

DOI:10.19651/j.cnki.emt.2211674

多重金字塔的轻量化遥感车辆小目标检测算法^{*}

赵倩 杨一聪

(上海电力大学电子与信息工程学院 上海 201306)

摘要: 针对遥感车辆检测任务中存在目标尺寸小、背景复杂等问题,提出一种基于多重金字塔和多尺度注意力的轻量级 YOLOv5 算法。在主干网络中减少下采样次数,提高小目标检测能力,实现轻量化;在颈部中通过重新设计的多重金字塔网络,充分利用不同特征层的信息,增强特征融合能力,并引入改进的多尺度注意力模块,为浅层特征图获得更大的感受野和感兴趣区域;最后使用 K-means++ 聚类算法对目标尺寸进行聚类分析,设计出适合目标的锚框尺度和宽高比。在自建遥感车辆数据集中不仅提升了目标检测精度,而且大大降低参数量。与 YOLOv5s 相比较,AP_{0.5}% 提高了 2.3%、AP_{0.5,0.75}% 提高了 4.3%;参数量降低了 65%、模型大小减少了 60%。在轻量化的同时有效地提高了小目标的检测精度。

关键词: 遥感车辆检测;特征融合;注意力机制;轻量化

中图分类号: P237 文献标识码: A 国家标准学科分类代码: 510.4050

Small object detection algorithm for lightweight remote sensing vehicles with multiple pyramids

Zhao Qian Yang Yicong

(School of Electronic and Information Engineering, Shanghai University of Electric Power, Shanghai 201306, China)

Abstract: Aiming at the problems of small target size and complex background in remote sensing vehicle detection tasks, a lightweight YOLOv5 algorithm based on multiple pyramids and multi-scale attention is proposed. In the backbone network, the number of downsampling is reduced, the small target detection ability is improved, and light weight is achieved; in the neck, the information of different feature layers is fully utilized through the redesigned multi-pyramid network to enhance the feature fusion ability. And introduce an improved multi-scale attention module to obtain a larger receptive field and area of interest for the shallow feature map; finally, the K-means++ clustering algorithm is used to cluster and analyze the target size, and an anchor frame scale suitable for the target is designed. and aspect ratio. In the self-built remote sensing vehicle dataset, the target detection accuracy is not only improved, but also the parameter quantity is greatly reduced. Compared with YOLOv5s, AP_{0.5}% is increased by 2.3%, AP_{0.5,0.75}% is increased by 4.3%; the number of parameters is reduced by 65%, and the model size is reduced by 60%. It effectively improves the detection accuracy of small targets while reducing weight.

Keywords: remote sensing vehicle detection; feature fusion; attention mechanism; lightweight

0 引言

近些年来,我国的机动车数量迅速增长。伴随而来的是交通阻塞日益严重,单靠人力疏导交通已经不能满足现代交通系统的需要。因此,智能交通建设是解决问题的关键,同时也加快智慧城市建设进度。通过对遥感图像中车辆目标位置进行精确定位、快速识别并传回智能交通系统,来缓解交通压力。因此车辆目标检测有很重要的现实

意义。

目前已有很好优秀的目标检测算法应用在对于自然场景的检测中。目前流行的目标检测方法大致分为两类,一类是两阶段目标检测方法,如典型的 Fast R-CNN^[1], Faster R-CNN^[2],它们通过区域建议网络生成候选框,并将特征提取、区域建议、边界回归和分类结成一个框架当中,可以显著提高检测性能,但是额外的区域建议网络会损失大量的计算资源,限制推理速度。第2类是以 YOLO 系

收稿日期:2022-10-12

^{*} 基金项目:国家自然科学基金(61802250)项目资助

列^[3-6]和 SSD^[7]为代表的单阶段目标检测算法,可以实现端到端的预测,并可以一步预测出所有目标的边界框,与两阶段检测方法相比,可以实现更快速的检测和较高的检测精度,实现实时的良好性能。

深度学习在自然场景中已取得重大进展,但直接应用在遥感车辆检测会存在很多问题^[8],比如目标尺寸小、尺寸大小不一、背景复杂等问题^[9],传统目标检测算法难以解决这个问题,导致在遥感车辆检测领域检测效果不佳。针对目标尺寸小问题,徐晓光等^[10]在网络中进行多尺度的改进,来扩大感受野提取多尺度信息。针对目标易受背景干扰问题,Hu 等^[11]融合主干网络用于提取的特征图,其次在融合的特征图上使用 SE 注意力机制消除冗余信息减弱噪声区域通道特征响应。张宏群等^[12]利用 DIOU_NMS 算法降低了遥感目标被遮挡后的影响,增强了检测效果。针对轻量化,李仁鹰等^[13]在 YOLO 网络中实现了轻量化处理。彭继慎等^[14]在 YOLO 网络中实现了模型压缩。以上方法有的增加了精度有的做了轻量化,但难以平衡模型大小和精度关系。

