

基于深度学习的遥感目标检测技术

章程军, 胡晓兵⁺, 魏上云, 郭 爽

(1. 四川大学 机械工程学院, 四川 成都 610065;

2. 四川大学 宜宾产业技术研究院, 四川 宜宾 644000)

摘 要: 针对遥感目标检测精度不足, 提出一种改进 YOLOv5s 遥感目标检测算法。主干网络采用 CSP-D 模块进行特征提取, 充分利用深层和浅层特征进行特征增强; 颈部网络采用 BiFPN 结构进行特征融合, 提高多尺度特征信息融合效率。实验结果表明, 针对遥感目标数据集 DIOR, 改进 YOLOv5s 网络平均准确率均值 (mAP) 提升 2.1%, 不同目标类别平均准确率 (AP) 均有提升, 缓解原网络检测存在的漏检误检问题, 改进网络检测速度仍能满足实时性要求, 具有更优的检测性能。

关键词: 遥感图像; 目标检测; YOLOv5s; 特征增强; 多尺度特征融合; 深度学习; 单阶段网络

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1000-7024 (2024) 02-0594-07

doi: 10.16208/j.issn1000-7024.2024.02.035

Remote sensing object detection based on deep learning

ZHANG Cheng-jun, HU Xiao-bing⁺, WEI Shang-yun, GUO Shuang

(1. School of Mechanical Engineering, Sichuan University, Chengdu 610065, China;

2. Yibin Institute of Industrial Technology, Sichuan University, Yibin 644000, China)

Abstract: Aiming at the lack of detection accuracy about remote sensing target detection, an improved YOLOv5s remote sensing target detection algorithm was proposed to improve the detection accuracy of remote sensing targets. CSP-D module was used for feature extraction in backbone network, and the feature information of the shallow and deep layers of the network was fully utilized for feature enhancement. The neck network used BiFPN structure for feature fusion to improve the efficiency of multi-scale feature information fusion. Experimental results show that for the remote sensing target dataset DIOR, compared with the original network, the mean average precision (mAP) is increased by 2.1% through the improved YOLOv5s network. The average precision (AP) of different object detection classes is improved. The problems of missed detection and false detection exist in the original network detection are alleviated. The improved network detection speed can still meet the real-time requirements and shows better detection performance.

Key words: remote sensing image; object detection; YOLOv5s; feature enhancement; multi-scale feature fusion; deep learning; one stage pipeline

0 引 言

相比于传统自然场景图像目标, 遥感目标具有尺度差异明显、成像视角特殊方向多样、目标小且分布密集、背景环境复杂等特点, 这给快速精准完成遥感目标检测造成困难^[1-3]。

传统目标检测算法主要分为感兴趣区域 (region of interest, ROI) 搜索、区域特征提取、分类器分类 3 个步骤^[2]。由于传统算法手工设计特征的局限性, 限制算法检

测精度以及迁移性。深度学习因其不需要人为设计特征, 能够自主实现检测目标特征识别, 具有更好的鲁棒性与迁移性。王金传等^[4]基于 Faster R-CNN 算法, 通过特征融合实现遥感图像中舰船等目标的识别, 虽然在检测精度上取得一定效果, 但检测效率不高。Lu 等^[5]基于 SSD 算法提出一种注意力特征融合结构, 通过将遥感目标多层特征融合增强网络检测能力, 但网络对于图像分辨率要求较高, 例如桥梁等某些特定目标检测效果不佳, 检测实时性存在不足。考虑到遥感目标复杂性, 本文基于深度学习技术, 对

收稿日期: 2022-04-18; 修订日期: 2024-01-28

作者简介: 章程军 (1998—), 男, 安徽池州人, 硕士研究生, 研究方向为机器视觉、图像处理; ⁺通讯作者: 胡晓兵 (1970—), 男, 湖北黄冈人, 博士, 教授, 研究方向为工业机器人、机器视觉、数字化车间、企业信息化; 魏上云 (1999—), 男, 四川宜宾人, 硕士研究生, 研究方向为智能控制、工业机器人; 郭爽 (1997—), 男, 湖北孝感人, 硕士研究生, 研究方向为企业信息化。E-mail: 2481379172@qq.com

遥感目标快速准确检测展开研究。

1 YOLOv5 网络结构

目前基于深度学习的目标检测算法主要有 R-CNN^[6,7]、YOLO^[6,8-10]、SSD^[6] 系列。综合各网络特点, 选用 YOLO 网络实现遥感目标检测。

