高技术通讯 2022 年 第32 卷 第5 期:528-534

doi:10.3772/j.issn.1002-0470.2022.05.010

改进的 PP-YOLO 网络遥感图像目标检测①

朱福珍②* 王 帅** 巫 红*

(* 黑龙江大学电子工程学院 哈尔滨 150080)

(** 中南大学物理与电子学院 长沙 410006)

摘 要 为了提高遥感图像目标检测精度,提出一种改进的 PP-YOLO 网络遥感图像目标检测方法。改进的 PP-YOLO 网络继承了骨干网络结构,改进了 PP-YOLO 网络的检测颈部分,在保持原检测颈框架基础上,在第 4 层和第 5 层加入了由低层网络向高层网络传达的通路,使得网络低层部分可以学习到高层部分的特征信息,加强高层网络学习的特征信息,在保证网络泛化能力的同时比未改进优化同骨干网络的 PP-YOLO 网络其平均精度均值(mAP)提高了 4.4%。同时,优化了 PP-YOLO 网络训练策略,即基于遥感数据集特点,以更优的 CutMix 数据增强算法替换掉原有的 Mixup 数据增强算法,加入 GridMask 算法增强网络特征的学习,实验取得了最高 89.3% 的 mAP,有效地提高了每一类目标实例的精度。

关键词 遥感图像; 目标检测; PP-YOLO 网络

0 引 言

遥感图像具有数据量大、数据复杂、目标分布广泛等特点,如何克服遥感图像目标繁多、种类密集、背景信息复杂等难点,快速精准地识别定位遥感图像中的目标,是当下遥感图像目标检测领域研究的热点和难点。遥感图像目标检测在遥感图像语义检索、智慧城市建设、资源调查、环境监测等领域具有重要意义[1]。本文对现阶段自然图像目标检测表现突出的 PP-YOLO 网络进行改进,并将该网络进行迁移,更好地用于遥感图像目标检测。

传统的遥感图像目标检测算法主要是基于人工提取特征的方法,对专家经验依赖程度很高,且工作量巨大。随着深度学习技术的成功应用,全球重要赛事和实际应用已证明了其在目标检测中的强大实力。深度学习领域的重大转折始于 2012 年的计算机视觉竞赛中由 Krizhevsky 等人[2] 改进设计的 AlexNet,该网络设计获得了卷积神经网络模型图像分类最高的准确率。随着硬件技术的快速发展,大规模的深度学习网络架构呈现井喷式发布。例如, 2014 年, Girshick 等人[3] 提出区域卷积神经网络(region convolutional neural network, R-CNN)算法,该算法将目标检测任务分割成了目标分类和定位两个任务,实现了检测精度的重大突破,但是选择性搜索算法产生了大量的冗余候选框,候选区域特征提取过程中存在大量重复运算,这些都限制了模型的速度和性能。随后,基于其改进的 Fast R-CNN[4]、Faster R-CNN 算法[5] 以及 Mask R-CNN[6] 都在深度学习发展史上留下了浓墨重彩的一笔。

2017 年,提出的 YOLO 系列算法成功地开创了单目标检测算法的时代[7-9],不再使用特定算法来产生候选框,而是巧妙地将回归问题与目标边框定位问题进行类比推理,极大地提升了目标检测器的速度和精度。同年,Lin 等人[10] 在特征金字塔网络

① 国家自然科学基金(61601174),黑龙江省级大学生创新创业训练计划(202010212101),黑龙江省博士后科研启动金(LBH – Q17150),黑龙江省普通高等学校电子工程重点实验室(黑龙江大学)开放课题及省高校科技创新团队课题(2012TD007)和黑龙江省自然科学基金(F2018026)资助项目。

② 女,1978 年生,博士,副教授;研究方向:图像超分辨,遥感图像处理,神经网络,深度学习;联系人,E-mail: zhufuzhen@hlju.edu.cn。

(收稿日期:2021-10-13)

(feature pyramid network, FPN)算法中提出了 bottom-up 与 top-down 的结构,利用深层特征图的高语义特征与浅层的高分辨率信息融合后的特征层独立预测,提升了目标检测算法对小目标检测的效果。2019 年,CornerNet 算法[11]的提出,开创了无锚框算法的时代,成功解决了基于锚框的算法依赖手动设计、训练和预测过程低效及正负样本不均衡的问题,并取得了与基于锚框的算法相媲美的精度。

