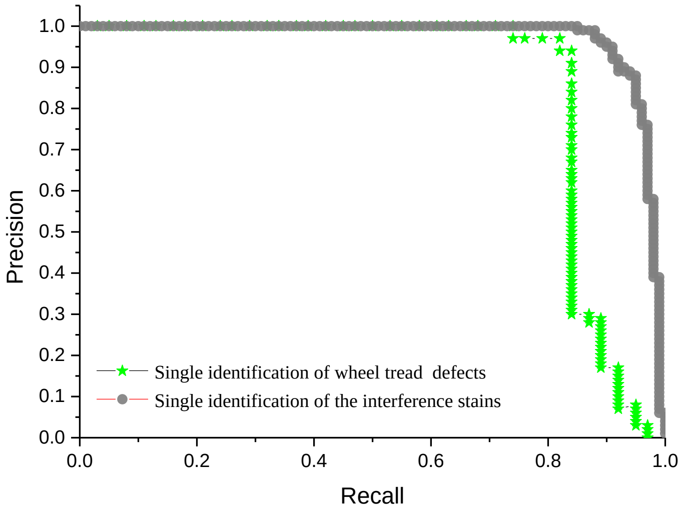

Figure 5: Individual inspection results when the ratio of stain interference to defect is 5:1

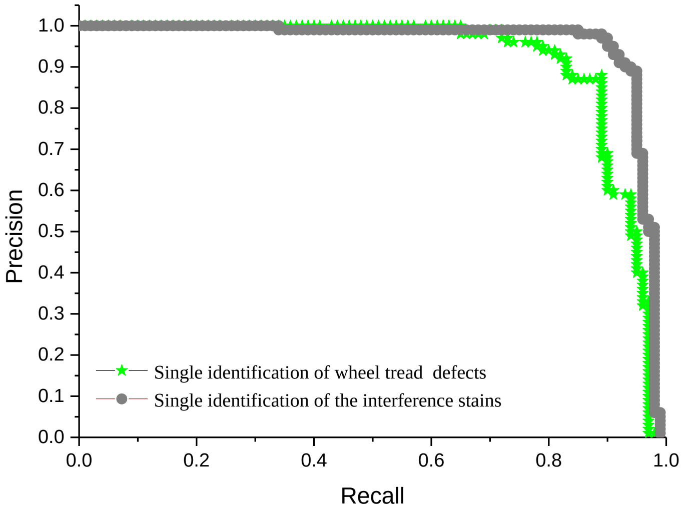

Figure 6: Individual inspection results when the ratio of stain interference to defect is 2.7:1

In Figures 5 and 6, the green line represents the PR curve of single identification of wheel tread defects, and the gray line is the PR curve of single identification of stain. Evidently, the detection accuracy of stains is much higher than that of the defect. Therefore, the defects and interference stains on the tread surface of the wheelset satisfy the applicable conditions of the twin strategy. Therefore, subsequent verification of the effectiveness of the twin strategy can be carried out.

3.2 Experimental results of EfficientDet network based on the twin strategy

The weights obtained after network training are loaded into the model, and the verification set is verified. Then, the AP value, recall rate, and precision rate can be obtained. When the defect–stain ratios are 1:5 and 1:2.7, the PR curves with and without twins are shown in Figures 7 and 8. The green line represents the PR curve of single identification of wheel tread defects, and the gray line is the PR curve of the result of defect detection after using twin strategy. The results of AP value, recall rate, and precision rate are shown in Table 1.

