

基于改进 DETR 模型的输电线路工程车辆检测研究

张林龙¹, 胡旭晓¹, 胡克轸²

(1.浙江理工大学机械工程学院, 浙江 杭州 310018;

2.悉尼大学计算机学院, 澳大利亚 悉尼 NSW2006)

✉ 19357388782@163.com; huxuxiao@zju.edu.cn; 13588266234@163.com



摘要:针对人工检测大型工程车辆频繁进出施工现场容易出现漏检或误检的问题,文章提出一种改进 DETR (基于 Transformer 的端到端目标检测网络)模型对输电线路工程车辆进行检测识别。首先在原始的 DETR 主干网络中,引入空洞卷积法获取更多深层次的特征,扩大感受野;其次加入特征金字塔网络(FPN),融合不同尺度的特征,增强特征的健壮性;最后将损失函数 GIOU 改为 CIOU,使模型在训练的过程中达到更快和更好的收敛效果。实验结果显示,改进后的 DETR 模型在自制数据集中 AP_{50} (IOU 阈值取 0.5) 和 AP_{50-95} (IOU 阈值取 0.5~0.95) 分别达到了 92.1% 和 61.3%,说明该改进模型在识别输电线路工程车辆场景中具有较高的应用价值。

关键词:空洞卷积;特征金字塔网络;DETR;损失函数

中图分类号:TP311 **文献标志码:**A

Research on Transmission Line Engineering Vehicle Detection Based on Improved DETR Model

ZHANG Linlong¹, HU Xuxiao¹, HU Kezhen²

(1.College of Mechanical Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China;

2.School of Computing University of Sydney, Sydney NSW2006, Australia)

✉ 19357388782@163.com; huxuxiao@zju.edu.cn; 13588266234@163.com

Abstract: In view of the problem of missing or false detection in manual detection of large engineering vehicles frequently entering and leaving the construction site, this paper proposes an improved DETR (End-to-End Object Detection with Transformers) model to detect and identify transmission line engineering vehicles. Firstly, in the original DETR backbone network, the dilated convolution method is introduced to obtain more deep features and enlarge the receptive field. Then, the Feature Pyramid Network (FPN) is added to fuse the features of different scales to enhance the robustness of the features. Finally, the loss function GIOU is changed to CIOU to make the model converge faster and better in the training process. The experimental results show that AP_{50} (IOU threshold of 0.5) and AP_{50-95} (IOU threshold of 0.5~0.95) of the improved DETR model reach 92.1% and 61.3% respectively in the self-made dataset, which indicates that the improved model has high value of application in identification of transmission line engineering vehicles.

Key words: dilated convolution; FPN; DETR; loss function

0 引言 (Introduction)

由于工程车辆频繁进出输电线路施工场地,采用人工检测车辆的方式不仅会消耗大量人力资源,而且无法做到实时检测,因此许多输电线路施工场地存在监控困难的问题。在这种

情况下,输电线路施工现场急需一种具有高精度且成本较低的大型工程车辆自动检测方法。

近年来,随着深度学习的快速发展,研究人员发现卷积神经网络可以有效地提取目标特征信息^[1],因此产生了许多能有

效提取视觉特征的目标检测模型,包括卷积神经网络(CNN)^[2]、长短时记忆网络(LSTM)^[3]等,并且经过不断地演化又产生了许多成熟的目标检测模型。目前,目标检测算法有Two-stage、One-stage和以Transformer为基础架构的检测算法。其中,Two-stage使用卷积神经网络对候选区域的目标进行分类定位,例如R-CNN^[4](Region Convolutional Neural Network)、Fast-RCNN^[5](Fast Region Convolutional Neural Network)、Faster-RCNN^[6](Faster Region Convolutional Neural Network)、Mask-RCNN^[7](Mask Region Convolutional Neural Network)系列模型。One-stage则不需要产生候选区域,可以直接通过神经网络进行特征提取,例如SSD^[8](Single Shot Multibox Detector)和YOLO^[9](You Only Look Once)系列模型。

本文选用的DETR模型以Transformer为架构,舍弃了以往算法中常用的anchor的先验,也取消了NMS后处理方法,真正实现了端到端的目标检测。目标检测对象的种类主要有卡车、挖掘机、推土机3种。实验结果表明,相比原始的DETR模型,改进后的DETR模型 AP_{50} 和 AP_{50-95} 值有了较大幅度的提高。