PP-YOLO 网络是基于百度编写的深度学习框架 PaddlePaddle 上编写的,在目标检测方面体现了非常好的性能,将 YOLOv3 模型的平均精度均值(mean average precision, mAP)在 COCO 数据集对象检测任务上从 38.9% 提升到 44.6%,并将推理速度(FPS)从 58 增加到 73。可见,与其他目标检测器相比 PP-YOLO 模型检测性能有了很大的提升。与 YOLO 系列算法不同,PP-YOLO 网络并没有探索不同的骨干网络和数据增强算法,也没有使用神经结构搜索(neural architecture search, NAS)超参数,而是采用一种改进的 ResNet50-vd-dcn 骨干网络[12]替换掉了原 YOLOv3 网络的骨干网络,并对现有的一些几乎不影响效率的有效技巧进行修改并叠加到原有网络中,以提高网络性能,做到有效性和效率之间的平衡。本文对 PP-YOLO[12]网络在训练策略和网络结构两方面进行了改进优化,使其能够更好地适应遥感图像小目标任务的检测。

1 改进的 PP-YOLO 网络遥感目标检测

PP-YOLO 网络是一种典型的单阶段目标检测算法,通常由骨干网络、检测颈(通常是一个特征金字塔网络)和检测头(物体分类和定位部分)组成[13]。本文在检测颈、数据增强策略和训练策略方面进行改进,用于提高遥感图像目标检测。

1.1 检测颈的改进

PP-YOLO 网络构建具有横向连接的特征金字塔,具体结构如图 1 所示。为了简洁起见,图中省略了激活层。与传统的特征金字塔不同的是,PP-YOLO 加入了许多不影响精度的方法提升 FPN 结构的性能,本文对检测颈的重要改进之一是对特征金字塔引入新的方法进行调整。在图 2 中使用不同颜色的圆角框和连接线进行了标注,其中棕色圆角框表示该卷积层加入 CoordConv 方法[14],紫色圆角框表示在该位置加入 DropBlock 方法[15],红色连接线表示加入空间金字塔池化[16]方法。骨干网络的最后 3 层的特征映射(C3,C4,C5)输入到 FPN 网络中,在此定义 l 层特征金字塔输出的特征映射为 P_l,本文实验时 $l=3,4,5$。当输入图像尺寸为 $W \times H$ 时,输出特征映射 P_l 为

$$P_l = \frac{W}{2^l} \times \frac{H}{2^l},\ l = 3,4,5 \tag{1}$$

本文在检测颈部分的另一处改进是加入了由低层网络向高层网络传达的通路,如图 2 所示。之所以做出这样的改进,是由于原 PP-YOLO 网络的检测颈只能实现低层网络学习到高层网络的语义信息,但是高层网络却没有办法学习低层网络丰富的空间特征信息。因此,本文在改进的检测颈结构中,保留了原网络的基本框架,即每一层的处理过程不变,在第 4 和第 5 层加入了低层网络语义信息传达通路,该通路通过名为 Downsample Block 的模块实现,Downsample Block 模块的具体信息在图 3 中给出。

