International Journal of New Developments in Engineering and Society
ISSN 2522-3488 Vol. 3, Issue 1: 284-301, DOI: 10.25236/IJNDES.190238

Damage detection of quayside crane structure based on improved Faster R-CNN

Dong Han, Gang Tang

Shanghai Maritime University, Shanghai, China

ABSTRACT. *In order to detect multi-damage types of quayside crane simultaneously, a visual detection method based on region proposal deep network is proposed. Aiming at the difference of damage types and size of quayside bridges and the large distribution of small targets, the feature pyramid network (FPN) method is used to increase the resolution of feature mapping and to detect larger cracks and smaller corrosion of quayside bridges by using high-level semantic information after top-down model. By using mixup data enhancement method, the generalization ability of the original Faster R-CNN model structure can be reduced, the memory of error labels can be reduced, and the robustness against samples can be increased. Experiments show that the improved Faster R-CNN improves the detection of mAP by 5.14%.*

KEYWORDS: *Faster R-CNN; Damage detection; Characteristic pyramid network; Mixing up.*

1. Introduction

Structural damage detection of quayside crane is an important part of health monitoring of large mechanical equipment. Cracks and corrosion are the main structural damage of quayside crane. Visual inspection is the main assessment method of bridge condition. In recent years, deep convolution neural network has been used in the field of health monitoring. In order to partly replace the manual on-site inspection of civil infrastructure, the paper [1] proposes a deep learning-based CNN crack detection method for concrete, which is compared with traditional IPT's Canny and Sobel edge detection algorithms. The results show that the CNN method has better performance and can find cracks in practical situations. Literature [2] Using one-dimensional CNN to classify railway defects and comparing with MLP and SAE methods, the classification method of railway state is improved. Reference [3] presents the existence and types of bridge cracks detected by CNN. The method is validated by comprehensive data of debonding damage of composite sandwich plate, and crack identification is carried out on actual concrete bridge crack images. Reference [4] proposes a fault diagnosis method based on CNN, which converts the signal into two-dimensional image, eliminates the influence of manual features, and improves the data-driven diagnosis method.

These methods have solved the shortcomings of traditional IPT using pretreatment and post-processing methods, avoided the tedious manual feature selection process, and improved the detection accuracy. However, it is difficult to find the optimal size of sliding windows by using sliding windows to locate damage. In order to provide multi-objective damage detection, the deep learning target detection method is applied to the field of structural health monitoring. Target detection algorithms are mainly divided into two categories, two-stage detection is represented by the region-based convolution neural network RCNN [5] series, mainly RCNN, Fast-RCNN [6], Faster-RCNN [7] and R-FCN [8]. Their detection accuracy is high but slightly slow; the other one is based on regression and single-stage detection algorithm represented by YOLO, such as YOLO [9], SDD [10], and so on, and their detection accuracy is high. It's slightly lower but faster. YOLO describes the detection task as a unified, end-to-end regression problem. It only processes one picture. Compared with the two-stage method, YOLO has obvious speed advantages, but its meshing is rough. The number of boxes generated by each grid limits the detection of small-scale objects and similar objects, which is not conducive to the detection of corrosion points of quayside crane. Compared with YOLO, SDD is improved to generate more anchor boxes. Each grid generates boxes of different sizes and length-width ratios, and the class prediction probability is based on box prediction, which is different from YOLO's prediction on the grid. At the same time, SDD generates multi-scale feature mapping to improve the detection accuracy of small objects. However, the accuracy of Faster-RCNN is still not as high as that of two-stage Faster-RCNN. so Faster R-CNN is used as the basic framework of this paper.