针对上述问题,为了提高车辆遥感图像小目标检测精度,本文在 YOLOv5^[15](you only look once version 5)基础上提出一种基于 FPF-FPN 和 MSAM(multi-scale attention mechanism)模块改进的轻量级 YOLOv5 算法,通过重新设计的多重金字塔网络,充分利用不同特征层的信息,增强特征融合能力;浅层特征层感受野小,不利于小目标的检测,为解决感受野问题参考 RFB 结构和注意力模块设计了多尺度注意力模块,以获得更大感受野和感兴趣区域;在预处理过程中使用 Mosaic 进行数据增强丰富数据表达能力;同时预先使用 K-means++ 聚类方法生成更加匹配的候选框参数,进一步提高目标检测的性能。

1 改进 YOLOv5 目标检测算法

1.1 YOLOv5 算法原理

自 2016 年以来 YOLO 算法经过数次更新迭代,其检测精度和速度不断优化,在目标检测领域有着广泛的应用和突出表现,YOLOv5 是 YOLO 系列最新版本,它融合了

很多先进技术,而 YOLOv5s 是 YOLOv5 最轻量化的版本,所以本文选用 YOLOv5s 为基准网络。

其主干网络使用 C3 模块,即有 3 次卷积的 CSPDarknet53^[16]结构,来提高算法的特征提取能力,也做到了轻量化。YOLOv5 的 5.0 版本以后将 SPP^[17]替换为 SPPF 模块,减少了指定卷积核的数量,提高了特征提取速度。Neck 中为进一步提升网络的特征提取能力使用 FPN^[18]、PAN^[19]构,实现了语义信息从深层到浅层的传递,和位置信息从浅层到深层的传递,两者结合,极大地增强了网络的特征提取能力。考虑了重叠面积、中心点距离、长宽比 3 个重要因素,Head 输出端的边界损失函数使用 CIOU_Loss,大大提高了预测框回归的精度和速度。

1.2 改进算法网络结构

本文在 YOLOv5s 的基础之上提出 FPF-FPN 融合模块和改进的多尺度注意力模块,网络结构如图 1 所示。算法具体步骤为:1)主干特征提取。浅层特征图包含更多的纹理信息,对小目标检测有益,而深层特征图包含更多的语义信息,含有更大的感受野,但由于 5 次的下采样操作不利于小目标检测,因此改进的网络,仅进行 4 次下采样操作(即 C1、C2、C3、C4 特征图),并选择 4、8、16 倍下采样的特征图(即 C2、C3、C4)参与特征融合,防止含有极少小目标信息的深层特征图在特征融合阶段带入冗余信息,影响小目标检测精度,同时也降低了模型大小和参数量实现了轻量化。2)对主干网络提取的特征进行增强,构建多尺度注意力模块。受 RFB 和注意力原理启发,使用包括残差连接、尺度不同的组卷积、尺度相同的带有膨胀率的组卷积构成多路结构以此横向拓宽网络、增加感受野,同时结合注意力机制聚焦感兴趣区域减少冗余信息(如图 1 模块 I 所示)。3)多重特征融合。针对小目标及对应特征图呈现的特性,本文再次对底层信息进行上采样融合到顶层特征图中,提出新融合结构,显著增加小目标特征(如图 1 中 FPF-FPN 结构)。4)输出端预测结果。基于由高低层级特征融合结构输出的两路特征图进行小目标检测,并经过非极大抑制处理得到最终小目标检测结果。