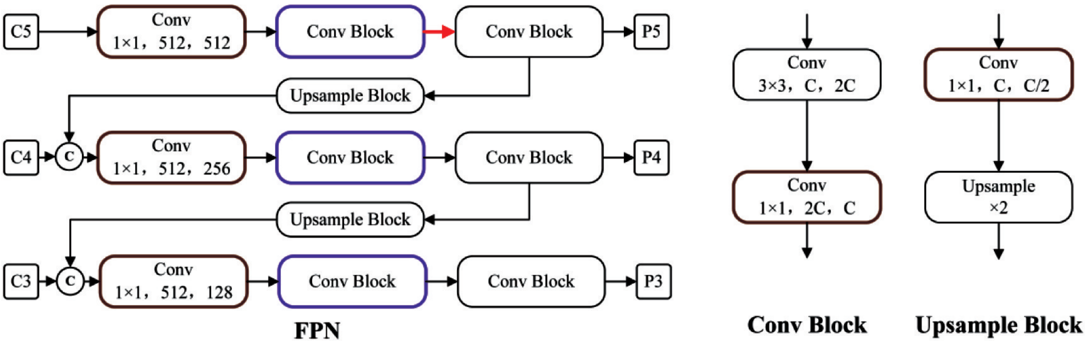


图 1 特征金字塔具体结构

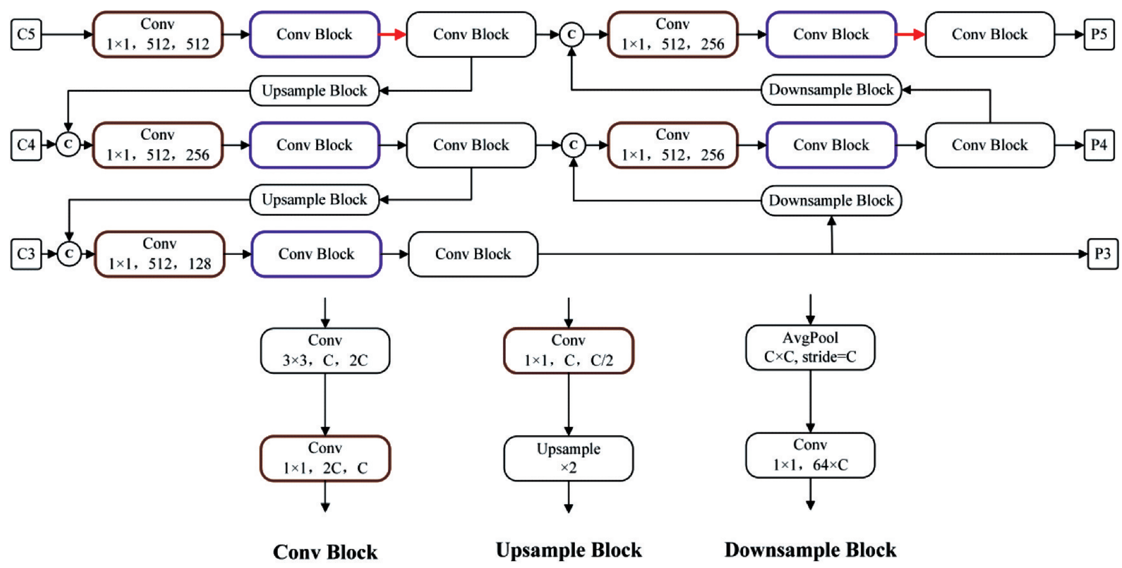


图 2　检测颈优化策略框图

经实验验证,改进后的检测颈设计在遥感图像目标识别过程中显示出了一定的优势,对提升检测精度很有帮助。

1.2　检测头

PP-YOLO 网络的检测头部分采用 3×3 卷积层和 1×1 卷积层进行最终预测,每个输出通道最终的输出是 $3\times(K+5)$,其中 K 是预测目标实例总类别。每一张输入图片的最终预测结果都会与 3 个锚框进行关联。每个锚框的输出结果都是由 K 个不同类别的预测概率加上 4 个预测锚框的位置坐标以及为某个类别的分数预测组成。针对分类和定位,分别采用了 cross entropy 损失和 L1 损失函数。检测头的结构以及损失函数的搭配使用如图 3 所示。

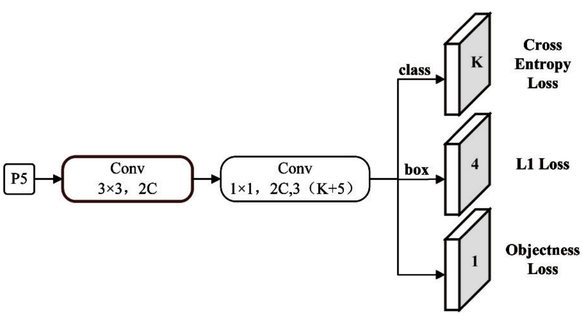


图 3　检测头的结构以及损失函数的搭配

Cross entropy 损失函数定义如式(2)所示,计算输入和标签之间的交叉熵,可用于训练分类器。如果提供权重参数,它是一个 1-D 的张量, 每个值对应每个类别的权重,损失函数为式(2);如果不提供权重参数,该损失函数为式(3)。

$$loss_j = weight[class](-input[class] + \log(\sum_{i=0}^{K}\exp(input_i))) \quad j = 1,2,\cdots,K \tag{2}$$

$$loss_j = -input[class] + \log(\sum_{i=0}^{K}\exp(input_i)) \quad j = 1,2,\cdots,K \tag{3}$$

