A UNIFIED FRAMEWORK FOR FAULT DETECTION OF FREIGHT TRAIN IMAGES UNDER COMPLEX ENVIRONMENT

Yang Zhang[a], *Kai Lin*[b], *Huiming Zhang*[a], *Yanwen Guo*[a,c] *and Guodong Sun*[b]

[a] National Key Laboratory for Novel Software Technology, Nanjing University, China
[b] School of Mechanical Engineering, Hubei University of Technology, China
[c] Science and Technology on Information Systems Engineering Laboratory, China

ABSTRACT

This paper proposes a novel unified framework for fault detection of the freight train images based on convolutional neural network (CNN) under complex environment. Firstly, the multi region proposal networks (MRPN) with a set of prior bounding boxes are introduced to achieve high quality fault proposal generation. And then, we apply a linear non-maximum suppression method to retain the most suitable anchor while removing redundant boxes. Finally, a powerful multi-level region-of-interest (ROI) pooling is proposed for proposal classification and accurate detection. The experimental results indicate that the proposed method can achieve high performance on four typical fault benchmarks, substantially outperforming the state-of-the-art methods.

***Index Terms*—**unified framework, freight train images, fault detection, convolutional neural network (CNN).

1. INTRODUCTION

Fault detection for the vehicle braking and steering systems is an important routine maintenance task to ensure the security of freight trains. For a long time, it has been performed by skilled workers, which has many drawbacks such as low detection probability and poor efficiency. For the efficiency and convenience of detection, vision-based fault detection of freight trains is getting more and more attention [1, 2, 3]. The vehicle braking and steering systems contain many important parts such as cut-out cocks, dust collectors, bogie block keys, and fastening bolts [4, 5], etc. Their technical status, performance and stability are directly related to the train operation safety.

Because an image acquisition device including several monochrome cameras is installed outdoors, the collected images are gray and easily affected by the weather, varying illumination, and other factors. As shown in Fig. 1, the image background is very complex and contains too much structural information. And common faults in the systems are mostly caused by the loss, the damage and position changes of the small parts. The differences between normal and fault images are not obvious. It is difficult to achieve the fast and accurate fault detection of freight train images, due to the diversity and complexity of faults.

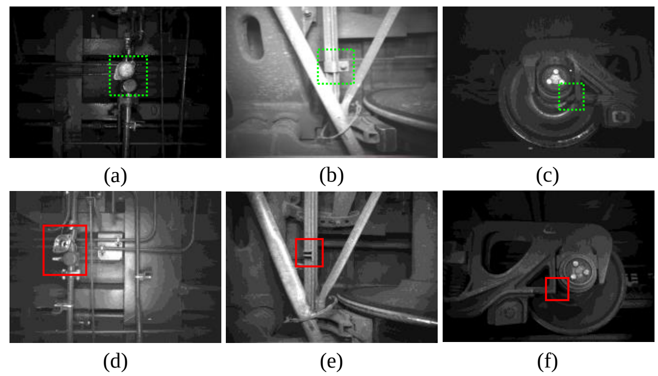

Fig. 1 Some typical samples of the vehicle brake system of freight trains. (a-c) Normal images. (d) Dirt collector damaged and cut-out cock handle closed (the handle of cut-out cock is not visible in a normal image). (e) The absence of fastening bolt. (f) Bogie block key missing.

To solve these problems, several methods have been implemented. Liu et al. [1], proposed a hierarchical fault inspection framework containing bearing cap detection, fault region localization, and bogie block key classification to detect the missing of BBK on freight trains with high speed and accuracy. Some inspection algorithms are extensively applied to inspect other components, such as brake cylinder [2], brake shoe [3], dust collector [4], brake shoe wear [5], fastening bolts [6], locking plate [7], angle cocks [8], and coupler yoke [9], etc. However, most of these detection algorithms aim at only one type of faults or even one fault.

Recently, some deep learning methods, especially the convolutional neural networks (CNN), have achieved remarkable successes in many computer vision tasks. In general, fault inspection can be considered as a special type of object detection task in computer vision. Sun et al. [10] proposed a CNN-based system for recognizing typical faults of freight trains, which can solve the problem of low quality images. However, the fault inspection system consists of two complex CNN-based models for target region detection and

This work was supported by the National Natural Science Foundation of China under Grants 61772257, 51775177, 51675166 and the Natural Science Foundation of Jiangsu Province under Grant BK20150016.