YOLO 目标检测算法在检测时, 会将输入图片划分为若干个 (SxS) 网格, 每个网格负责预测中心点在其内部的目标, 这包括目标边界框位置、置信度、目标类别信息。作为 YOLO 系列中最新版本的检测算法, YOLOv5 中每个网格预测 3 个边界框 (bounding box) 的坐标信息 (x,y,w,h)、置信度和条件类别概率信息。目前 YOLOv5 包含 4 种不同宽度和深度的网络结构模型, 见表 1。

| 表 1 YOLOv5 不同宽度深度模型 | | |
|---------------------|----------------|----------------|
| Model | Width multiple | Depth multiple |
| YOLOv5s | 0.50 | 0.33 |
| YOLOv5m | 0.75 | 0.67 |
| YOLOv5l | 1.00 | 1.00 |
| YOLOv5x | 1.25 | 1.33 |

YOLOv5 网络相比于前 4 个版本, 在目标检测的精度与速度上均得到提升, 这与其网络结构的改进以及合适训练技巧的应用密不可分。其网络结构由 4 部分组成: ① 输入端 (Input): 对图像进行统一尺寸、计算最优锚框等预处理

理操作, 同时进行数据增强以丰富数据集提高网络检测能力; ② 主干网络 (Backbone): 主干网络是实现目标特征提取的关键部分, 其主要有 Focus、CBL、CSP、SPP 这 4 种模块组成。Focus 模块是 YOLOv5 相较于 YOLOv1-v4 所独有的模块, 其将输入主干网络的图片进行 2 倍下采样的切片操作, 每个特征图被分成 4 份后在空间通道进行拼接, 图片 W、H 信息转成空间信息并不会造成目标特征信息减少, 相比于通过卷积实现下采样操作, 网络能够提取更多目标特征信息。CSP 模块借鉴跨阶段局部网络^[11] (cross stage partial network, CSPNet) 网络思想对残差结构进行改进, 从网络结构设计角度出发, 解决网络模型在推理过程中梯度信息重复问题, 在一定程度上减少网络计算量的同时增强了网络目标特征提取能力, SPP 模块是通过空间金字塔池化实现多尺度特征融合; ③ 颈部网络 (Neck): 用于生成特征金字塔实现对主干网络提取的目标特征进行增强, 提高网络对多尺度目标的检测精度。YOLOv5 颈部网络为 FPN-PAN 结构, FPN^[12] 对主干网络提取特征图进行上采样, 将深层语义信息与浅层目标位置信息进行融合, 在此基础上融合 PAN^[13] 结构, 对上采样得到的特征图进行下采样, 这种自顶向下和自底向上方式全面提升网络对深层语义信息和浅层定位信息的学习性能; ④ 输出层 (Head): 由 3 个检测头组成, 分别对应 20×20、40×40、80×80 大小特征图, 每个检测头会生成对应大小的特征向量, 用于预测目标边界框位置, 目标类别以及对置信度信息。YOLOv5 网络结构如图 1 所示。