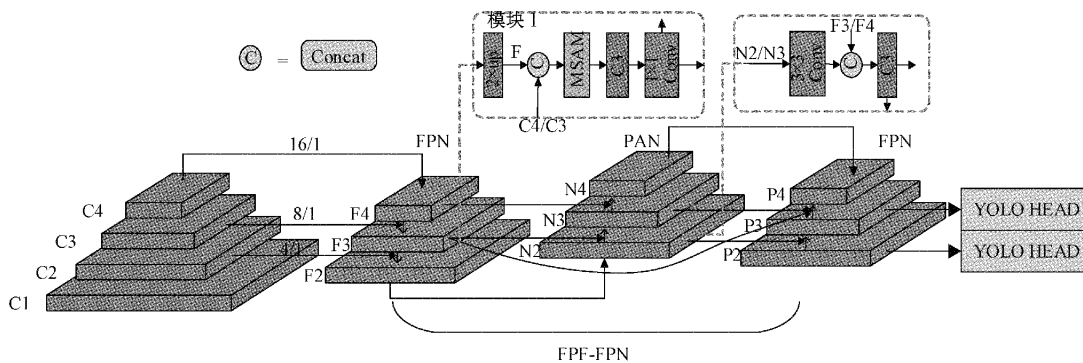


图 1 FPF-YOLOv5 网络结构

1.3 多重特征融合 FPF-FPN

原 YOLOv5 颈部网络结构采用 FPN+PAN 的结构,如图 2 所示。为了增加小目标的检测能力,加强不同尺度特征融合,设计了新的特征融合结构 FPF-FPN(即 FPN+PAN+FPN 多重特征融合结构)。在经过又一次上采样以及特征融合之后,小目标检测能力显著提升。

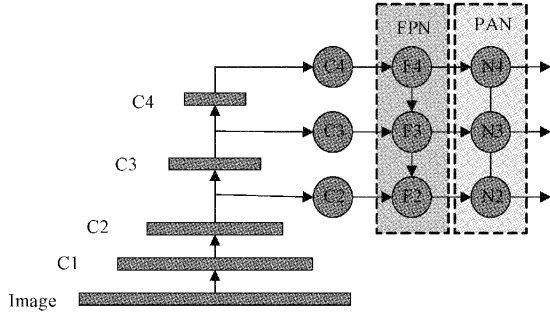


图 2 FPN+PAN 结构

同时参考 BiFPN^[20] 结构,为了减少参数运算量,删掉了部分节点,对 FPN+PAN 输出的 P4 特征图再进行上采样得到 P3,进一步增加对小目标检测的精度,同时为了更好地加强不同特征图的融合,将其与 F3 特征层采用级联方式融合,如图 3 所示。

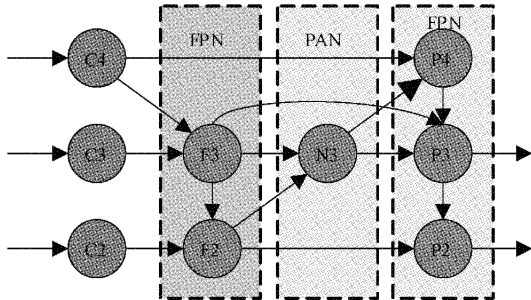


图 3 FPF-FPN 结构

C4 层特征图为输入图像的 1/16,经过 FPN+PAN 后并不用来输出以检测大物体,因为本数据集中绝大多数为小物体,经过实验得知如果将 P4 层也作为输出并不利于小物体的检测,反而因为 P4 拥有比 P3、P2 更加丰富的语义信息,将其上采样后注入浅层特征图,能够丰富浅层特征图的语义信息,更有利于进行小目标检测。

1.4 多尺度注意力模块

小目标检测主要是使用低层特征图,因为其包含更多小目标形状等纹理信息。但主干网络特征提取能力有限,不能与带有丰富语义信息的深层特征信息充分融合,同时浅层特征图感受野有限,并不能有效检测小目标,因此不能一味仅使用浅层特征图来检测小目标。

所以,参考 RFB^[21]和注意力结构设计出多尺度注意力模块 MSAM。加入到颈部特征图融合中,在获得大感受野的同时也能够关注更感兴趣区域,整体结构如图 4 所示。

C4/C3 为主干网络输出特征图,F 为颈部网络与 C4/

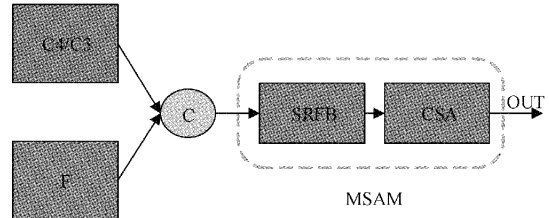


图 4 MSAM 结构

C3 对应层上一层经上采样后的特征图。将两个特征图拼接后依次经过 SRFB 和 CSA 注意力模块后输出。为了提高特征图的感受野和小目标的敏感性和适应性,提出的 SRFB 结构,如图 5 所示。经拼接后的特征图经过 SRFB 后,在不改变分辨率的情况下增强了对小目标提取的鲁棒性,增强了感受野。