其中,$weight$ 为指定每个类别的权重,$input$ 为指定类别的输入。

L1 损失函数用于计算输入和标签之间的损失,根据指定应用于输出结果的计算方式 L1 损失函数有绝对值、均值和求和 3 种表达,本文中采用求和 L1 损失,如式(4)所示。

$$out = \mathrm{SUM}(|input - label|) \tag{4}$$

其中,$input$ 为指定类别的输入,$label$ 为与 $input$ 在维度和类别上相同的标签。

1.3　数据增强优化策略的改进

数据增强的目的是扩充原训练数据集,产生与原图像相似但又不完全相同的样本,防止网络过拟合,以提升模型的泛化能力。PP-YOLO 首先使用 Image Mixup(服从 Beta(1.5,1.5)分布)训练,使训练的前 250 轮每幅图像都是任意 2 张图像混合,最后 20 轮不做混合,这样做的优点是提高了网络对空间扰动的泛化能力。但 Mixup 数据增强算法会存在

图像混合不自然的情况,从而出现非信息像素,降低训练效率。因此,本文对数据增强策略进行改进,改进策略框图如图 4 所示,以更优的 CutMix 数据增强算法[17]替换掉原有的 Mixup 数据增强算法。这种 CutMix 数据增强算法采用裁剪区域然后以补丁的方式混合图像,不存在以上问题,同时训练和推理代价不变。

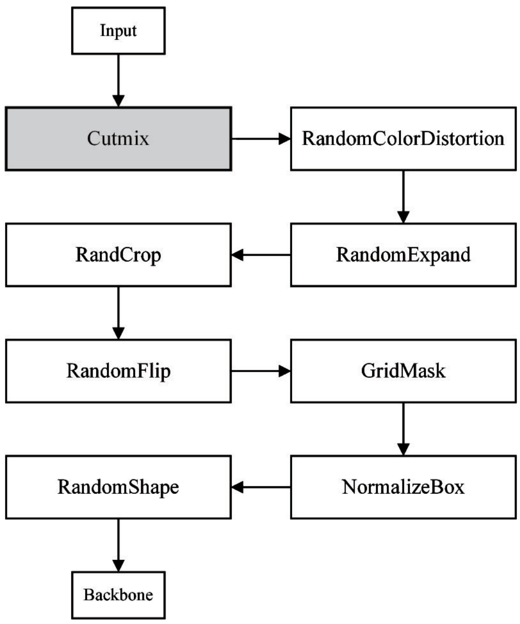


图 4　改进的数据增强策略框图

1.4　训练策略的优化

本文实验数据集为公开的 DIOR 和 DOTA 遥感图像目标检测数据集,这类数据集目标总数分布不均匀(车辆的图片高达 6000 张,而包含烟囱的图片却不到 1000 张)、类别总数分布不均匀(船只总数高出了车辆总数的 1/3,是火车站数量的 60 倍)。在模型训练过程中,这会导致模型的差别学习或偏向性学习,使得相同识别难度下某些在数据集中分布更广泛的类别识别精度远高于相同识别难度的其他类别。同时,具有类间相似性的不同类别还存在相互抑制性,在分布差距较大的情况下,可能会导致对某一类目标的过度学习,抑制具有类间相似性的其他类别的学习。这都给目标检测算法带来了巨大的挑战。即使采取了适当的数据增强策略,距离满足网络训练的需求还有一定差距,因此,实验中对 DOTA 数据集进行了分析,取该数据集与 DIOR 数据集中相同且难以识别的类别,修改其数据标注文件为 VOC 数据集标注格式,然后将该部分数据混合入 DIOR 数据集的训练集中完成扩充。由于 DIOR 数据集中包含类别众多,类间相似性明显,训练过程中难免会出现部分类别训练不充分的问题。为了解决该问题,本文提出了训练策略的优化并编写了相应的程序来实现该功能。