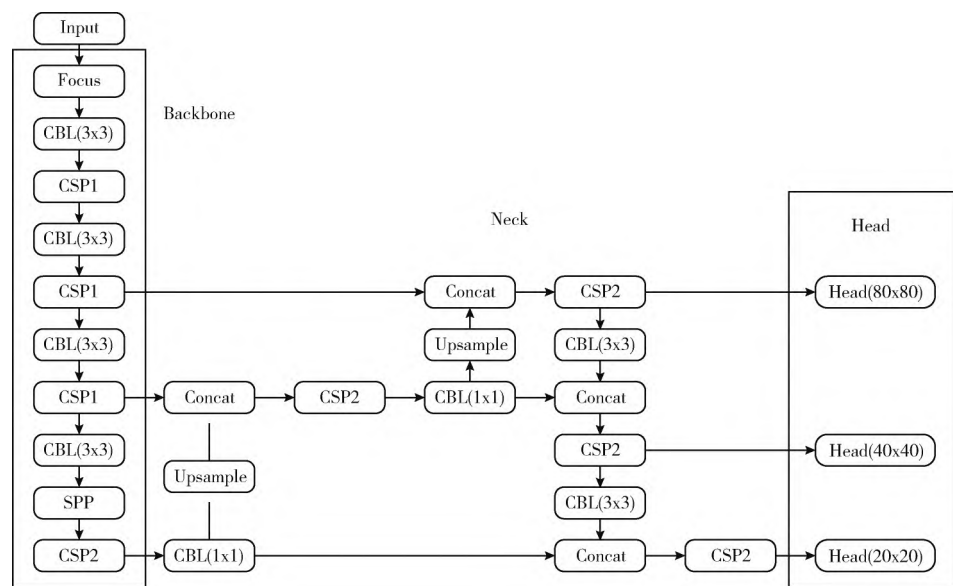


图 1 YOLOv5 网络模型结构

2 YOLOv5 模型改进

综合比较单阶段和双阶段网络, 虽然单阶段网络在检

测精度上低于双阶段网络, 对于网络检测速度却有明显优势, YOLOv5 在速度和精度方面得到较好的平衡, 网络对于目标检测的效果较好。由于 YOLOv5 在提出来时主要以

COCO 数据集为研究对象, 针对特定领域的目标检测虽然能够达到较高的检测精度满足实时性要求, 但仍有一定提升空间。考虑到遥感目标尺度差异大, 小目标占比多。本文对原有网络进行改进, 提出一种基于密集连接的 YOLOv5 遥感目标检测网络。将网络主干部分用于提取特征信息的 CSP 模块替换为采用 DenseNet^[14] 网络结构思想的

CSP-D 模块, 在不同尺度上捕获目标特征信息, 从而获取更有效的目标特征; 采用 BiFPN^[15] 结构对颈部网络进行改进, 充分考虑浅层位置信息和深层语义信息, 实现特征信息多尺度融合, 避免网络过多注重深层次信息进行目标预测, 提高网络对于遥感目标检测能力。改进 YOLOv5 模型如图 2 所示。

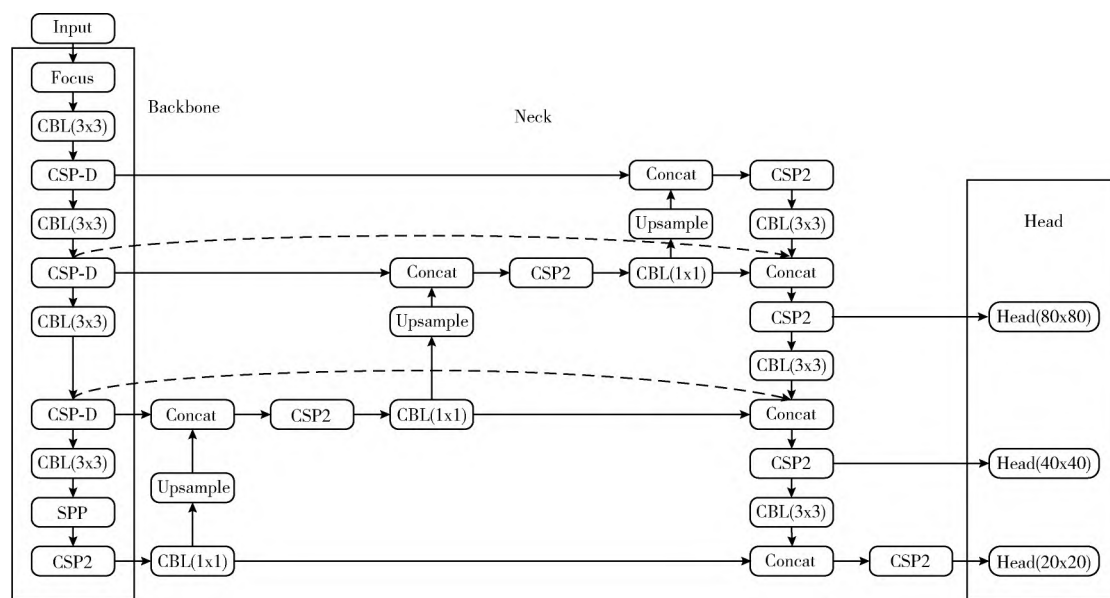


图 2 改进 YOLOv5 网络模型结构

2.1 主干网络 (Backbone) 改进

目前基于深度学习的目标检测网络都是利用卷积神经网络 (CNN) 进行目标特征提取, 本文针对 YOLOv5 提取目标特征的主干网络, 对网络中 CSP 模块进行改进, 借鉴 DenseNet 网络思想, 以提高网络对目标特征提取能力。