Faster-RCNN proposes a regional proposal network (RPN) to replace the selective search (SS) [11] algorithm. The task can be completed end-to-end by the neural network, which can be regarded as the combination of RPN and Fast R-CNN. The shared convolution feature of RPN and R-CNN makes RPN introduce a very small amount of computation, which makes Faster R-CNN run at 5 FPS speed on a single GPU and achieve SOTA (State of the Art) in accuracy. In order to better detect small target objects of quayside crane, such as tiny corrosion points and tiny cracks, improve the detection of different damage sizes by models, reduce the missed detection rate, feature pyramid feature extraction network (FPN) [12] is added to the Faster-RCNN basic framework, and features are transmitted downward in high-level feature mapping, and feature pyramids are constructed backwards, which can simultaneously utilize low-level feature pyramids. The high resolution and high semantic information of high-level features can be fused to achieve the prediction effect, and each fused feature layer can be predicted separately. Different from the conventional feature fusion method, the resolution of feature mapping is increased. The combination of RPN and FPN improves the accuracy of small target recognition. In order to reduce the memory of the wrong label of the damage image of quayside crane, increase the robustness of the countermeasure sample and stabilize the training process of generating the countermeasure network, the data enhancement method of mixup [13] is added. This method is independent of data and is a form of domain risk minimization. Training in virtual samples can not only be well integrated into the existing training pipeline, but also be used in training. There is almost no computational overhead, which improves the generalization ability of the original Faster R-CNN. The method of multi-scale training, mixup data enhancement and characteristic pyramid network are used to solve the problems of large size difference and small target recognition of quayside crane damage targets, overcome the uneven distribution of positive and negative samples in quayside bridge damage images, and improve the accuracy of quayside bridge damage detection.

2. Damage detection method of quayside crane structure

In order to make the Faster R-CNN model better integrate the multi-scale damage detection of quayside bridges, this paper combines the data enhancement methods of feature pyramid (FPN), mixup and multi-scale training methods to enhance the performance of the model, and solve the detection problem of small damage points of quayside bridges.

2.1 Feature Pyramid Network (FPN)

Detecting targets of different scales is challenging, especially for small targets. Feature pyramid network is a feature extractor designed to improve accuracy and speed. It can replace feature extractors in Faster R-CNN to generate higher quality feature pyramids. FPN consists of top-down and bottom-up paths. The bottom-up path is the common CNN for feature extraction. The spatial resolution decreases from bottom to top. When higher structure is detected, the semantics of each layer increases. The schematic diagram is shown in Figure 1.

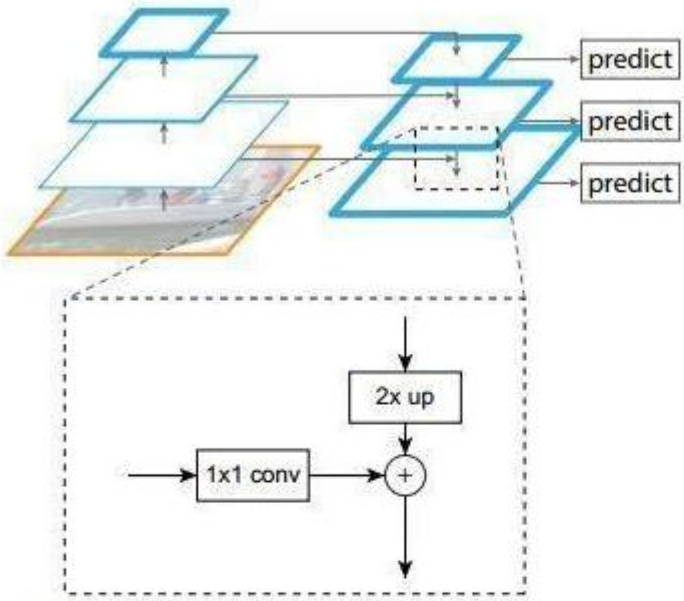


Fig.1. Feature Pyramid Network

Firstly, deep convolution is performed on the input image, and then dimensionality reduction is performed on the features above the second layer, that is, adding a convolution layer of layer 1 x1, and down sampling the features above the fourth layer, which have the corresponding size. Then, additive operation is performed on the fourth layer after processing, and the obtained results are input into the fifth layer. The purpose is to obtain a strong semantic information, which can improve the detection performance. We use deeper layers to construct pyramids in order to use more robust information. In addition, we add the processed bottom features to the processed high-level features. The purpose is that the underlying features can provide more accurate location information. However, multiple down sampling and up-sampling make the position information of deep network errors and combine them. A deeper feature pyramid is constructed, which combines the bottom and top information of the damage, and outputs different features.

