

轻量化的光学遥感影像目标检测方法

王浩^{1,2*}, 尹增山^{1,2}, 刘国华^{1,2}, 胡登辉¹, 高爽^{1,2}

¹中国科学院微小卫星创新研究院, 上海 201203;

²中国科学院大学, 北京 100049

摘要 针对深度学习目标检测算法由于模型体积过大、参数数量过多而在星上部署困难的问题,在YOLOv5检测模型的基础上,提出了一种轻量化的光学遥感影像目标检测算法——LW-YOLO。首先,针对YOLOv5网络中的特征信息冗余所带来的计算开销,引入轻量化Ghost模块替换网络中的普通卷积以减少参数数量;接着,设计了一种空间和通道融合的注意力模块Fusion Attention(FA),并在此基础上重构了网络的瓶颈层FABottleneck,进一步减少参数数量,并提高算法对于光学遥感影像目标的定位能力;最后,提出了一种稀疏参数自适应的网络剪枝方法对网络进行剪枝,进一步压缩模型大小。在DOTA数据集上的实验表明,LW-YOLO算法相比于YOLOv5s参数数量下降了64.7%,模型大小下降了62.7%,推理时间降低了3.7%,同时平均精度均值仅仅下降了6.4%。该算法以较小的精度损失为代价实现了网络模型的轻量化,为星上光学影像在轨目标检测提供了理论基础。

关键词 图像处理; 光学遥感影像; YOLOv5; 轻量化模型; 注意力机制; FABottleneck; 网络剪枝

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP202259.2210004

Lightweight Object Detection Method for Optical Remote Sensing Image

Wang Hao^{1,2*}, Yin Zengshan^{1,2}, Liu Guohua^{1,2}, Hu Denghui¹, Gao Shuang^{1,2}

¹Innovation Academy for Microsatellite, Chinese Academy of Sciences, Shanghai 201203, China;

²University of Chinese Academy of Sciences, Beijing 100049, China

Abstract In this paper, a lightweight optical remote sensing image target detection algorithm LW-YOLO is proposed based on the YOLOv5 detection model to solve the difficulty of deploying the deep learning target detection algorithm on the satellite due to the large volume of the model and too many parameters. First, a lightweight Ghost module is introduced to replace the ordinary convolution in the network to reduce the number of parameters and solve the computational overhead caused by feature information redundancy in the YOLOv5 network. Then, a space and channel Fusion Attention (FA) module is designed, and the bottleneck layer FABottleneck of the network is reconstructed to further reduce the parameters and improve the positioning ability of the algorithm for optical remote sensing image targets. Finally, a sparse parameter adaptive network pruning method is proposed to prune the network and further compress the model size. Experiments on the DOTA dataset show that compared with YOLOv5s, the LW-YOLO algorithm reduces 64.7% of parameters, 62.7% of model size, 3.7% of reasoning time, and only 6.4% of mean precision. The algorithm achieves the lightweight of the network model at the cost of small accuracy loss and provides a theoretical basis for on-orbit target detection in spaceborne optical images.

Key words image processing; optical remote sensing image; YOLOv5; lightweight model; attention mechanism; FABottleneck; network pruning

1 引言

光学遥感影像目标检测一直以来是人们关注的热

点,被广泛应用于渔业海事、交通监管、应急减灾、智慧城市等多个领域。传统的光学遥感影像目标检测方法主要包括基于特征判别^[1]的方法、基于Hough投票^[2]

收稿日期: 2021-07-29; 修回日期: 2021-08-30; 录用日期: 2021-09-28

基金项目: 中国科学院国防科技创新重点部署项目(KGFZD-135-20-03)、上海人才发展资金(2020014)、科技部国家重点研发计划(2016YFB0501100)

通信作者: *xidianwhgood@163.com

的方法、基于模板匹配^[3]的方法等。这些传统的方法主要通过人工设计的特征来对图像中的目标进行检测,不能完整地表达目标的抽象语义特征,因此存在泛化性能差、鲁棒性较低的缺点,难以快速、精准地识别目标。

随着深度学习算法在计算机视觉领域的广泛应用,基于卷积神经网络(CNN)^[4]的目标检测算法获得了极大的发展。典型的基于CNN的目标检测算法大致可以分为双阶段(two-stage)目标检测和单阶段(one-stage)目标检测^[5]两种。双阶段目标检测算法的代表是R-CNN系列算法,这类算法通过“建议区域”加“目标分类与边框回归”这样双阶段的方式来进行目标检测,例如R-CNN^[6]、Fast-RCNN^[7]及后续的Faster-RCNN^[8]、Cascade-RCNN^[9]等目标检测模型。虽然以上双阶段检测算法模型在检测精度上相比于传统方法取得了极大的提升,但仍然存在检测速度较慢的缺点。另一大目标检测算法是以YOLO^[10-13]、SSD^[14]、RetinaNet^[15]等为代表的单阶段目标检测算法,其中的YOLO算法通过在原始图像中划分网格,预设初始锚框的方式直接回归出目标类别信息与目标边界信息,虽然在检测精度上略差于双阶段目标检测算法,但极大提高了模型的检测速度。

