Progressive Fusion Network for Safety Protection Detection

Futian Wang, Lugang Wang, Jin Tang, and Chenglong Li(✉)

Anhui University, Hefei 230601, China
lcl1314@foxmail.com

Abstract. In recent years, it leads to the occurrence of many accidents and huge economic losses, because the construction personnel do not wear safety protective equipment normatively. Therefore, safety protection detection becomes an important problem in the computer vision community. It is a challenging problem because the targets are usually very small, the background is usually very complex at construction site image. To solve these problems, we propose a progressive fusion network PFNet. In PFNet, we use a progressive fusion module to enrich semantic information and a feature enhancement module to enhance detailed information in feature learning. Therefore, we can obtain effective features for safety protection detection. To provide an evaluation platform, we create an image dataset, with 5430 images and careful annotations for safety protection detection. PFNet achieves detection accuracy of 63.7% mAP in our dataset, which is 3.6% higher than the baseline method. PFNet also achieves great detection performance on other datasets.

Keywords: Object detection · Safety protection detection · Progressive fusion · Feature enhancement

1 Introduction

With the continuous expansion of the scale of engineering construction, safety accidents of construction projects often occur. Safety accidents not only affect normal production but also bring enormous impact on people's lives and property safety. And a large part of the reason for the occurrence of safety accidents is people's unsafe behaviors. If the construction personnel can wear safety protective equipment, the probability of accidents can reduce to a minimum. Therefore, it is particularly important to supervise the safety protective equipment of workers.

In recent years, deep learning becomes one of the hot research directions of scholars. Many experts propose a series of deep learning object detection algorithms, such as YOLO [20], SSD [15], Faster R-CNN [22], Cascade R-CNN [1], etc. Safety protection detection based on deep learning aims to realize intelligent supervision of construction personnel and find people's unsafe behaviors, such as not wearing safety protective equipment. Once defects are found, people can make adjustments to greatly improve the safety of the construction site.

Y. Peng et al. (Eds.): ICIG 2021, LNCS 12888, pp. 293–304, 2021.
https://doi.org/10.1007/978-3-030-87355-4_25

As a subtask of object detection, safety protection detection has many challenges. Firstly, the environment of the construction site is very complex, there are lots of construction equipment, buildings, trees, and so on. It causes target occlusion and illumination change. Secondly, because the construction site is large and the shooting is usually far away from the construction site, there are many small targets. These difficulties also exist in general object detection. Different from general object detection, safety protection detection also identifies the normal and abnormal of safety protective equipment. There are many similarities between some classes, such as the normal and abnormal wearing of safety helmet. This situation can easily lead to misclassification in the detection task. Therefore, safety protection detection has a very high detection difficulty.

In order to solve these problems, we propose a progressive fusion network for better feature extraction, as shown in Fig. 1. Firstly, we progressively fuse features of two identical backbones at each stage to obtain features with richer semantic information. Then, we fuse the features of different networks to enhance the detailed information, which is more conducive to the detection of safety protective equipment.

The main contributions of this paper are as follows:

- We create a dataset and call it PSPD. We label the normal and abnormal wearing of safety protective equipment in PSPD. The dataset covers a variety of real-world challenges, such as complex background, small targets, illumination change, occlusion and small differences between different classes. PSPD will be made available to the public.
- We propose a progressive fusion network for safety protection detection, named PFNet. It uses two backbones to extract features and fuses features from different networks. Experimental results show that PFNet can improve the detection performance greatly. The detection accuracy of PFNet is 63.7% mAP on PSPD, which is 3.6% higher than the baseline and higher than some existing advanced detection algorithms. We also conduct experiments on other datasets and obtain great detection results.

2 Related Work

According to the relevance of our work, we review the relevant work from three research directions: object detection, safety protection detection and feature fusion.

2.1 Object Detection

Object detection is an important task in computer vision. Before 2014, the most effective method of object detection is the Deformable Part Model [5]. However, the detection performance of DPM is far inferior to the deep learning methods in recent years.