具体过程是将公开的 DIOR 和 DOTA 遥感图像目标检测数据集经特定的筛选规则组成混合训练集,当模型训练达到某个瓶颈,总体平均 mAP 值在取得最优值的 10 个批次内不再提升时,停止训练并测试输出每一个类别的 AP 值;将单一类别的 AP 值与总体平均 mAP 值进行对比;根据统计出的各单一类别的平均像素大小,取所有包含低于总体平均 mAP 值且像素点在平均像素以下类别的图片;根据单一类别 AP 值升序排列,分别在混合数据集中取原训练集大小的 9.75% 依次递减 0.5% 的数量(小数部分按四舍五入处理),然后再扩充 2 倍写入新的训练集。这样做的目的是为了使网络能对那些难以识别的图片进行学习,并且尽可能地保证不会因为引入更多的某一类图片而导致具有类间相似性类别的识别精度降低。

2　实验结果与分析

具体实验中,选择开源 DIOR 和 DOTA 数据集,确定训练集图像 5861 张,测试集图像 11 738 张,共 20 类目标。为了表述方便,分别用 H1 ~ H20 依次表示飞机、飞机场、棒球场、篮球场、桥、烟囱、大坝、高速公路服务区、高速公路收费站、高尔夫球场、地面田径场、港口、立交桥、船、体育场、储存罐、网球场、火车站、车辆和风车。实验的环境为 GPU: Tesla V100, Video Mem:16 GB; CPU: 2 Cores, RAM:16 GB; Disk: 100 GB。实验过程主要步骤可概括为:首先设定训练模型路径和评估模型路径,模型分类数设定为 20,依次设定模型的 backbone、检测颈、检测头,损失函数、学习率设定为 0.0003;然后训练数据增广和优化;最后设置预测模型参数。整个网络训练约 30 h收敛。详细参数设置及具体实现过程已开源,

开源网址为 https://aistudio. baidu. com/aistudio/projectdetail/1480790。

实验采取平均精度均值(mAP)作为评判指标，mAP 是 n 个不同类别平均精度(average precision，AP)的平均值，其定义为

$$mAP = \frac{1}{n}\sum_{i=0}^{n} AP_i \tag{5}$$

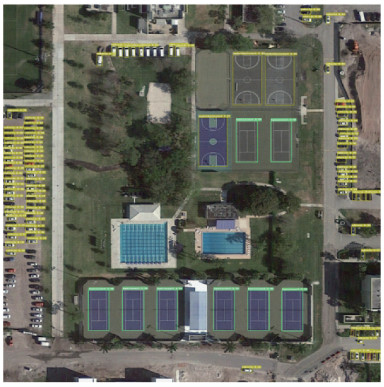
式中，AP 是对数据集中的一个类别的精度进行平均，其定义为

$$AP = \sum_{k=1}^{N} P(k)\Delta r(k) \tag{6}$$

式中，P 和 Δr 为查准率-查全率(precision-recall，P-R)曲线中的查准率和查全率[13,18]。

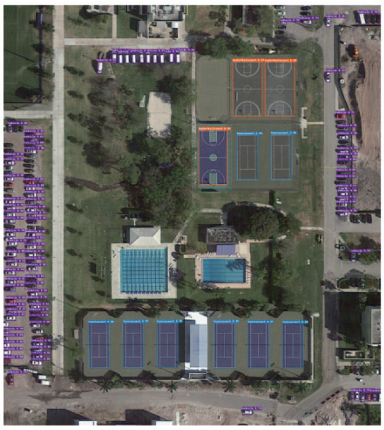
为了验证本文改进网络的有效性，并在实验过程中提供具有参考意义的对照组，本文提出的每一个改进点，均是在单独的实验环境中进行测试，以保证不引入非修改因素对实验结果造成影响。为了便于描述，本文将 ResNet101-vd-dcn 骨干网络的整体网络架构实验称为方法 1，将训练策略优化的方法在 ResNet50-vd-dcn 以及 ResNet101-vd-dcn 骨干网络的实验，分别称为方法 2 和方法 3。将优化的检测颈结构在 ResNet50-vd-dcn 骨干网络的实验称为方法 4。实验结果如图 5 所示。