随着CNN的不断发展,网络层数不断加深,类似CSPDarkNet53^[16]、ResNet101^[17]等大型骨干网络被提出,这极大地提升了网络的特征提取能力,但随之而来的问题是深度学习模型参数量与模型体积不断增加,所需的计算资源也不断增加,这让深度学习模型在边缘设备部署时变得非常困难。特别是在卫星等移动设备上,有限的计算资源、存储能力和功耗的限制^[18]进一步阻碍了模型体积较大的目标检测算法的运行,从而无法进行精准的在轨目标检测。而复杂在轨任务往往要求星上处理系统能够支持信息共享与协同操作,提高卫星系统利用率^[19],同时随着遥感图像分辨率的不断提高,不断增大的数据量导致星地之间数据传输时间大大增加,传统的在地面进行目标检测与跟踪处理的模式面临时效性低的问题^[20],无法满足实际的在轨任务需求。

为了解决在卫星等移动设备上部署目标检测模型的难点,学者们针对网络轻量化提出了一系列解决方案。一种思路是通过优化卷积模块,调整局部网络结构,在尽可能不影响性能的情况下实现网络参数量与模型大小的下降。Iandola等^[21]在SqueezeNet中提出了fire模块,在Squeeze层用 1×1 卷积减小通道数、在Expand层使用 1×1 卷积和 3×3 卷积将输入进行合并,通过这些操作减小参数量,缩减模型大小。针对过多的卷积层带来的巨大参数量,MobileNet^[22]使用深度可分离卷积代替普通卷积,大大减少了卷积层的参数量。MobileNetV2^[23]引入了线性瓶颈结构和反向残差结构优化网络,使得网络模型的体积进一步减小。

MobileNetV3^[24]则使用神经架构搜索(NAS)生成最优网络结构。ShuffleNet^[25]对分组卷积后的特征图进行通道重组,减小参数量的同时保证了特征图中各通道之间的信息交流。ShuffleNetV2^[26]针对内存使用量对网络结构进一步优化,提升模型运行速度。GhostNet^[27]使用较少的卷积核生成同样大小的特征图,保证网络精度的同时大大减少了参数量。除了设计轻量化网络,另一种思路是通过网络剪枝^[28]、知识蒸馏^[29]、低秩分解^[30]等算法对模型进行整体优化,压缩模型体积,降低计算开销。

针对星上计算资源、功耗、内存等受限的现状^[31-33],本文基于YOLOv5目标检测模型提出了一种轻量化的光学遥感影像目标检测算法——LW-YOLO。首先引入轻量化卷积Ghost替换YOLOv5s网络中的普通卷积单元,减少特征冗余,降低参数量;接着针对光学遥感影像中存在小目标众多难以识别与定位的问题,提出了一种融合空间注意力与通道注意力的Fusion Attention(FA)模块,该模块可以有效提高网络对于小目标的空间定位能力。以结合FA模块与轻量化卷积设计的FABottleneck重构网络的瓶颈层,进一步降低模型大小与参数量。对重构后的模型进行稀疏参数自适应的网络剪枝以减少模型层数,提高前向推理速度,从而整体上提高光学遥感影像在轨目标检测的处理效率。

2 YOLOv5目标检测算法

YOLOv5目标检测模型,相比于SSD、YOLOv3、YOLOv4等其他的单阶段目标检测算法,收敛速度更快,网络结构更加灵活,可以依据检测速度、模型大小、检测精度等实际需要选择不同版本。YOLOv5共有s,m,l,x等4个版本,网络深度与模型大小依次增加,可通过控制配置文件中的depth_multiple和width_multiple参数进行调整,其中depth_multiple参数通过控制C3模块的数量来控制网络的深度,width_multiple参数通过控制卷积核个数来控制网络的宽度。图1是简单的YOLOv5s模型网络结构示意图,表1展示了不同版本YOLOv5各模块的数量。

由图1可知,YOLOv5与一般的单阶段目标检测模型一样,具有backbone、neck、head等结构,其中backbone负责提取图像特征,neck进行特征融合,head作为检测头负责预测输出。

通过backbone进行特征提取及不同尺度特征的融合,可以输出3个不同尺度的特征,分别针对小、中、大等3类目标进行预测,其中小尺度特征图预测大目标,中尺寸特征图预测中等目标,大尺度特征图预测小目标。输入图像尺寸为 1024×1024 ,下采样步长stride分别为8、16、32,3个尺寸的特征图分别为 128×128 、 64×64 、 32×32 。以 32×32 的特征图为例,在预测时输入图像被划分为 32×32 个网格,每个网格对应输出

