Multi-Scale Feature Pair Based R-CNN Method for Defect Detection

1st Zihao Huang
Department of Computer Science
Guangdong University of Technology
Guangzhou, China
zihaolalala@163.com

2nd Hong Xiao
Department of Computer Science
Guangdong University of Technology
Guangzhou, China
wh_red@163.com

3rd Rongyue Zhang
Department of Computer Science
Guangdong University of Technology
Guangzhou, China
flowet@163.com

4th Hao Wang
Department of Computer Science
Norwegian Univ. of Sci. & Tech.
Norway
hawa@ntnu.no

5th Cheng Zhang
Department of Computer Science
Guangdong University of Technology
Guangzhou, China
2529091204@qq.com

6th Xiucong Shi
Department of Computer Science
Guangdong University of Technology
Guangzhou, China
xiuc_shi@163.com

***Abstract*—The traditional defect detection algorithms based on image registration, image contrast and other image processing algorithms are only limited to a single defect. Though deep-learning-based object detection algorithms can be used to detect a variety of different defects, the state-of-the-art deep-learning-based object detection algorithms still have low detection accuracy on small size defects. Basing on Cascade R-CNN in this paper, a new multi-scale feature extraction method—-the Multi-Scale Feature Pair—-is proposed and is used to establish a defect detection model for metal can products of an enterprise. Experimental results show that the accuracy (AP@0.5) of our improved model is 6.1% higher than Cascade R-CNN.**

***Keywords*-Cascade R-CNN; Multi-Scale Feature Pair; ROI Boundary Extension; Defect Detection**

I. Introduction

Product defect detection consists of locating defect area and identifying defect type. Take the defect detection of metal can products in an enterprise as an example, there are thousands of kinds of metal cans, and the purpose is to mark the defect areas and the types of defects with rectangular boxes in the product images. The defects which need to be detected include scratches, poor printing and uneven surface. It is suitable for using deep-learning-based object detection method to detect multitype defects in these kinds of metal cans. However, different defect sizes vary greatly, where the area of scratches is generally small and the area of uneven printing is generally large, as a result the current deep learning detection algorithm cannot meet the requirements.

Deep learning object detection algorithm mainly includes one-stage algorithm [5, 7. 12, 28] and two-stage algorithm [4, 10, 11, 15, 16]. The one-stage algorithm extracts image features through the multi-layer convolutional neural network (CNN). Then, in the image features, the sliding window and anchor [6, 7, 8, 9, 18] are used to set the default bounding box (Bbox). Then the image features are input into two branches of networks, which respectively classify the object and regress the bbox for features of each anchor box. Meanwhile, the first stage of a two-stage algorithm uses selective search [1, 2, 3, 23] or region proposal network(RPN) [4, 10, 11, 13, 14] to predict the region proposals(ROI) in the image, and then extracts the features of each ROI through ROI pooling, and inputs the ROI features into two branches of networks respectively for object classification and Bbox regression. As R-CNN, SPP-Net, Fast R-CNN, Faster R-CNN, SSD series, YOLO series, RON and other algorithms are proposed, the object detection method has gradually become mature, which provides many successful ideas and conclusions for the object detection field.

Almost all the state-of-the-art object detection algorithms use the low layer and high layer features of CNN at the same time for object detection. The difference only lies in the different combination of features from different layers. After multiple layers of CNN, CNN will extract the features that are favorable to the task of the image, and constantly filter the features that have little impact on the task in the sampling steps such as pooling. For example, for object classification, the image features are extracted from higher layers which are more abstract features with more semantic information, where Bbox regression requires the features to be extracted from lower layers. In this way, the extracted features will have more detail feature of the object, and it is easier to predict the object boundary. In addition, because the receptive field of high layer features is larger than that of low layer features, high layer features are beneficial to the detection of big object and low layer features are beneficial to the detection of small object.

In this paper, basing on Cascade R-CNN method, a new combination method of high layer CNN features and low layer CNN features is proposed to meet the detection needs of defects at different scales. Experiments are carried out in the metal can defect detection task. And the experimental

results show that our improved method is superior to other object detection methods.

II. Related Work

Because of the proposal of basic CNN architecture such as Resnet [17], great progress has been made in Faster R-CNN [4] which uses Resnet as RPN. Therefore, more and more algorithms have been improved on its basis.