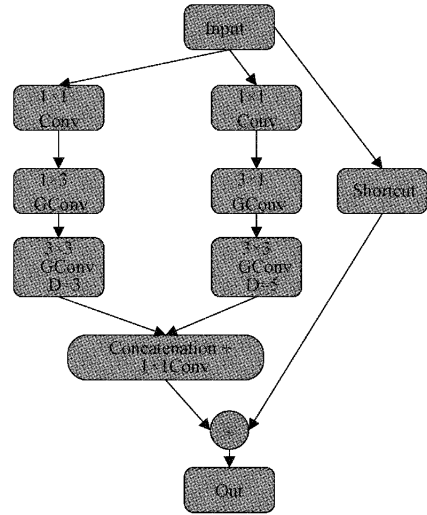


图 5 SRFB 结构

因 RFB 结构支路繁多,参数量巨大,不满足轻量化要求,所以减少支路并引入组卷积操作提出了改进的 SRFB 结构。该改进结构为 3 条支路,参考 Inception^[22]可以获得不同尺度感受野。首先都进行 1x1 卷积操作来调整通道数,以初步处理输入的特征图,为后续操作做准备。左边两条支路首先分别经过 1x3 和 3x1 卷积操作,同时为了降低参数量都使用了组卷积操作(如图 5 中 GConv),以不同尺度卷积操作来提取更加细腻度的小目标特征同时兼顾了参数量。为了扩大感受野,在左边两条支路中的 3x3 卷积分别使用了空洞率为 3 和 5 的空洞卷积,同时也引入了组卷积来降低参数量(如图 5 中 GConv, D=3/5)。第 3 条支路为残差结构,在保留小目标同时可以防止梯度消失。改进 RFB 模块计算过程可表示为:

$$X1 = \int_{D1GConv}^{3 \times 3} \left\{ \int_{GConv}^{1 \times 3} \left[\int_{Conv}^{1 \times 1} M \right] \right\} \quad (1)$$

$$X2 = \int_{D1GConv}^{3 \times 3} \left\{ \int_{GConv}^{3 \times 1} \left[\int_{Conv}^{1 \times 1} M \right] \right\} \quad (2)$$

$$Y = \int_{Conv}^{1 \times 1} Cat(X1, X2) \oplus \int_{Conv}^{1 \times 1} M \quad (3)$$

式中: $\int_{Conv}^{1 \times 1}$ 表示为 1×1 的常规卷积操作, $\int_{GCov}^{1 \times 3}$ 、 $\int_{GCov}^{3 \times 1}$ 为 1×3 、 3×1 的组卷积操作, $\int_{DICCov}^{3 \times 3}$ 为扩张率分别为 3、5 的空洞卷积和组卷积操作。Cat 表示拼接操作, \oplus 表示像素相加操作, M 为输入特征图, X_1 、 X_2 为前两个支路经过常规卷积和组卷积空洞卷积之后的特征图。 Y 为最后输出的新特征图。

为了提高特征图的特征表达能力,将注意力模块添加到特征融合层中。如图 6 所示,CSA 注意力模块为 Coordinate Attention^[23] 和 SAM^[24] 的组合。通道注意力来自 Coordinate Attention,空间注意力来自 SAM。Coordinate Attention 将位置信息嵌入到通道注意力中,与通过二维全局池化将张量转换为单个特征向量的通道注意力不同,Coordinate Attention 将通道注意力分解为两个一维特征编码过程,分别沿两个空间方向聚合特征。一个方向用来捕获远程依赖关系,同时可以沿另一个空间方向保留精确的位置信息,最后将生成的特征图分别编码为一维方向感知和位置敏感的特征图,可以互补地应用于输入特征图,以增强关注对象的表示。与广泛使用的 SE Block 不同,它只考虑通过建模通道关系来重新衡量每个通道的重要性,而忽略了位置信息,但是位置信息对于生成空间选择性的特征图很重要。所以可以通过 Coordinate Attention 对通道关系和长期依赖性进行编码。