Since AlexNet [7] shows great results in image classification, various deep learning methods are used in visual tasks. At present, object detection methods are divided into one-stage detectors and two-stage detectors. The one-stage detectors aim to directly classify the predefined anchors and further refine them without generating the suggested steps. There are mainly algorithms such as YOLO [20], SSD [15], RetinaNet [13], CornerNet [9] and FSAF [28]. The two-stage detectors detect objects by generating region suggestions and the region classifier. The two-stage detectors includes Faster R-CNN [22], R-FCN [2], FPN [12], Cascade R-CNN [1], SNIPER [23], TridentNet [10], and so on. In general, the accuracy of the two-stage detector is higher, while the speed of the one-stage detector is faster.

2.2 Safety Protection Detection

Due to the occurrence of safety accidents, people are gradually concerned about the safety protection detection. Long et al. [17] propose a new detection method based on SSD [15] to detect the safety helmet of substation personnel, but the detection accuracy is only 78.3% AP. Marco Di Benedetto et al. [3] create a virtual dataset of safety protective equipment and use Faster R-CNN [22] for training. However, the detection accuracy of virtual data much higher than actual data and the detection performance in real complex scenes is not good. Fatih Can Ksafetyurnaz et al. [8] create a dataset of tools. It includes two types of safety protective equipment: helmet and gloves. Faster R-CNN [22], Cascade R-CNN [1] and other algorithms are used to detect the dataset, Cascade R-CNN [1] with the highest detection accuracy achieves 33.4% mAP.

2.3 Feature Fusion

For object detection, feature fusion is an important means to improve performance. Low-level features have higher resolution and contain more location information. High-level features have much semantic information.

Through top-down connection and horizontal connection, FPN [12] fuses the adjacent features of the backbone to construct the feature pyramid. It enhances the expression of shallow features and significantly improves detection performance. On the basis of FPN [12], PANet [14] performs one more feature fusion from the bottom to the top to further enhance the fusion information of FPN [12]. It has good performance on detection and segmentation. Golnaz Ghiasi et al. [6] create a new feature pyramid structure called NAS-FPN. Unlike the previous method of designing feature fusion, it uses the neural architecture search [29] to select the optimal model architecture in a given search space. It fuses features across a range by top-down and bottom-up connections. Recently, Qiao et al. [19] propose the Recursive Feature Pyramid in DetectoRS. It adds the additional feedback of the feature pyramid network into the backbone network and achieves great detection accuracy.

3 Progressive Fusion Network

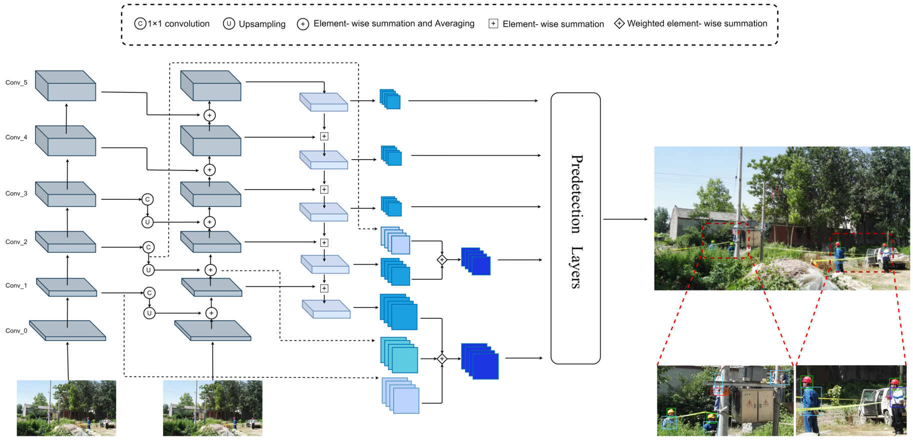
C 1×1 convolution
U Upsampling
+ Element- wise summation and Averaging
+ Element- wise summation
Weighted element- wise summation
Conv_5
Conv_4
Conv_3
Conv_2
Conv_1
Conv_0
C
U
Predetection Layers

Fig. 1. Illustration of the PFNet architecture, including a progressive fusion module and a feature enhancement module.

