Enhanced Feature Fusion and Multiple Receptive Fields Object Detection

Hailong Liu, Jinrong Cui(✉), Haowei Zhong, and Cheng Huang

College of Mathematics and Informatics, South China Agricultural University,
Guangzhou 510642, China
tweety1028@163.com

Abstract. CenterNet is a widely used single-stage anchor-free object detector. It only uses single feature map to detect all size objects, and does not effectively use different levels of feature maps. We present an enhanced feature fusion and multi receptive field object detector, named EM-CenterNet. Our detector first fuses different levels of feature maps, and then enhances feature fusion through semantic information transfer path. Besides, we design another key component, which is composed of continuous several dilated convolutions and shortcut connections, so that our detector can cover all object's scales. We compare the EM-CenterNet method with the baseline on the Pascal VOC and COCO datasets. Experiments show that our method increases the AP by 12.2% on the Pascal VOC dataset, and increases the AP by 5.9% on the COCO dataset.

Keywords: Deep learning · Object detection · Receptive field · Feature fusion

1 Introduction

Computer vision technology has played a powerful role in many fields [1, 2]. At present, object detection technology is widely used. The accuracy of object detector has a great influence on the accuracy of subsequent tasks based on object detection. In recent years, more and more excellent detectors have been proposed. According to different detection processes, the types of object detectors generally include one-stage detectors [3, 4] and two-stage detectors [5, 6]. Two-stage detectors predict boxes proposals and one-stage detectors slide several specific bounding boxes over the image and classify them directly. Whether one-stage detector or two-stage detector, most of them are anchor-based detectors. Specifically, they need to place many carefully set anchors on the image in advance. And it is very complex and needs to adjust the hyper-parameters of the anchors according to different datasets.

Recently, anchor-free detector, as an emerging framework for object detection with a simple strategy and excellent performance, has received increasing attention. FCOS [7] is a widely used anchor-free object detector. It treats all samples in the ground truth as positive samples, and then obtains the distance from the center point to the four sides directly through regression. At the same time, FCOS adds a center-ness branch to

S. Yang and H. Lu (Eds.): ISAIR 2022, CCIS 1700, pp. 118–129, 2022.
https://doi.org/10.1007/978-981-19-7946-0_11

obtain higher quality detection boxes. Another anchor-free object detector, CornerNet [8], detects two bounding box corners as keypoints.

Different from the above two anchor-free detector, CenterNet [9] only regards the center point of the object as the positive sample, and other points are regarded as positive samples, and then regresses at the center point to obtain the size of the object. CenterNet is simpler and faster because it only needs to detect all the objects of different sizes on a single high resolution feature map. However, CenterNet only obtain the feature map by upsampling the smallest high-level feature map for detection, but does not effectively use the low-level feature maps. The research of Li et al. shows that objects with different scales have different needs for the best receptive field of the network [10]. In addition, our investigation found that the receptive field of CenterNet is not enough to cover all objects' scales.

In this paper, two key components, Enhanced feature fusion and Residual dilated convolution are proposed, which bring considerable improvements. First, we propose a semantic information transmission path to enhance feature fusion. Specifically, we fuse feature maps at different levels, and we use semantic information transfer path to transfer the semantic information, so as to significantly enhance feature fusion. Then we design a continuous dilated convolution module with shortcut connections. Experiments on two datasets which are challenging, Pascal VOC and COCO datasets, demonstrate the effectiveness of our method. The main contributions of our work are as follows:

1) We fusing the low-level feature maps of the backbone to improve the performance of CenterNet. And we propose a semantic information transfer path to enhance feature fusion.
2) We verify the influence of the receptive field of CenterNet on the detection performance of objects with different scales, and the current receptive field of CenterNet is not enough to cover targets with all scales. And we propose a continuous dilated convolutions module with shortcut connections, which can generate a feature with multiple receptive fields.
3) Sufficient experiments show the advantages of our proposed detector.