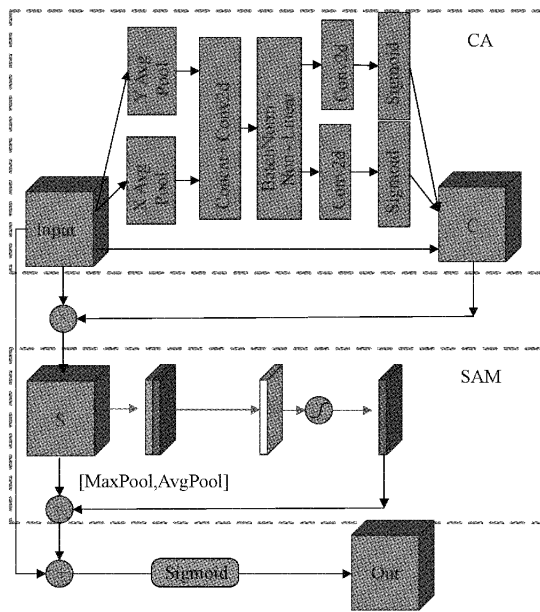


图 6 CSA 注意力模块

Coordinate Attention 生成的特征图与 Input 相乘得到 S 特征图作为 SAM 空间注意力模块的输入,将 SAM 注意力生成的特征图与原始特征图求和,最后再通过激活函数输出最终特征图。

2 实验过程与结果分析

2.1 实验环境及数据集准备

本实验在 Win10 操作系统下进行,CPU 为 i512400F, GPU 为 RTX3060,显存为 12 G,实验环境为 Python3.6, Pytorch1.8.1,在 YOLOv5 中,根据网络的宽度和深度不同,分为 YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l 及 YOLOv5x 这 5 个模型,深度和宽度依次增加,本文基于 YOLOv5s 为基础模型,在其上进行网络改进和优化。

虽然目前有一些用于小目标检测的数据集,但其数据集中小目标物体所占比例很低,大中目标比例很高,因此数据集样本分布很不均匀。因此,为了实现遥感车辆小目标检测,从公开的 DIOR^[25] 遥感图像数据集和 DOTA^[26] 遥感数据集中筛选车辆图像,并对其进行图像分割等操作。

DIOR 数据集包含 20 种类别,共包括 23 463 张图片,原图像尺寸为 800×800 。DOTA 数据集包含 15 个类别,图片大小 800×800 到 $4\,000 \times 4\,000$ 大小不等,从两个数据集中选出车辆图片组成新数据集。由于 DOTA 数据集中原图像尺寸过大且不一,因此对于 DOTA 数据集中选出的车辆数图像进行裁剪分割操作并将原图中的标注信息导入分割后的特征图中。最终得到新构建的图片共计 5 262 张,按照 8 : 2 比例分配,得到 4 147 张训练集图像和 1 115 张验证集图像,以此来构建遥感图像车辆小目标数据集。构建的遥感车辆数据集示例如图 7 所示。

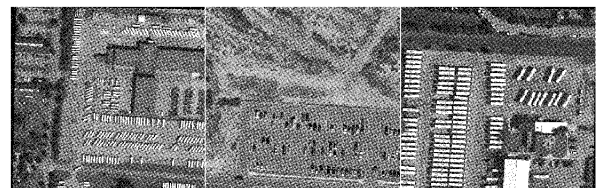


图 7 自建数据集示例

为验证所构建遥感车辆小目标数据集合理性,对标签框宽高进行统计,在物体检测算法中一般认为目标尺寸小于 32×32 的物体为小目标,如图 8 所示,该数据集中标签框的尺寸绝大部分都小于图像的 5% (即 32×32),且根据类别目标数量分布图如图 9 所示,small-vehicle 数量为 63 575 个,远高于 large-vehicle 的 27 041,可以证明本构建数据集对于小目标检测的有效性。

2.2 模型训练及评价指标

图像输入图片的大小为 640×640 pixels 并采用 Mosaic 数据增强操作,初始学习率设为 0.001,使用 Adam 优化器进行优化训练,训练批次大小设为 20,动量参数为 0.937,权重衰减系数为 0.000 5,训练 100 轮次。

本文使用准确率的评价指标为平均精度值 (mean average precision, mAP),相关公式如下表示。

精确率 Precision 计算公式为:

