PDD-NET: A PRECISE DEFECT DETECTION NETWORK BASED ON POINT SET REPRESENTATION

Miaoju Ban, Runwei Ding[*]*, Jian Zhang, Tianyu Guo, Tao Wang*

Key Laboratory of Machine Perception, Shenzhen Graduate School, Peking University

{miaoju.ban, zhangjian, levigty, taowang}@stu.pku.edu.cn, dingrunwei@pku.edu.cn

ABSTRACT

Defect detection has been widely studied in computer vision and used in industrial production. However, most existing methods for defect detection mainly suffer three drawbacks: i) Low-contrast problem between defects and background. ii) Large scale changes in defects size. iii) Extreme imbalance problem between defects and background classes during training. To address these issues, we propose a novel anchor-free defect detection network named PDD-Net. Specifically, a global-context FPN (GC-FPN) is designed to capture long-range dependency between defects and background. Simultaneously, to enhance feature extraction of defects at different scales, a receptive field pyramid block (RFPB) is proposed to provide various receptive field sizes. Furthermore, an equipped adaptive positive and negative samples allocation (APNSA) mechanism is built with statistical characteristics of defects, thus can select training samples automatically. We conduct experiments on MPSD dataset, DAGM2007 dataset, and NEU-DET dataset. Extensive experimental results on the three challenging datasets show that our PDD-Net achieves superior detection accuracy over the state-of-the-art methods.

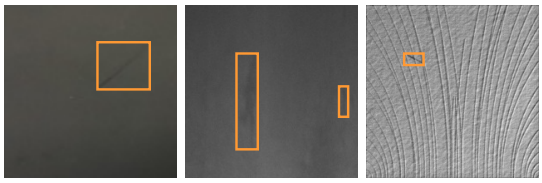
***Index Terms*—** Defect Detection, Global Context, Pyramid Receptive Field, Adaptive Samples Allocation

1. INTRODUCTION

Defect detection is a fundamental but essential task in the computer vision field, aiming to locate and classify defects in a defect image. The mainstream defect detection methods can be categorized into object detection method and semantic segmentation method. In this paper, we focus on the object detection method. With the development of object detection, the object detectors can be categorized into anchor-based and anchor-free by whether they use anchor boxes or not.

Relation to prior work: Currently, the popular defect detection methods are mainly based on anchor-based detectors. Cha et al. [1] directly used Faster R-CNN [2] with ZF-Net as the backbone [3] to detect damages on bridges. Xue et al. [4] proposed position-sensitive RoI pooling to replace the RoI pooling in Faster R-CNN [2] to detect tunnel lining defects.

* Corresponding author. This work is supported by National Key R&D Program of China (2018YFB1308600, 2018YFB1308602).



(a) (b) (c)

Fig. 1. Main challenges in defect detection. (a) Low-contrast problem between defects and background. (b) Large changes in defects size. (c) Imbalance between positive and negative samples in training.

Chen et al. [5] designed an improved SSD [6] by cascading three DCNN-based detection stages in a coarse-to-fine manner to detect the fasteners defect. Zhang et al. [7] proposed a detector based on YOLOv3 [8] for detecting multiple concrete bridge damages. However, all these works only focus on a specific defect, and there are still many challenges in defect detection. As shown in Fig. 1, there are mainly three challenges in defect detection. Fig. 1 (a) shows the low-contrast problem between defects and background. Fig. 1 (b) gives an example of large-scale changes in defects size. Fig. 1 (c) illustrates that much background causes a severe imbalance between positive and negative samples in training.