2 Related Work

2.1 Anchor-Based Detector

Two-Stage Method. R-CNN [11] innovatively uses convolutional neural network to detect objects. The detection steps of R-CNN are complex. Firstly, it obtains redundant candidate boxes through selective search, and then classifies the candidate boxes and obtains the size of the objects through regression in the second stage. R-CNN has good detection performance, but its slow inference speed limits its application. Faster R-CNN [5] uses the Regional Proposal Network (RPN) to speed up the generation of candidate boxes. At present, the two-stage detectors have the most advanced accuracy.

One-Stage Method. Different from the above object detectors, the earliest one-stage object detector YOLOv1 [12] based on deep learning does not need to form excessive candidate boxes, but directly divides the image into many regions, classifies each region and predicts the size of the objects. This method significantly shortens the inference speed, but it is less accurate than two-stage detectors. SSD [13] uses multi-scale feature maps to predict the location and category of objects, so it has better detection performance for small objects. Thereafter, many object detectors have followed this approach [14, 15]. YOLOv3 [3] draws on the multi-scale feature maps of SSD and introduces feature pyramid network (FPN) [16] to improve the detection accuracy of small objects. RetinaNet [4] is a relatively new detector, which uses focal loss to alleviate the problem of category imbalance in the process of network training.

2.2 Anchor-Free Detector

The first successful universal anchor-free detector is YOLOv1 [12]. The inference speed of YOLOv1 is surprising, but it is not as accurate as the anchor-based object detectors. Therefore, its successor, YOLOv2, abandons the anchor-free design. Recently, the proposal of CornerNet [8] has turned the attention of the academic to the anchor-free object detectors. CornerNet does not need to regress the size of the objects, but only needs to predict two key points of the objects, and determine the category and location through the key points. Another successful anchor-free detector, FCOS [7], introduces FPN to detect objects with different scales and has achieved competitive performance. CenterNet [9] determines the location and category by predicting the center point of the objects, and then predicts the size of the objects at the center point. CenterNet detects objects at all scales only through a high-resolution feature map, thus its reasoning speed is very fast.

2.3 Feature Fusion

Different levels of feature maps contain different semantic information or spatial information. Feature pyramid network (FPN) [16] fuses different level feature maps, and significantly improves the detection effect of small objects. Zhang et al. added semantic information to the low-level feature maps to enhance the fusion effect, which slightly improved the performance of the instance segmentation method [17]. We use a similar way to improve the object detector. The experimental results show that the low-level feature map containing more semantic information can be better fused with the high-level feature map.

2.4 Dilated Convolutions

Dilated convolution is a common component in many semantic segmentation methods [18, 19]. It increases the receptive field without losing information. Now, many object detectors [20, 21] also use dilated convolution to improved accuracy. In this paper, we stack convolution layers with different dilated rates to obtain feature map that can cover objects of different scales.

3 Receptive Field of CenterNet

In this section, we will introduce CenterNet at length, and then we design a scientific experiment to investigate the influence of receptive field on CenterNet object detector.

3.1 CenterNet

Different from other target detectors, CenterNet's approach is similar to key point detection. [22, 23], which represents objects by a single center point. In the reasoning stage, we only need to input the image into the network to get the heatmap representing the location and category of the objects, and the size of the objects.

The construction of CenterNet is very sample, the backbone network generates a low-resolution feature map, and then obtains a high-resolution feature map after three consecutive up sampling. After the image is input into the network, four feature maps with width and height gradually reduced by half are obtained from the bottom-up pathway, that is, the backbone network. Then the smallest high-level feature map is upsampled for three consecutive times to obtain a high-resolution feature map for subsequent detection.

Specifically, for ResNets [24] we choose the last four output feature maps as the selection and mark them as C2, C3, C4, and C5. The top-down pathway generates feature maps with higher resolution by upsampling the high-level feature maps, and we mark them as P2, P3, P4, and P5. It should be noted that P5 is produced applying one 1×1 convolutional layer on C5. CenterNet determines the location and category by detecting the center point of the objects, and directly predicts the width and height of the objects. At the same time, to compensate for the center point offset caused by downsampling, CenterNet predicts the center point offsets.