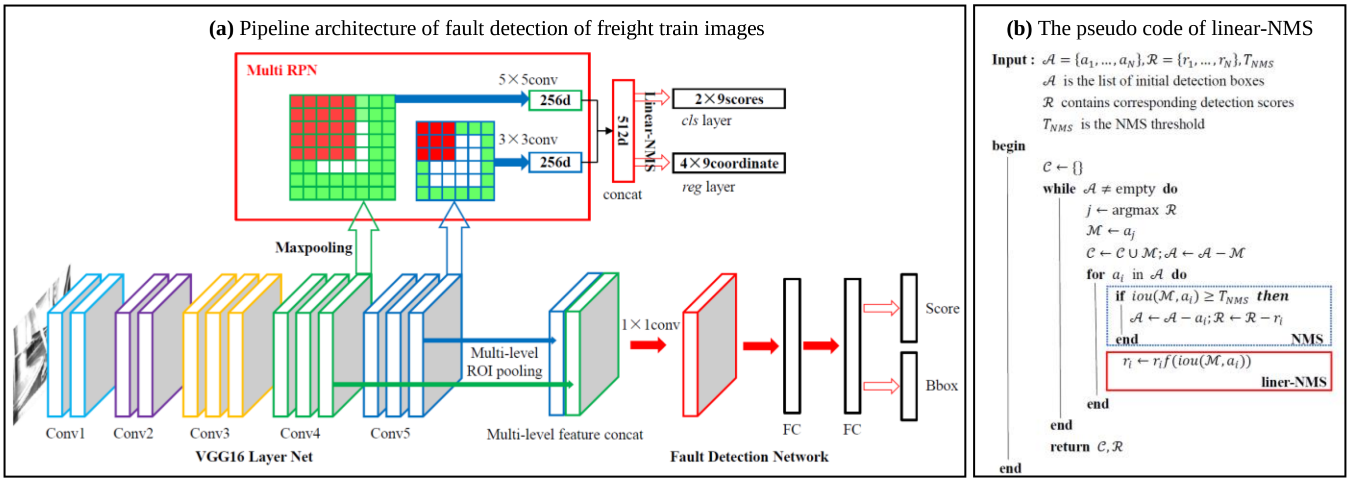

Fig. 2 (a) Pipeline architecture of fault detection of freight train images. Our approach takes an image as input, generates hundreds of fault region proposals via multi RPN, and then scores and refines each proposal using the fault detection network. (b) The pseudo code of liner-NMS [14]. The code in blue (NMS) is replaced with the one in red (liner-NMS). We propose to revise the detection scores by scaling them as a linear function of overlap.

fault recognition, respectively. The portability and generality of this method is not sufficient to meet the requirements of fault inspection.

For an object detection task, region-based CNN (RCNN) detection methods [11] such as the Faster R-CNN [12] and the region-based fully convolutional networks (R-FCN) [13] are now main paradigms with increasingly better accuracy. Inspired by Faster R-CNN, our motivation is to design a unified framework for fault region proposal generation (RPN) and classification of freight train images. To avoid the sequential error accumulation of bottom-up fault candidate extraction strategies, we propose the novel MRPN and design a set of prior anchors to achieve high-quality fault region proposals. Subsequently, a linear non-maximum suppression (NMS) [14] method is used to remove the redundant boxes. Finally, a powerful fault detection network is introduced by incorporating multi-level region of interest (ROI) pooling into the optimization process. Experiments show that the proposed method can be effectively applied to the fault detection of freight train images with high detection accuracy and fast speed.

The rest of this paper is organized as follows. The proposed fault detection method is introduced in Section 2. The databases, experimental results and analysis are showed in Section 3. This paper finally presents the conclusion and future work in Section 4.

2. FAULT DETECTION

The detailed architecture of the proposed unified framework is shown in Fig. 2(a). In the original Faster RCNN, the RPN and the ROI pooling are performed on the final feature map layer (e.g. Conv5_3 in the VGG16 model [15]) for feature prior regions generation and object detection. Such approaches on final feature map are not always optimal and may omit some effective features, as features in deeper convolutional layer output have wider reception fields, resulting in a grosser granularity. Therefore, in order to capture much more fine-grained details of the different convolutional layers, we propose to improve the RPN and the ROI pooling by combining the feature maps of multiple convolutional layers, including lower-level and higher-level features.