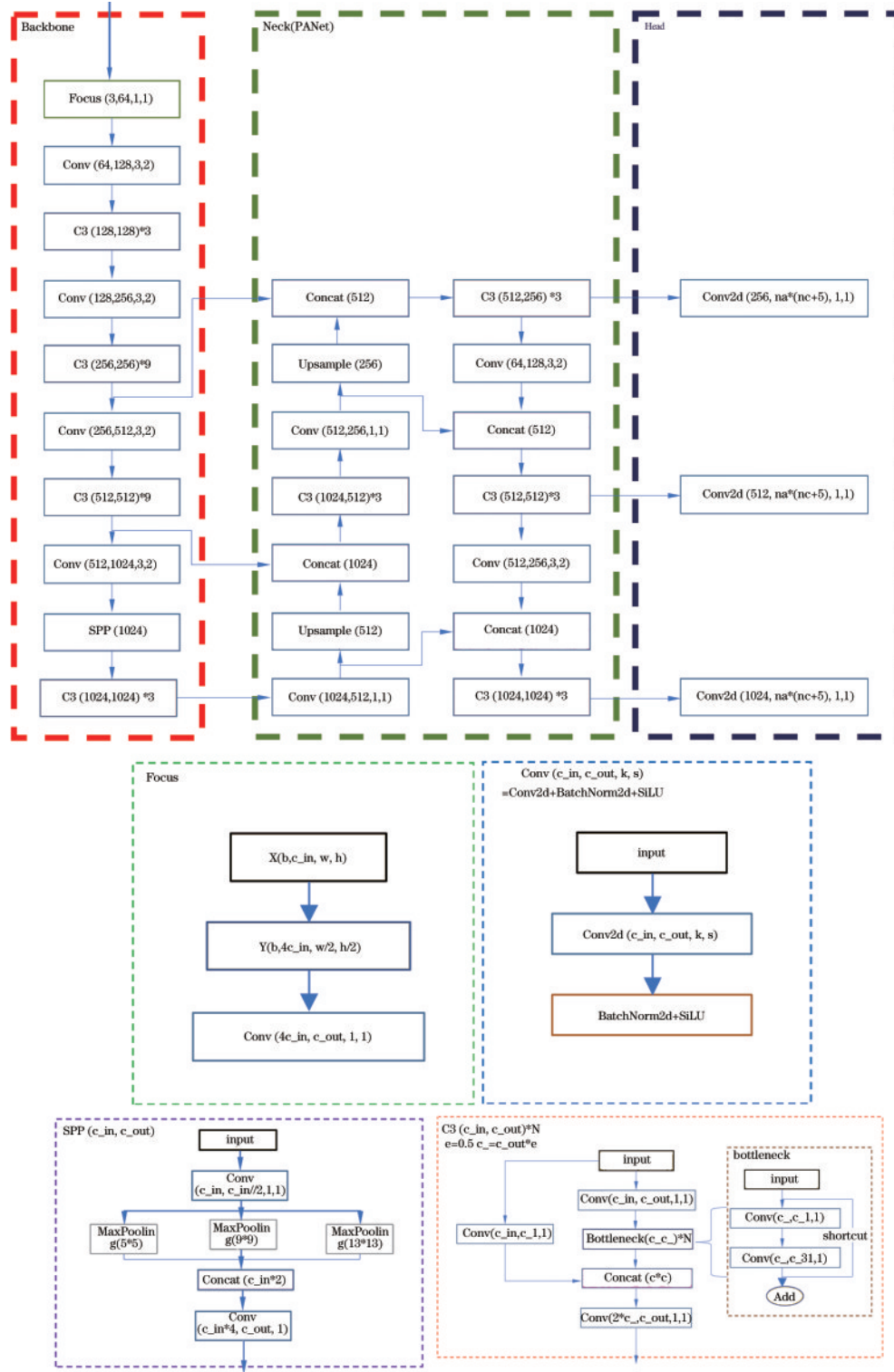


图 1 YOLOv5s 网络结构图

Fig. 1 YOLOv5s network structure

特征图张量 $32 \times 32 \times [n_a \times (4 + 1 + n_c)]$ 中的一个子张量 $1 \times 1 \times [n_a \times (4 + 1 + n_c)]$, 其中 n_a 代表每个网格预测边界框的数量 (默认为 3), n_c 代表所需要预测的类别数, 4 代表边界框位置信息的预测值 t_x, t_y, t_h, t_w , 1 表示每个边界框的置信度 (confidence)。置信度的公式为

$$V_{\text{confidence}} = P_r(\text{Object}) \times R_{\text{IOU}}^{\text{truth}_{\text{pred}}}, \quad (1)$$

式中: $P_r(\text{Object})$ 代表该网格中包含目标的可能性; $R_{\text{IOU}}^{\text{truth}_{\text{pred}}}$ 代表预测框与真实框的交并比。与之前的 YOLO 系列算法不同, YOLOv5 并没有采用 max IOU 的 anchor 匹配机制 (即选择 IOU 最大的 anchor 微调作为网络的输出预测框), 而是在目标落入的网格基础

表 1 不同版本 YOLOv5 模块数量

Table 1 Number of YOLOv5 modules of different versions

Module	YOLOv5s	YOLOv5m	YOLOv5l	YOLOv5x
depth_multiple	0.33	0.67	1.0	1.33
width_multiple	0.50	0.75	1.0	1.25
Number of C3 in backbone	1,3,3	2,6,6	3,9,9	4,12,12
Number of C3 in neck	1	2	3	4
Number of Conv	32,64,128,256,512	48,96,192,384,768	64,128,256,512,1024	80,160,320,640,1280

上,利用最近邻的原则再找出两个网格,并认为这 3 个网格都负责预测该边界框。这样的设计引入了更多的高质量正样本,加速了网络的收敛,如图 2 所示,当边界框中心落入网格左下方,则认为该网格与其左边的网格及下方的网格都负责预测该边界框。