2.2 Construction of Faster R-CNN Detector Using FPN

1) Select a shore bridge damage picture and perform pre-processing on the picture;

2) Send the processed image into the feature network of the pre-training, and use ResNet [14] to pre-train the network to start building the bottom-up network;

3) As shown in Figure 1, construct the corresponding top-down network, that is, perform upsampling operation on layer 4, first perform dimensionality reduction on layer 2 with 1x1 convolution, then add the two, and finally perform 3x3 Rolling machine operation;

4) Perform RPN operations on layers 4, 5, and 6 above, that is, a 3x3 convolution is divided into two paths, which are respectively connected to a 1x1 convolution for classification and regression operations;

5) Then, the candidate RoI (Region of Interest) obtained in the previous step is input to the 4th, 5th, and 6th layers respectively to perform the RoI Pooling operation, and is fixed to the 7x7 feature;

6) Finally, connect two 1024-layer fully-connected network layers on the basis of the previous step, and then divide the two branches to connect the corresponding classification layer and regression layer. The improved Faster R-CNN network structure is shown in Figure 2

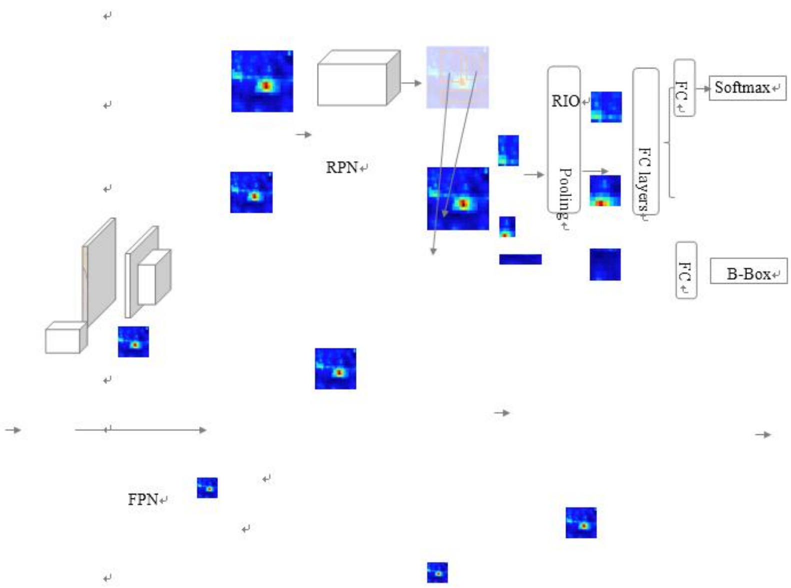


Fig. 2 An improved Faster R-CNN schematic diagram

2.3 Mixup Training

In order to improve the generalization error of Faster R-CNN model in this paper and improve the performance of quayside bridge classification task, a conventional data enhancement method, mixup, is adopted. This method has nothing to do with data, and a new training sample and label are constructed by linear interpolation. Ultimately, the label is treated as equation (1):

$$\begin{cases} \tilde{x} = \lambda x_i + (1-\lambda) x_j \\ \tilde{y} = \lambda y_i + (1-\lambda) y_i \end{cases} \quad (1)$$

Among them, (x_i, y_i) is two samples randomly extracted from training data, $\lambda \epsilon [0,1]$. So mixup expands the training distribution by combining prior knowledge, that is, linear interpolation of eigenvectors should lead to linear interpolation of related labels, which is convenient to implement and introduces minimum computational overhead. The general domain distribution of mixup is shown in Notice (2):

$$\mu\left(\tilde{x}, \tilde{y} \middle| x_i, y_i\right) = \frac{1}{n} \sum_j^n \mathbb{E} \left[\delta(\tilde{x} = \lambda x_i + (1-\lambda) x_j, \tilde{y} = \lambda y_i + (1-\lambda) y_j\right] (2)$$

Which $\lambda \sim \mathrm{Beta}(\alpha, \alpha), \alpha \epsilon (0, \infty)$. The mixup hyperparameter α controls the interpolation strength between feature and target vectors. With the increase of α, the training error of the network will increase, but the generalization ability of the network will also increase. When α approaches zero, the model will degenerate into the most original training strategy. According to the characteristics of quayside bridge data, the classification performance of the model will be improved by setting α to 0.2 pairs through experimental comparison. It is also the numerical value used in the original paper.