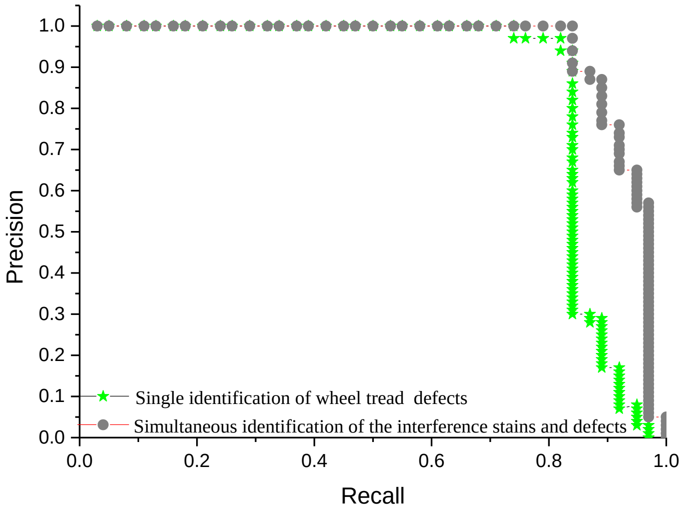

Figure 7: Under the defect:stain ratio of 1:5; test results using separate test and twin strategy

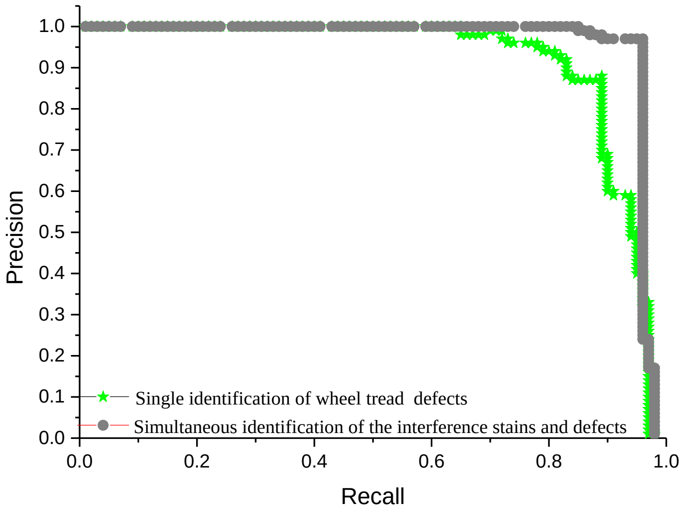

Figure 8: Under the defect:stain ratio of 1:2.7; test results using separate test and twin strategy

Experimental results	Defect:Stain=1:5		Defect:Stain=1:2.7	
	Single detection	Twin strategy	Single detection	Twin strategy
Precision	86.49%	94.12%	93.75%	97.65%
Recall	84.21%	84.21%	79.79%	88.30%
AP value	86.11%	91.15%	91.85%	95.90%

Table 1:Recall rate, precision rate, and AP value of the EfficientDet network based on the twin strategy

Figures 7 and 8 show that the EfficientDet network based on the twin strategy can improve the detection accuracy of the target object in the case of similar interference when the sample imbalance degree is different; it confirms the effectiveness of the twin strategy. After using the twin strategy, when the defect:stain ratio is 1:5, the defect detection accuracy is increased from 86.11% to 94.01%, and the accuracy is increased by 7.9%, which is a significant improvement. When the defect:stain ratio is 1:2.7, after using the twin strategy, the defect detection accuracy is increased from 91.85% to 95.90%, and the accuracy is increased by 4.05%. At the same time, Table 1 suggests that this method not only improves the detection accuracy, but also enhances the recall rate and accuracy rate; it mainly improves the accuracy rate, that is, it can detect as many target objects as possible.

IV. CONCLUSION

The similar interference in the target detection interference problem is studied. The interference is similar to the target object and has a greater impact on the detection result. A twin strategy is proposed to deal with similar interference. Combining the twin strategy with the target detection network EfficientDet has proven that this architecture can effectively deal with similar interference problems and improve the detection accuracy of target objects. The twin strategy can also be applied to other different applicable objects and can be combined with different target detection networks for target recognition.

REFERENCE

[1]. Xiongwei W, Sahoo D, Hoi S C H. Recent advances in deep learning for object detection[J]. Neurocomputing. 2020, 396: 39-64.

[2]. Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. science. 2006, 313(5786): 504-507.