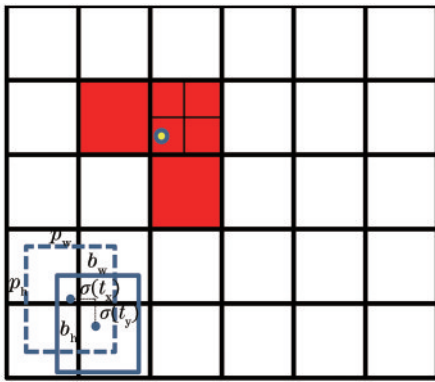


图 2 YOLOv5 预测过程

Fig. 2 Prediction process of YOLOv5

YOLOv5 网络预测的位置信息 t_x, t_y, t_h, t_w 需要进一步解码才可以得到预测框在图片上的位置。图 2 中,先验锚框为虚线框,边界框为实线框, p_w 和 p_h 为先验锚框的宽和高, b_w 和 b_h 为预测得到的边界框的宽和高, b_x 和 b_y 表示边界框相对于网格点左上角的偏移量,其解码过程可表示为

$$\begin{cases} b_x = 2 \times \sigma(t_x) - 0.5 \\ b_y = 2 \times \sigma(t_y) - 0.5 \\ b_w = 4 \times \sigma(t_w)^2 \times p_w \\ b_h = 4 \times \sigma(t_h)^2 \times p_h \end{cases} \quad (2)$$

3 基于 YOLOv5 改进的轻量化 LW-YOLO 算法

3.1 基于 GhostNet 的轻量化卷积

当下基于深度学习的目标检测算法往往使用大量的卷积核对图像进行特征提取,随着网络层数的不断加深与扩展,卷积核的数目也随之增加,这不仅带来了大量的参数与计算量,不利于网络的轻量化与边缘部署,同时也存在着特征信息冗余的现象^[27]。为了解决大量卷积核带来的特征冗余问题,本实验组在 YOLOv5 中引入了轻量化的 Ghost 模块来替换主干特征提取网络中的普通卷积层。Ghost 模块通过重新构建卷积过程,使用线性操作来生成一部分的特征图,可以有效减少模型参数与计算量。

图 3 是 Ghost 模块的示意图,其中 identity 是恒等映射, Φ 为逐通道卷积。从图中可以看到,针对原本的特征图,首先使用原输出特征图通道数量 $\frac{1}{s}$ 的卷积核 s 对输入特征图进行卷积生成较小的特征图,对中间特征图中的每一个通道进行逐通道卷积生成另外 $(1 - \frac{1}{s})$ 的通道,并与之前的特征图进行连接,以获得与原输出特征图尺寸相同大小的特征图。对于一个大小为 $c \times h \times w$ 的特征图,使用卷积核大小为 $k \times k$ 的普通卷积,输出大小为 $c' \times h' \times w'$ 的特征图,其计算量为

$$c' \times h' \times w' \times c \times k \times k, \quad (3)$$

而采用 Ghost 模块输出特征图,逐通道卷积核大小为 $d \times d$,其计算量为

$$\frac{c'}{s} \times h' \times w' \times c \times k \times k + (s-1) \times \frac{c'}{s} \times h' \times w' \times d \times d, \quad (4)$$

其计算量之比可表示为

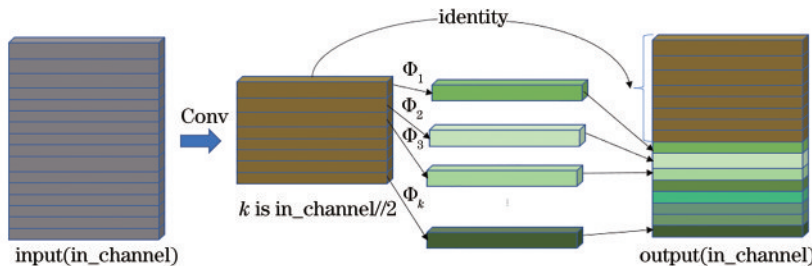


图 3 Ghost 模块示意图

Fig. 3 Schematic diagram of Ghost module

$$r = \frac{c' \times h' \times w' \times c \times k \times k}{\frac{c'}{s} \times h' \times w' \times c \times k \times k + (s-1) \times \frac{c'}{s} \times h' \times w' \times d \times d} = \frac{c \times k \times k}{\frac{1}{s} \times c \times k \times k + \frac{s-1}{s} \times d \times d} \quad (5)$$

逐通道卷积核大小 d 一般与 k 相等, 而 c 远大于 1, 所以

$$r \approx \frac{s \times c}{s + c + 1} \approx s \quad (6)$$

由此可见, 使用 Ghost 卷积与普通卷积相比, 计算量大约下降为原来的 $\frac{1}{s}$ 。

3.2 FABottleneck 设计

光学遥感影像与自然图像中的目标存在很大差异, 其目标尺度变化大、背景复杂、小目标众多, 因此对于目标检测算法的识别能力、空间定位能力等提出了更高的要求。为了解决这个问题, 本实验组参考坐标注意力^[34]提出了一种轻量化的注意力模块 FA, FA 模块可以有效融合特征提取过程中的空间信息与通道信息, 加强对小

目标的特征提取能力与定位能力。FA 模块的具体结构如图 4 所示: 对输入特征图进行水平和垂直方向的一维全局池化, 以此沿着一个空间方向捕获远程依赖关系, 同时保存另一个方向上的空间位置信息; 接着将具有不同方向空间位置信息的一对注意力图沿通道维进行堆叠与压缩得到融合水平方向与垂直方向信息的空间注意力图。将上述分别代表水平、垂直、融合方向的 3 个注意力图进行拼接并进行类似 Squeeze-and-Excitation (SE) 注意力的通道重组以捕获不同通道之间的依赖关系, 增强通道注意力。对处理后的特征图进行分割, 输出 3 个融合注意力图。这 3 个注意力图分别具有水平、垂直与全局空间注意力, 并且也对原特征图不同通道进行了建模, 因此融合了空间注意力与通道注意力, 可以有效增强感兴趣区域的目标表示, 提高定位能力。