3. Experiment

The experimental environment is based on the deep learning framework Tensorflow. The hardware device used is 64G 220GB SSD + 3TB hard disk with GTX 1080 Ti 40 core and 12G GPU memory. The software environment is Python

3.5 CUDA 9.0 and opencv. The network framework Faster R-CNN chose VGG16 [15] and ResNet101.

3.1 Quayside crane dataset

In order to develop data sets of quayside bridges containing corrosion and cracks, 587 original damage maps of quayside bridges were collected with intelligent cameras. After cutting, stretching and random transformation, the original damage maps were expanded to 4364 and the data format of Pascal VOC 2007 was made, which contained various background and different damage targets of quayside bridges. ImgLabel was used for manual labeling. At the same time, 156 original images of quayside bridges were collected to test the method, and the effectiveness of the proposed method was verified.

3.2 Model evaluation index

In order to evaluate the effectiveness of the proposed method for shore bridge damage detection, the experimental evaluation indicators of Accuracy, Precision and Recall are used as formula (3)(4)(5)(6) (7):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3);$$

$$Precision = \frac{TP}{TP+FP} \quad (4);$$

$$Recall = \frac{TP}{TP+FN} \quad (5);$$

$$mAP = \int_0^1 P(R)d(R) \quad (6);$$

$$\mathrm{MR} = \frac{FN}{TP+FN} \quad (7)$$

Which TP represents the number of samples of true positive; FP represents the number of samples of false positive cases; FN represents the number of samples of false negative cases; TN represents the number of samples of true and negative cases; mAP represents the average accuracy rate (P and R represent the accuracy rate and recall rate respectively); MR represents the missed detection rate.

3.3 Analysis of experimental results

The experiment compares the performance and test results of the improved model with that of the improved model, the test results of the network under different strategies, the use of different pre-training models mAP, and the comparison of the single-stage detector.

3.3.1 Comparison of detection effect before and after improvement of Faster R-CNN

Using 4364 self-made VOC2007 format damage data of quayside crane, the original Faster R-CNN network combined with feature pyramid network, mixup data enhancement, multi-scale training were used respectively. 156 on-site test pictures of quayside bridges were used to evaluate the effects of the two models, as shown in Table 1.

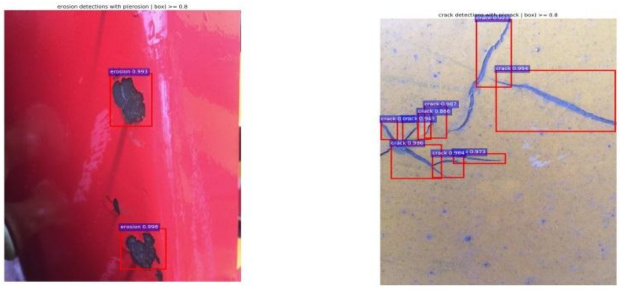
Table 1 Comparison of Model Effect before and after Improvement

Method	Crack(AP/%)	Erosion(AP/%)	mAP/%	MR/%

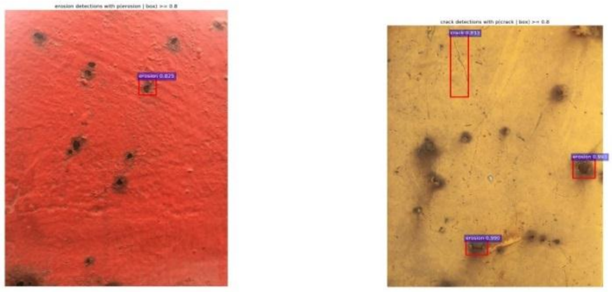
ISSN 2522-3488 Vol. 3, Issue 1: 284-301, DOI: 10.25236/IJNDES.190238

Faster RCCN	90.23	80.64	85.44	7.68
Ours	91.40	89.76	90.58	6.19

Table 1 shows that the average accuracy of the improved mAP is increased by 5.14%, the corrosion accuracy is increased by 9.14%, compared with 1.17% of the cracks. This shows that the crack detection effect before and after the improvement is not much improved. The original Faster R-CNN network can accurately detect the crack damage, but the corrosion detection ability is poor, so the leakage detection rate of the improved Faster R-CNN is reduced by 1.49%. The main reason is that the high-level semantic features obtained by FPN are more robust than before. It shows that the feature pyramid network has a strong ability to detect small targets. As shown in Fig. 3 and 4, it proves that the improved Faster R-CNN has a good detection effect on small targets of quayside bridges and a strong multi-scale adaptability.