为了提高网络检测精度, 通过加深加宽网络结构被证明为一种有效的手段, 但随着网络的加深, 梯度消失成为限制网络精度的重要问题。残差网络^[16] (ResNet) 的提出有效解决了梯度消失问题, 常用于目标检测领域中作为主干网络进行目标特征的提取, 但残差网络提取特征信息不具有细粒度特征, 对于小样本数据集容易产生过拟合。密集连接网络 (DenseNet) 网络的提出为目标特征提取提供一种新思路, 其不同于通过加深加宽网络结构以提高网络预测精度的思想, 而是直接从特征图 (feature) 入手, 更加充分利用特征图信息。相比于 ResNet 网络, DenseNet 网络不仅减轻梯度消失问题; 它加强了 feature 传递, 更加有效地利用 feature, 特征提取能力更强, 特征表达能力更多样化; Dense Block 结构的设计使得 DenseNet 网络更窄, 网络参数和计算量 (FLOPs) 更少; 在一定程度上抑制过拟合问题, 对于小样本数据集目标检测效果更好。ResNet 与 DenseNet 网络差异主要是由于两者对特征聚合方式不同, ResNet 网络聚合特征信息采用求和方式, 这种方式虽

然增加了来自上一层的输入, 但随着网络结构的加深, 浅层特征信息会逐渐稀释, 网络预测结果更加注重深层信息; 而 DenseNet 网络聚合特征信息采用密集连接 (concat) 方式, 当前层会连接前面所有层作为输入, 这种方式充分考虑浅层和深层特征信息, 浅层特征信息不会随着网络的加深逐渐稀释, 每层都具有不同大小感受野信息, 可以捕获不同尺度目标特征, 使得输出可以综合考虑浅层和深层特征信息对目标进行检测, 更加有利于目标特征提取。Dense Block 网络结构如图 3 所示。

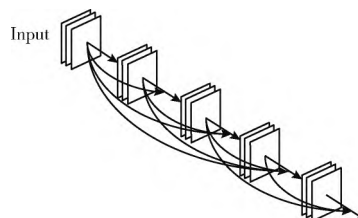


图 3 Dense Block 结构

DenseNet 所采用的这种密集连接方式随着网络的加深, 聚合前面层特征逐渐增多, 会造成特征冗余, 增加运行成本, 一定程度上降低网络效率, 使得目标预测速度变慢。为了能够保证网络特征提取能力的同时解决特征冗余造成的问题, 我们对网络密集连接方式进行改进, 每层

不再连接前方所有层特征信息, 而是直接在最后一层对所有层特征进行融合, 这种方式在利用深层和浅层信息同时, 降低运行成本, 避免因密集连接导致目标检测速度降低。本文针对 YOLOv5 改进的 CSP-D 模块, 将输入特征图分为两部分, 一部分直接聚合到模块末层, 另一部分则通过改进的 Dense Block 模块; 将两部分特征信息在模块末层进行融合, 加强特征图信息传递, 充分利用不同层之间特征信息, 实现网络不同梯度信息组合。CSP-D 模块结构如图 4 所示。

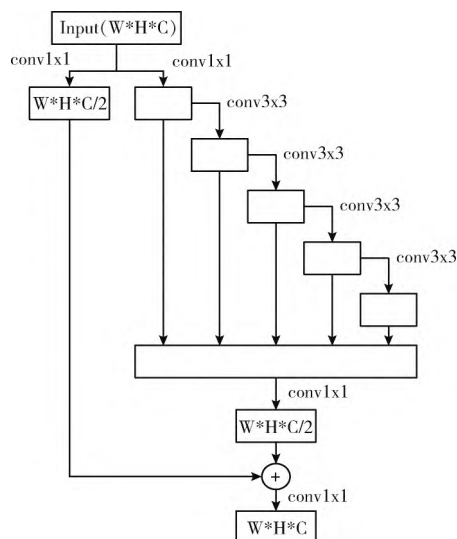


图 4 CSP-D 模块结构

2.2 颈部网络 (Neck) 结构改进

作为主干网络 (Backbone) 与输出层 (Head) 之间的过渡层, 颈部 (Neck) 网络会进行特征增强以提高网络检测精度。在深度学习领域早期的研究中, 目标检测算法主要结构都是采用自底向上的前向过程, 这种方法直接利用深层特征信息对目标进行预测, 浅层信息没有直接参与网络检测结果预测, 且随着网络加深, 浅层信息逐渐稀释, 导致预测结果只注重深层语义信息, 对于浅层位置信息利用不充分, 目标检测时定位误差较大。随着研究的进一步发展, 特征增强模块 FPN (特征金字塔) 的提出有效解决了深层语义信息与浅层位置信息利用不平衡的问题, 成为目前目标检测和实例分割等领域重要组成模块。相比于早期纯粹的利用前向过程进行目标预测, FPN (特征金字塔) 网络在原有网络基础上添加了自顶向下结构即将深层特征进行上采样获得新的特征图, 通过横向连接与原有网络特征信息进行融合。这种特征增强结构将深层语义信息连接到浅层特征中, 实现融合后不同大小特征层与输出层互相对应, 对多尺度目标检测以及小目标检测能力有所提升, 同时提高目标检测时目标定位精度。由于 FPN 结构对深层的小尺度特征图进行上采样会导致一些目标详细信息的丢