To address the challenges mentioned above, we propose PDD-Net, a precise defect detection network. The framework of PDD-Net is shown in Fig. 2. Unlike traditional defect detection networks, our method is anchor-free, which can detect defects at different scales and shapes. To be specific, the contrast information of defects in an input defect image is first enhanced by the global-context FPN (GC-FPN). Then after the receptive field pyramid block (RFPB), the features of defects at different scales are fused effectively. Finally, the adaptive positive and negative samples allocation (APNSA) mechanism in the detector head automatically selects positive samples to get the detection results. The main contributions of our work can be summarized as follows:

- Aiming at the different characteristics between defect detection and traditional object detection, a precise anchor-free defect detection network named PDD-Net is proposed to pursue both precise localization and accurate classification.
- Specifically, the GC-FPN is proposed to build long-

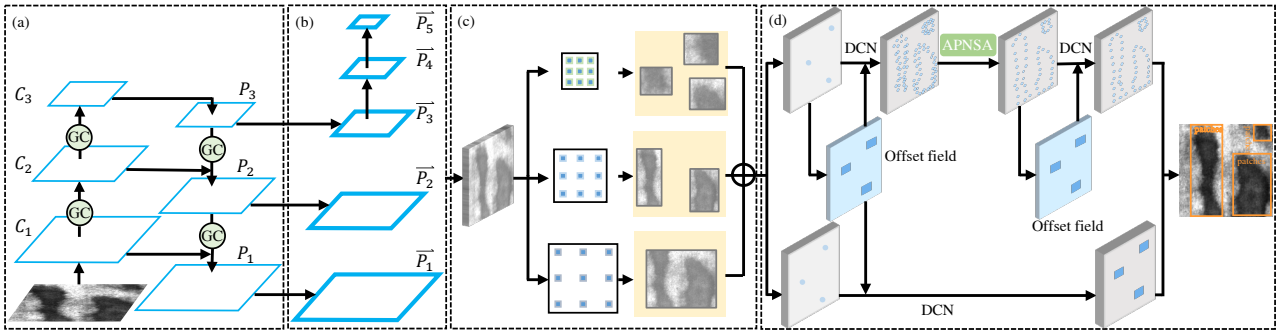


Fig. 2. An overview of PDD-Net. (a) GC-FPN backbone. (b) Pyramid feature maps. (c) Receptive Field Pyramid Block (RFPB). (d) Adaptive Positive and Negative Samples Allocation (APNSA) RepPoints Head.

range dependency between defects and background to make the model more sensitive to the location of defects. Then the RFPB is designed to effectively fuse defect features at different scales by using various receptive field sizes. In order to eliminate a large number of negative samples during training, the APNSA mechanism is equipped in the detector head.

- We propose a new dataset named MPSD for mobile phone surface defect detection, which provides valuable research data for this field. Our PDD-Net achieves state-of-the-art defect detection accuracy on the MSPD and DAGM2007 datasets, and the performance is also competitive on the NEU-DET dataset.

2. PROPOSED METHOD

2.1. Global Context FPN

It is well known that FPN [9] has been widely used in many detection tasks due to its good performance in dealing with different sizes of objects. We know that the original FPN has a bottom-up pathway and a top-down pathway. First, the bottom-up pathway computes a feature hierarchy consisting of feature maps at several scales with a scaling step of 2. Next, in the top-down pathway, it upsamples the spatial resolution of the nearest feature map by a factor of 2. Then the network fuses the corresponding features provided by the two pathways to generate the final feature for detecting. However, the details of objects will be lost during the subsampling and upsampling. It affects detect performance for the objects that are not rich in detail, especially for the defect images similar to the background. Cao et al. [10] propose an attention mechanism that is effective for context modeling. Hence, as shown in Fig. 2 (a), we propose GC-FPN, which takes full advantage of the long-range dependency between defects and background, to help us enhance edge information of low-contrast defects and locate them more accurately.

The GC module can be divided into three parts to calculate. They are context modeling, capture channel-wise dependencies, and feature fusion. So the GC module can be formulated as:

$$GC\left(x_{in}\right) = x_{in} + W_v \sum_{j=1}^{H\times W} \frac{\exp\left(W_k x_j\right)}{\sum_{n=1}^{H\times W} \exp\left(W_k x_n\right)} x_j, \quad (1)$$

where x denotes the input feature map of GC module, $H \times W$ is the number of positions in the feature map, in is the index of query positions, and j represents all possible positions. W_v and W_k denote the linear transformation matrices.