(a) 方法 1 目标检测测结果



(b) 方法 2 目标检测测结果

(c) 方法 3 目标检测测结果



(d) 方法 4 目标检测测结果

图 5　各方法预测检测效果对比图

根据表 1 和表 2 的统计结果，可以看出使用优化训练策略后网络各类别预测 AP 值均有显著的提升，充分训练后的网络其 mAP 值高达 89.3%。通过对比方法 4 和方法 1，可以看到优化后的检测颈结构对网络的精度提升是有效的，相比于使用更大骨干网络的方法 1，方法 4 仍取得了 2.5% 的 mAP 值提升。故整体考虑，该修改结构引入的参数增多是可以忽略的。即使造成了推理速度的降低，但精度的提升弥补了该缺陷。通过检测效果图 5，可以看到优化的训练策略存在一些矛盾，由于 DIOR 数据集中汽车等目标存在为数众多的特小目标，中等以上目标较少，而网络更多地对小目标以及特小目

表 1 不同方法在 DIOR 数据集上取得的 AP 对比表

方法	优化	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11
方法 1 (ResNet101-vd-dcn)	未使用	89.3	81.4	75.9	89.4	49.7	80.7	61.4	72.3	63.2	80.3	76.8
方法 2 (ResNet50-vd-dcn)	训练策略优化	90.5	89.8	90.4	90.1	86.9	86.4	87.6	90.5	90.5	90.1	89.7
方法 3 (ResNet101-vd-dcn)	训练策略优化	90.5	90.2	89.4	89.3	90.7	90.0	90.4	90.1	89.2	88.1	90.5
方法 4 (ResNet50-vd-dcn)	检测颈优化	89.4	84.3	83.3	89.7	52.0	75.6	68.4	74.0	64.9	78.9	78.9

表 2 不同方法在 DIOR 数据集上取得的 AP 对比表续

方法	优化	H12	H13	H14	H15	H16	H17	H18	H19	H20	mAP
方法 1 (ResNet101-vd-dcn)	未使用	63.7	62.2	70.3	64.1	67.8	89.6	64.4	63.0	82.1	72.4
方法 2 (ResNet50-vd-dcn)	训练策略优化	88.9	88.7	88.1	90.2	89.0	90.4	89.4	80.5	88.2	88.8
方法 3 (ResNet101-vd-dcn)	训练策略优化	88.8	90.4	88.5	89.7	89.5	88.1	90.5	81.5	88.9	89.3
方法 4 (ResNet50-vd-dcn)	检测颈优化	63.9	63.1	78.1	66.5	75.5	89.7	68.9	66.3	85.9	74.9

标的特征进行了学习,反而对中等以上的目标特征学习造成了影响。

3 结 论

本文提出一种改进的 PP-YOLO 网络遥感图像目标检测方法,分别进行了训练策略的优化和检测颈的优化。优化训练方法基于遥感数据集特点,使用更优的 CutMix 数据增强算法替换掉原有的 Mixup 数据增强算法,加入增强网络特征学习的 GridMask 算法,生成增广数据集对模型进行训练,充分且均衡地学习每一类目标的特征。训练优化策略均能有效地提高每一类目标实例的精度,解决了遥感图像目标实例尺寸变化大、类间相识性和类内相似性大,从而造成不同目标实例精度差距较大、部分目标精度不高的问题,最高取得了89.3%的 mAP。

对 PP-YOLO 网络的检测颈部分进行了改进优化,保留了原基本框架,在第 4 和第 5 层加入了由低层网络向高层网络传达的通路,不仅使网络低层部分可以学习到高层部分的特征信息,同时加强高层网络学习到的特征信息。检测颈优化方法提高了检测精度,保证了极佳的泛化能力,即使是在识别完全不同的数据集时,仍然可以取得优异的检测效果,其 mAP 比未改进优化同骨干网络的 PP-YOLO 网络提高了4.4%。