失, 从而限制大中型目标的检测精度的提升。在 YOLO 系列中, YOLOv3 已经融合 FPN 结构, YOLOv5 为进一步改进网络性能, 在此基础上借鉴 PANet 网络结构进行改进, 在 FPN 基础上为了使浅层丰富的位置信息与深层丰富的语义信息进行充分融合, 添加了下采样的金字塔结构, 自底向上将底层信息传递到高层。这种方式加强了在相邻层之间信息的融合, 对于非相邻层间的语义信息和位置信息并没有过多注重。无论是 FPN 结构还是 PANet 结构, 在进行特征融合时都是平等对待不同尺度的特征信息, 为了能够使特征信息高效进行多尺度融合, 我们借鉴 BIFPN (加权双向特征金字塔) 结构, 引入权重以对不同尺度的特征信息进行融合。颈部网络 (Neck) 结构如图 5 所示。

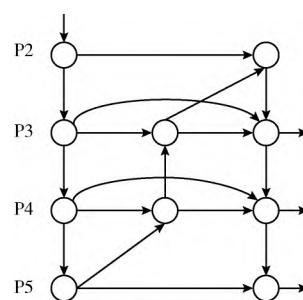


图 5 颈部网络 (Neck) 结构

本文颈部网络 (Neck) 包含两条通路, 一条为自顶向下进行上采样通路, 将高层特征的语义信息传递到底层; 另一条是自底向上进行下采样通路, 将底层特征位置信息传递到顶层。颈部网络删除只有一条输入的节点, 确保每个节点至少有两条输入边以保证特征融合, 对于原始输入节点和输出节点, 若两者处于同一层具有相同大小特征信息, 则会额外添加一条连接, 使输出结果充分考虑原始输入特征信息, 在不增加过多运算成本的情况下融合更多特征信息。这种引入权重以平衡不同尺度信息的融合方式提高具有多尺度特征信息数据集的检测精度。

3 实验过程与分析

3.1 实验数据集

本文采用的是西北工业大学公开的 DIOR 遥感数据集^[17], 其包含 23 463 张图片和 192 472 个实例, 覆盖 20 种目标类别: Airplane、Airport、Baseball field、Basketball court、Bridge、Chimney、Dam、Expressway service area、Expressway toll station、Golf course、Ground track field、Harbor、Overpass、Ship、Stadium、Storage tank、Tennis court、Train station、Vehicle、Wind mill。数据集目标类间和类内尺寸变换范围大, 具有较高的类间相似性和类内多样性, 涉及不同天气、季节等环境因素。实验时将数据集按 6:2:2 比例划分训练集、验证集和测试集。

3.2 实验环境与评价指标

本实验采用操作系统为 Linux 系统,使用软件为 python3.7+Torch1.6.0;实验硬件配置为处理器(CPU)为英特尔 Core(TM) i9-9900k 3.60 GHz、显卡(GPU)为 NVIDIA GeForce RTX 2080Ti、内存 11 G。实验训练模型时采用随机梯度下降法(SGD)法更新网络训练参数,以减少算法收敛时间。实验设置网络超参数见表 2。

表 2 网络训练超参数

| 训练参数名称 | 参数值 |
|--------------|---------|
| 初始学习率 | 0.01 |
| 学习率下降参数 | 0.0001 |
| 权重衰减系数 | 0.0005 |
| 动量(momentum) | 0.937 |
| 批处理大小(batch) | 16 |
| 训练批次(epoch) | 300 |
| 图像输入尺寸 | 800×800 |

为衡量改进前后网络性能优劣,实验结果从召回率(recall)、查准率(precision)、平均准确率均值(mean average precision, mAP)、各个类别的平均准确率(average precision, AP)、推理时间(s)以及浮点运算量(FLOPs)6个方面来对比网络性能。相关计算公式如下

$$recall = \frac{TP}{TP + FN} \quad (1)$$

$$precision = \frac{TP}{TP + FP} \quad (2)$$

$$AP = \int_0^1 P(R) dR \quad (3)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (4)$$