$$P = \frac{TP}{TP + FP} \quad (4)$$

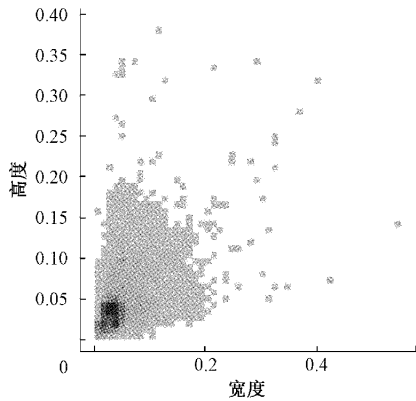


图 8 标签框尺寸分布

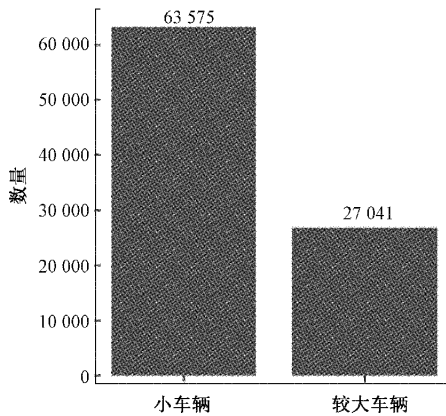


图 9 类别目标数量分布

召回率 Recall 计算公式为：

$$R = \frac{TP}{TP + FN} \quad (5)$$

平均精度 (average precision, AP) 计算公式为：

$$AP_{IoU} = \int_0^1 P(R) dR \quad (6)$$

式中：P、R 分别表示为准确率 (Precision)、召回率 (Recall)，其中 TP (true positive) 为真正例，FN (false negative) 为反例，FP (false positive) 为假正例。AP 为单一类别的准确率-召回率曲线下方形成的面积积分，N 为类别数。IoU 为检验框与真实框之间的交并比阈值，本实验采用 $AP_{0.5}$ 及 $AP_{0.5,0.95}$ 作为评判标准。

2.3 消融实验

为验证本文提出的 FPF-YOLOv5 算法中各个模块对于小目标检测的影响，采用消融实验逐一对不同模块在相同实验条件下目标检测算法的性能进行评估。

各个模块实验属于递进关系，基于 Ultralytics 6.0 版本的 YOLOv5s 不断改进优化，最终得出最好的模型。这 7 个模型的描述在表 1 中介绍，体现了本算法对于小目标检测的优化过程和优化思路。

消融实验采用 $AP_{0.5}$ 、 $AP_{0.5,0.95}$ 、Params (参数量)、

表 1 模型介绍

模型	模型描述
1	YOLOv5s
2	YOLOv5s 去掉主干中 C5(out C5)
3	YOLOv5s+outC5+FPF
4	YOLOv5s+outC5+FPF+级联
5	YOLOv5s+outC5+FPF+去掉 P4
6	YOLOv5s+outC5+FPF+级联+去掉 P4
7	YOLOv5s+outC5+FPF+级联+去掉 P4+SRFB
8	YOLOv5s+outC5+FPF+级联+去掉 P4+CSA
9	YOLOv5s+outC5+FPF+级联+去掉 P4+MSAM

M(模型大小)为模型评价指标。消融实验结果如表 2 所示。

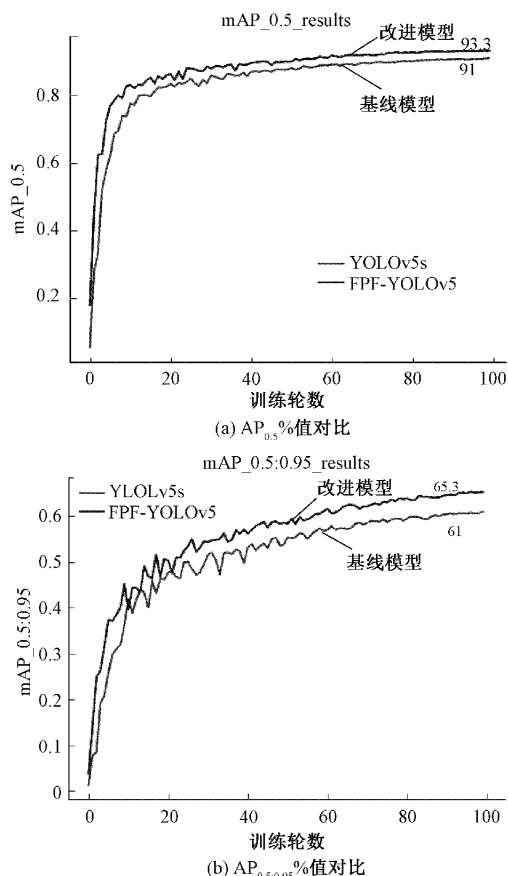
表 2 消融实验

模型	$AP_{0.5} \%$	$AP_{0.5,0.95} \%$	Params	M
1	91.0	61.0	7.02	13.7
2	92.0	62.5	2.03	5.1
3	92.5	63.5	2.19	5.5
4	92.4	63.6	2.20	5.5
5	92.8	63.8	2.18	5.1
6	92.9	64.0	2.20	5.2
7	93.2	65.3	2.45	5.7
8	92.8	63.9	2.20	5.2
9	93.3	65.3	2.46	5.7