3.1 Overal Architecture

In this work, we propose a progressive fusion network PFNet, as shown in Fig. 1. Firstly, we propose a progressive fusion module to extract features with more semantic information. The progressive fusion module fuses the adjacent high-level and low-level features of two backbones. Then we use the second backbone to extract the fused features and obtain rich semantic information. The feature fusion method we used can also reduce the noise in the features. But the features have less detailed information due to lower resolution. Therefore, we design a feature enhancement module to increase more detailed information. We perform the feature pyramid operation on the features from the second backbone. Then the features from the two backbones and the feature pyramid are fused for enriching detailed information.

3.2 Backbone

In PFNet, we use DetNet59 [11] as our backbone. It designs new bottlenecks using dilated convolution [25]. The dilated convolution can increase the range of the receptive field while the feature map is unchanged. Therefore, DetNet59 [11] is more powerful at locating large targets and finding small targets, and it can be well applied in object detection. During training, the classification loss function in the network is CrossEntropyLoss, and the bounding box regression loss function is SmoothL1Loss.

3.3 Progressive Fusion Module

Firstly, we propose a progressive fusion module based on CBNet [16]. We believe that the feature fusion method of direct element-wise summation introduces lots of noise when training complex data. It is not conducive to detect small targets. Therefore, we propose a progressive fusion module to better extract features.

As shown in Fig. 1, the progressive fusion module consists of two identical backbones and they have the same input. We fuse the high-level features of the first backbone with the adjacent low-level features of the second backbone. In detail, the input for each stage of the second backbone is the fusion of the features of the previous stage with the adjacent higher-level features of the first backbone. In this way, we gradually introduce the high-level information in the first backbone into the second backbone. Finally, we obtain the features with rich semantic information. The method of feature fusion is the element-wise summation and averaging. This method can not only ensure the increase of useful information and the averaging operation can also avoid the introduction of excessive noise. Adjacent higher-level and lower-level features are less different, and feature fusion between them allows for the better introduction of higher-level information. And the subsequent convolution operation can reduce the information difference caused by the feature fusion of different layers. The fusion method can be described as:

$$F_2^i = G_i[\frac{F_2^{i-1} + C(F_1^i)}{2}] \tag{1}$$

where F_1^i refers to the output of the i-th stage in the first backbone and F_2^{i-1} is the output of the $i-1$-th stage in the second backbone. C consists of 1×1 convolution, batch normalization and upsampling. In this way, we can change the size and channels of the features for subsequent fusion. G_i is the i-th stage in the second backbone.

3.4 Feature Enhancement Module

After progressive fusion, we obtain the features containing rich semantic information, but the low resolution of the features leads to insufficient detailed information. Therefore, we design a feature enhancement module to enhance the detailed information.

Firstly, we use FPN [12] to obtain features with different resolutions. Then we do feature enhancement on them to enhance the detailed information.

As shown in Fig. 1, it can be seen that PFNet consists of three parts from left to right: the first backbone, the second backbone and the feature pyramid. Relatively speaking, high-level semantic information gradually increases while shallow detailed information gradually decreases. So we can integrate the features from multiple networks to further enhance the detailed information. The feature enhancement method is to fuse the same level features of different networks. Using features at the same level can ensure that the differences between features are small and can better enhance detailed information. The feature enhancement method is as follows:

$$F_2 = P_2 + \alpha \times F_{12} + F_{22} \tag{2}$$

$$F_3 = P_3 + \beta \times F_{13} \tag{3}$$

where F_{12} and F_{13} are the output features of the second and third stages in the first backbone. In the second backbone, F_{22} is the input of the third stage. And P_2 and P_3 are the features in FPN. In this paper, $\alpha = \beta = 0.3$.