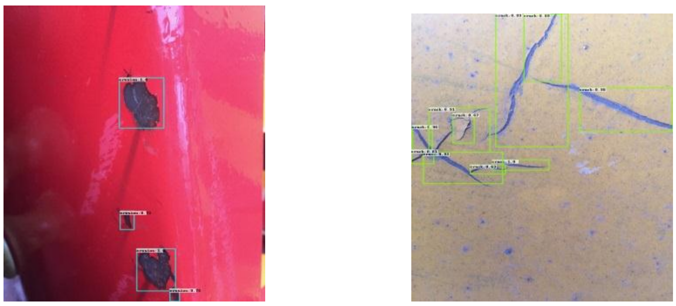


(a) (b)

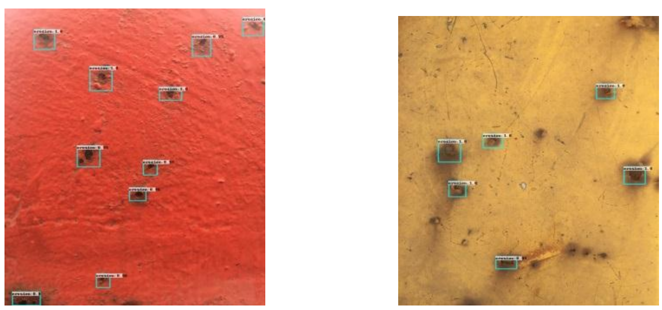


(c) (d)

Fig. 3 Effect of damage detection before improvement



(e) (f)



(g) (h)

Fig. 4 Effect of damage detection after improvement

Compared with Fig. 3 and Fig. 4, some minor damage points can be well detected after improvement, including the missing detection and misdetection of corrosion points and some cracks in Fig. 3, such as the missing detection of small pieces of corrosion in (a); (b) the missing detection of splitting lines in the middle and repeated detection of individual cracks; (c) the large area missing detection of corrosion points; (d) the misdetection of middle cracks, which can incorrectly detect the scratches on the surface into cracks. Grain. Compared with the improved algorithm, the improved algorithm solves the above problems well, reduces the rate of missed detection and false detection, and improves the detection accuracy of the algorithm. The biggest contrast is that Figure 4 detects more corrosion points with high confidence. There is no redundant target frame in crack detection, which improves the repeated detection of cracks. Experiments show that the improved Faster R-CNN framework can effectively optimize the model, adapt to different scales of object detection, and improve the detection of small size objects. It has practical value.

3.3.2 Contrast one-stage detection algorithm

Finally, the improved algorithm is compared with the one-stage detection algorithm YOLO and SSD. The results are as follows: Table 2:

Table 2 and Comparison of Single-stage Detection Algorithms

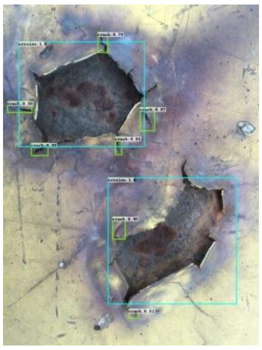
methods	Crack(AP/%)	Erosion(AP/%)	mAP/%	MR/%
YOLO	79.68	74.89	77.29	13.89

SSD	82.99	79.69	81.34	10.18
Ours	91.40	89.76	90.58	6.19

Experiments show that SDD improves the detection accuracy of some small targets, trying to reach the two-stage detection effect, but there is still a gap between the text and the improved Faster R-CNN algorithm. The one-stage detection method YOLO and SSD have low average precision. In the method of this paper, the missed detection rate table is higher, and the method of this paper is superior to YOLO and SSD in both indicators. All the above comparisons can be seen: after the mixup data enhancement, the feature pyramid network, and the ResNet101 feature extraction network, the number of anchor points is improved, so the improved Faster R-CNN detection algorithm improves the average accuracy of damage detection and reduces the damage. The missed detection rate is suitable for simultaneous detection of multi-target damage of the shore bridge.