Table 1. Results with different receptive fields using CenterNet [9] evaluated on the Pascal VOC dataset [26].

Dilation rate	AP	AP_S	AP_M	AP_L
1	38.6	**6.2**	25.0	48.5
2	40.3	5.2	**25.1**	51.3
4	**41.2**	5.7	25.0	**52.6**

3.2 Investigation of Receptive Field

The receptive field of the network is one of the key factors affecting the performance of object detectors [25]. To investigate the relationship between receptive field and CenterNet detection effect, we add a dilated convolution layer between the backbone network and the upsampling structure. We use three different dilation rates to generate networks with three different receptive fields.

We conduct our experiment using the CenterNet with the ResNet18 backbone on the VOC dataset. The dilation rates used in the experiment are 1, 2 and 4. And we report Average Precision (AP) on object of small (AP_S), medium (AP_S) and large sizes (AP_L).

We can find that the detection performance of objects with different scales is positively correlated with the dilation rate from Table 1. In other words, larger receptive fields are better for detecting large objects. This phenomenon strongly shows that the receptive field of CenterNet is not enough to cover all objects' scales. These findings inspire the following improvements to the CenterNet object Detector.

4 EM-CenterNet

This section will describe the main components of our proposed EM-CenterNet detector in detail. The proposed EM-CenterNet consists of Enhanced feature fusion and Residual dilated convolution. The brief structure of EM-CenterNet as shown in Fig. 1. First, we describe the proposed components of EM-CenterNet. Then, we also introduce several loss functions for training in detail.

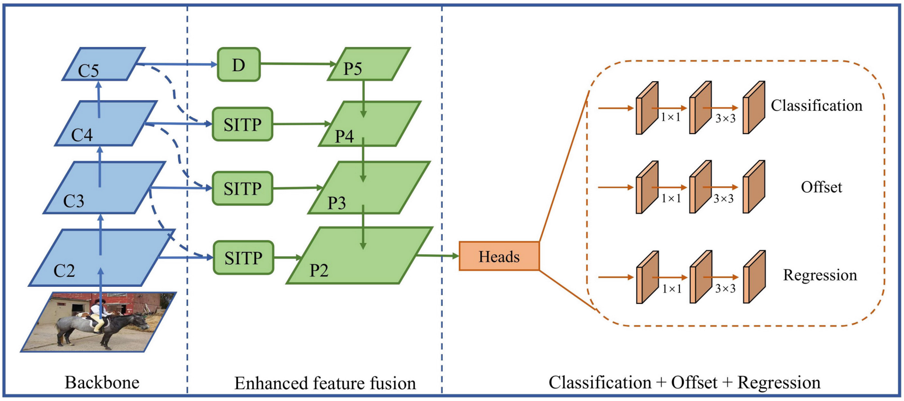

Fig. 1. The brief structure of EM-CenterNet detector, in which C2, C3, C4 and C5 are the feature maps of different scales in the backbone network respectively, D is residual dilated convolution, the bule dotted line is the path of semantic information transmission, SITP means semantic information transmission path, and P2 to P5 are the feature levels.

4.1 Enhanced Feature Fusion

To effectively use the low-level feature map rich in spatial information, we improve the structure of CenterNet. After sampling on the feature maps, we fuse them with the feature maps in the backbone, and finally get a feature map containing rich information. And to enhance the effect of feature fusion, we introduce semantic information transmission path. We will describe it in detail later.

Feature Fusion. The construction of our feature fusion as like FPN [16]. We upsample the high-level feature maps and add them pixel by pixel with the feature maps in the backbone to get the final high-resolution feature map for detection. The detailed design of feature fusion is illustrated in Fig. 2.

Semantic Information Transmission Path. In object detection, semantic information is conducive to the classification of objects, and spatial information is more conducive to the positioning of objects. Previous studies [17] believe that due to the large difference of semantic information between them, it is not the best way to directly fuse them.