In terms of the feature extraction network, MSCNN [11] and HyperNet [26] use different layers of CNN features in feature extraction network to solve the problem of multi-scale object detection. Low layer features are used to detect small objects, while high layer features are used to detect big objects, and meanwhile add deconvolution layer before Bbox regression layer to improve the resolution of the ROI features. It helps to improve the accuracy of object localization. PVANET[13] adopts C.ReLU + Residual structure in the low layer of the RPN and Inception + Residual structure in the high layer of the RPN, and uses a HyperNet structure to combine the high layer features and low layer features to predict the ROI, so as to solve the problem of multi-scale object detection.

In terms of RPN, FPN [15] uses deep CNN to construct multi-layer feature pyramid, and creates a structure connecting from high layer features to low layer features to extract multi-scale features of images. CRAFT [27] adds a Fast R-CNN binary classifier between RPN and the second stage to further select the RPN-generated proposals and leave some high-quality proposals. In the second stage, a binary classifier for N categories (excluding the background category) was cascaded after the original classifier for more refined object detection.

In terms of ROI processing, Cascade R-CNN [10] points out the disadvantage of using a single IoU threshold to filter ROI, and cascades two object classification and Bbox regression structure (the same structure as the second stage) after the second stage of Faster R-CNN, which is equivalent to input ROI features serially into three classification and Bbox regression structure. IoU threshold used by three classification and bbox regression structure are gradually increased instead of using a single IOU threshold to filter ROI to alleviate the problem of large gap between the positive and negative samples quantity, which helps to improve the accuracy of object detection.

In terms of loss function design, RetinaNet [18] proposes the fundamental reason why the one-stage algorithm is faster and less accurate than the two-stage object detection algorithm. To solve this problem, a new loss function called Focal Loss is proposed, and a new network called RetinaNet is created to verify the ability of Focal Loss.

At present, deep learning object detection algorithm has been applied in the field of product defect detection. Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences has developed a deep learning based detection method for stainless steel surface defects of auto parts (based on YOLO [5] object detection algorithm), which can effectively improve the accuracy of detection. ViDi Systems SA has developed a commercial software for defect detection of industrial image using deep learning. Kuang K [29] applies SSD deep learning object detection algorithm to defect detection of aero-engine, and the accuracy of defect detection reaches 89.36%. The results of defect detection in these applications are difficult to compare with each other since they do experiments in different datasets, but the performance of these algorithms is roughly the same as they do in public datasets. When using deep learning object detection algorithm for defect detection, it has high accuracy with simple background and big object size, but it is difficult to detect small object defect.

III. Multi-Scale Feature Pair Based R-CNN Method for Defect Detection

A. Cascade R-CNN

Cascade R-CNN is an improvement of Faster R-CNN, and the defect detection method in this paper is an improvement of Cascade R-CNN algorithm. Both of the algorithms are two-stage object detection algorithms, and include the region proposed network (RPN), ROI pooling, Bbox regression, and object classification.

Regional proposal network (RPN) is the first step of two-stage object detection algorithm. RetinaNet [18] points out that two-stage algorithm has a higher detection accuracy than one-stage algorithm, mainly because two-stage algorithm has greatly alleviated the problem of "unbalanced proportion of positive and negative sample number of proposals" after it passes through the RPN. Therefore, RPN is the most critical step in two-stage algorithm. RPN network is generally composed of several convolution layers and two branches of fully connected network which respectively used for region proposal prediction and object classification (binary classification, background and not background). As shown in the yellow part in figure 1, there is a feature extraction network (basic CNN architecture such as Resnet [17], Inception [25], etc.) before RPN. To get the proposals, combine the CNN features of different layers in the feature extraction network (multi-scale features) and input them into RPN.

ROI pooling is to extract fixed size feature maps of ROI on image features according to different size ROI areas. As shown in figure 2 (a), the size of the proposal area obtained by the RPN network is different, so it cannot be directly input into the second step of the network. ROI pooling can be carried out in a way similar to spatial pooling pyramid in SPPNet [2], so that feature map in basic CNN can be reused and detection speed can be improved.