We only perform feature enhancement on the second and third layers. This is because higher features are used to detect large targets and adding too much detailed information will affect the detection effect. That is worth mentioning that the feature enhancement just performs the addition operation on the features of the same size and adds almost no computation.

4 Dataset

To provide an evaluation platform, we create a dataset and call it PSPD in this paper. This section introduces the details of PSPD, including dataset collection, annotation and statistics.

4.1 Data Collection

The images in PSPD are high-resolution images taken at the construction sites and cover various scene changes such as weather changes, day and night alternations and seasonal changes. They can effectively reflect the actual complex situation of the construction site.

4.2 Data Annotation

In the actual construction process, if construction personnel can wear safety helmet, work clothes and safety belt normatively, it can make the incidence of construction accidents greatly reduced. Therefore, we label them in detail and divide them into six classes in our dataset.

For safety helmet, we label three classes: standard wearing(aqm), without safety helmet (aqmqs), and nonstandard (aqmyc). Aqmyc refers to the helmet strap that does not lace-ups correctly and this situation also has safety hazards. For work clothes, it is divided into wearing standard (gzf) and nonstandard (gzfyc). After discussing with the relevant staff, if the construction personnel have naked arms, their clothes are labeled as nonstandard. Besides, we label the safety belt (aqd) of the construction personnel at height.

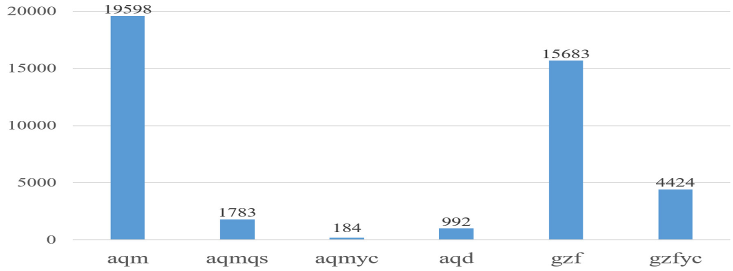

Fig. 2. The number of instances each class in PSPD.

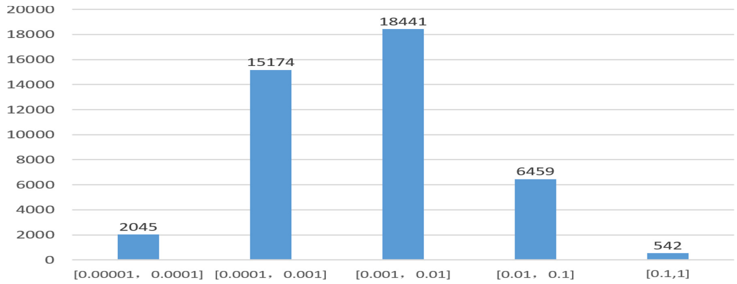

Fig. 3. The number of instances in the proportion of the image.

4.3 Data Statistics

After filtering the data, we obtain 5430 images and 42661 target instances. Then we divide the training set and test set according to the ratio of 9:1. The dataset is shot in real construction scenes. So there are numerous targets of standardized wearing of safety helmet and work clothes, the number of aqmqs, aqmyc and aqd is relatively small. As shown in Fig. 2.

The distribution of pixel size about the images is inconsistent, and the size of most images ranges from 480 × 640 to 14272 × 3968. Therefore we make statistics on the proportion of the targets in the image, as shown in Fig. 3. Then, we use relative size to distinguish between large and small targets, because the small target is less than one hundredth of the image. We find that small targets account for about 83% of PSPD.

The characteristics of our dataset are that it completely conforms to the actual construction scene, the classes are more detailed, the proportion of small targets is relatively high, the background is very complex and the difference between different classes is small. Therefore, this dataset has high research value and is of great help to practical engineering applications.