1 DETR模型简介(Introduction to DETR model)

DETR模型^[10]是由Facebook AI团队在2020年提出的基于Transformer^[11]的端到端目标检测,利用Transformer结构的自注意力机制为各个目标编码,取消了非极大值抑制(NMS)处理步骤,舍弃了anchor等先验知识,没有了约束,也没有多尺度特征融合及复杂的数据增强,大大简化了目标检测的流程,并且在COCO数据集上的大目标检测效果比Faster RCNN好,迁移性好,对识别图像类别具有较好的适应性。

DETR模型主要由4个部分构成,首先是CNN的backbone部分,其次是Transformer的Encoder部分,再次是Transformer的Decoder部分,最后是预测层FFN。在backbone部分,以CNN卷积神经网络为主体,提取图像的主要特征信息,得到输入图片的特征矩阵,然后使用 1×1 的卷积核对通道数进行降维,并将图像的位置编码信息一同传入Encoder中;在Encoder中学习图像的全局信息,得到一些候选特征,然后Decoder将得到的一些候选特征以及在Decoder中的Object Query进行独立解码,其中Object Query的作用是从原始特征中找到物体的位置;在FFN中得到预测框的准确坐标以及类标签。

2 改进后的DETR模型(Improved DETR model)

2.1 空洞卷积法

DETR模型以Resnet-50^[12]作为主干网络,对输入图像进行特征图的提取。Resnet-50系列网络通过引入残差结构有效缓和了神经网络在训练时产生的梯度消失问题,但经过降维后也丢失了很多特征信息,难以对特征信息较少的目标或者遮挡的小目标进行有效识别,也无法为后续编码网络提供高效的特征图,所以在Resnet-50中引入空洞卷积,通过对传统图像卷积添加空洞完成对传统图像卷积的感受野扩张^[13],空洞卷积法

如公式(1)所示:

$$y[i] = \sum_{j=1}^n x[i + r \times n]w[j] \quad (1)$$

其中: x 为原始图像的向量值; w 为计算时的权值; y 为输出的图像向量值; r 代表扩张率,它是调节感受野大小,输出图像特征的一个重要参数, r 改变,代表卷积核得到不同程度的扩张。引入空洞卷积后,通过不同的扩张率不仅能扩大检测图像的感受野,而且能在保持相同计算量的同时获取更多工程车辆的特征信息,有效地提高了特征的健壮性。

空洞卷积核和感受野对应的关系式如公式(2)和公式(3)所示:

$$T_m = t + (t-1)(r-1) \quad (2)$$

$$R_{m+1} = R_m + (T_m - 1) \prod_{i=1}^m S_i \quad (3)$$

其中: t 表示原始卷积核的大小,本研究选择 1×1 的卷积核大小; R_m 为经过 m 层卷积后,感受野的大小; r 为扩张率; T_m 为经过 m 层空洞卷积后,卷积核的实际尺寸的大小; S_i 表示在 i 层时的步长。

本研究使用扩张率 r 为1、2、4、8四种空洞卷积核对输入图像进行特征提取,输出4种不同的特征图,部分不同扩张率的卷积核示意图如图1所示。利用上述公式(2)和公式(3)计算可得每一层感受野的大小和卷积核的大小。即:

$$T_1 = 3 + (3-1) \times (1-1) = 3, \text{感受野 } R_1 = 1 + (3-1) \times 1 = 3;$$

$$T_2 = 3 + (3-1) \times (2-1) = 5, \text{感受野 } R_2 = 3 + (5-1) \times 1 = 7;$$

$$T_3 = 3 + (3-1) \times (4-1) = 9, \text{感受野 } R_3 = 7 + (9-1) \times 1 = 15;$$

$$T_4 = 3 + (3-1) \times (8-1) = 17, \text{感受野 } R_4 = 15 + (17-1) \times 1 = 31.$$