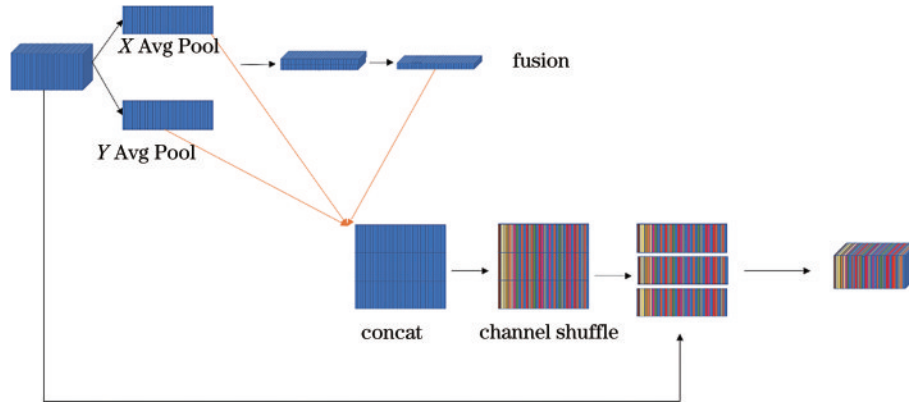


图 4 FA 模块示意图

Fig. 4 Schematic diagram of FA module

为了进一步减少模型大小与参数量, 同时一定程度上提高网络对于小目标的特征提取能力, 结合上述融合注意力机制, 设计了新的瓶颈层结构 FABottleneck, 其示意图如图 5 所示。

参考 ResNet 的设计思路, 将一条残差边与经过两次 Ghost 卷积的输入特征相加, 获得输出, 其中第 1 个 Ghost 模块负责提升特征维度, 后一个 Ghost 模块负责对齐 shortcut 的维度。图 5 为步长为 2 的 Bottleneck, 在两次 Ghost 模块之间使用一个 3×3 的深度可分离卷积层 (DWConv) 进行下采样, 而当步长为 1 时则取消该深度可分离卷积层直接进行第 2 个 Ghost 模块的堆叠。在两个 Ghost 卷积之间, 引入上述 FA 模块对中间特征图进行重构以提高网络对于小目标的定位能力。相比于原网络中的瓶颈结构, 新设计的 FABottleneck 由于引入了 Ghost 模块与深度可分离卷积而大大减少了参数量, 同时 FA 模块可以有效加强网络对于小目标的特征提取能力与定位能力。

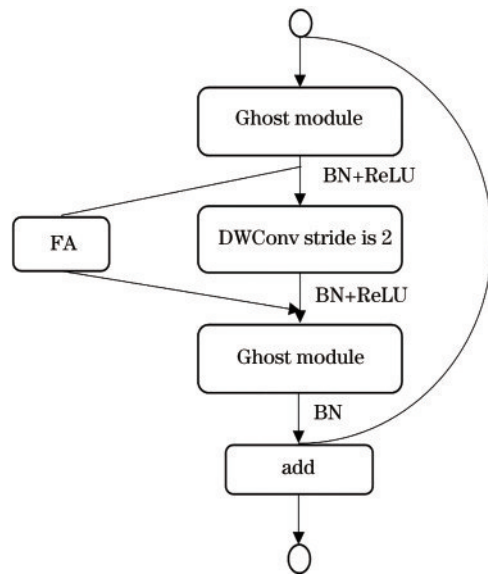


图 5 FABottleneck 设计

Fig. 5 FABottleneck design

3.3 稀疏参数自适应的网络剪枝

批标准化(BN)^[35]是现代CNN架构中广泛采用的一种数据处理方法,用以对输入数据进行归一化,加速网络收敛,提升泛化性能。

基于BN层的通道剪枝算法^[36]将BN层的 γ 参数作为网络剪枝的缩放因子来衡量每个通道的重要性,将 γ 与每个通道的输出相乘,接着联合训练原网络的网络权重与 γ 缩放因子,对 γ 进行统计并排序,设定一个全局阈值,将小于阈值的通道直接移除。目标函数为

$$L = \sum_{(x,y)} l[f(x, W), y] + \lambda \sum_{\gamma \in \Gamma} g(\gamma), \quad (7)$$

式中:第1项是YOLOv5模型预测所产生的损失;第2项为对缩放因子 γ 的约束项,一般使用L1正则化; λ 稀疏参数为平衡两项的超参数。

稀疏参数 λ 控制着神经网络的稀疏程度:较大的 λ 会使网络获得较大程度的稀疏,但与此同时也会使

精度损失得较多;较小的 λ 稀疏程度较低,模型压缩不够充分,但精度不会过多下降。文献^[36]采用搜索以获得最佳单一 λ ,本实验组提出了一种稀疏参数自适应的方法,在稀疏训练的初始阶段采用较大的 λ ,随着训练次数的增长,当训练次数超出总训练轮次的1/2后,让稀疏参数 λ 自适应衰减,从而在全局训练中获得较好的稀疏效果。设置初始稀疏参数 λ_0 为0.001,训练总轮次为 N_{epochs} ,训练中稀疏参数 λ 与训练次数 n 的关系为

$$\lambda = \begin{cases} \lambda_0, & n \leq \frac{1}{2} N_{\text{epochs}} \\ \frac{1}{2} \lambda_0, & n > \frac{1}{2} N_{\text{epochs}} \end{cases} \quad (8)$$