The second step of two-stage object detection algorithm is the object classification and Bbox regression network, as shown in figure 2(b). The ROI feature is input into the

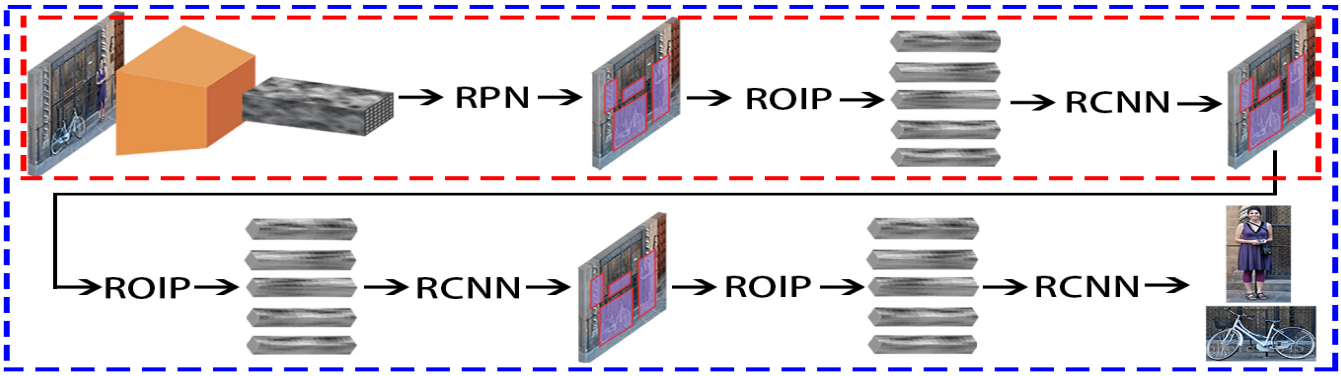

Figure 1: Faster R-CNN structure(red box), Cascade R-CNN structure(blue box)

object classification and Bbox regression layer after passing through several convolution layers. The loss function for Bbox regression is smoothed L1, and the object classification loss function is a classical cross entropy function.

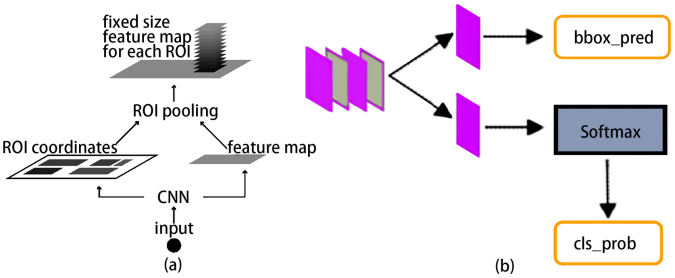

Figure 2: (a)ROI pooling, (b)Object classification and bbox regression network

B. Multi-Scale Feature Extraction Analysis and Improvement

Multi-scale feature extraction is used to detect different size object. As shown in figure 3, the multi-scale feature extraction methods generally include image pyramid (figure 3(a)) and feature pyramid (figure 3(b)). The image pyramid: resize the original image to a variety of different sizes, and then respectively extract the features of each image. Feature pyramid: the original image is processed through multiple layers of CNN, and then extract the CNN features of different layers and combine them into multi-scale features. These two multi-scale features extraction methods can be combined or improved to form a variety of feature extraction networks.

For different object, top layer features are good for big object detection, while low layer features are good for small object detection. For the same object, high layer features are advantageous to object classification, while low layer features are advantageous to Bbox regression. In most state-of-the-art object detection algorithms, only different layers of CNN features are considered for object detection at

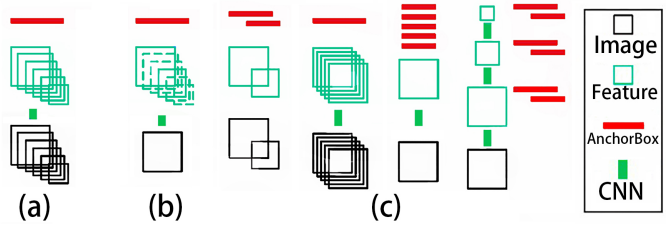

Figure 3: (a)Image pyramid, (b)Feature pyramid, (c)Other

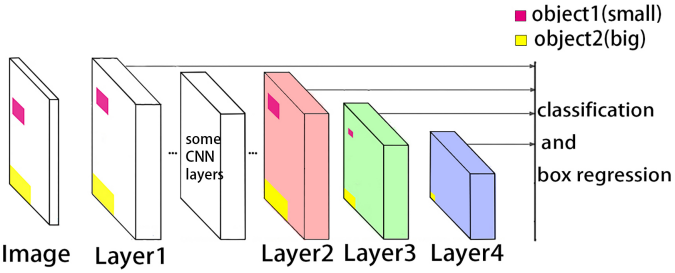

Figure 4: Multi-Scale feature extraction methods in most papers

different scales, but how to use different layers of CNN features for classification and Bbox regression of the same object is not considered. Take the feature extraction network of figure 4 as an example, the yellow and purple blocks represent different objects and their multiscale features can be extracted from different layers of the CNN architecture. In many state-of-the-art object detection algorithms, the features in yellow block of layer4 (high layer) is used for big object (object2) classification and Bbox regression, and the features in purple block in layer2 (low layer) is used for small object (object1) classification and Bbox regression. This method solves the problem that object detection at different scales requires different layer features. However, classification and Bbox regression of the same object use the same layer of features, which does not take into account the problem that the classification and Bbox regression of