To enhance the effect of feature fusion, we introduce three semantic information transmission paths. Specifically, we first sample the three high-level feature maps in

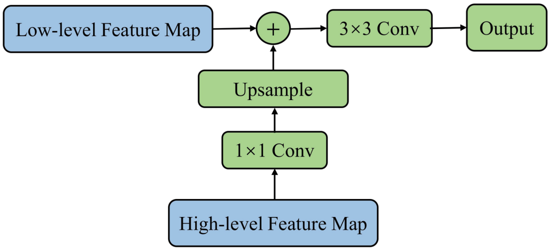

Fig. 2. The network architecture of feature fusion. The " +" sign means element-wise addition.

the backbone, then multiply them pixel by pixel with the feature maps in the previous stages in the backbone, and finally add the obtained feature maps pixel by pixel with the feature maps in the top-down path. The detailed design of the semantic information transmission path is illustrated in Fig. 3.

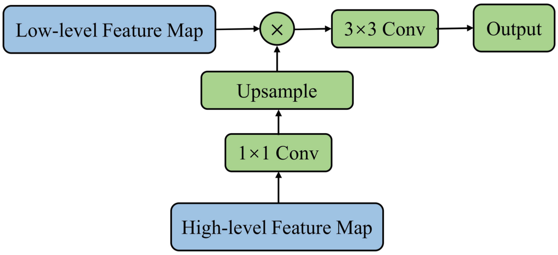

Fig. 3. The network architecture of Semantic information transmission path. The " ×" sign means element-wise multiplication.

4.2 Residual Dilated Convolution

From Sect. 3, we can see that the detection effect of CenterNet object detector on objects of different sizes is closely related to its receptive field. And the receptive field of CenterNet is not enough to cover all objects' scales.

To increase the receptive field of the CenterNet, we first designed a continuous dilated convolution structure. We add four continuous convolution layers with different dilation rates between the backbone network and the upsampling structure. At the same time, we reduce the channel dimension of the feature map by applying one 1 × 1 convolution layer and then add a 3 × 3 convolution layer. The dilation rates are 2, 4, 6, and 8, respectively.

The above structure is very simple, but continuous dilated convolution will lead to a large receptive field, which is not friendly for small objects. To solve this problem, we add a shortcut connection after each dilated convolution layer. The residual dilated convolution structure is shown in Fig. 4.

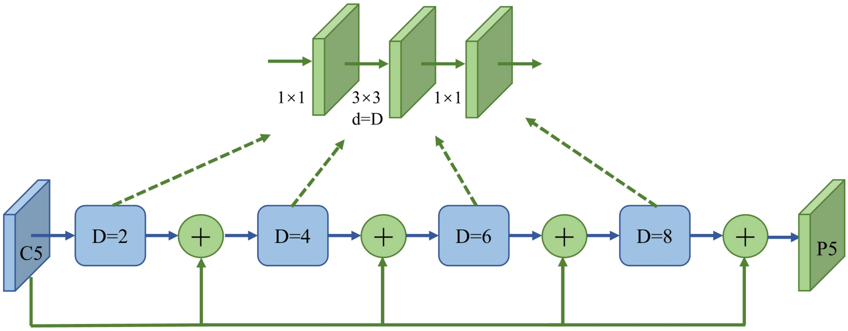

Fig. 4. The network architecture of Residual Dilated Convolution. The D means dilation rate.

4.3 Loss Function

The training process of our EM-CenterNet is consistent with CenterNet. For center localization, we use Gaussian kernel to produce a heat-map. For the prediction σ and the target y, we have:

$$L_K = \frac{-1}{N} \sum_{xyc} \begin{cases} (1-\sigma)^{\alpha} \log \sigma & \text{if } y = 1 \\ (1-y)^{\beta} (\sigma)^{\alpha} \log(1-\sigma) & \text{otherwise} \end{cases} \tag{1}$$

where $\alpha = 2$ and $\beta = 4$, N is the number of keypoints, following CenterNet [9].