5 Experimental Results

In this section, we present the experimental results of PFNet on PSPD, PASCAL VOC [4] and SHWD. SHWD is a dataset for safety helmet detection.

5.1 Implementation Details

We re-implement Cascade RCNN [1] with the FPN [12] as our baseline and the backbone is DetNet59 [11]. For the PSPD dataset, we set the image size to 600 × 800 during training and 1000 × 1000 during testing. Our experiments train and test on an NVIDIA Titan XP GPU and develop with PyTorch 0.3.0. The learning rate is initialized to 0.001, and learning rate decay is set to 0.1. In addition, the batchsize is set to 1 on PSPD.

5.2 Results

In order to prove the effectiveness of PFNet, we use the most advanced detectors to carry out a series of experiments, including Faster R-CNN [22], R-FCN [2], YOLOV3 [21], SSD [15], Cascade R-CNN [1], Libra R-CNN [18], ATSS [27], Dynamic R-CNN [26] and FCOS [24]. The detection results are shown in Table 1. It can be seen that PFNet achieves a detection accuracy of 63.73% mAP, which is higher than many existing advanced algorithms. PFNet has good detection performance for aqm, aqmyc and aqd. It can prove that our algorithm can extract more effective features and achieve good detection for small targets and different classes with high similarity.

Table 1. Detection accuracy comparisons in terms of mAP percentage on the PSPD test set.

Methods	aqm	aqmqs	aqmyc	aqd	gzf	gzfyc	mAP
Faster R-CNN [22]	62.10	31.80	38.72	25.21	78.40	73.70	51.66
SSD512 [15]	80.88	38.63	30.07	43.67	76.09	69.96	56.55
RFCN [2]	79.50	40.02	26.07	43.65	82.90	79.56	58.62
YOLOV3 [21]	92.38	54.68	11.91	42.42	85.26	76.32	60.50
Libra R-CNN [18]	84.90	41.40	11.30	68.80	83.80	78.90	61.52
ATSS [27]	88.60	44.10	7.400	52.10	83.90	76.50	58.77
Dynamic R-CNN [26]	81.00	39.90	26.20	50.40	83.70	77.30	59.75
FCOS [24]	89.50	43.70	10.20	54.50	85.30	77.30	60.08
Cascade R-CNN [1] w FPN [12]	80.66	43.33	39.95	43.64	78.70	74.69	60.16
PFNet	80.62	44.25	42.32	55.00	83.21	76.96	63.73

Moreover, we conduct experiments on PASCAL VOC [4] and SHWD. For PASCAL VOC, we reduce the image to 500 × 500 during training and 600 × 600 during testing. For SHWD, the image size is set to 600 × 600. The experimental results are shown in Table 2. Compared with the baseline, the detection accuracy of PFNet improves 0.64% mAP and 1.65% mAP respectively. Therefore, PFNet achieves great detection performance on these datasets.

Table 2. Comparison between PFNet and Baseline on PASCAL VOC and SHWD.

Methods	Dataset	mAP	Hat	Person
Baseline	PASCAL VOC	79.64	–	–
	SHWD	83.13	88.07	78.19
PFNet	PASCAL VOC	80.28	–	–
	SHWD	84.78	89.35	80.22

5.3 Ablation Study

Since PFNet consists of two components, we need to verify the impact of each component on the final performance.

Progressive Fusion Module. As shown in Table 3, to demonstrate the effectiveness of the progressive fusion module, we conduct a set of experiments. On the one hand, we add this module to the baseline and compare it with the baseline. It can be seen that the detection accuracy of the algorithm is 61.74%mAP, which is 1.6% higher than the baseline. On the other hand, we prove the effectiveness of the feature fusion method in the progressive fusion module. In addition to the method we used, we also try the fusion methods of concat, element-wise summation and max operation. They obtain 58.77%, 60.60% and 60.20% mAP respectively, which are lower than our method. It concludes that our method can increase the useful information effectively and has better detection performance.