在完成稀疏训练后,需要对每个通道的BN参数 γ 进行统计并排序,设置剪枝阈值,去除小于剪枝阈值的通道。对剪枝后的模型进行微调训练以获得最终的模型。根据剪枝后的模型大小、参数量等的具体情况,上述步骤可重复进行。具体流程如图6所示。

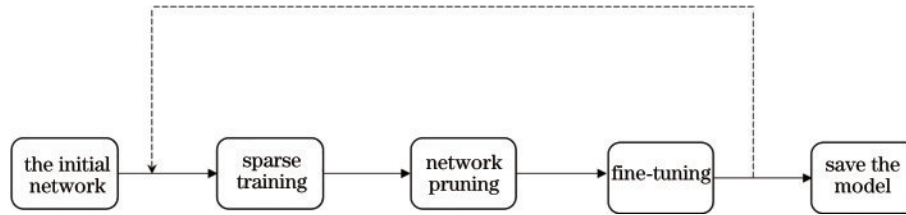


图6 稀疏参数自适应的通道剪枝算法流程图

Fig. 6 Flow chart of sparse parameter adaptive channel pruning algorithm

4 实验结果与分析

4.1 实验数据集与实验设置

在DOTA数据集^[37]上进行网络的训练与测试。DOTA数据集是武汉大学在2018年正式提出的一个大型光学遥感影像目标检测基准数据集。该数据集共包含2806幅遥感图像,来自Google Earth、GF-2卫星和JL-1卫星。实验使用DOTA数据集v1.5,每张图像尺寸在 $800 \times 800 \sim 4000 \times 4000$ 之间,其中包含不同方向与不同尺度的共40万个带注释的实例对象,分为16个类别,包括飞机、船舶、储油罐、棒球场、网球场、篮球场、田径场、港口、桥梁、大型车辆、小型车辆、直升机、环形交叉路口、足球场、游泳池和集装箱,每个实例均由四边形边界框进行标注。图7展示了DOTA数据集的样例图片。

由于图片尺寸过大,难以直接送入网络进行训练。参考YOLT^[38]算法,将DOTA数据集中的图片先进行分割,分割为大小为 1024×1024 的图像。为了避免被分割区域有目标被截断,设置中间overlap重叠区域大小为512 pixel。图8展示了分割前后的图像。

实验环境及设置如下:在Ubuntu18.04系统下训练,使用Pytorch框架搭建网络模型;CPU为Intel(R) Core(TM) i9-9900K CPU,显卡为NVIDIA 2080ti。

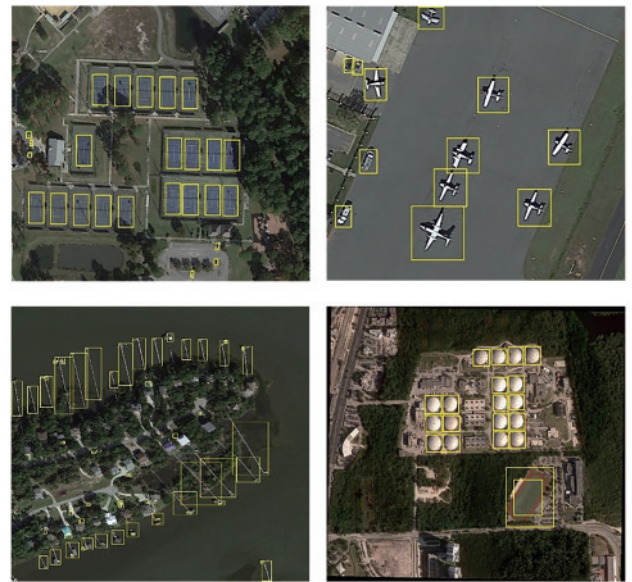


图7 DOTA数据集样例

Fig.7 Samples in DOTA dataset

训练过程中输入图像尺寸为 1024×1024 ,训练次数设为100,batch-size设为8,采用了余弦退火学习率衰减,初始学习率为0.01。

4.2 评价标准

目标检测问题需要综合考虑准确率(P)与召回率



图 8 分割前后的图像。(a)分割前的图像;(b)分割后的图像
Fig. 8 Image before and after segmentation. (a) Image before segmentation; (b) image after segmentation

(R),其表达式分别为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (9)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (10)$$

式中: N_{TP} 为真正例的数量; N_{FP} 为假正例的数量; N_{FN} 为假负例的数量; $N_{TP} + N_{FP}$ 为全部被分为正样本的数量; $N_{TP} + N_{FN}$ 为全部正样本的数量。在此基础上引入平均精确率(AP),可以综合衡量精确率与召回率:

$$P_{AP} = \int_0^1 P(R) dR. \quad (11)$$

表 2 YOLOv5s 与 Ghost-YOLOv5s 对比

Table 2 Comparison between YOLOv5s and Ghost-YOLOv5s

Model	Parameters /MB	Model size /MB	mAP	Precision	Recall
YOLOv5s	7.23	14.2	0.692	0.93	0.82
Ghost-YOLOv5s	5.9	11.6	0.676	0.929	0.82