Center point offset due to downsampling, CenterNet predict the offset of the center of the objects. The offset loss:

$$L_{off} = \frac{-1}{N} \sum |p - q| \tag{2}$$

where p is the true offset and q is the offset of the network output.

For size regression, we use an L1 loss:

$$L_{size} = \frac{-1}{N} \sum |S - \hat{S}| \tag{3}$$

where S is the true width and height, and $\hat{\mathrm{S}}$ is the predicted width and height.

The total loss L includes localization loss L_K, offset loss L_{off} and size loss L_{size}, weight by two scalars. The total loss is:

$$L = L_K + w_1 L_{off} + w_2 L_{size} \tag{4}$$

where $w_1 = 1$ and $w_2 = 0.1$ in our setting as in CenterNet.

5 Experiments

We evaluate our EM-CenterNet on the Pascal VOC [26] and MS COCO datasets [27]. We first introduce our experimental setting, including datasets and training details. Then we compare the results of EM-CenterNet on the test-dev set of MS COCO dataset with CenterNet and other methods. Finally, we provide detailed ablation experiments of each component of our proposed EM-CenterNet object detector and provide quantitative results and analysis.

5.1 Experimental Setting

Datasets. In this section, we first describe the datasets and experimental settings we used, then we show the main results, and finally we show the ablation results. We experimented on Pascal VOC and COCO datasets. Pascal VOC datasets include VOC 2007 and VOC 2012 datasets. VOC 2007 includes 5011 training images and 4952 test images. VOC 2012 has 11540 training images, and the annotation of the test images is not disclosed. We test the performance of the detector on the test images of the VOC 2007 dataset.

Training Details. ResNet-18 is the backbone in our method for experiments and we resize the images to 512 × 512. The initial learning rate is 0.001, and the mini-batch size is 16. We reduced the learning rate by a factor of 10 at epoch 90 and 120 for 140 total epochs. Our optimizer is SGD, and weight decay is 0.0005. Warm-up is applied for the first epoch.

5.2 Main Results

We evaluate our EM-CenterNet on the COCO dataset and VOC dataset, and we adopt ResNet-18 as the backbone. As shown in Table 2, the AP of our detector is improved by 5.9% compared with baseline on the VOC dataset. As shown in Table 3, the AP of our detector is improved by 5.9% compared with baseline on the COCO dataset. And compared with other popular detectors, our method also has strong competitiveness with the same backbone.

Table 2. The experimental results on VOC dataset are compared with several newer detectors.

Method	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
FCOS [7]	R50	44.1	73.5	46.2	14.9	32.2	51.9
RetinaNet [4]	R50	44.8	73.1	46.9	13.3	32.2	52.6
YOLOF [28]	R50	49.6	**76.6**	54.1	11.2	35.4	59.6
EM-CenterNet	R50	**51.4**	74.6	**56.3**	**17.1**	**36.7**	**61.2**

5.3 Visualization Results

Figure 5 visualizes the detection results of CenterNet and EM-CenterNet. We can see that the detection result of our proposed method is better than the baseline. Specifically, the detection boxes of our method are more accurate and have higher classification confidence.

Table 3. Comparison with CenterNet and other popular object detectors on COCO test-dev. Using the same backbone resnet-18, EM-CenterNet outperforms the baseline counterpart CenterNet by 5.9% in AP. EM-CenterNet also outperforms the CenterNet with ResNet-18-DCN as its backbone.

Method	Backbone	AP	AP$_{50}$	AP$_{75}$	AP$_S$	AP$_M$	AP$_L$
SSD [13]	VGG16	25.7	43.9	26.2	6.9	27.7	**42.6**
YOLOv3 [3]	D53	28.2	-	-	-	-	-
FCOS [7]	R18	26.9	43.2	27.9	**13.9**	28.9	36.0
CenterNet [9]	R18-DCN	28.1	44.9	29.6	-	-	-
CenterNet	R18	23.9	41.6	25.0	9.1	26.1	33.4
EM-CenterNet	R18	**29.8**	**47.3**	**31.6**	11.3	**30.7**	**42.6**