[3]. Zeng D, Zhu M, Kuijper A. Combining Background Subtraction Algorithms with Convolutional Neural Network[J]. Journal of Optical Technology. 2018, 1(88): 25-29.

[4]. C. H M, J. C W. Replacing Missing Data Between Airborne SAR Coherent Image Pairs[J]. IEEE Transactions on Aerospace and Electronic Systems. 2017, 53(6): 3150-3158.

[5]. Pavlov N I, Rezunkov Y A. Influence of laser interference on the detection capabilities of an infrared optoelecronic surveillance system[J]. Journal of Optical Technology. 2021, 88(1): 25-29.

[6]. Z. W, D. X, S. H, et al. Background-Driven Salient Object Detection[J]. IEEE Transactions on Multimedia. 2017, 19(4): 750-762.

[7]. P. G, H. R M. Muscle Activity Map Reconstruction from High Density Surface EMG Signals With Missing Channels Using Image Inpainting and Surface Reconstruction Methods[J]. IEEE Transactions on Biomedical Engineering. 2017, 64(7): 1513-1523.

[8]. J. B, B. K, S. L, et al. Bistatic ISAR image reconstruction using sparse-recovery interpolation of missing data[J]. IEEE Transactions on Aerospace and Electronic Systems. 2016, 52(3): 1155-1167.

[9]. Xiang G, Jing H, Lijun R, et al. Method of quality assessment based on convolution feature similarity for laser disturbing image[C]. 2020.

[10]. Cristiane N S, Sophie C, Lorenz M, et al. Visible and near-infrared laser dazzling of CCD and CMOS cameras[C]. 2018.

[11]. WANG Yuntao，SHENG Qinghua，GUO Chenjie，LI Zhu. Surface defect recognition method based on multiple Siamese neural network [J]. Journal of Computer Applications. 2020, 40(S2): 225-229.(in Chinese)

[12]. Perera P, Oza P, Patel V M. One-Class Classification: A Survey[J]. arXiv preprint arXiv:2101.03064. 2021.

[13]. Tan M., Pang R., Le Q.V. Efficientdet: scalable and efficient object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 10781-10790.

[14]. Tan M., Le Q.V. Efficientnet: rethinking model scaling for convolutional neural networks[C]. Proceedings of the 36th International Conference on Machine Learning. 2019: 6105-6114.

[15]. Girshick R. Fast r-cnn[C]. Proceedings of the IEEE International Conference on Computer Vision. 2015: 1440-1448.

[16]. Lin T., Goyal P., Girshick R., et al. Focal loss for dense object detection[C]. Proceedings of the IEEE International Conference on Computer Vision. 2017: 2980-2988.

[17]. Sun X, Gu J, Huang R, et al. Surface defects recognition of wheel hub based on improved faster R-CNN[J]. Electronics. 2019, 8(5): 481.

[18]. Liu J, Liu L, He J, et al. Wheel/Rail Adhesion State Identification of Heavy-Haul Locomotive Based on Particle Swarm Optimization and Kernel Extreme Learning Machine[J]. Journal of advanced transportation. 2020, 2020: 1-6.

[19]. He J, Yang B, Zhang C, et al. Integrated cooperative braking algorithm of non-linear electric multiple units with external disturbance[J]. The Journal of Engineering. 2019, 2019(2): 8937-8941.

[20]. He J, Zuo X, Zhang C, et al. Anti-slip control based on optimal slip ratio for heavy-haul locomotives[J]. The Journal of Engineering. 2019, 2019(6): 9069 – 9074.

[21]. Frhling R, Spangenberg U, Reitmann E. Root cause analysis of locomotive wheel tread polygonisation[J]. Wear. 2019, 432-433.

[22]. Zhang C, Xiang C, Liu J, et al. Deep Sparse Autoencoder for Feature Extraction and Diagnosis of Locomotive Adhesion Status[J]. Journal of Control Science and Engineering. 2018, 2018: 1-9.