International Core Journal of Engineering Volume 10 Issue 5, 2024
ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0013

A Lightweight Defect Detection Network for Transmission Conductors

Xueming Zhai, Guangyao Hou

North China Electric Power University, Baoding 071003, China

Abstract

Due to the limited area of defects in pictures captured by UAVs, there is a high misdetection and leakage rate when using machine learning and other detection methods. This paper proposes a lightweight defect detection method that improves detection rate while reducing computational requirements of the model, speeds up detection, and is suitable for deployment on mobile platforms. The architecture of ShuffleNetV2's lightweight network was adapted for use in CspDarknet's backbone network. This was achieved by replacing the standard convolution with a new convolutional approach and modifying the structure of CspDarknet. The resulting model improves the ability to model small features while reducing the amount of computation and parameter size of the network. By utilizing the AnchorFree method to modify the allocation of positive and negative samples, the model can acquire a greater number of high-quality positive samples, thereby enhancing the final detection accuracy. Ultimately, on the transmission conductor defect detection dataset, the proposed method achieves a detection accuracy of 93.6%map, with a model size of only 5.2M and a photo detection speed of 124fps. This approach realizes a joint improvement of speed and accuracy.

Keywords

Aerial Photography; Electric Power Transmission; Defect Detection; Small Object Detection; Light Weight Structures; Convolutional Neural Networks; Deep Learning.

1. Introduction

In recent years, the demand for electricity has led to the widespread distribution of transmission lines, resulting in a rapid increase in their total length. These lines are erected in various natural environments and are exposed to year-round weather conditions such as wind, sun, and rain, as well as occasional extreme weather events like ice and snow. As a result, power equipment is often subject to damage or failure. Regular inspection is crucial to ensure a continuous supply of electricity and the safe operation of transmission lines [1].

The manual inspection method, traditionally used, is inefficient, high-risk, and not suitable for real-time monitoring [2]. Moreover, many substations are located in remote areas with poor geographic conditions, which poses a significant challenge to the intelligent operation and maintenance of power transmission and distribution. In recent years, the power industry has increasingly relied on Unmanned Aerial Vehicles (UAVs) for intelligent inspections. This technology has become an indispensable and important means of operation and maintenance. Despite its widespread use, a new problem has arisen: UAV inspection often generates a large number of pictures, resulting in massive inspection results. Adopting manual inspection methods would require significant manpower and financial resources. Additionally, continuous engagement in repetitive and tedious work can not only affect the efficiency of manual work but also increase the risk of errors and missed detections.

The field of computer vision technology has seen continuous development, resulting in the successful application of a series of deep learning algorithms based on Deep Convolutional Neural Network (DCNN) for object detection and classification. Deep learning technology has achieved remarkable results in object detection, image processing, and unmanned driving. Additionally, transmission line inspection is moving towards automation and intelligence [3].

2. Related Work

In power line defect detection, Wang Wangguo [4] and others used Gestalt perception theory to quantitatively calculate the identified parallel transmission lines in terms of proximity, continuity, and covariance. They analyzed the presence of broken strands or attached foreign object defects on the conductor based on the detected mutation region. However, the transmission line may not be parallel to the line due to the shooting perspective. Therefore, spatial transformation and other processing are necessary for practical applications. Liu Changyin [5] and others improved the Canny algorithm to extract image edges and combined it with the Hough algorithm to obtain the wire's edge. They determined the presence of defects by analyzing the angle between the broken strand and the wire.

Since AlexNet[6] won the ImageNet 2012 competition with a significantly better result than the runner-up, neural networks have expanded and thrived in various fields. Faster RCNN is the first end-to-end deep learning detection algorithm that achieves closest to real-time performance. Wang [7] improved the Faster-RCNN model to achieve 98.3% accuracy on the insulator detection task. SSD [8] is a typical one-phase detection model that performs multiple branch detection at different depths, allowing for multi-scale object detection and much better detection of small objects. The SSD[9] model achieved 85.2% accuracy in detecting foreign objects on transmission lines, which is a task-heavy process. Dong Zhaojie [10] utilized yolov3 as the foundation for power component recognition and achieved a speed of 57 frames/s for three types of power components. The component recognition mAP reached 90.80%, which is almost 3% higher than that of the Faster R-CNN algorithm.