召回率(recall)和查准率(precision)指所检测到正样本数量分别占标记真实样本数量和检测到的所有样本数量比率,为平均准确率(average precision, AP)与平均准确率均值(mean average precision, mAP)计算基础。在式(1)、式(2)中 TP、FP、FN 关系见表 3。

表 3 TP、FP、FN 关系

| 真实情况 | 预测情况 | |
|------|------|----|
| | 正例 | 反例 |
| 正例 | TP | FN |
| 反例 | FP | TN |

3.3 模块改进对比分析

为验证各个模块改进是否提高网络检测精度,进行消融实验,见表 4。

表 4 改进网络消融实验

| 网络模型 | mAP.5 | mAP.5:.95 |
|---------------------|-------|-----------|
| YOLOv5s | 0.869 | 0.609 |
| YOLOv5s+BiFPN | 0.875 | 0.633 |
| YOLOv5s+BiFPN+CSP-D | 0.890 | 0.656 |

借鉴 BiFPN 结构,颈部网络的改进能够更高效融合特征信息,主干网络采用改进后的 CSP-D 模块可以更加充分提取特征信息。通过表 4 数据可以看出,相比于与原网络模型 YOLOv5s,颈部网络改进(YOLOv5s+BiFPN) mAP.5 值提升 0.6%, mAP.5:.95 值提升 2.4%;最终改进模型(YOLOv5s+BiFPN+CSP-D) mAP.5 值提升 2.1%, mAP.5:.95 值提升 4.7%。各个模块改进对于网络检测精度均有提升。

3.4 YOLOv5 改进前后对比分析

为验证改进对网络性能影响,改进 YOLOv5s 与原网络训练结果对比分析见表 5、表 6。

表 5 改进前后网络性能对比

| 网络模型 | Precision | Recall | mAP | 推理时间/s |
|------------|-----------|--------|-------|--------|
| YOLOv5s | 0.885 | 0.825 | 0.869 | 0.007 |
| 改进 YOLOv5s | 0.902 | 0.844 | 0.890 | 0.011 |

由表 5 可以看出,改进 YOLOv5s 相比于原网络查准率(precision)、召回率(recall)、平均准确率均值(mAP)均有所提高,其中查准率提高 1.7%、召回率提高 1.9%、mAP 值提高 2.1%。改进后模型推理时间为 0.011 s,相比于原网络 0.007 s 虽有升高,仍能达到毫秒级别,满足实时性要求。

表 6 为网络模型改进前后数据集 20 个目标类平均准确率(AP)对比表,相比原 YOLOv5s,改进网络对于每个目标类检测 AP 值均有所提升,其中 Train station 目标类 AP 值提升最高,达到 6.60%;bridge 目标类 AP 值也提升 4.2%。

已有研究表明,通过增加网络深度和宽度可以提高网络对目标检测精度,但此方法同时增加网络复杂度,浮点运算量(FLOPs)作为衡量网络模型复杂度指标,表 7 为不同大小 YOLOv5 模型与改进后模型对比分析。

表 7 表明,YOLOv5m 模型通过增加网络深度宽度,相比于 YOLOv5s 模型 mAP 值提升 1.4%,但网络浮点运算量(FLOPs)大幅升高,为 YOLOv5s 网络浮点运算量 3 倍多。改进 YOLOv5s 网络相比于 YOLOv5m 网络 mAP 值高出 0.7%,并且浮点运算量更少,网络复杂度低。