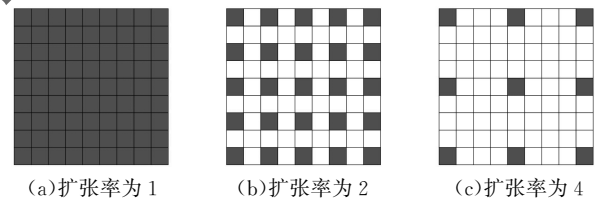


图1 空洞卷积示意图

Fig. 1 Schematic diagram of dilated convolution

2.2 加入空洞卷积后的Resnet-50

Resnet-50主要由3个部分构成,第一部分为一个 7×7 的卷积层,通道数为64,步长为2。第二部分(卷积层 $2_x \sim$ 卷积层 5_x)为残差块的堆叠,残差块采用瓶颈结构,“瓶颈结构”由多组大小为 $1 \times 1, 3 \times 3, 1 \times 1$ 的卷积核组成。第三部分由一个全局平均池化、1 000维的全连接层和激活函数组成,主要作用是防止参数过拟合,提高输入和输出空间变换的鲁棒性。

本研究主要对Resnet-50的第二部分,即残差网络进行改进,在卷积层 $2_x \sim$ 卷积层 5_x 的残差模块中引入不同扩张率的空间卷积,改进后的Resnet-50残差模块如图2所示,图2中的Conv 1×1 表示卷积核的大小为 1×1 ,64和256代表通道数,Conv3-R中的3代表卷积核大小为 3×3 ,R代表不同的扩张率。在保持 1×1 卷积核大小和通道数量不变的情况下,将残差模块中的 3×3 大小的卷积核改为扩张率为1、2、4、8的空洞卷积核。即在卷积层 2_x 中的空洞卷积核扩张率为1,在卷

积层 3_x 中空洞卷积核扩张率为 2, 以此类推。通过不同的空洞卷积扩张率, 得到不同的特征图, 以此提高对目标检测的准确率。

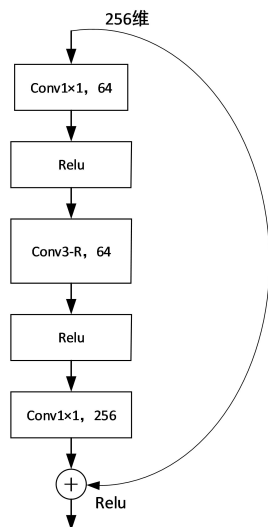


图 2 改进后的 Resnet-50 残差模块

Fig. 2 Improved Resnet-50 residual module

2.3 特征金字塔网络

由于 DETR 模型中的 backbone 模块只输出单一的特征图层进行检测,但是单一的特征图层经过多次的卷积以及池化过程会丢失目标的许多关键特征信息,而且当图像分辨率不高或者目标较小时,则无法有效识别,所以只使用单一的特征图是无法达到精确识别效果的。特征金字塔网络通过简单的网络连接,以极小的计算量对不同尺度的特征图进行调整融合,使得融合后的特征具有更丰富的表达能力,大大提高了模型对小目标或者特殊情况下的检测准确率,也获得了更高的鲁棒性。

本次研究的 FPN 主要由 3 个部分组成,分别为 C1~C5 自下而上、M5~M2 自上而下及特征相加 3 个部分,其结构图如图 3 所示。

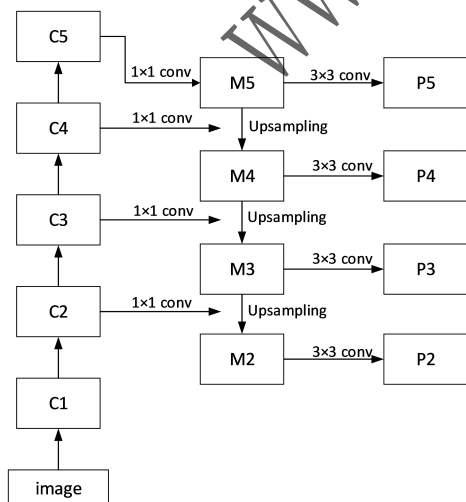


图 3 FPN 结构图

Fig. 3 Structure diagram of FPN

FPN^[14]结构首先为 C1~C5 自下而上的过程,即空洞卷积改进后的 Resnet-50 正向传播过程,在 Resnet-50 的每个阶段最后输出基层特征图为{M5,M4,M3,M2}。其次为 M5~M2 自上而下的过程,通过 2 倍上采样操作,让高层次语义更强的特征与低层次特征融合。最后通过 1×1 卷积核改变不同特征图的 channel,横向连接至基层特征图进行特征相加,得到最终的 FPN 特征图{P2,P3,P4,P5},此时的特征图具有更丰富的特征信息。