As illustrated in Fig. 2 (a), based on the FPN module and GC module we introduced above, we build the GC-FPN. We put the GC module after each of the downsampling and upsampling. It can efficiently enhance contrast information by modeling the long-range dependency between defects and other positions. Taking the P_2 as an example, the final fused feature of the proposed GC-FPN can be obtained as:

$$P_2 = Conv(C_2) + GC(Resize(P_3)), \quad (2)$$

where $Conv$ denotes 3×3 convolution of the lateral connection. $Resize(\cdot)$ represents upsample feature map by a factor of 2.

Finally, as Fig. 2 (b) shows, the output of GC-FPN P_i transform into $\vec{P}_i$ for detecting by the following equations:

$$\vec{P}_i = \begin{cases} Conv(P_i) & i = 1, 2, 3 \\ Maxpool(\vec{P}_{i-1}) & i = 4, 5. \end{cases} \quad (3)$$

$\vec{P}_4$ and $\vec{P}_5$ are maxpooled by $\vec{P}_3$ and $\vec{P}_4$, respectively.

2.2. Receptive Field Pyramid Block

Different defects usually have different scales in size, whether they belong to the same category or not. This character of defects often makes it difficult to detect. Considering the human visual system, we can distinguish different scales of objects because we have multi-scale perceptive fields. Actually, there are many works to discuss the receptive fields in CNNs, such as Inception families [11, 12], Deformable CNN [13], SRF-Net [14], and so on. Based on the discussion above, as Fig. 2 (c) shows, we design a natural and intuitive Receptive Field Pyramid Block (RFPB), which can fuse the feature of defects in different scales effectively.

As shown in Fig. 3, RFPB consists of a multi receptive field (MRF) convolutions branch and a shortcut branch. Specifically, the MRF convolutions branch comprises three 3×3 dilated convolutions with dilated rate of 1, 3, 5. Just as shown in Fig. 2 (c) right part, this branch aims at extracting defect features in different scales respectively. The whole

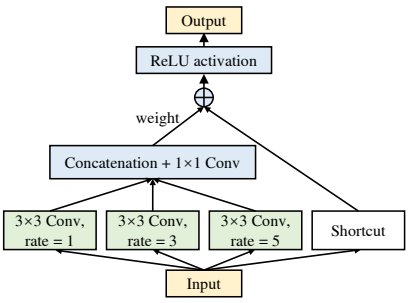


Fig. 3. The architecture of RFPB. It consists of multi receptive filed (MRF) convolutions branch and shortcut branch.

MRF convolutions branch can be formulated as:

$$P_{MRF}^{out} = Conv(Concat(\sum_{i=1,3,5} DConv_i(P^{in}))), \quad (4)$$

where P^{in} and P_{MRF}^{out} denote the input and output feature maps of RFPB. $DConv_i$ represents dilated convolution with dilated rate i.

The shortcut branch is to preserve the global features extracted by the GC-FPN. Finally, the output fusion feature of RFPB can be obtained by the following equation:

$$P_{RFPB}^{out} = ReLU(P^{in} + \omega * P_{MRF}^{out}), \quad (5)$$

where ω denotes the weight that multiplies to the fusion feature of the MRF output.

2.3. APNSA RepPoints Head

Different from common objects in detection, the defects often only take up a small part of a whole image, which usually causes a serious imbalance between positive and negative samples for training. Besides, our method is anchor-free for better adapting different defects, which further aggravates this situation. In the previous object detection methods, many strategies attempt to solve the problem of imbalance between positive and negative samples, such as focal loss [15] and ohem [16]. However, these strategies can not select positive samples well in the training process with many negative samples, especially in the defect detection fields. Zhang et al. [17] propose a strategy to adaptive select training samples according to statistical characteristics of objects. Therefore, based on [18] and [17], we improve the bounding box samples selection strategy in the original RepPoints head to eliminate the impact brought by a large number of background samples in defect detection.