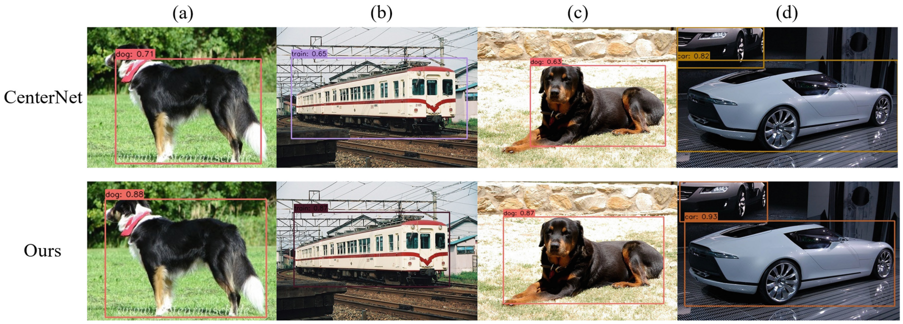

Fig. 5. Visualization of CenterNet and EM-CenterNet detection results.

5.4 Ablation Experiments

Enhanced Feature Fusion. As shown in Table 4, we first add feature pyramid network to CenterNet, and the result of ablation experiment result show that feature fusion can significantly improve the performance of CenterNet (36.8 vs. 41.3). Then we add the semantic information transfer path to feature pyramid network. The result of ablation experiment show that the semantic information transfer path significantly enhances feature fusion (41.3 vs. 42.4). After adding semantic information transfer path, the performance of large objects is significantly improved, and the performance of medium objects and small objects is slightly increased.

Residual Dilated Convolution. Based on the enhanced feature fusion component, we further add the residual dilated convolution component. As shown in Table 4, residual dilated convolution also significantly improves the performance of baseline (41.3 vs. 48.3). Finally, EM-CenterNet achieve significant improvements (12.2 AP increase) on the baseline.

Table 4. Results on the Pascal VOC dataset with ResNet-18. Starting from our baseline, we gradually add Enhanced feature fusion, Residual dilated convolution in our EM-CenterNet for ablation studies. FPN means feature pyramid network.

FPN	Enhanced feature fusion	Residual dilated convolution	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
-	-	-	36.8	60.0	39.9	9.1	27.8	44.6
√	-	-	41.3	65.8	44.7	13.1	30.5	49.4
√	√	-	42.4	66.7	46.0	13.3	31.0	50.7
√	-	√	48.3	73.2	52.8	**14.2**	32.2	58.3
√	√	√	**49.0**	**73.4**	**53.5**	13.7	**32.8**	**59.1**

6 Conclusion

In this work, we find that fusing the low-level features of the backbone can significantly improve the performance of CenterNet. And we propose a semantic information transfer path to enhance feature fusion. In addition, to make the receptive field of our detector can cover all objects of different sizes, we propose residual dilated convolution. We conducted experiments on two challenging data sets. The performance of our proposed detector, EM-CenterNet, is significantly improved compared with the baseline. We hope that our EM-CenterNet object detector can provide insights for designing anchor-free detectors.

References

1. Gao, G., Yang, J., Jing, X.Y., et al.: Learning robust and discriminative low-rank representations for face recognition with occlusion. Pattern Recogn. **66**, 129–143 (2017)
2. Gao, G., Yu, Y., Yang, M., et al.: Cross-resolution face recognition with pose variations via multilayer locality-constrained structural orthogonal procrustes regression. Inf. Sci. **506**, 19–36 (2020)
3. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
4. Lin, T.Y., Goyal, P., Girshick, R., et al.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
5. Ren, S., He, K., Girshick, R., et al.: Faster r-cnn: Towards real-time object detection with region proposal networks. Adv. Neural Inf. Processing Syst. **28** (2015)
6. He, K., Gkioxari, G., Dollár, P., et al.: Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)
7. Tian, Z., Shen, C., Chen, H., et al.: Fcos: fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9627–9636 (2019)
8. Law, H., Deng, J.: Cornernet: detecting objects as paired keypoints. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 734–750 (2018)
9. Zhou, X., Wang, D., Krähenbühl, P.: Objects as points. arXiv preprint arXiv:1904.07850 (2019)