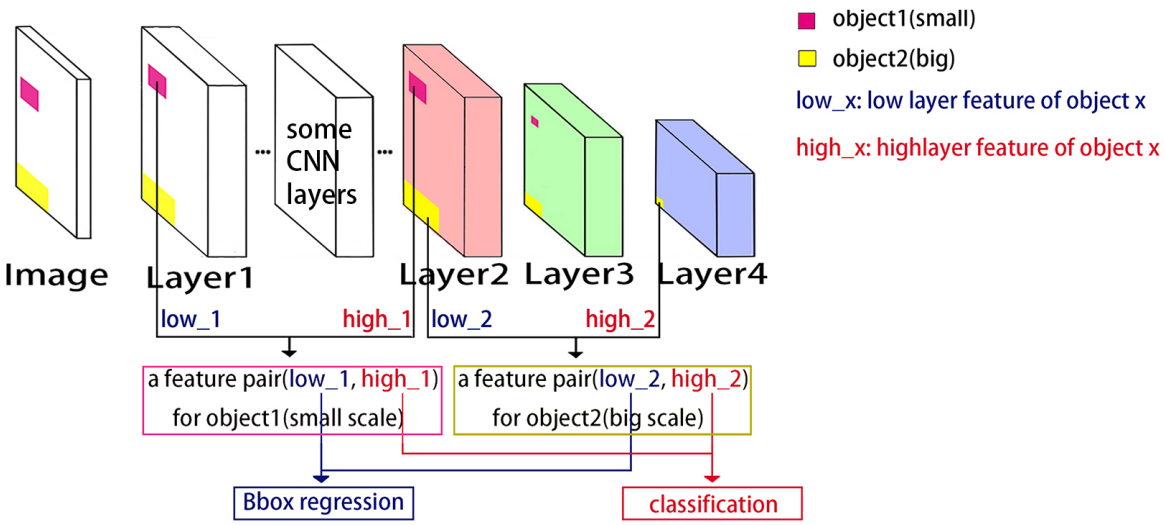

Figure 5: Improvement: Multi-Scale Feature Pair

the same object need different layer features.

In order to use different layers of CNN features for object classification and Bbox regression for the same object and improve the accuracy of object detection, we propose a new multi-scale feature combination method, called Multi-Scale Feature Pair.

Take the feature extraction network of figure 5 as an example, the features in yellow blocks of layer4 and layer2 are combined into a pair of large-scale object feature pairs, and the features in purple blocks of layer2 and layer1 are combined into a pair of small-scale object feature pairs. These two pairs of features constitute multi-scale feature pairs. In each pair of features, high and low layer features belonging to the same object are included. Low layer features are used for object Bbox regression while high layer features are used for object classification, which makes detection more accurate.

C. ROI Boundary Extension

Extension of ROI boundary means that the central point of the ROI area predicted by RPN remains unchanged, and the width and height of the ROI area are increased to 1.5 times the original size. Then the ROI Pooling and the second stage will be carried out after the extension. This improvement refers to the MRCNN [24] algorithm. In MRCNN, the features sampled around the ROI area and the features of image segmentation are also used to improve the object detection accuracy. After extending the boundaries of the ROI, the ROI features contain more detailed information, which is conducive to the accuracy of the object positioning.

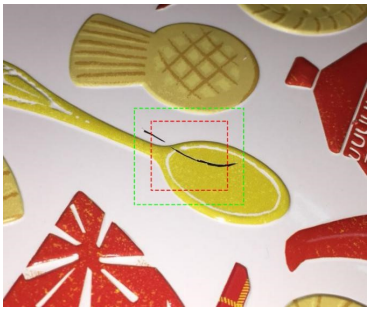

Figure 6: Before the extension of the ROI boundary(red box), After the extension of the ROI boundary(green box)