The original RepPoints [18] head has two kinds of sample selection strategies. One is for points, the other is for bounding boxes. As shown in Fig. 2 (d), we improve the bounding boxes samples selection from a simple MaxIoU mechanism to adaptive positive and negative samples allocation (APNSA). Firstly, for each ground-truth bounding box g, we select k candidate proposals whose center are closest to the center of g based on L_2 distance for each pyramid level as initial positive samples $PS^{'}$. Then we calculate the mean m_k and standard deviation v_k of the IoU between $PS^{'}$ and g with the equations $m_k = Mean(IoU(PS^{'}, g))$ and $v_k = Std(IoU(PS^{'}, g))$, respectively. Based on the m_k and v_k, we set a new IoU threshold $t_k = m_k + v_k$ to select final positive samples. The selection can be depicted as:

$$PS = \{S_i | S_i \in PS^{'}, IOU(S_i, g) \geq t_k\}, \quad (6)$$

where i is the index of a positive sample. PS represents the final positive samples. S denotes the whole samples. $NS = \{S_i | S_i \in S, S_i \notin PS\}$ represents the final negative samples.

3. EXPERIMENTS

3.1. Datasets

MSPD Dataset (Mobile Phone Surface Defect Dataset). We collect 8919 mobile phone surface defect images from the real phones we used. It includes 5 common defects, which are jag, pit, stain, oil, and scratch. All of these defect images are with the size around 512 × 256 pixels. In our experiment, we randomly select 7134 defect images for training and 1785 defect images for testing.

DAGM2007 Dataset. This dataset [19] is related to various surface defects under textured background. Although the provided data is artificially generated, it is very close to the real world. It contains 2100 defect images, which are divided into 10 categories, and each image is 512 × 512 pixels. In this experiment, we choose 1046 images for training and 1054 images for testing.

NEU-DET Dataset. This dataset [20] is provided by the Northeastern University (NEU) to detect the surface defects of the hot-rolled steel strip. There are 1800 defect images. It includes 6 kinds of typical surface defects, which are crazing, inclusion, patches, pitted surface, rolled-in scale, and scratches. The resolution of each image is 200 × 200 pixels. Finally, we pick 1440 and 360 defect images at random for training and testing, respectively.

3.2. Implementation Details

We implement the PDD-Net with MMDetection [21]. Our model is trained on one NVIDIA 1080Ti GPU with CUDA 10.2 and Pytorch 1.6. The batch size is set to 8 for all datasets. The network is optimized with SGD [22] with learning rate of 1e-3, weight decay of 1e-4 and momentum of 0.9. We use different epochs for different datasets. For the MPSD dataset and NEU-DET dataset, the model is trained using 32 epochs. For the DAGM2007 dataset, the model is trained using 80 epochs. To make a fair comparison, the compared models are trained with the corresponding epochs in performance comparison.

3.3. Comparison with State-of-the-art methods

We compare our PDD-Net with the state-of-art methods [2, 23, 24, 8, 6, 15, 25, 26, 27, 28, 18] in object detection filed on the three datasets we introduced in 3.1.

Table 1 shows the performance on MSPD, DAGM2007, and NEU-DET datasets. On the MSPD dataset, our method with backbone ResNet-50 achieves 60.0% mAP, outperforming all the other anchor-based and anchor-free methods with the same backbone.

As for DAGM2007 dataset, our method achieves 99.4% and 80.1% accuracy of AP_{50} and AP_{75} respectively, which are