Table 3. Ablation study. PFM is the progressive fusion module and FEM is the feature enhancement module.

Methods	aqm	aqmqs	aqmyc	aqd	gzf	gzfyc	mAP
Baseline	80.66	43.33	39.95	43.64	78.70	74.69	60.16
+PFM (max)	80.85	44.94	27.89	49.90	78.87	75.66	60.20
+PFM (concat)	80.63	41.36	33.17	42.91	78.87	75.66	58.77
+PFM (element-wise summation)	80.76	47.73	30.81	50.15	79.16	74.97	60.60
+PFM (ours)	80.67	49.77	30.76	54.17	79.08	75.96	61.74
+PFM + FEM ($\alpha, \beta = 1$)	80.86	44.37	35.22	53.39	83.23	76.60	62.28
+PFM + FEM (ours)	80.62	44.25	42.32	55.00	83.21	76.96	63.73

Feature Enhancement Module. We use the baseline with the progressive fusion module to verify the effectiveness of the feature enhancement module. We use the directly element-wise summation to fuse the features from different networks and obtained 62.28% mAP. It improves 0.54% compared with the original. Therefore, a good detection effect can be obtained by fusing the features of different networks. Finally, we fuse the different features in the weighted method and obtain 63.73% mAP, which is 2% higher than the method with the progressive fusion module. Therefore, it can be proved that the feature enhancement module can better enhance detailed information by fusing the features.

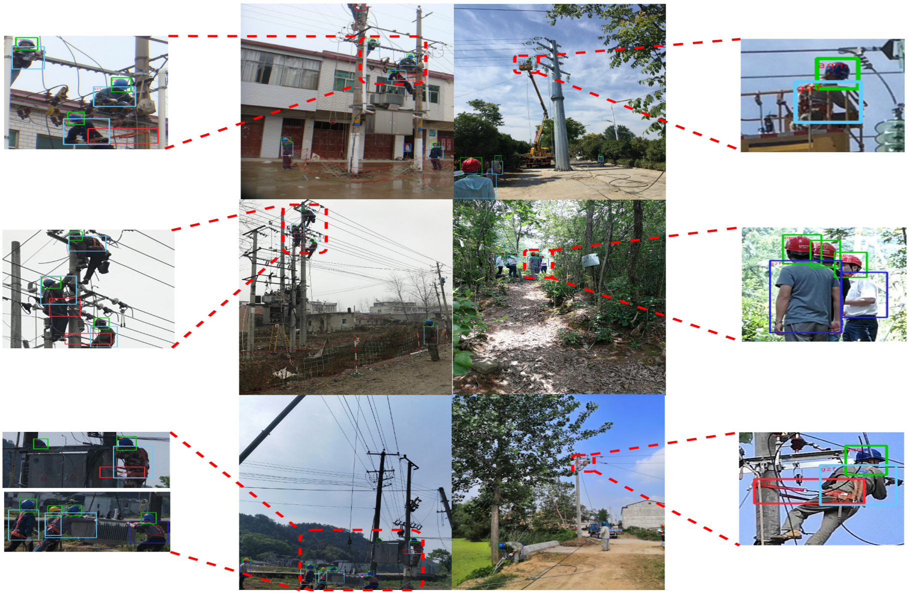

Fig. 4. Detection results on PSPD. The green box is aqm, the black box is aqmqs, the sky blue box is gzf, the purple box is gzfyc, and the red box is aqd. (Color figure online)

6 Conclusion

In this paper, we propose a dataset PSPD and a progressive fusion network PFNet for safety protection detection. PSPD includes lots of images of construction personnel taken at the construction site, and we label the safety protective equipment of the construction personnel. It can be used as an evaluation platform for safety protection detection. PFNet includes a progressive fusion module and a feature enhancement module. These modules can enrich semantic and detailed information in feature learning and achieve great detection performance on multiple datasets.