IV. Experimental Result and Analysis

A. Model Evaluation

In this paper, the commonly used AP@0.5 in object detection is used to evaluate the accuracy of our model,

$$Accuracy = AP@0.5 = \frac{I_{correct}}{I} \tag{1}$$

where $I_{correct}$ is the number of objects correctly predicted and I is the number of all predicted objects. $I_{correct}$ is computed as:

$$I_{correct} = \sum_{i=1}^{I} \begin{cases} 1, & Max_{j=1}^{J} IoU(b_i, g_j) \geq 0.5 \quad and \quad c_i = c_j^{'} \\ 0, & otherwise \end{cases} \tag{2}$$

Where (b_i, c_i) denotes (Bbox, class) of the object predicted by detection model, (g_j,$c_j^{'}$) denotes (ground truth Bbox, ground truth class of the object) of the input image. As shown in figure 7, IoU(Intersection over Union) means:

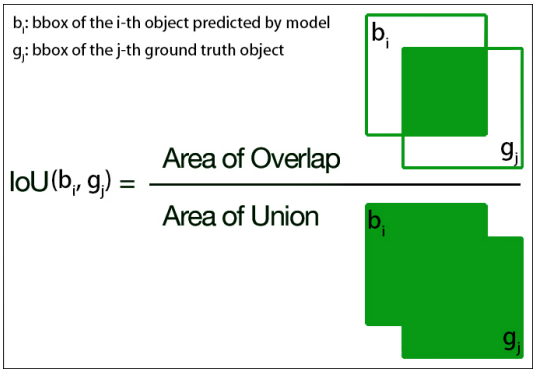

Figure 7: IoU(Intersection over Union)

B. Experimental Results

In this paper, 500 images of metal cans are used for the experiment. 400 of them are for training and 100 of them are for verification. The image size is 2048*2448, and the defects include scratches, poor printing and uneven surface. The detection results are shown in figure 8.

As shown in figure 8, most areas of the defects only take up a very small part of the image, and factors such as the printing pattern and texture of the product and the influence of ambient light during data collection have great interference on defect detection, making it much more difficult to detect defects in metal cans than the commonly used public datasets.

Figure 8: Detection results

C. Comparison with State-of-the-art Algorithm

In this paper, four current state-of-the-art object detection algorithms and the base Cascade R-CNN algorithm are used for experiments to verify the performance of our improvement. Table 1 shows the comparison of various algorithms:

Table I: Comparison with the state-of-the-art object detection algorithms on metal can dataset

	backbone	AP@0.5
Faster R-CNN[4]	Resnet101	30.90%
SSD[7]	Resnet101	27.02%
FPN[15]	Resnet101	30.51%
R-FCN[16]	Resnet101	30.62%
Cascade R-CNN[10]	**Resnet101**	**32.94%**

The AP@0.5 of the current state-of-the-art object detection algorithm in the metal can defect detection dataset in this paper is lower than that when it is applied to COCO dataset (in COCO2015, Faster R-CNN_Resnet101_AP@0.5 = 59.0%, SSD_VGG16_AP@0.5 = 43.1%). Therefore, it is very difficult to detect defects in the dataset used in this paper. As shown in table 1, Cascade R-CNN algorithm achieves the highest detection accuracy, so this paper chooses Cascade R-CNN as the prototype of defect detection model.

Table II: Comparison between the improved Cascade R-CNN and the original algorithm

	Multi-Scale Feature Pair	ROI Boundary Extension	AP@0.5
Cascade R-CNN			32.94%
Cascade R-CNN+	✓		38.27%
Cascade R-CNN+		✓	32.23%
Cascade R-CNN++	✓	✓	**39.04%**

As shown in table 2, the Multi-Scale Feature Pair proposed in this paper greatly improves the performance of defect detection, while the ROI Boundary Extension improves little. When these two improvements are combined, the AP@0.5 of our detection method is 39.04%, 6.1% higher than the original Cascade R-CNN method.

V. Conclusion

Based on Cascade R-CNN algorithm, a defect detection model for metal can is established in this paper. Two improvements are made on this basis: (1) aiming at the multi-scale object detection problem, a new multi-scale feature extraction method, called Multi-Scale Feature Pair, is proposed and (2) before ROI Pooling, expand the ROI area output by RPN by 1.5 times to enhance the detailed information of the object and improve the detection accuracy. Compared to other state-of-the-art object detection algorithms, the method in this paper greatly improves the performance of object detection. Due to the great success of Focal loss in one-stage object detection algorithm, in future work, we plan to combine multi-scale feature pairs and Focal loss to improve the performance of one-stage object detection algorithm.