2.1. Multi region proposal generation

In general, lower and higher-level layers can complement each other by up-sampling higher level feature maps to match the size of its lower counterpart [16, 17]. In this way, small anchors can catch important details in the lower layer, whereas in previous work all anchors reside in the same layer and thus small anchors can be missed. In this paper, we split the set of anchor candidates.

Inspired by the idea of the inception module in GoogLeNet [16] and HyperNet [17], we propose a novel MRPN, which applies multi-scale sliding windows over multi-level convolutional feature maps and associates a set of prior anchors with each sliding position to generate fault region proposals. To search for fault region proposals, an inception network is slid over two feature maps (Conv4_3 and Conv5_3) in the VGG16 model. That is, a 5×5 convolution is applied to extract local feature over a 2×2 max pooling layer employed on Conv4_3 feature maps. Meanwhile, a 3×3 convolution is used to extract local feature over Conv5_3 feature maps at each sliding position.

Next, we concatenate each feature along the channel axis and a 512-d concatenated feature vector is entered two output layers: a classification layer that predicts the score of the fault region and a regression layer that refines the region location for each kind of prior anchor. In addition, our prior anchors are like the bounding boxes defined in RPN, which contain three scales (128, 256, and 512) and three aspect ratios (0.5, 1, and 2), for a total of N = 9 prior anchors at each sliding position. In the learning stage, we assign a positive label to a prior box that has an intersection over union (IoU) overlap greater than 0.5 with a ground-truth anchor, while assigning a background label to a prior box with an IoU overlap less than 0.3. After that, we apply a linear NMS [14] (as shown in Fig. 2(b)) with a threshold of T_{NMS} = 0.7 to retain the highest score anchor and rapidly suppress the lower scoring boxes in the neighborhood. We finally select the top-2000 candidate region proposals for the fault detection network.

2.2. Multi-level fault detection network

Previous state-of-the-art object detection models such as SPP-Net [18], fast-RCNN [19], and faster-RCNN [12], all simply apply ROI pooling over the last convolutional layer. However, to better utilize the multi-level convolutional features and enrich the differentiate information of each anchor, we perform multi-ROI pooling over the Conv4_3 as well as Conv5_3 convolutional feature maps of the VGG16 network and obtain two $512\times H\times W$ pooled features (both H and W are set to 7 in practice [12, 20]). And then, we apply concatenation on each feature and encode the concatenated feature with $512\times1\times1$ convolutional layer to combine the multi-level pooled features and match the first fully-connected layer of the VGG16 network.

The MRPN and the fault detection network are trained in an end-to-end manner via back-propagation and stochastic gradient descent (SGD). The shared convolutional layers are initialized by a pre-trained VGG16 model for ImageNet [15]. The base learning rate is 0.001 and is divided to 10 for each 40K mini-batch until convergence. Other parameters are same as [12]. All experiments are conducted in Caffe [21].

3. EXPERIMENTS AND ANALYSIS

To verify the performance of the proposed fault detection method, we first introduce four databases and evaluation metric using in this section, and then we present the experimental results followed the previous approaches with some state-of-the-art methods. The experiment is developed under the PC condition of 3.60GHz of Intel Core i7-7700 processor, 16G RAM, a K40 GPU, and Ubuntu 16.04 OS.

3.1. Databases and evaluation metric

The databases [22] contain two kind of data sets: training set and testing set, as shown in Table 1. And all images are the size of 700×512 pixels. Each training set image is labeled according to the format of the PASCAL VOC dataset [23]. To evaluate the effectiveness of the proposed fault detection algorithm, there are four indexes [1, 22]: correct detection rate (CDR), missing detection rate (MDR), false detection rate (FDR), and detection speed. For example, if a testing set contains m fault images and n no-fault images, by detection of the proposed method, a images are inspected as fault, among them b images are inspected by error, meanwhile, c images are inspected as no-fault, among them d images are inspected by error. So, the above indexes can be defined as: $CDR = (a\text{-}b)/(m+n)$, MDR = 1- CDR, $FDR = d/(m+n)$. In the task of the fault detection, the index of FDR is less crucial than MDR, because the impact is not serious if a no-fault region is detected by error.