Table 1. Performance on MPSD Dataset & DAGM2007 Dataset & NEU-DET Dataset

Method	Backbone	MPSD						DAGM2007		NEU-DET
		mAP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AP_{50}	AP_{75}	mAP
Anchor-Based:										
Faster R-CNN [2]	ResNet-50	53.7	91.5	53.9	52.2	52.2	39.7	97.4	68.0	74.6
Mask R-CNN [23]	ResNet-50	55.1	93.5	56.0	53.8	50.7	43.0	93.6	56.8	73.7
Cascade R-CNN [24]	ResNet-50	57.5	92.5	59.3	56.2	53.9	35.1	91.3	50.3	74.5
YOLOv3 [8]	DarkNet-53	50.4	91.4	48.5	48.6	47.3	46.3	97.9	65.3	61.2
SSD512 [6]	VGG-16	56.2	93.5	56.0	54.2	55.8	51.3	98.9	75.4	69.4
RetinaNet [15]	ResNet-50	55.4	92.9	56.6	53.9	53.4	42.7	94.3	64.0	69.3
Anchor-Free:										
CornerNet [25]	Hourglass-104	44.1	84.3	35.9	43.6	43.0	25.0	91.9	60.0	51.3
FCOS [26]	ResNet-50	54.7	92.4	55.0	53.4	52.0	42.6	96.1	67.7	68.2
FSAF [27]	ResNet-50	56.2	93.8	57.7	54.1	54.7	45.6	95.9	64.6	71.0
FoveaBox [28]	ResNet-50	57.0	**94.4**	58.6	55.4	52.7	45.2	96.8	63.7	71.9
RepPoints [18]	ResNet-50	55.6	92.9	56.1	53.8	54.7	37.0	98.0	67.9	74.5
PDD-Net(ours)	**ResNet-50**	**60.0**	93.8	**64.3**	**58.3**	**55.6**	**55.9**	**99.4**	**80.1**	**76.5**

[1] The evaluation of MPSD and DAGM2007 is under MS COCO's standard metrics.
[2] The evaluation of NEU-DET is under Pascal VOC2007's metrics.

Table 2. The Ablation Experiments on MPSD Dataset

GC-FPN	RFPB	APNSA	mAP	AP_{50}	AP_{75}
			55.6	92.9	56.1
✓			57.2	94.2	60.1
	✓		56.4	94.1	56.2
		✓	57.9	93.6	59.9
✓	✓		59.4	94.4	62.6
✓		✓	59.1	**94.5**	62.7
	✓	✓	58.3	94.3	60.7
✓	✓	✓	**60.0**	93.8	**64.3**

superior results compared with other state-of-the-art methods. As for the NEU-DET dataset, our method also gets a competitive result with 76.5% mAP. Noting that the evaluations of these two datasets are different because we want to make a fair comparison under the format of original datasets. So we evaluate DAGM2007 with AP_{50} and AP_{75} in COCO metric, NEU-DET with mAP in Pascal VOC2007 metric.

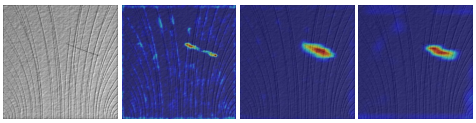
3.4. Ablation Study

To analyze the effectiveness of different components in our approach, we perform the ablation experiments on MSPD dataset. The baseline is RepPoints [18] with mAP of 55.6%, AP_{50} of 92.9%, AP_{75} of 56.1%. All the ablation experiments follow the default parameter setting in 3.2.

As shown in Table 2, firstly, we conduct a series of experiments based on the baseline with a single module to verify the effectiveness of our proposed GC-FPN, RFPB, APNSA, respectively. The gains they provide in detection performance at mAP shows the impact of our modules towards achieving precise localization and accurate classification. Besides, we give the results of any combination of two modules. They give stable gains of 2%∼4% mAP, which proves complementary of methods. Finally, our PDD-Net equipped with the three modules gains 4.4% mAP, 0.9% AP_{50}, and 8.2% AP_{75}. It further shows that solving the three main challenges in defect detection can significantly improve detection accuracy.

3.5. Qualitative Results

In order to give an intuitive understanding of the performance of our model, we show the heatmaps generated by our model and final detection results in Fig. 4 and Fig. 5, respectively.



(a) (b) (c) (d)

Fig. 4. Visualization of heatmaps of DAGM. (a) Input image. (b) (c) (d) are the heatmaps of Fig. 2 (b) from bottom to up.

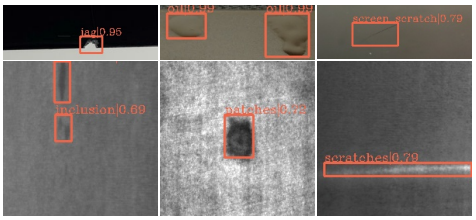


Fig. 5. PDD-Net defect detection results. MPSD detect results (top row). NEU-DET detect results (bottom row).