2.1 Problems and Difficulties in Detecting Defects in Transmission Conductors

When relying on UAVs for transmission line inspections, a large amount of data must be inspected. The majority of the image data captured by these UAVs shows normal and intact transmission lines, while the portions that show broken strands and wear and tear only account for a small fraction. Additionally, the worn and broken strand portion is relatively small in the image, making it difficult for the network to extract efficient features. On the other hand, misclassifying fewer positive samples as defects results in minimal loss during model training. This is because the parameters tend to favor a high number of fits, which can ultimately lead to model failure. Additionally, high-precision models often have high computational complexity and are not suitable for deployment on lightweight devices with limited arithmetic power. Most detectors widely used in industry have inherent shortcomings and flaws in various aspects, such as model structure design, positive and negative sample handling, and loss functions. It is important to note that subjective evaluations should be excluded unless clearly marked as such.

In response to the characteristics listed above, we present our work:

(1) To address issues with the conventional dataset, this paper employs traditional image processing algorithms to simulate various weather conditions and enhance the background of the dataset's images. In cases where the proportion of positive samples is low, the paper utilizes random copying to add more objects for detection in each image.

(2) The yolov5s model is used as a benchmark due to its ability to combine accuracy and speed. A new convolutional approach, known as local convolution, is introduced instead of the standard convolution, and the network architecture is reconfigured.

(3) The static positive and negative sample allocation method based on anchor frames is less effective in small targets and datasets with imbalanced positive and negative samples. Therefore, we use a

Volume 10 Issue 5, 2024
DOI: 10.6919/ICJE.202405_10(5).0013

dynamic positive and negative sample allocation algorithm without anchor frames to filter out positive samples with higher quality.

The paper achieves a 93.6% map accuracy in the transmission conductor defect detection task, surpassing the benchmark networks yolov5s, yoloxs, and yolov7-tiny. The model size is only 5.2M, and the detection speed (FPS) reaches 124, which is higher than yolov5s and yoloxs by 26.5% and 20.4%, respectively, and is comparable to yolov7-tiny. The detection speed (FPS) is 124, which exceeds the speed of yolov5s and yoloxs by 26.5% and 20.5%, respectively.

3. Methods

Yolov5 is an end-to-end object detection model that improves on the modeling and feature extraction of yolov3 and yolov4 through enhanced network architecture. Figure 1 illustrates that yolov5 comprises three components: the backbone network, the neck network, and the detection head. The backbone network is responsible for initial modeling and information extraction from the image. The neck network is primarily responsible for fusing features at various scales. YOLOv5 utilizes the PAN structure [11] for information transfer and fusion between deep and shallow features. The PAN includes both top-to-bottom and bottom-to-top paths, which are designed to enable better utilization and transfer of semantic information in the network. The detection head comprises three different detection heads corresponding to different object detection scales.

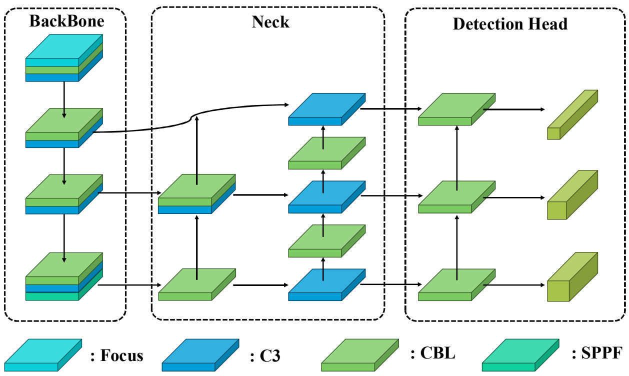


Figure 1. YOLOv5 Network Architecture

There are four different configurations of the Yolov5 model, each corresponding to a different model size. This paper is based on the Yolov5s model, which aims to create a lightweight network model. Yolov5s adjusts the size of the network model by controlling its depth and width. However, it is important to note that the depth and width of the model have a significant impact on its final accuracy. Therefore, this paper introduces a more effective lightweight module to replace the simple method of controlling the depth and width of the network. This achieves a more lightweight model while still improving accuracy.