3.2. Experimental results

To illustrate the superiority of the proposed method for fault detection, we also carry out experiments using hand-crafted

Table 1 Databases

Databases	Training set images	Testing set images		
		No-fault	Fault	Total
Cut-out cock handle	815	671	179	850
Dust collector	815	798	52	850
Fastening bolts	1724	445	1257	1902
Bogie block key	5440	2530	367	2897

Table 2 Detection results of different databases

Methods	Cut-out cock handle			Dust collector			Fastening bolts			Bogie block key			Detection speed /s
	CDR/%	MDR/%	FDR/%	CDR/%	MDR/%	FDR/%	CDR/%	MDR/%	FDR/%	CDR/%	MDR/%	FDR/%	
Cascade detector(LBP)	92.12	7.88	15.29	98.12	1.88	8.82	96.79	3.21	4.73	97.89	2.11	1.31	0.036
HOG+Adaboost+SVM	97.41	2.59	9.41	99.53	0.47	2.59	98.58	1.42	2.89	99.1	0.90	2.14	0.049
FAMRF+EHF	98.71	1.29	5.41	98.94	1.06	2.82	99.11	0.89	6.41	**99.24**	0.76	1.52	0.725
SSD(VGG16)	**99.88**	0.12	23.06	100	0	26.71	97.69	2.31	0.05	98.07	1.93	0	0.153
R-FCN(ResNet-50)	99.17	0.83	2.59	100	0	19.41	99.89	0.11	0.05	96.41	3.59	0	0.177
+Soft NMS	99.88	0.12	29.88	100	0	26.82	99.74	0.26	0	64.45	35.55	0.03	0.179
Faster-RCNN(ZF)	98.82	1.18	4.00	100	0	14.94	99.42	0.58	0.05	98.86	1.14	0	0.073
Faster RCNN(VGGM)	98.82	1.18	7.41	100	0	13.53	99.79	0.21	0.05	97.45	2.55	0	0.079
Faster RCNN(VGG16)	99.06	0.94	1.41	100	0	3.65	99.95	0.05	0	95.76	4.24	0.10	0.238
+Soft NMS	99.17	0.83	0.82	100	0	4.12	99.95	0.05	0.05	77.98	22.02	0	0.243
Our method	99.18	0.82	**0.47**	**100**	**0**	**0.35**	**100**	**0**	**0**	98.76	1.24	**0**	0.244

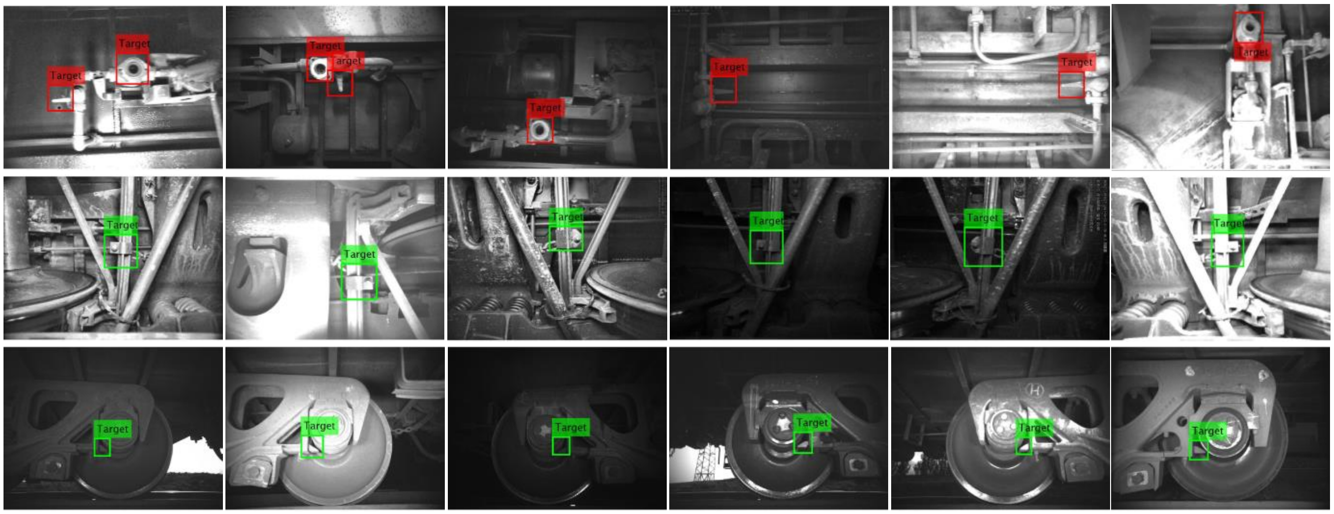

Fig. 4 Results of railway equipment detection.