由消融实验可以看出，在加入各个单模块后，均能获得比基线模型更好的性能。但是模型的性能并不随着模块的增加而线性提升。在数据集上改进的 YOLOv5 算法 $AP_{0.5} \%$ 提升了 2.3%， $AP_{0.5,0.95} \%$ 提升了 4.3%，证明所改进模型的有效性。

由模型 2 可以看出去掉主干中 32 倍下采样层以及颈部相应层结构对于小目标检测提升明显，同时参数量和模型大小相对于基线分别减少 71% 和 62%。在颈部网络添加 FPF 结构对性能贡献也较大，因为其增加了特征融合。从模型 4、5、6 可以看出，只有在去掉 P4 输出头后加入级联结构才能有效果，因为 P4 并不利于小目标检测反而影响 P3 中跨层级联的效果。最后由模型 7、8、9 可以看出，单独使用 MSAM 模块并不比与 SRFB 结合使用有效果，因为浅层特征图缺少感受野，经过 SRFB 增加感受野后再经过注意力结构能够更好的发挥其作用。

图 10 为改进网络与 YOLOv5s 的 $AP_{0.5} \%$ 、 $AP_{0.5,0.95} \%$ 对比曲线。由图可以看出改进模型的 $AP_{0.5} \%$ 、 $AP_{0.5,0.95} \%$ 值明显优于 YOLOv5s 模型。

图 10 YOLOv5s 与 $AP_{0.5}$ %、 $AP_{0.5:0.95}$ % 值对比

2.4 对比实验

为进一步探究 FPF-YOLOv5 算法的性能,使用相同小目标数据集进行验证,将本文算法与 YOLOv5m 对比,同时与其他主流算法 SSD、YOLOX、YOLOv3SPP 在检测精度、参数量和模型大小方面进行比较,如表 3 所示。

表 3 不同目标检测算法对比实验

算法	$AP_{0.5}$ %	$AP_{0.5:0.95}$ %	Params	M
SSD	22.29	—	—	—
YOLOX ^[27]	87	62.7	8.94	68.5
YOLOv3SPP	90.7	63.7	9.57	19.5
YOLOv5m	90.2	60	20.8	42.3
本文	93.3	65.3	2.46	5.7

从这几种算法中可以看出,SSD 作为比较早期的单阶段目标检测算法,相较于表中其他算法效果最差。YOLOx 算法是 YOLO 新推出的高性能目标检测算法,其主要在无锚框、标签分配策略、解耦头 3 个方面较 YOLOv5 有大创新,因此其检测精度也较高,但其在精度不如本算法同时模型大小远高于本算法,不满足轻量化要求。YOLOv5m 作为 YOLOv5s 的网络加深版本,可以看出在检测精度上并没有比 YOLOv5s 高,原因是因为加深

的网络宽度深度过高对小目标检测反而起到相反作用。总体上,本文所提出的改进算法不管在精度、参数量还是模型大小上均有明显优势。

2.5 与其他轻量化网络对比实验

为验证 FPF-YOLOv5 在轻量化方面的精度优势,将其与 GhostYOLOv5 (G-YOLO)、MobileYOLOv5 (M-YOLO)、YOLOv3tiny (YOLOv3-t) 以及 ShuffleYOLOv5 (S-YOLO) 等轻量化网络进行比较。如表 4 所示。

表 4 不同轻量化网络对比

算法	$AP_{0.5}$ %	$AP_{0.5:0.95}$ %	Params	M
G-YOLO	79.1	42.2	3.68	7.54
M-YOLO	84.2	51	3.52	7.15
S-YOLO	84.6	51	3.18	6.45
YOLOv3-t	63.8	29.3	2.17	4.23
本文	93.3	65.3	2.46	5.7

从几种轻量化算法的对比中可以看出 YOLOv3tiny 作为使用最广泛的 YOLO 系列算法,虽然其参数量和模型大小比本文提出的 FPF-YOLOv5 算法低,但其整体检测精度比本算法分别低 29.5% 和 36%。基于 YOLO 改进的几个轻量化网络 Ghost、Mobile 和 Shuffle,其检测精度、模型大小、参数量均差不多,相比于 YOLOv3tiny 有很大改善,但是比本文算法精度值仍有大约 9% 的差距,并且在参数量和模型大小方面也高于本算法。因此,本文提出的 FPF-YOLOv5 算法在对比各轻量化算法中最佳。