3.1 Lightweight Convolution Module

The use of grouped convolution and deep convolution is crucial in the development of lightweight networks, as demonstrated in related works such as Xception [12], MobileNet [13], MobileNetV2 [14], ShuffleNet [15], and CondenseNet [16]. However, recent studies (ShuffleNetv2, FasterNet) have found that excessive use of grouped convolutions can actually decrease model speed. ShuffleNetv2 [17] notes that the current common measure of model complexity is FLOPs, specifically the number of multiple-additions, but this is an indirect indicator since it is not exactly equivalent to speed. Shufflenetv2 proposes practical guidelines for efficient network design, including ensuring an equal number of channels whenever possible, avoiding overuse of grouped convolution to maintain speed, minimizing network fragmentation to preserve parallelism, and acknowledging the importance of element-level operations. Faster Net also highlights the impact of

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0013

memory accesses on model speed. The article notes that feature maps exhibit high similarity between different channels, a finding that is also supported by GhostNet [18]. To address this issue, the authors propose a new technique called Partial Convolution (PConv), which enables more efficient extraction of spatial features by reducing redundant computations and memory accesses simultaneously.

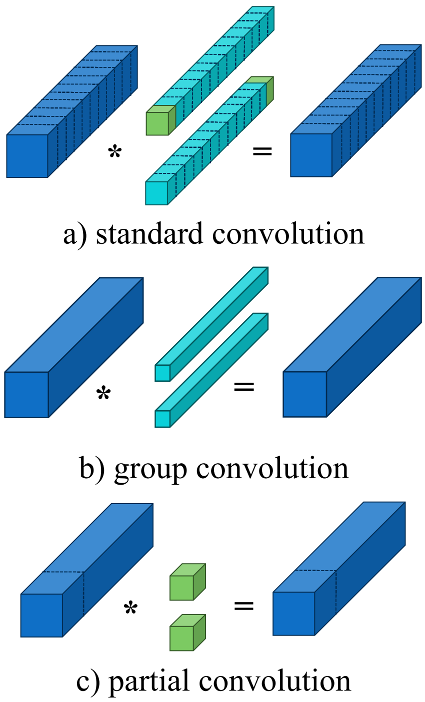


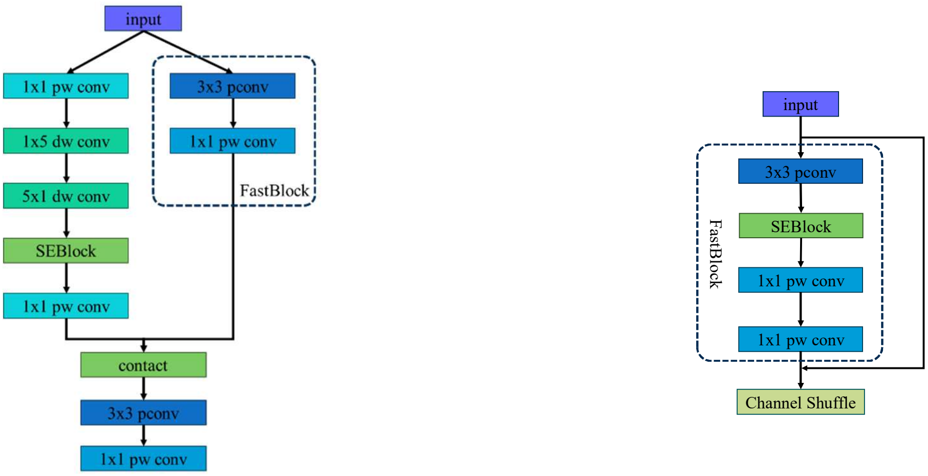
Figure 2. Comparison of three types of convolutions