Acknowledgment. This work was supported by University Synergy Innovation Program of Anhui Province (No. GXXT-2019-007), the National Natural Science Foundation of China (No. 62076003), Anhui Provincial Natural Science Foundation (No. 1908085MF206).

References

1. Cai, Z., Vasconcelos, N.: Cascade R-CNN: delving into high quality object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6154–6162 (2018)
2. Dai, J., Li, Y., He, K., Sun, J.: R-FCN: object detection via region-based fully convolutional networks. arXiv preprint arXiv:1605.06409 (2016)
3. Di Benedetto, M., Carrara, F., Meloni, E., Amato, G., Falchi, F., Gennaro, C.: Learning accurate personal protective equipment detection from virtual worlds. Multimedia Tools Appl. **80**(15), 23241–23253 (2020). https://doi.org/10.1007/s11042-020-09597-9
4. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (VOC) challenge. Int. J. Comput. Vis. **88**(2), 303–338 (2010)
5. Felzenszwalb, P., McAllester, D., Ramanan, D.: A discriminatively trained, multiscale, deformable part model. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE (2008)
6. Ghiasi, G., Lin, T.Y., Le, Q.V.: NAS-FPN: learning scalable feature pyramid architecture for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7036–7045 (2019)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105 (2012)
8. Kurnaz, F.C., Hocaog̃lu, B., Yılmaz, M.K., Sülo, İ, Kalkan, S.: ALET (Automated Labeling of Equipment and Tools): a dataset for tool detection and human worker safety detection. In: Bartoli, A., Fusiello, A. (eds.) ECCV 2020. LNCS, vol. 12538, pp. 371–386. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-66823-5_22
9. Law, H., Deng, J.: CornerNet: detecting objects as paired keypoints. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 734–750 (2018)
10. Li, Y., Chen, Y., Wang, N., Zhang, Z.: Scale-aware trident networks for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6054–6063 (2019)
11. Li, Z., Peng, C., Yu, G., Zhang, X., Deng, Y., Sun, J.: DetNet: a backbone network for object detection. arXiv preprint arXiv:1804.06215 (2018)
12. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125 (2017)
13. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
14. Liu, S., Qi, L., Qin, H., Shi, J., Jia, J.: Path aggregation network for instance segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8759–8768 (2018)
15. Liu, W., et al.: SSD: single shot MultiBox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
16. Liu, Y., et al.: CBNet: a novel composite backbone network architecture for object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 11653–11660 (2020)

17. Long, X., Cui, W., Zheng, Z.: Safety helmet wearing detection based on deep learning. In: 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), pp. 2495–2499. IEEE (2019)
18. Pang, J., Chen, K., Shi, J., Feng, H., Ouyang, W., Lin, D.: Libra R-CNN: towards balanced learning for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 821–830 (2019)
19. Qiao, S., Chen, L.C., Yuille, A.: DetectoRS: detecting objects with recursive feature pyramid and switchable atrous convolution. arXiv preprint arXiv:2006.02334 (2020)
20. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
21. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
22. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497 (2015)
23. Singh, B., Najibi, M., Davis, L.S.: SNIPER: efficient multi-scale training. arXiv preprint arXiv:1805.09300 (2018)
24. Tian, Z., Shen, C., Chen, H., He, T.: FCOS: fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9627–9636 (2019)
25. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)
26. Zhang, H., Chang, H., Ma, B., Wang, N., Chen, X.: Dynamic R-CNN: towards high quality object detection via dynamic training. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12360, pp. 260–275. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58555-6_16
27. Zhang, S., Chi, C., Yao, Y., Lei, Z., Li, S.Z.: Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9759–9768 (2020)
28. Zhu, C., He, Y., Savvides, M.: Feature selective anchor-free module for single-shot object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 840–849 (2019)
29. Zoph, B., Le, Q.V.: Neural architecture search with reinforcement learning. arXiv preprint arXiv:1611.01578 (2016)