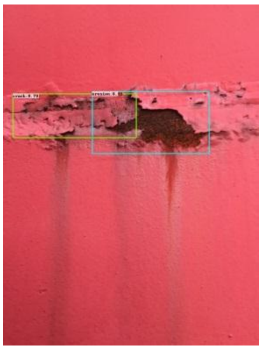
3.4 Quayside crane special damage state test result

Figure 5 illustrates the test of unconventional corrosion and cracking in different backgrounds of the quayside crane. In each background, there are two types of damage, including large corrosion blocks and inconspicuous cracks, which exist in the same target detection area. In order to test the generalization ability of the improved Faster R-CNN, four images were collected from different parts of the quayside. From Fig. 5, it can be seen that the four pictures from three different background parts represent three different damage parts. There is a relatively high degree of corrosion in a); (b) in the cracks with inconspicuous block corrosion; (c) the corrosion boundary is not obvious; (d) there are many noises such as some pollutants and scratches and corrosion analogs and the like. The four states a, b, c, and d have no identical or similar label annotations in the training set. They can be used to test the performance of the model. The test results are as follows:

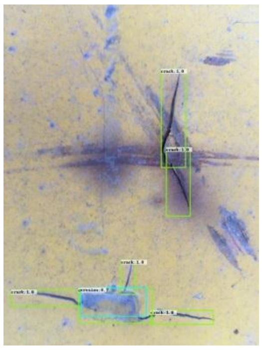


（a）

(b)



(c)



(d)

Fig. 5 Test results of different damage states of quayside bridges

It can be seen from Fig. 5(a) that the algorithm not only detects large-scale block corrosion, but also identifies cracks with small scale around the corrosion, which proves that the improved algorithm can adapt to multi-scale and small-target damage detection; There is no redundant boundary frame generated and accurate positioning; (c) the boundary of corrosion is relatively accurate, which basically meets the accuracy of human eye recognition; some pollutants in (d) and pseudo-cracks do not affect the model. Correct detection, that is, no missed detection and no false detection. Experiments show that the improved Faster R-CNN not only has multi-scale and small-target damage detection capability, but also has strong generalization ability. It has strong adaptability to some boundary blur, irregular shape irregularity and serious noise pollution. The ability to illustrate the algorithm of this paper has practical significance.

4. Conclusion

Firstly, the original Faster R-CNN network model is optimized by using the strategy of feature pyramid network, mixup data enhancement, and increasing the number of anchor points. Then, the improved Faster R-CNN model is trained by using the self-made VOC2007 and TFrecord data sets of quayside bridge damage, the information of crack and corrosion characteristics is extracted, the feature pyramid network is constructed, and the training techniques and feature extraction techniques are used to improve the quayside crane damage data set. Finally, the performance of models in different environments is tested and different detection algorithms are compared. The experimental results show that the method used in this paper improves the detection accuracy of quayside crane damage significantly, and has a good detection effect on multi-scale and small targets. It can not only detect cracks with large scale difference at the same time, but also improve the accuracy of detection of small corrosion points. However, the environment of quayside crane is complex, and there are many unobvious damages that can not be defined as cracks. Stripe or corrosion, and large area distribution, no fixed features, can not define the type of label, and there are difficulties in labeling, so it does not involve the undefined damage type detection of this part.

Reference

[1] Cha Y J, Choi W, Büyüköztürk O(2017). Deep Learning‐Based Crack Damage Detection Using Convolutional Neural Networks. Computer‐aided Civil & Infrastructure Engineering, vol. 32, no. 5, pp.361-378.

[2] Zhang X, Wang K, Wang Y, et al. (2017). An improved method of rail health monitoring based on CNN and multiple acoustic emission events. 2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC).

[3] Modarres C, Astorga N, Droguett E L, et al. (2018). Convolutional neural networks for automated damage recognition and damage type identification. Structural Control and Health Monitoring, vol. 25, no.10, pp. e2230.

[4] Wen L, Li X, Gao L, et al. (2018). A new convolutional neural network-based data-driven fault diagnosis method. IEEE Transactions on Industrial Electronics, vol.65, no. 7, pp.5990-5998.

[5] Girshick R , Donahue J , Darrelland T , et al. (2014). Rich feature hierarchies for object detection and semantic segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition.

[6] Girshick R(2015). Fast R-CNN. Computer Science.

[7] Ren S, He K, Girshick R, et al. (2015). Faster R-CNN: towards real-time object detection with region proposal networks. International Conference on Neural Information Processing Systems.

[8] Dai J, Li Y, He K, et al. (2016). R-FCN: object detection via region-based fully convolutional networks. NIPS'16 Proceedings of the 30th International Conference on Neural Information Processing Systems, pp. 379-387.