10. Li, Y., Chen, Y., Wang, N., et al.: Scale-aware trident networks for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6054–6063 (2019)
11. Girshick, R., Donahue, J., Darrell, T., et al.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)
12. Redmon, J., Divvala, S., Girshick, R., et al.: 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)
13. Liu, W., Anguelov, D., Erhan, D., et al.: Ssd: Single shot multibox detector. In: European Conference on Computer Vision. Springer, Cham, pp. 21–37 (2016). https://doi.org/10.1007/978-3-319-46448-0_2
14. Fu, C.Y., Liu, W., Ranga, A., et al.: Dssd: deconvolutional single shot detector. arXiv preprint arXiv:1701.06659 (2017)
15. Yi, J., Wu, P., Metaxas, D.N.: ASSD: Attentive single shot multibox detector. Comput. Vis. Image Underst. **189**, 102827 (2019)
16. Lin, T.Y., Dollár, P., Girshick, R., et al.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125 (2017)
17. Zhang, Z., Zhang, X., Peng, C., Xue, X., Sun, J.: ExFuse: enhancing feature fusion for semantic segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11214, pp. 273–288. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01249-6_17
18. Yu, F., Koltun, V., Funkhouser, T.: Dilated residual networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 472–480 (2017)
19. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)
20. Li, Z., Peng, C., Yu, G., et al.: Detnet: A backbone network for object detection. arXiv preprint arXiv:1804.06215 (2018)
21. Liu, S., Huang, D., Wang, Y.: Receptive field block net for accurate and fast object detection. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11215, pp. 404–419. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01252-6_24
22. Cao, Z., Simon, T., Wei, S.E., et al.: Realtime multi-person 2d pose estimation using part affinity fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7291–7299 (2017)
23. Zhou, X., Karpur, A., Luo, L., Huang, Q.: StarMap for category-agnostic keypoint and viewpoint estimation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11205, pp. 328–345. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01246-5_20
24. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
25. Cai, Z., Fan, Q., Feris, R.S., et al.: A unified multi-scale deep convolutional neural network for fast object detection. In: European Conference on Computer Vision. Springer, Cham, pp. 354–370 (2016). https://doi.org/10.1007/978-3-319-46493-0_22
26. Everingham, M., Van Gool, L., Williams, C.K.I., et al.: The pascal visual object classes (voc) challenge. Int. J. Comput. Vision **88**(2), 303–338 (2010)
27. Lin, T.Y., Maire, M., Belongie, S., et al.: Microsoft coco: Common objects in context. In: European Conference on Computer Vision. Springer, Cham, pp. 740–755 (2014). https://doi.org/10.1007/978-3-319-10602-1_48

28. Chen, Q., Wang, Y., Yang, T., et al.: You only look one-level feature. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13039–13048 (2021)
29. Huimin, L., Zhang, M., Xing, X.: Deep fuzzy hashing network for efficient image retrieval. IEEE Trans. Fuzzy Syst. (2020). https://doi.org/10.1109/TFUZZ.2020.2984991
30. Huimin, L., Li, Y., Chen, M., et al.: Brain Intelligence: go beyond artificial intelligence. Mobile Networks Appl. **23**, 368–375 (2018)
31. Huimin, L., Li, Y., Shenglin, M., et al.: Motor anomaly detection for unmanned aerial vehicles using reinforcement learning. IEEE Internet Things J. **5**(4), 2315–2322 (2018)
32. Huimin, L., Qin, M., Zhang, F., et al.: RSCNN: A CNN-based method to enhance low-light remote-sensing images. Remote Sensing **13**(1), 62 (2020)
33. Huimin, L., Zhang, Y., Li, Y., et al.: User-oriented virtual mobile network resource management for vehicle communications. IEEE Trans. Intell. Transp. Syst. **22**(6), 3521–3532 (2021)