网络改进前后总损失值和平均准确率(mAP)对比曲线如图 6 所示,从图 6(a)、图 6(b)中看出,网络在训练

表 6 目标平均准确率 (AP) 对比

| 目标类别 | YOLOv5s | 改进 YOLOv5s | 提升量/% |
|-------------------------|---------|------------|-------|
| airplane | 0.974 | 0.981 | 0.70 |
| airport | 0.921 | 0.937 | 1.60 |
| Baseball field | 0.973 | 0.981 | 0.80 |
| Basketball court | 0.931 | 0.949 | 1.80 |
| bridge | 0.616 | 0.658 | 4.20 |
| chimney | 0.926 | 0.936 | 1.00 |
| dam | 0.802 | 0.833 | 3.10 |
| Expressway-Service-area | 0.947 | 0.959 | 1.20 |
| Expressway-toll-station | 0.918 | 0.940 | 2.20 |
| Golf course | 0.818 | 0.857 | 3.90 |
| Ground track field | 0.926 | 0.937 | 1.10 |
| harbor | 0.695 | 0.734 | 3.90 |
| overpass | 0.757 | 0.793 | 3.60 |
| ship | 0.946 | 0.954 | 0.80 |
| stadium | 0.928 | 0.941 | 1.30 |
| Storage tank | 0.892 | 0.914 | 2.20 |
| Tennis court | 0.966 | 0.970 | 0.40 |
| Train station | 0.758 | 0.824 | 6.60 |
| vehicle | 0.736 | 0.768 | 3.20 |
| windmill | 0.949 | 0.961 | 1.20 |
| all | 0.869 | 0.890 | 2.10 |

表 7 模型复杂度对比

| 网络模型 | 浮点运算量 (FLOPs) | mAP |
|------------|---------------|-------|
| YOLOv5s | 16.5 G | 0.869 |
| YOLOv5m | 50.6 G | 0.883 |
| 改进 YOLOv5s | 33.3 G | 0.890 |

批次迭代到 60 次时基本达到稳定。随着迭代次数的增加, 网络趋于稳定后改进网络 mAP 值稳定高于 YOLOv5s 网络; 总损失值不断降低, 整体低于 YOLOv5s 网络。

图 7 为网络改进前后实际检测效果对比图, 其中左侧为原始网络实际检测效果, 右侧为改进后网络实际检测效果。与改进 YOLOv5s 网络相比, YOLOv5s 网络针对不同密度目标虽能检测出大部分目标, 但仍存在漏检、误检问题, 且已检目标置信度要低于改进网络。如前 4 幅图中, 针对 harbor 类目标, YOLOv5s 网络存在漏检情况; 后两幅图中, YOLOv5s 网络将非 ship 类目标检测为 ship 类目标, 出现误检情况。改进 YOLOv5s 网络模型缓解原网络漏检、误检问题, 提高网络检测性能。

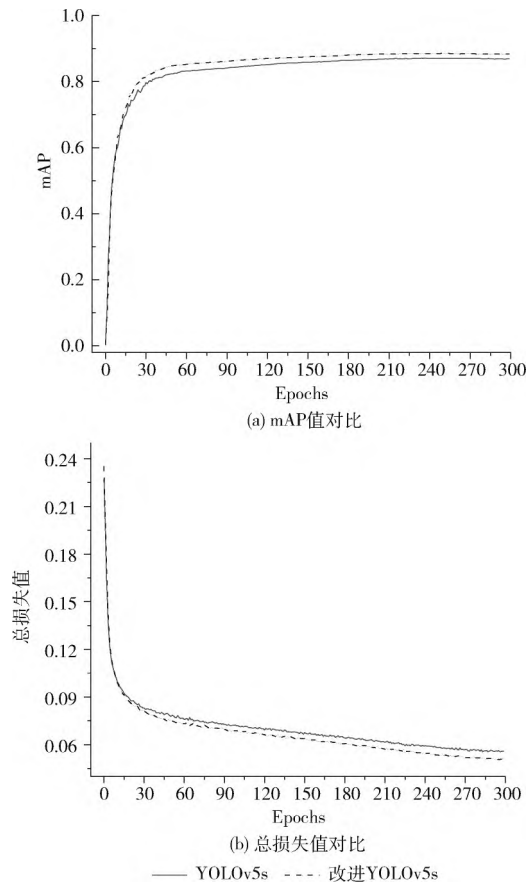


图 6 YOLOv5s 改进前后总损失值和 mAP 值对比

4 结束语

本文选用单阶段网络 YOLOv5s 作为遥感目标检测模型, 在原网络基础上, 对主干网络采用新设计 CSP-D 模块替代原网络中 CSP 模块, 提高主干网络对目标信息提取能力; 借鉴 BiFPN 结构改进颈部网络, 使特征信息能够进行更高效的多尺度特征融合, 提高网络对不同目标检测精度。在遥感目标数据集 DIOR 进行实验后表明, 改进 YOLOv5s 相比于原网络平均准确率均值 (mAP) 有所提高, 总损失低于原网络模型, 提升各个类目标检测精度, 缓解 YOLOv5s 网络漏检误检问题; 虽然增加网络推理时间以及模型复杂度, 仍能满足实时性要求, 并且相比于通过加深加宽网络以提升检测精度方法, 相对降低网络模型复杂度。改进 YOLOv5s 相比于原网络整体性能有所提升, 但对于实际中包含各种复杂因素的遥感目标检测, 还具有一定局限性, 后续将进一步研究以提升深度学习领域对遥感目标检测性能。