2.4 损失函数的改进

DETR 模型的边框损失函数采用 GIOU^[15] 和 L1 损失结合的方法进行运算,如公式(4)所示:

$$L_{\text{box}}(b_i, \hat{b}_{s(i)}) = \lambda_{\text{IOU}} L_{\text{GIOU}}(b_i, \hat{b}_{s(i)}) + \lambda_{L_1} L_{L_1}(b_i, \hat{b}_{s(i)}) \quad (4)$$

其中: λ_{IOU} 和 λ_{L1} 代表 GIOU 和 $L1$ 的权重系数, b_i 和 $\hat{b}_{s(i)}$ 分别是第 i 个需要检测目标的真实框坐标值以及预测的第 $s(i)$ 个的预测框的坐标值, L_{GIOU} 和 L_{L1} 分别为 GIOU 和 $L1$ 的损失函数, 其中 GIOU 损失函数由公式(5)和公式(6)所示:

$$IOU(b, b^{gt}) = \frac{b \cap b^{gt}}{b \cup b^{gt}} \quad (5)$$

$$L_{\text{GIOU}} = \begin{cases} \text{IOU}(b, b^{st}) - \frac{c - (b \cup b^{st})}{c}, & \text{IOU} \neq 0 \\ 1 + \frac{(b \cup b^{st})}{c}, & \text{IOU} = 0 \end{cases} \quad (6)$$

GIOW 是一种距离度量,它的取值范围 $(-1,1]$,其中 c 为关注非重合区域引入的最小闭包区面积, $b \cup b^{st}$ 为预测框和真实框的重叠部分,相比 IOU,GIOW 虽然能更好地反映两者的重合度,但是它有如下缺点:当预测框和真实框处于包含关系的特殊情况下,GIOW 和 IOU 计算出的 $loss$ 值就会一样,两者的相对位置关系无法有效的区分,这样会导致边界框回归收敛得很慢,训练时间也会大大地延长,无法达到有效回归边界框的目的。

CIOU^[16]可以很好地解决上述问题,它在原有的基础上引入了对角线距离和中心点距离的比值关系,解决了 GIOU 在预测框和真实框包含关系下无法进行收敛的问题,同时加入了预测框和真实框的长宽比,更好地反映了两者的重合度,其中 CIOU 损失函数如公式(7)至公式(9)所示:

$$v = \frac{4}{\pi^2} \left[\arctan\left(\frac{\omega^{g_I}}{h^{g_I}}\right) - \arctan\left(\frac{\omega}{h}\right) \right]^2 \quad (7)$$

$$a = \frac{v}{(1 - \text{IOU}(b, b^{gt})) + v} \quad (8)$$

$$L_{\text{CIU}} = 1 - \text{IOU}(b, b^{gt}) + \frac{m^2(b, b^{gt})}{c^2} + av \quad (9)$$

其中: w^{gt} 、 h^{gt} 、 w 、 h 分别代表真实框的宽、高和检测框的宽、高, $m(b, b^{gt})$ 为预测框和目标框中心点的欧式距离。改进后的 DETR 模型的边框损失函数如公式(10)所示:

$$L_{\text{box}}(b_i, \hat{b}_{s(i)}) = \lambda_{\text{IOU}} L_{\text{ClOU}}(b_i, \hat{b}_{s(i)}) + \lambda_{L1} L_{L1}(b_i, \hat{b}_{s(i)}) \quad (10)$$

3 实验分析 (Experimental analysis)

3.1 实验数据集制作和实验环境

本文使用的数据集来源于电力公司实际拍摄的图像,主要

通过装在杆塔上的摄像头进行实时拍摄,拍摄目标主要为卡车、推土机、挖掘机 3 类,由于采集时间较短,所以数据集总共只有 4 000 张目标图像,将工程车辆目标数据集以 8 : 2 的比例划分为训练集和测试集,并使用 LabelImg 图像标注工具对数据集进行统一标注,由于 DETR 模型通常采用的数据集格式为 COCO 数据集格式,所以需要将标注的 PascolVOC 格式转化为 COCO 数据集格式。本文采用的数据集图片由于目标检测数量较多,背景较复杂等,导致模型较难收敛,检测精度不高,所以使用图像增广技术对数据集进行图像增强,即通过旋转、增强亮度、增强对比度、增强色度、锐化等操作(图 4),让模型学习到更多的目标图像的特征。