feature descriptors and conventional machine learning tools. With some widely used object detection methods such as FAMRF + EHF [23], cascade detector [23] based on the local binary patterns (LBP), histogram of oriented gradient (HOG) (to descript features) + Adaboost (to select features) with the linear SVM classifier [24], single shot multi-box detector (SSD) [25] based on the VGG16 model, R-FCN [13] based on a ResNet-50, Faster RCNN [12] based on different models, R-FCN and Faster RCNN with the soft NMS [14], the same freight train images are trained for fault detection. The parameters of above nine methods are all default values. The detection results are detailed in Table 2.

For the dust collector database, our method has 100% correct detection rate, 0% missing detection rate, and 0.35% false detection rate. And for the fastening bolts database, our method has 100% correct detection rate, 0% missing detection rate, and 0% false detection rate. The proposed detection method outperforms other related methods in the above two databases. For the cut-out cock handle and bogie block key databases, the performances of our method are slightly inferior to others, but our method has lower false detection rate.

Some typical detection results are shown in Fig. 4. The main reason for the high missing detection rate of RCNN-based method especially Faster RCNN in the bogie block key database is that the bogie block key is too small compared with other components of the freight train. For the traditional Faster RCNN, the ROI pooling is performed on the final feature map layer to generate features of the region. However, such an approach omit some important features of the fault. Our method uses the MRPN and multi-level ROI pooling to learn more effective and comprehensive features for distinguishing faults from complex backgrounds.

Table 2 also lists a comparison of the detection speed. To perform a freight train image with a size of 700×512 pixels, its detection speed is up to 4 fps including all steps with a K40 GPU. In Table 2, the computation speed of the cascade of LBP features and HOG+Adaboost+SVM method are faster, but their accuracy are the lowest. The accuracy of the FAMRF+EHF method is satisfactory, but its speed is too low because of the computational complexity of the Markov random field [26]. In addition, the performance of the SSD and R-FCN methods are comparable with our method. However, the SSD and R-FCN are sensitive to the noise, and their false detection rate are high. The experiment results confirm that our method is preferable for the unified fault detection framework.

There are three main reasons why the unified framework has high accuracy and fast speed. First, the proposed MRPN enables achieving the more accurate fault region proposals. Second, the linear NMS algorithm is able to keep the most suitable anchor and remove the remainders. Third, the multi-level ROI pooling is applied into the fault detection network, which helps it to learn more effective information for distinguishing faults from backgrounds.

4. CONCLUSION

In this paper, we presented a novel unified framework for fault detection of freight train images of the vehicle braking and steering system with a powerful deep learning method in an end-to-end manner. The proposed framework consists of a MRPN with a set of characteristic prior anchors for high quality fault proposal generation and a powerful multi-level fault detection network for proposal classification and accurate localization. Specially, a linear NMS method is applied to effectively remove redundant boxes. Experiments on four benchmarks including cut-out cock handle, dust collector, bogie block key and fastening bolts show that the proposed method is able to achieve high performance with a fast detection speed up to over 4 fps (including all steps), substantially outperforming previous methods. In the future, we plan to further enhance accuracy and computation speed of the fault detection.

5. REFERENCES

[1] L. Liu, F. Zhou, and Y. He, "Automated visual inspection system for bogie block key under complex freight train environment," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no.1, pp.2-14, 2015.

[2] Y. Cao, B. Zhang, J. Liu, and J. Ma, "Weighted margin sparse embedded classifier for brake cylinder detection," *Neurocomputing*, vol.120, no.10, pp. 560-568, 2013.

[3] H. C. Kim, and W. Y. Kim. "Automated inspection system for rolling stock brake shoes," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no.8, pp. 2835-2847, 2011.