ACKNOWLEDGMENT

The work described in this paper was partially supported by: Research on Self-Organizing Elasticity Enhancement Strategy of Intelligent Manufacturing IoT Network, National Natural Science Foundation of China(No.61672170)

The Core Technology Research and Application Demonstration of Transformer Core Intelligent Lamination Robot, Guangdong Provincial Science and Technology Plan Project(No.2016B090918017)

Development and Industrialization of Intelligent Patrol Robot for Quality Defects of Exquisite Cans, Ministry of Education "Blue Fire Plan" Industry-University-Research Joint Innovation Fund Project(No.CXZJHZ201730)

REFERENCES

[1] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2014:580-587.

[2] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Trans Pattern Anal Mach Intell, 2014, 37(9):1904-1916.

[3] Girshick R. Fast R-CNN. Computer Science, 2015.

[4] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks// International Conference on Neural Information Processing Systems. MIT Press, 2015:91-99.

[5] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection// Computer Vision and Pattern Recognition. IEEE, 2016:779-788.

[6] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger. 2016:6517-6525.

[7] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector. 2015:21-37.

[8] Jeong J, Park H, Kwak N, et al. Enhancement of SSD by concatenating feature maps for object detection. british machine vision conference, 2017.

[9] Fu C, Liu W, Ranga A, et al. DSSD : Deconvolutional Single Shot Detector. arXiv: Computer Vision and Pattern Recognition, 2017.

[10] Cai Z, Vasconcelos N. Cascade R-CNN: Delving Into High Quality Object Detection. computer vision and pattern recognition, 2018: 6154-6162.

[11] Cai Z, Fan Q, Feris R S, et al. A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection. european conference on computer vision, 2016: 354-370.

[12] Najibi M, Rastegari M, Davis L S, et al. G-CNN: An Iterative Grid Based Object Detector. computer vision and pattern recognition, 2016: 2369-2377.

[13] Kim K, Cheon Y, Hong S, et al. PVANET: Deep but Lightweight Neural Networks for Real-time Object Detection. arXiv: Computer Vision and Pattern Recognition, 2016.

[14] Zhu Y, Zhao C, Wang J, et al. CoupleNet: Coupling Global Structure with Local Parts for Object Detection. international conference on computer vision, 2017: 4146-4154.

[15] Lin T, Dollar P, Girshick R B, et al. Feature Pyramid Networks for Object Detection. computer vision and pattern recognition, 2017: 936-944.

[16] Dai J, Li Y, He K, et al. R-FCN: Object Detection via Region-based Fully Convolutional Networks. neural information processing systems, 2016: 379-387.

[17] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. computer vision and pattern recognition, 2016: 770-778.

[18] Lin T, Goyal P, Girshick R B, et al. Focal Loss for Dense Object Detection. international conference on computer vision, 2017: 2999-3007.

[19] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement. arXiv: Computer Vision and Pattern Recognition, 2018.

[20] Huang L, Yang Y, Deng Y, et al. DenseBox: Unifying Landmark Localization with End to End Object Detection. arXiv: Computer Vision and Pattern Recognition, 2015.

[21] Bell S, Zitnick C L, Bala K, et al. Inside-Outside Net: Detecting Objects in Context with Skip Pooling and Recurrent Neural Networks. computer vision and pattern recognition, 2016: 2874-2883.

[22] Peng C, Xiao T, Li Z, et al. MegDet: A Large Mini-Batch Object Detector. computer vision and pattern recognition, 2018: 6181-6189.

[23] Jeong J, Park H, Kwak N, et al. Enhancement of SSD by concatenating feature maps for object detection. british machine vision conference, 2017.

[24] Gidaris S, Komodakis N. Object Detection via a Multi-region and Semantic Segmentation-Aware CNN Model. international conference on computer vision, 2015: 1134-1142.

[25] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. computer vision and pattern recognition, 2015: 1-9.

[26] Kong T, Yao A, Chen Y, et al. HyperNet: Towards Accurate Region Proposal Generation and Joint Object Detection. computer vision and pattern recognition, 2016: 845-853.

[27] Yang B, Yan J, Lei Z, et al. CRAFT Objects from Images. computer vision and pattern recognition, 2016: 6043-6051.

[28] Li J, Liang X, Wei Y, et al. Perceptual Generative Adversarial Networks for Small Object Detection. computer vision and pattern recognition, 2017: 1951-1959.

[29] Kuang K. Research on Deep Learning and Its Application in Defect Detection of Aero-Engine. South China University of Technology, 2017.