表 3 YOLOv5s 与 Ghost-YOLOv5s 中不同目标种类的 AP 值

Table 3 AP for different types of targets in YOLOv5s and Ghost-YOLOv5s

Type of targets	AP for different types of targets (YOLOv5s)	AP for different types of targets (Ghost-YOLOv5s)
plane	0.942	0.939
baseball_diamond	0.791	0.795
bridge	0.543	0.518
ground_track_field	0.653	0.589
small_vehicle	0.600	0.586
large_vehicle	0.806	0.799
ship	0.878	0.871
tennis_court	0.943	0.926
basketball_court	0.683	0.657
storage_tank	0.795	0.774
soccer_ball_field	0.547	0.570
roundabout	0.719	0.691
harbor	0.819	0.825
swimming_pool	0.711	0.718
helicopter	0.638	0.532
container_crane	0.012	0.019
All classes mAP	0.692	0.676

对于多目标检测问题,需要衡量算法对多个种类目标的检测能力,因此需要对多个目标的 AP 求均值来求得 mAP:

$$P_{mAP} = \frac{\sum_{n=0}^N P_{AP,n}}{N}. \quad (12)$$

4.3 实验与分析

4.3.1 Ghost 模块分析

为了分析 Ghost 模块重构 YOLOv5s 特征提取网络带来的效果,将重构后的模型称为 Ghost-YOLOv5s。需要注意的是,此时的 Bottleneck 中并未使用 FABottleneck,而仍然保持原网络的瓶颈结构。表 2 展示了 YOLOv5s 与 Ghost-YOLOv5s 模型的对比。相比于 YOLOv5s, Ghost-YOLOv5s 的模型大小下降了 18.3%,参数量下降了 18.4%,mAP 由原来的 0.692 下降到了 0.676。模型的召回率并没有下降但准确率下降了 0.1 个百分点,导致 mAP 下降了约 1.6 个百分点。可能是相比于普通的卷积,Ghost 模块直接舍弃了一部分的冗余特征而用简单线性操作(逐通道卷积)来生成该部分特征,在一定程度上削弱了特征提取的能力,造成准确率有小幅下降,从而降低了 Ghost-YOLOv5s 模型的性能,这一点从各个种类目标对象的 AP 值(表 3)也能够看出来,Ghost-YOLOv5s 模型对几乎所有目标的 AP 值相比于 YOLOv5s 均有小幅下降。

3.3.2 FABottleneck

为了分析所提 FABottleneck 的效果,使用 FABottleneck 替换网络中的 Bottleneck,称为 FA-YOLO。表 4 展示了使用 FABottleneck 前后网络模型大小、参数量及 mAP 对比。由于 Ghost 模块和

DWConv 的引入,模型的参数量与大小进一步下降,mAP 的损失仅为 0.3 个百分点,并且由于 FA 注意力模块的引入,FA-YOLO 的准确率相比 Ghost-YOLOv5s 提高了约 1.8 个百分点,甚至优于 YOLOv5s 的准确率 0.93。

表 4 Ghost-YOLOv5s 与 FA-YOLO 对比

Table 4 Comparison between Ghost-YOLOv5s and FA-YOLO

Model	Parameters /MB	Model size /MB	mAP	Precision	Recall
Ghost-YOLOv5s	5.9	11.6	0.676	0.929	0.82
FA-YOLO	3.75	7.64	0.673	0.947	0.79

为了进一步验证所提 FA 模块的效果,对插入 FA 模块前后的网络进行了对比分析,结果如表 5 所示。加入 FA 模块后,网络的参数量仅仅增加了 0.02 MB,

模型大小增加了 0.1 MB,但是模型的召回率提升了大约 1 个百分点,准确率提高了 0.7 个百分点,mAP 提升了 1.1 个百分点。

表 5 FA 模块添加前后对比

Table 5 Comparison before and after adding FA module

Model	Parameters /MB	Model size /MB	mAP	Precision	Recall
FA-YOLO (without FA)	3.73	7.54	0.662	0.94	0.78
FA-YOLO (with FA)	3.75	7.64	0.673	0.947	0.79

图 9 展示了添加 FA 模块前后对同一幅遥感影像进行检测的效果图。加入 FA 模块后,FA-YOLO 可以检测出未添加之前无法检测出的目标。在 [图 9(a)] 中,FA-YOLO(with FA)检测出了图像左

上角的大型车辆,[图 9(b)] 中,FA-YOLO(with FA)检测出了更多的储油罐与小型车辆。这表明 FA 模块有效提升了网络对于小目标的特征提取能力与识别能力。

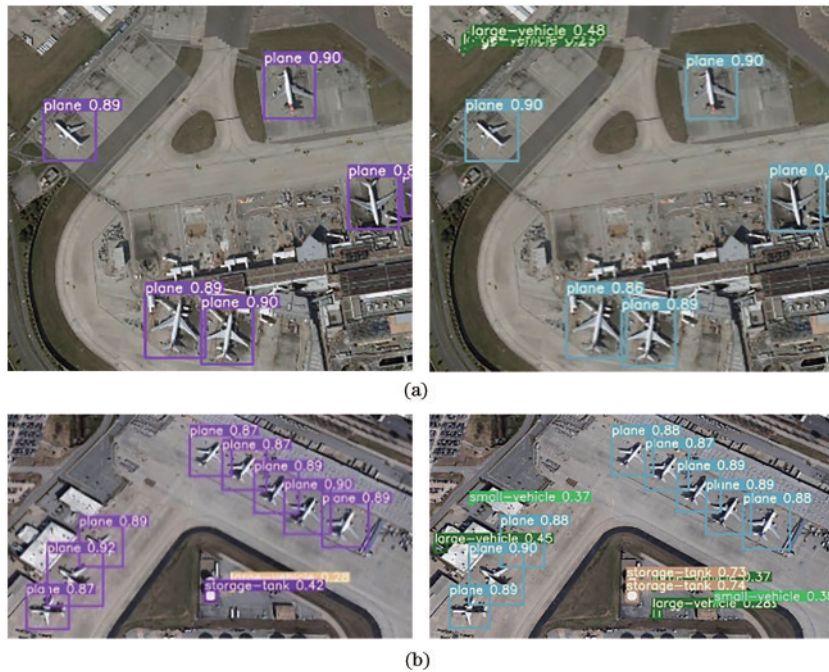


图 9 添加 FA 模块前后检测效果对比。(a)添加 FA 模块前;(b)添加 FA 模块后

Fig. 9 Comparison of detection effect before and after adding FA module. (a) Before adding FA module; (b) after adding FA module