3 结 论

遥感车辆目标检测具有重要的研究意义,本文在 YOLOv5 基础上提出一种基于多重金字塔 (FPF-YOLOv5) 和 MSAM 模块的遥感车辆目标检测算法。其中,为减少参数量降低模型大小同时有利于小目标检测,将主干中 32 倍下采样层及其颈部相应结构去除;为加强颈部网络特征融合,设计多重金字塔结构,并通过级联方式,加强跨层特征融合更有利于目标检测;针对遥感车辆图像中背景复杂、目标小的问题,在颈部还创新提出 MSAM 结构,通过增强感受野和引入注意力机制,来提高感知力,增强目标检测效果;最后在输出层,将三层特征层输出改为两层特征层输出,在减少参数量的同时,能够使输出层更关注于小目标物体。实验表明,该算法能够有效提升对于遥感车辆检测的精度,在自建遥感车辆数据集上 $AP_{0.5}$ % 达到 93.3%, $AP_{0.5:0.75}$ % 达到 65.3%, 分别相较于 YOLOv5s 算法,提高了 2.3%、4.3%。

参考文献

- [1] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015:1440-1448.

- [2] 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, 2016, 39(6): 1137-1149.
- [3] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [4] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. IEEE Conference on Computer Vision & Pattern Recognition, IEEE, 2017:6517-6525.
- [5] REDMON J, FARHADI A. YOLOv3: An incremental improvement [EB/OL]. [2021-11-21]. <http://arxiv.org/abs/1804.02767>.
- [6] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: Optimal speed and accuracy of object detection EB/OL]. [2021-11-21]. <https://arxiv.org/abs/2004.10934>.
- [7] LIU W, ANGELOV D, ERHAN D, et al. SSD: Single shot multiBox detector [C]. Proceedings of European Conference on Computer Vision, 2016: 21-37.
- [8] JI H, GAO Z, MEI T, et al. Vehicle detection in remote sensing images leveraging on simultaneous super-resolution[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(4):676-680.
- [9] PENG Y, ZHAO X, TAO H, et al. Hand gesture recognition against complex background based on deep learning[J]. Robot, 2019, 41(4): 534-542.
- [10] 徐晓光,李海.多尺度特征在 YOLO 算法中的应用研究[J].电子测量与仪器学报,2021, 35,246(6):96-101.
- [11] HU J, LI S, SAMUEL A, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8):2011-2023.
- [12] 张宏群,班勇苗,郭玲玲,等.基于 YOLOv5 的遥感图像舰船的检测方法[J].电子测量技术,2021,38(8): 87-92.
- [13] 李仁鹰,钱慧芳,郭佳豪,等.基于 M-YOLOv4 模型的轻量化目标检测算法[J].国外电子测量技术,2022, 41,329(4):15-21.
- [14] 彭继慎,孙礼鑫,王凯,等.基于模型压缩的 ED-YOLO 电力巡检无人机避障目标检测算法[J].仪器仪表学报,2021,42(10):161-170.
- [15] GLENN J. YOLOv5 [EB/OL]. [2020-01-20] <https://github.com/ultralytics/YOLOv5>.
- [16] WANG C Y, LIAO H, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2020: 1571-1580.
- [17] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [18] LIN T Y, DOLLÁR, PIOTR, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:936-944.
- [19] WANG C, XU C, WANG C, et al. Perceptual adversarial networks for image-to-image transformation [J]. IEEE Trans. on Image Processing, 2018:4066-4079.
- [20] TAN M, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[C]. Proc. of the International Conference on Machine Learning, 2019.
- [21] LIU S, DI H, WANG Y. Receptive field block net for accurate and fast object detection[C]. Proc. of the European Conference on Computer Vision, 2018: 404-419.
- [22] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [J]. ArXiv Preprint, 2014, ArXiv:1409.4842.
- [23] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [24] PARK J, WOO S, LEE J Y, et al. Bam: Bottleneck attention module [J]. ArXiv Preprint, 2018, ArXiv: 1807.06514.
- [25] 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.
- [26] 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, 159:296-30.
- [27] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021 [J]. ArXiv Preprint, 2021, ArXiv:2107.08430.

作者简介

赵倩,博士,副教授,主要研究方向为图像检测与识别、视频图像处理。

E-mail:zhaoqian@shiep.edu.cn

杨一聪(通信作者),硕士研究生,主要研究方向为目标检测。

E-mail:yyc1232021@163.com