参考文献:

[1] Xu D, Wu Y. MRFF-YOLO: A multi-receptive fields fusion network for remote sensing target detection [J]. Remote Sen-

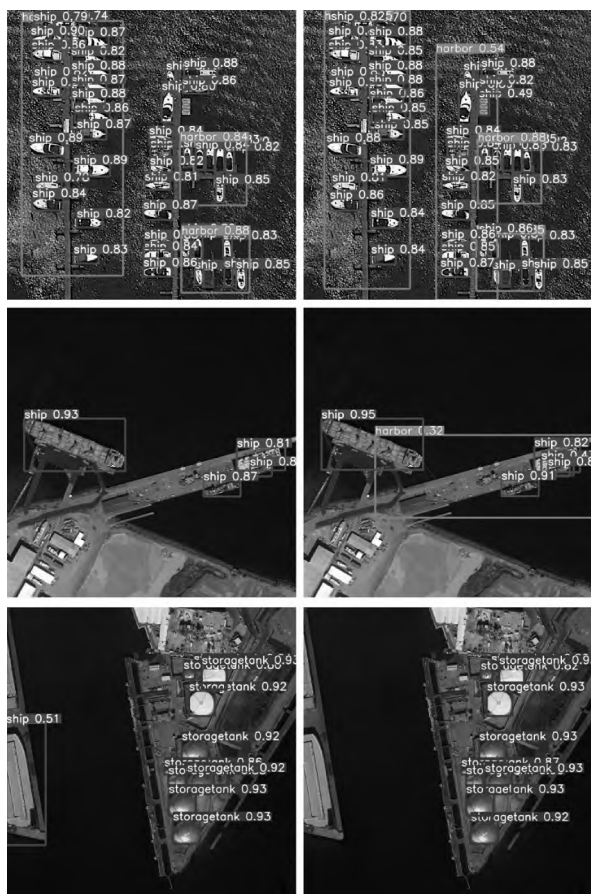


图7 实际检测效果对比

sing, 2020, 12 (19): 3118-3129.

- [2] Xu D, Wu Y. FE-YOLO: A feature enhancement network for remote sensing target detection [J]. Remote Sensing, 2021, 13 (7): 1311-1320.
- [3] Xu H, Wang X, Rui T, et al. Exploiting full-scale feature for remote sensing object detection based on refined feature mining and adaptive fusion [J]. IEEE Access, 2021, 9: 145422-145434.
- [4] WANG Jinchuan, TAN Xicheng, WANG Zhaohai, et al. Faster R-CNN deep learning network based object recognition of remote sensing image [J]. Journal of Geo-Information Science, 2018, 20 (10): 1500-1508 (in Chinese). [王金传, 谭喜成, 王召海, 等. 基于 Faster-RCNN 深度网络的遥感影像目标识别方法研究 [J]. 地球信息科学学报, 2018, 20 (10): 1500-1508.]
- [5] Lu X, Ji J, Xing Z, et al. Attention and feature fusion SSD for remote sensing object detection [J]. IEEE Transactions on

Instrumentation and Measurement, 2021, 70: 1-9.

- [6] Aziz L, Salam MSBH, Sheikh UU, et al. Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review [J]. IEEE Access, 2020, 8: 170461-170495.
- [7] 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.
- [8] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [9] Redmon J, Farhadi A. Yolov3: An incremental improvement [J]. arXiv preprint arXiv, 2018, 1804: 02767-02778.
- [10] Bochkovskiy A, Wang CY, Liao HYM. Yolov4: Optimal speed and accuracy of object detection [J]. arXiv preprint arXiv, 2020, 2004: 10934-10945.
- [11] Wang CY, Liao HYM, Wu YH, et al. CSPNet: A new backbone that can enhance learning capability of CNN [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 390-391.
- [12] Lin TY, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [13] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [14] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
- [15] Tan M, Pang R, Le QV. Efficientdet: Scalable and efficient object detection [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [16] Shrestha A, Mahmood A. Review of deep learning algorithms and architectures [J]. IEEE Access, 2019, 7: 53040-53065.
- [17] Li K, Wan G, Cheng G, et al. Object detection in optical remote sensing images: A survey and a new benchmark [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296-307.