The paper adheres to the aforementioned provisions in constructing a new lightweight module, FS Block. SE Block [19] is a channel attention module that corrects the relationship of each channel by weighting the individual channels of the feature layer to improve the network's expressive power. Therefore, SE Block is added to FS Block in this paper. Like MobileNetV3, SEBlock uses ReLU and H-Sigmoid as its two-layer activation functions. However, to balance its impact on speed and accuracy, SEBlock is only added at the end of FSBlock. This paper avoids adding any new content beyond the original text.

To minimize memory access during model inference, FasterNet [20] uses local convolution instead of grouped convolution. FasterNet convolves the features of consecutive channels localized in the feature map, taking advantage of the great similarity between the feature maps of different channels. To make the most of all channels, a point-by-point convolution (Point-Wise Conv) is typically used after the local convolution to combine the information from all channels. However, relying solely on one layer of point-by-point convolution to fuse all channel information is insufficient. A more direct approach, without increasing computational effort, is to perform channel blending. Therefore, in this paper, we perform a channel blending operation after the point-by-point convolution. The distinctions between the three convolutions are illustrated in Figure 2.

ShuffleNetV2 has demonstrated excellent performance in many tasks. However, in order to adapt to the task of transmission line defect detection, which is rich in small objects to be detected, as well as to further improve the performance of the ShuffleNet unit in ShuffleNetV2. The work in this thesis improves the ShuffleNet unit as follows. First, for small targets, the model needs sufficient background information for detection and localization. This requires the model to have a sufficient sensory field. For the unit with stride=2, this paper replaces the 3x3 convolution in a path with a 5x5 convolution. To avoid increasing the computational complexity, this paper orthogonally decomposes the 5x5 convolution into a 1x5 convolution and a 5x1 convolution. Due to the parallelism in model inference, the additional convolutional layers do not cause too much performance loss.

On the other hand, although the channel shuffling operation brings about the exchange of information between channels, the random disturbance of the channels will lead to the loss of fusion features. Therefore, in this paper, for the transition unit between two units, i.e., the downsampling unit, the channel shuffling operation is canceled, and a short-circuit connection is added to balance the effects of channel shuffling and network depth. For the stride=1 unit, we cancel the point-by-point convolution and depth convolution and introduce a local convolution to reconstruct the original structure. For the two units with different step sizes, this paper introduces the channel attention mechanism to further improve their performance. Finally, the final structure of the FS block designed in this paper is shown in Figure 3:



a) stride = 2 b) stride = 1

Figure 3. Two different settings for Fs Block

Since the PAN structure has good bidirectional feature fusion characteristics, the neck structure of PAN is still retained in this paper. The difference is that to further reduce the computational cost of the network, this paper makes the following adjustments to the PAN structure. In this paper, we uniformly reduce the number of channels for each layer of features fed into the PAN network, while following the guideline discussed in ShuffleNetV2 to fix the number of channels in the subsequent computation. Due to the stacking of convolutional blocks in the C3 structure, the computational complexity is large, but due to the simplicity of the structure, in this paper, only the deep convolution is used to replace its regular convolution.

3.2 Adaptive Positive and Negative Label Assignment Algorithm

Previous one-stage detection networks, including Anchor Free and Anchor Base detectors, relied on human a priori knowledge for positive and negative sample delineation. However, yolov3[21], RetinaNet[22], and yolov4 [23] use the gt box and the iou of the preset anchors to delineate positive and negative samples. Yolov1[24] and Fcos[25], along with other anchor-free methods, eliminate the need for anchor box design and determine positive and negative samples based on whether the center of the ground truth falls within the bounding box or not.