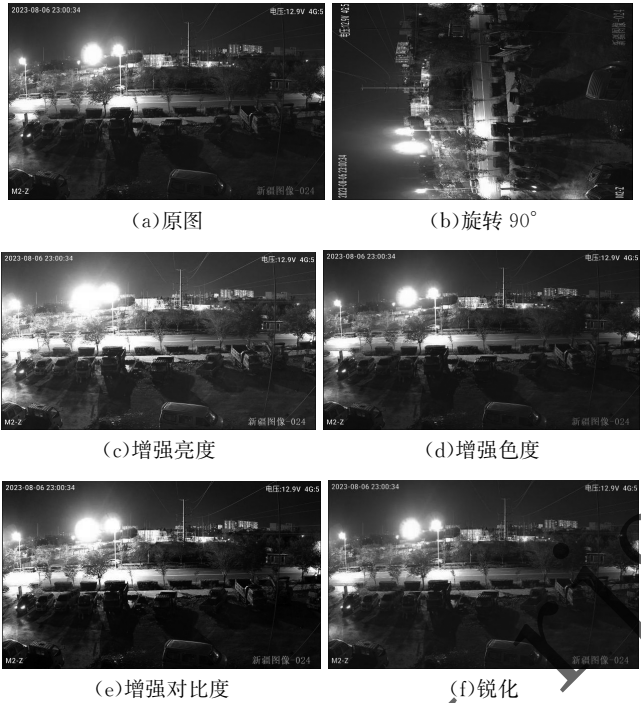


图 4 图像增广处理
Fig. 4 Image augmentation processing

本次研究实验使用 Adam 作为优化器,在图像增广处理后的数据集上进行实验,实验平台环境为 Windows 11 操作系统, NVIDIA GeForce RTX 3060 显卡,i7-10870 处理器,Pytorch1.70 深度学习框架,CUDA 版本为 11.0,CUDNN v8.0.4 加速训练,Python 版本为 3.7,选取的 batch_size 为 8,训练的 epoch 设置为 200 轮,使用 Python 为主要编程语言。

3.2 实验结果分析

为了验证改进后的 DETR 模型应用在输电线路工程车辆检测场景的准确性及适用性,从电网数据库中随机选取包含卡车、挖掘机、推土机的图片一共 800 张,用于实验测试。

将测试样本放于 DETR 文件夹下,分别通过原始的 DETR、改进主干网络的 DETR、损失函数改进后的 DETR 以及将主干网络和损失函数结合改进后的 DETR 模型进行测试,结果如表 1 所示,本文采用平均精准度(AP)对模型的检测精度进行验证,采用指标为 AP_{50} 和 AP_{50-95} ,分别表示 IOU 阈值取 0.5 和 0.5~0.95,通过对 2 个 AP 值的并比较,更有效地反映改进后的 DETR 模型测试效果。

表 1 加入不同改进策略的 DETR 模型测试效果

Tab.1 Test results of DETR model with different improvement strategies

空洞卷积+FPN	改进损失函数	$AP_{50}/\%$	$AP_{50-95}/\%$
×	×	86.5	58.1
✓	×	88.5	60.1
×	✓	87.1	59.3
✓	✓	92.1	61.3

表 1 中,第一行为原始的 DETR 模型在测试集上的测试结果, AP_{50} 和 AP_{50-95} 分别为 86.5%和 58.1%,在此基础上的第一步是在主干网络中加入空洞卷积及特征金字塔网络进行实验,结果显示 AP_{50} 和 AP_{50-95} 均提高了 2 百分点。第二步是对损失函数进行改进,实验结果显示,模型收敛效果较之前有明显的进步, AP_{50} 和 AP_{50-95} 分别提高了 0.5 百分点和 1 百分点左右。第三步加入所有改进策略进行实验,结果显示 AP_{50} 和 AP_{50-95} 都得到了明显的提升,分别达到了 92.1%和 61.3%,表明改进后的 DETR 模型在识别输电线路工程车辆场景方面具有良好的适用性。

为了进一步验证改进后的 DETR 模型在输电线路工程车辆场景下的算法性能,将测试集与其他具有代表性的模型(SSD,Faster-RCNN,YOLOv3,YOLOv5)进行测试比较,其结果如表 2 所示。