[4] F. Zhou, R. Zou, and H. Gao, "Dust collector localization in trouble of moving freight car detection system," *Journal of Zhejiang University Science C: Computer and Electronics*, vol.14, no.2, pp.98-106, 2013.

[5] R. Zou, Z. Xu, J. Li, and F. Zhou, "Real-time monitoring of brake shoe keys in freight cars," *Frontiers of Information Technology and Electronic Engineering*, vol.16, no.3, pp. 191-204, 2015.

[6] L. Liu, F. Zhou, and Y. He, "Automated status inspection of fastening bolts on freight trains using a machine vision approach," in *Proceedings of the Institution of Mechanical Engineers Part F Journal of Rail and Rapid Transit*, 2015, pp.1159-1166.

[7] G. Nan and J. E. Liu, "A real-time visual inspection method of fastening bolts in freight car operation," in *Applied Optics and Photonics China*, 2015, pp.96752G.

[8] F. Zhou, R. Zou, Y. Qiu, and H. Gao, "Automated visual inspection of angle cocks during train operation," in *Proceedings of the Institution of Mechanical Engineers Part F Journal of Rail and Rapid Transit*, 2013, pp.794-806.

[9] C. Zheng and Z. Wei, "Automatic online vision-based inspection system of coupler yoke for freight trains," *Journal of Electronic Imaging*, vol.25, no.6, pp.061602, 2016.

[10] J. Sun, Z. Xiao, and Y. Xie, "Automatic multi-fault recognition in TFDS based on convolutional neural network," *Neurocomputing*, vol.222, pp.127-136, 2017.

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," In *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2014, pp.580–587.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," In *Advances in Neural Information Processing Systems(NIPS)*, 2015 pp.91–99.

[13] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," *arxiv preprint arXiv:1605.06409*, 2016.

[14] N. Bodla, B. Singh, R. Chellappa, and L.S. Davis, "Soft-nms-improving object detection with one line of code," in *IEEE International Conference on Computer Vision(ICCV)*, 2017, pp.5562-5570.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," In *Processing. International Conference on Learning Representations (ICLR)*, 2015.

[16] C. Szegedy, W. Liu, and Y. Jia, *et al*., "Going deeper with convolutions," In *IEEE International Conference on Computer Vision and Pattern Recognition(CVPR)*, 2015, pp.1-9.

[17] T. Kong, A. Yao, Y. Chen, and F. Sun, "HyperNet: Towards accurate region proposal generation and joint object detection," In *IEEE International Conference on Computer Vision and Pattern Recognition(CVPR)*, 2016, pp.845-853.

[18] K. He, X. Zhang, S. Ren, and J. Sun. "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.37, no.9, pp.1904-1916, 2015.

[19] R. Girshick, "Fast R-CNN," in *IEEE International Conference on Computer Vision(ICCV)*, 2015, pp.1440-1448.

[20] Z. Zhong, L. Jin, and S. Huang. "DeepText: A new approach for text proposal generation and text detection in natural images," in *IEEE International Conference on Acoustics, Speech and Signal Processing(ICAASSP)*, 2017, pp.1-18.

[21] Y. Jia, E. Shelhamer, and J. Donahue, *et al*., "Caffe: Convolutional architecture for fast feature embedding," *arxiv preprint arXiv:1408.5093*, 2014.

[22] Y. Zhang, "Hierarchical feature matching of fault images in TFDS based on improved Markov random field and exact height function," M.S. thesis, Department of Mechanical Engineering, Hubei University Technology, Wuhan, China, 2017.

[23] M. Everingham, S. M. A. Eslami, and L. V. Gool, *et al.*, "The pascal visual object classes challenge: a retrospective," *International Journal of Computer Vision*, vol.111, no.1, pp.98-136, 2015.

[24] P. Dollar, R. Appel, Belongie S, and P. Perona, "Fast feature pyramids for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.36, no.8, pp.1532-1545, 2014.

[25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. Reed, "SSD: single shot multibox detector," in *European Conference on Computer Vision(ECCV)*, 2016, pp.21-37.

[26] Y. Duan, F. Liu, L. Jiao, P. Zhao, L. Zhang, "SAR Image segmentation based on convolutional-wavelet neural network and Markov random field," *Pattern Recognition*, vol.64, pp.255-267, 2017.