ATSS [26] points out that the fundamental difference between anchor-free and anchor-based detectors is the division of positive and negative samples.ATSS proposes an adaptive training sample selection method that automatically selects positive and negative training samples based on the statistical properties of the objects.SimOTA is a dynamic label assignment strategy, and unlike ATSS, SimOTA's assignment process changes continuously with the training process.SimOTA was first used in YoloX [27] and has been widely used by detection models for various tasks.

To improve the selection of positive samples of higher quality, this paper enhances the computation of the cost matrix in the original SimOTA algorithm. The final loss function uses Giou as a measure between two anchor frames, as discussed in VarifocalNet[28]. To account for different IOU values in anchor frames, this paper employs Varifocal cal Loss instead of CELoss as the categorization loss.

$$\cos t = loss_{vfl} + \lambda \cdot loss_{giou} \tag{1}$$

4. Experiments

4.1 Experimental Settings

This paper presents the results of a 5-fold cross-validation using 6000 images of damaged transmission lines in a 4:1 allocation. The training process includes warm-up training for the first 3 epochs, with an initial learning rate of 0.01 and a cosine decay learning strategy. The optimizer used is a momentum optimizer with a momentum factor of 0.973. To reduce overfitting, L2 regularization is employed with a weight decay factor of 5e-4.

4.2 Results

As a method for detecting defects in objects, it is important to have an accurate index for detecting all types of defects. Therefore, this paper's experiments use the average precision (AP) and mean average precision (mAP) of all categories as the precision evaluation index. Unless otherwise specified, AP in the following results represents map50:90.

Table 1. Experiments on the dataset of transmission line defects

Model	Worn AP(%)	Break AP(%)	mAP(%)	Fps
Faster-Rcnn	81.8	83.7	83.4	12
FCOS	86.7	88.2	87.8	56
yolov3-mobileNet	87.9	89.6	89.0	105
yoloxs	90.5	91.8	91.4	103
yolov5s	89.2	90.6	90.3	98
yolov7-tiny	92.4	93.2	92.8	129
Ours	92.2	93.8	93.6	124

Table 1 shows that the method proposed in this paper outperforms the classical FasterRcnn[29], FCOS, and yolov3 detectors in terms of both accuracy and speed. This is due to the introduction of the dynamic positive and negative sample allocation algorithm, which allows the model to focus on high-quality small samples while ignoring lower-quality positive samples, resulting in faster model fitting during training. However, the network architecture in this paper has been improved in a lightweight manner, allowing the model to maintain its accuracy advantage while also leading in speed and model size metrics. In comparison to the latest object detection models, this paper achieves higher working accuracy than the yolov7-tiny[30] model, albeit with a slightly slower speed. The faster speed of yolov7 is presumed to be due to the use of reparameterization techniques and complex model scaling methods. However, these techniques also increase the complexity during model training.

The detection results of our proposed model on the validation set are as follows:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202405_10(5).0013

Figure 4. Inference results

4.3 Ablation Experiment

To assess the effectiveness of each module and its impact on the final model, this paper outlines the following design for ablation experiments:

(1) The C3 module in the backbone network CSPDarkNet is replaced with a shuffle block and a FS block to explore the performance of each sub-module.

(2) The C3 module in the backbone network CSPDarkNet is replaced with a shuffle block and a FS block, respectively. This step aims to evaluate the impact of lightweight necking networks on the model. To achieve this, the PAN is used to unify the number of channels of the feature layer at different depths and replace the ordinary convolution in the C3 module with a depth-separable convolution. The paper explores the performance of each of the three different submodules.

(3) Verify the impact of the new positive and negative sample allocation algorithm. The algorithm improves model accuracy and speeds up model convergence.