表 2 不同模型对比实验

Tab.2 Model comparison experiment

模型	$AP_{50}/\%$	$AP_{50-95}/\%$
DETR	86.5	58.1
SSD	85.6	45.3
Faster-RCNN	79.1	49.2
YOLO-v3	87.5	54.0
YOLO-v5	89.3	58.1
改进后的 DETR	92.1	61.3

由表 2 可知,改进后的 DETR 模型应用在施工现场多目标场景下的检测精度均高于其他 YOLO 系列和 Faster-RCNN 等常规目标检测模型,并且对比原始的 DETR 模型,改进后的 DETR 模型取得了更高的检测精度。图 5 为改进后的 DETR 模型的检测效果示意图。

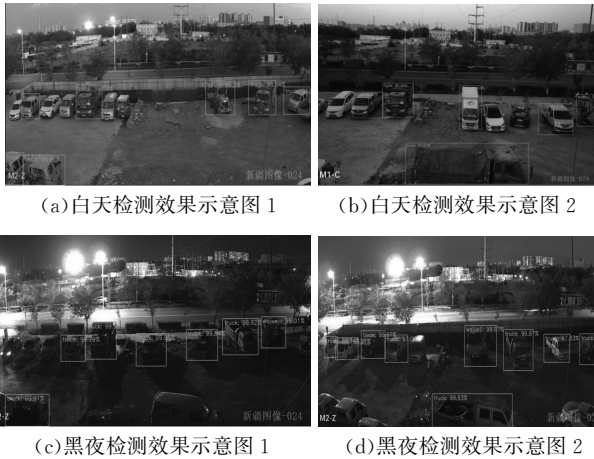


图 5 改进后的 DETR 模型检测效果示意图

Fig. 5 Detection effect diagram of the improved DETR model

4 结论(Conclusion)

本文针对人工检测大型工程车辆频繁进出施工现场容易出现漏检或误检的问题,提出了一种基于改进 DETR 模型的输电线路工程车辆检测模型,构建了 4 000 张包含卡车、挖掘机和推土机 3 种目标类型的数据集。通过对原始的 DETR 模型的主干网络以及损失函数进行改进,取得了较原始 DETR 模型以及其他经典目标检测模型更好的检测效果。实验结果表明,改进后的 DETR 模型在输电线路工程车辆数据集中获得的 AP 值都得到了明显的提升,其 AP_{50} 和 AP_{50-95} 分别达到了 92.1% 和 61.3%,能在节省大量人力资源的同时,更好地完成工程车辆的检测工作。下一步的工作是进一步提高 DETR 模型在小目标或者有雾天气等特殊场景下的检测准确率,以及解决 DETR 模型参数量较大导致训练速度较慢的问题,尝试把本文模型运用到需要检测输电线路工程车辆的实际场景中,辅助工作人员更好地完成对工程车辆的检测。

参考文献(References)

- [1] 汪志勇. 多视角 SAR 目标识别与观测路径自主规划方法研究[D]. 成都:电子科技大学,2022.
- [2] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [3] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//IEEE. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 580-587.
- [5] GIRSHICK R. Fast R-CNN[C]//IEEE. Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2015: 1440-1448.
- [6] 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 and machine intelligence, 2017, 39(6): 1137-1149.
- [7] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[J]. IEEE transactions on pattern analysis and machine intelligence, 2020, 42(2): 386-397.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//Springer. Proceedings of the European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]// IEEE. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 779-788.
- [10] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Vedaldi A, Bischof H, Brox T, et al. European Conference on Computer Vision. Cham: Springer, 2020: 213-229.
- [11] HAN K, WANG Y H, CHEN H T, et al. A survey on vision transformer[J]. IEEE transactions on pattern analysis and machine intelligence, 2023, 45(1): 87-110.
- [12] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//IEEE. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.
- [13] 马利, 刘新宇, 李皓宇, 等. 应用空洞卷积的神经网络轻量化方法[J]. 计算机工程与应用, 2022, 58(5): 85-93.
- [14] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//IEEE. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 936-944.
- [15] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression [C] // IEEE. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 658-666.
- [16] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]// AAAI. Proceedings of the AAAI Conference on Artificial Intelligence. Texas: AAAI Press, 2020, 34(7): 12993-13000.

作者简介:

张林龙(1999-),男,硕士生。研究领域:图像处理。

胡旭晓(1965-),男,博士,教授。研究领域:图像处理,深度学习。

胡克轸(2004-),男,本科生。研究领域:图像处理。