参考文献

[1] 赵文清,孔子旭,周震东,等. 增强小目标特征的航空遥感目标检测[J]. 中国图象图形学报,2021,26(3):644-653

[2] KRIZHEVSKY A, SUTSKEVER I, HINTON G. Image net classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems, New York, USA, 2012: 1097-1105

[3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, USA, 2014: 580-587

[4] GIRSHICK R B. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, Piscataway, USA,2015:1440-1448

[5] REN S, HE K, GIRSHIEK R B, 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):91-99

[6] HE K, GKIOXARIi G, DOLLAR P, et al. Mask R-CNN [C] // Proceedings of the IEEE International Conference on Computer Vision, Piscataway, USA, 2017: 2961-2969

[7] JOSEPH R, SANTOSH D, ROSS G, et al. You only look once: unified, real-time object detection [C] // IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016:779-788

[8] WEI Y H, WEN Y L. Ratio-and-scale-aware YOLO for pedestrian detection [J]. *IEEE Transaction on Image Processing*, 2021,30:934-947

[9] VENKAT A A, HUAN C C, CHI C H, et al. Applying deep learning to defect detection in printed circuit boards via a newest model of you-only-look-once[J]. *Mathematical Biosciences and Engineering*, 2021, 18(4): 4411-4428

[10] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, USA, 2017: 2117-2125

[11] LAW H, DENG J. Cornernet: detecting objects as paired key points [EB/OL]. https://arxiv. org/abs/1808. 01244: arXiv,(2019-03-18),[2021-09-13]

[12] LONG X, DENG K, WANG G, et al. PP-YOLO: an effective and efficient implementation of object detector [EB/OL]. https://arxiv. org/abs/2007. 12099: arXiv, (2020-08-03),[2021-09-13]

[13] 马骏,姚震,许翠峰,等. 基于改进 PP-YOLO 和 DeepSORT 的多无人机实时跟踪算法[J]. 计算机应用, doi:10. 11772/j. issn. 1001-9081. 2021071146

[14] LIU R, LEHMAN J, MOLINO P, et al. An intriguing failing of convolutional neural networks and the coordconv solution. [EB/OL]. https://arxiv. org/abs/1807. 03247: arXiv, (2018-12-03),[2021-09-13]

[15] WANG Y, ZHEN P B, HOU J H, et al. Convolutional neural networks with dynamic regularization[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021,32(5):2299-2304

[16] LIU Z H, ZHANG Y D, CHEN Y Z, et al. Detection of algorithmically generated domain names using the recurrent convolutional neural network with spatial pyramid pooling[J]. *Entropy*, 2020,22(9):1-20

[17] YUN S, HAN D, OH J S, et al. CutMix: regularization strategy to train strong classifiers with localizable features [C]//2019 IEEE/CVF International Conference on Computer Vision (ICC-V), Piscataway, USA, 2019: 6022-6031

[18] 李柯泉,陈燕,刘佳晨,等. 基于深度学习的目标检测算法综述[J]. 计算机工程, doi: 10. 19678/j. issn. 1000-3428. 0062725

Improved PP-YOLO network for remote sensing image target detection

ZHU Fuzhen*, WANG Shuai**, WU Hong*

(*College of Electrical Engineering, Heilongjiang University, Harbin 150080)

(**College of Physics and Electronics, Central South University, Changsha 410006)

Abstract

To improve the accuracy of remote sensing images target detection, an improved PP-YOLO network for remote sensing image target detection is proposed. In the improved PP-YOLO network, the backbone network structure is adopted, the detection neck network is improved, and a path communicated from the low-level network to the high-level network is added between the fourth layer and fifth layer. In this way, features from the low-level network can be passed to the high-level net, which can improve the accuracy of target detection effectively and ensure the generalization ability of the network. Experiment results show that its mAP is 4.4% higher than that of the unimproved net. At the same time, the training strategy of PP-YOLO network is optimized, which is based on the characteristics of the remote sensing data set. The original Mixup data enhancement algorithm is replaced by CutMix data enhancement algorithm, and the GridMask algorithm is added to enhance the learning of network features. Experiments are done and the mAP is 89.3%, which improves the accuracy of target instances detection effectively.

Key words: remote sensing image, target detection, PP-YOLO network