Table 2. The results of the ablation experimen

Index	Model	mAP(%)	parameters (M)	FPS
Model1	BaseLine	90.3	7.2	98
Model2	Shuffle Block	90.5	4.6	112
Model3	FS Block	91.4	5.6	118
Model4	Light PAN	90.2	6.8	106
Model5	Anchor Free+SimOTA	91.5	7.2	98
Model6	FS Block+Anchor free+SimOTA+Light PAN	93.6	5.2	124

Table 2 shows that, when comparing Model1, Model2, and Model3, it is more efficient to reduce the network architecture rather than simply scaling the number of model channels. The resulting lightened network outperforms the yolov5s benchmark network in terms of final speed and accuracy. By comparing the results of Model2 and Model3, this paper has increased the SE channel attention mechanism, but due to the introduction of Faster Conv, the access to memory in the model inference process has been reduced. As a result, the accuracy has improved while maintaining speed.

Models 4 and 1 investigate the impact of Light PAN on the model's final performance. The experimental results show that the depth-separable convolution reduces the model's parameters and weakens the feature fusion ability of the PAN layer, resulting in a slight decrease in final accuracy compared to the benchmark model. However, despite the speed improvement brought by the lightweight, this paper still aims to address the accuracy loss resulting from lightweight in other parts of the model.

As analyzed in ATSS, the experiments for Model5 and Model1 show that the positive and negative sample allocation algorithms significantly impact the final accuracy of the detection model. By changing the positive and negative sample allocation of the baseline network and improving the SimOTA algorithm without altering the network architecture, the accuracy improved by 1.2%. The final experimental results of the model are shown in Model 6. Compared to the benchmark model, the work presented in this paper achieves a 3.3% increase in accuracy while reducing the model size by 28%.

5. Conclusion

For the task of detecting defects in transmission conductors, the dataset quality is poor, making small object detection difficult. Additionally, the model speed is slow. In this paper, we address these challenges by designing a lightweight convolution module and improving the original network architecture with YoloV5. We also introduce new dynamic positive and negative sample allocation algorithms and train the model on an expanded dataset. As a result, we achieve a 93% success rate. The model achieves 6% map accuracy, surpassing the benchmark networks yolov5s, yoloxs, and yolov7-tiny, while maintaining a small size of only 5.2M. Additionally, it has a detection speed (FPS) of 124, which is 26.5% and 20.4% faster than yolov5s and yoloxs, respectively, and comparable to yolov7-tiny. This model effectively addresses the challenges and issues in the transmission line defect detection task.

An end-to-end object detection network can avoid the tedious steps of traditional methods that require manual feature extraction. Allowing the network to learn the features itself reduces the difficulty and complexity of the task, making the object detection process more automated and efficient.Future research can further expand and deepen on this basis.

References

[1] KANG Zhongjian, LI Changchao,YU Hongguo,et al c . Identification method of key transmission lines in power system [J]. Electric Power Automation Equipment. ,2020,40(04):63-70.

[2] CHEN Meilin, WANG Yizhou, DAI Yan, et al. Small and Strong: Power Line Segmentation Network in Real Time Based onSelf-supervised Learning[J]. Proceedings of the CSEE,2022,42(04):1365-1375.

[3] LIU Zhiying, MIAO Xiren, CHEN Jing,et al. Review of Visible Image Intelligent Processing for Transmission Line Inspection[J]. Power System Technology,2020,44(03):1057-1069.

[4] WANG Wanguo, TIAN Bing,LIU Yue1,et al. Study on the Electrical Devices Detection in UAV Images based on Region Based Convolutional Neural Networks[J]. Journal of Geo-information Science,2017, 19(02):256-263.

[5] LIU Changyin, WANG Yichun, HOU Yanquan, et al. Detection and location of broken-strand transmission lines in UAV power inspection. Heilongjiang Electric Power[J], 2020,42(04):314-318.

[6] Krizhevsky A , Sutskever I , Hinton G .ImageNet Classification with Deep Convolutional Neural Networks[J].Advances in neural information processing systems, 2012, 25(2).

[7] Wang S , Liu Y , Qing Y ,et al.Detection of Insulator Defects With Improved ResNeSt and Region Proposal Network[J].IEEE Access, 2020, 8:184841-184850.

[8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[OL]//Computer Vision – ECCV 2016,Lecture Notes in Computer Science. 2016: 21-37. http://dx.doi.org/10.1007/978-3-319-46448-0_2.