From Fig. 4, we can see that our network has an accurate perception of defects location. Fig. 5 shows detection results of MSPD dataset (top row) and NEU-DET dataset (bottom row). We can see that no matter how much the defect varies in scale or how similar it is to the background, our model can detect them all with high confidence.

4. CONCLUSION

In this paper, a precise anchor-free defect detection network called PDD-Net is proposed to solve three main challenges in defect detection, achieving both precise localization and accurate classification. Firstly, the GC-FPN is proposed to capture the long-range dependency between defects and background to solve the low-contrast problem. Besides, to address the issue of large changes in defects size, the RFPB is designed to extract features of defects at different scales effectively. Furthermore, we improve the detector head with the APNSA mechanism, which makes full use of the statistical characteristics of defects, to solve the severe imbalance problem of positive and negative samples in defect detection. Experimental results on three different datasets show our method achieves better performance in detection accuracy compared with other state-of-the-art methods.

5. REFERENCES

[1] Y. J. Cha, W. Choi, G. Suh, S. Mahmoudkhani, and O. Büyüköztürk, "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types," *COMPUT-AIDED CIV INF*, vol. 33, no. 9, pp. 731–747, 2018.

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. NIPS*, 2015, pp. 91–99.

[3] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. ECCV*, 2014, pp. 818–833.

[4] Y. Xue and Y. Li, "A fast detection method via region-based fully convolutional neural networks for shield tunnel lining defects," *COMPUT-AIDED CIV INF*, vol. 33, no. 8, pp. 638–654, 2018.

[5] J. Chen, Z. Liu, H. Wang, A. Núñez, and Z. Han, "Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network," *IEEE TIM*, vol. 67, no. 2, pp. 257–269, 2017.

[6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. ECCV*, 2016, pp. 21–37.

[7] C. Zhang, C. Chang, and M. Jamshidi, "Concrete bridge surface damage detection using a single-stage detector," *COMPUT-AIDED CIV INF*, vol. 35, no. 4, pp. 389–409, 2020.

[8] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[9] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. CVPR*, 2017, pp. 2117–2125.

[10] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "GCNet: Non-local networks meet squeeze-excitation networks and beyond," in *Proc. ICCVW*, 2019, pp. 1971–1980.

[11] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. ICML*, 2015, pp. 448–456.

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. CVPR*, 2016, pp. 2818–2826.

[13] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proc. ICCV*, 2017, pp. 764–773.

[14] R. Ning, C. Zhang, and Y. Zou, "SRF-Net: Selective receptive field network for anchor-free temporal action detection," in *Proc. ICASSP*, 2021, pp. 2460–2464.

[15] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. ICCV*, 2017, pp. 2980–2988.

[16] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proc. CVPR*, 2016, pp. 761–769.

[17] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," in *Proc. CVPR*, 2020, pp. 9759–9768.

[18] Z. Yang, S. Liu, H. Hu, L. Wang, and S. Lin, "Reppoints: Point set representation for object detection," in *Proc. ICCV*, 2019, pp. 9657–9666.

[19] M. Wieler and T. Hahn, "Weakly supervised learning for industrialoptical inspection," Accessed Jul. 21, 2021, [Online]. Available: https://hci.iwr.uni-heidelberg.de/node/3616.

[20] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE TIM*, vol. 69, no. 4, pp. 1493–1504, 2020.

[21] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, et al., "MMDetection: Open mmlab detection toolbox and benchmark," *arXiv preprint arXiv:1906.07155*, 2019.

[22] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, pp. 400–407, 1951.

[23] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. ICCV*, 2017, pp. 2961–2969.

[24] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. CVPR*, 2018, pp. 6154–6162.

[25] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *Proc. ECCV*, 2018, pp. 734–750.

[26] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. ICCV*, 2019, pp. 9627–9636.

[27] C. Zhu, Y. He, and M. Savvides, "Feature selective anchor-free module for single-shot object detection," in *Proc. CVPR*, 2019, pp. 840–849.

[28] T. Kong, F. Sun, H. Liu, Y. Jiang, L. Li, and J. Shi, "FoveaBox: Beyound anchor-based object detection," *IEEE TIP*, vol. 29, pp. 7389–7398, 2020.