[9] WANG Y, GAO H, LIU Y, et al. Obstacle Detection for Power Transmission Line Based on Deep Learning[C/OL]//2019 Prognostics and System Health Management Conference (PHM-Qingdao), Qingdao, China. 2019. http://dx.doi.org/10.1109/phm-qingdao46334.2019.8942868. DOI:10.1109/phm-qingdao46334.2019.8942868.

[10] Dong Zhaojie. Real-time detection of power transmission line key components based on YOLOv3[j].E lectronic Measurement Technology,2019,42(23):173-178.

[11] LIU S, QI L, QIN H, et al. Path Aggregation Network for Instance Segmentation[C/OL]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT. 2018. http://dx.doi.org/10.1109/cvpr.2018.00913. DOI:10.1109/cvpr.2018.00913.

[12] CHOLLET F. Xception: Deep Learning with Depthwise Separable Convolutions[C/OL]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI. 2017. http://dx.doi.org/10.1109/cvpr.2017.195. DOI:10.1109/cvpr.2017.195.

[13] HOWARD A, ZHU M, CHEN B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. arXiv: Computer Vision and Pattern Recognition,arXiv: Computer Vision and Pattern Recognition, 2017.

[14] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[C/OL]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT. 2018. http://dx.doi.org/10.1109/cvpr.2018.00474. DOI:10.1109/cvpr.2018.00474.

[15] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices[C/OL]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT. 2018. http://dx.doi.org/10.1109/cvpr.2018.00716. DOI:10.1109/cvpr.2018.00716.

[16] HUANG G, LIU S, MAATEN L van der, et al. CondenseNet: An Efficient DenseNet using Learned Group Convolutions[C/OL]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT. 2018. http://dx.doi.org/10.1109/cvpr.2018.00291. DOI:10.1109/cvpr.2018.00291.

[17] MA N, ZHANG X, ZHENG H T, et al. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design[M/OL]//Computer Vision – ECCV 2018,Lecture Notes in Computer Science. 2018: 122-138. http://dx.doi.org/10.1007/978-3-030-01264-9_8. DOI:10.1007/978-3-030-01264-9_8.

[18] HAN K, WANG Y, TIAN Q, et al. GhostNet: More Features from Cheap Operations[C/OL]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA. 2020. http://dx.doi.org/10.1109/cvpr42600.2020.00165. DOI:10.1109/cvpr42600.2020.00165.

[19] Hu J , Shen L , Albanie S ,et al.Squeeze-and-Excitation Networks.[J].IEEE transactions on pattern analysis and machine intelligence, 2020, 42(8):2011-2023.

[20] Chen J , Kao S H , He H ,et al.Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).2023.

[21] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement.[J]. arXiv: Computer Vision and Pattern Recognition,arXiv: Computer Vision and Pattern Recognition, 2018.

[22] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[J].

[23] BOCHKOVSKIY A, WANG C Y, LIAO H Y. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. Cornell University - arXiv,Cornell University - arXiv, 2020.

[24] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C/OL]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. 2016. http://dx.doi.org/10.1109/cvpr.2016.91.

[25] TIAN Z, SHEN C, CHEN H, et al. FCOS: Fully Convolutional One-Stage Object Detection[C/OL]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South). 2019. http://dx.doi.org/10.1109/iccv.2019.00972.

[26] hang S , Chi C , Yao Y ,et al.Bridging the Gap Between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).IEEE, 2020.

[27] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO Series in 2021[J]. 2021.

[28] ZHANG H, WANG Y, DAYOUB F, et al. VarifocalNet: An IoU-aware Dense Object Detector[C/OL]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA. 2021. http://dx.doi.org/10.1109/cvpr46437.2021.00841.

[29] Ren S , He K , Girshick R ,et al.Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J].IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.

[30] Wang C Y , Bochkovskiy A , Liao H Y M .YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//arXiv.arXiv, 2022.