3.3.3 网络剪枝效果分析

为了追求进一步的网络压缩从而更好地满足星上在轨目标检测需求,对 FA-YOLO 进行网络剪枝,采用所提稀疏参数自适应稀疏训练,图 10 展示了稀

疏训练后的 BN 层 γ 参数分布。从图中可以看出,经过稀疏训练后,存在大量 γ 参数较小的通道。设定剪枝阈值对网络进行剪枝,图 11 展示了剪枝后的 γ 参数分布。

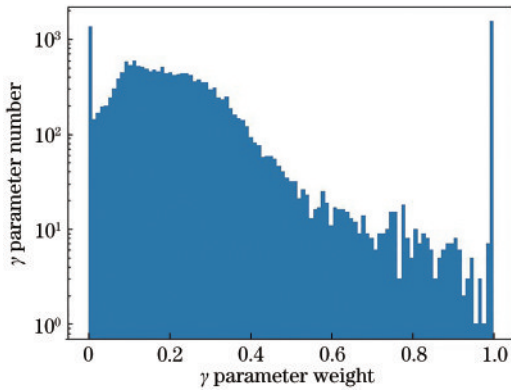


图 10 剪枝前 γ 参数分布

Fig. 10 Gamma parameter distribution before pruning

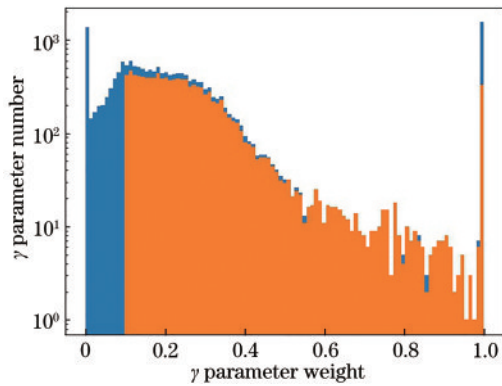


图 11 剪枝后 γ 参数分布

Fig. 11 Gamma parameter distribution after pruning

为了分析不同剪枝阈值对剪枝后算法的影响,在稀疏参数与模型微调次数相同的情况下设置了不同的剪枝阈值进行实验。表 6 展示了不同剪枝阈值下微调网络得到的最终模型大小及参数。从表中可以看出,经过稀疏训练后,存在大量 γ 参数权重较小的通道,这些通道的去除对于网络性能的影响可以忽略不计。例如剪枝阈值为 0.05、0.01、0.15 时,虽然随着剪枝阈

表 6 不同剪枝阈值下网络模型对比

Table 6 Comparison of network models under different pruning thresholds

mAP	Number of finetune	Pruning threshold	Parameters / MB	Model size / MB
0.647	20	0.05	3.43	6.99
0.646	20	0.1	3.18	6.49
0.648	20	0.15	2.55	5.29
0.577	20	0.2	1.41	3.11
0.424	20	0.25	0.85	2.05

表 7 LW-YOLO 与 YOLOv5s 对比

Table 7 Comparison between Ghost-YOLOv5s and LW-YOLO

Model	Precision	Recall	mAP	Parameters / MB	Model size / MB	Mean inference time / s
YOLOv5s	0.93	0.82	0.692	7.23	14.2	5.4
LW-YOLO	0.943	0.79	0.648	2.55	5.29	5.2

值的增加剪除了更多的通道,但网络的性能并没有因此下降,经过 20 轮的微调 mAP 分别为 0.647、0.61、0.648。但随着剪枝阈值进一步提升,模型的精度急剧下降。这主要是因为此时剪除的通道中已经存在相当一部分权重较大的通道,这部分通道对于模型特征提取能力、识别与分类能力等都起着重要的作用,因此模型的 mAP 会大幅下降。具体应用时,可以根据精度与算力需求进行权衡,选择适合的网络模型进行后续的部署。

4.3.4 整体分析

综合检测性能与压缩情况,选取剪枝阈值为 0.15,微调剪枝后的模型称为 LW-YOLO,其与 YOLOv5s 模型的对比如表 7 所示。表 8 展示了 LW-YOLO 算法在 DOTA 数据集上的 mAP 及各个类别目标的 AP 值,表 9 展示了 DOTA 数据集训练集中不同种类目标的数量分布。从表 8 和表 9 可以看到,LW-YOLO 对于飞机、舰船等目标的 AP 值仍然 0.85 以上,保持着不错的检测性能。原因在于由于训练集中目标实例对象数量较多,因此网络学到了不同背景下更加丰富的特征。同时,由于 DOTA 数据集中的光学遥感影像关注的区域大都是机场、港口、沿海城市等,这些区域背景下的目标数量较多,目标数量的差异集中表现为区域范围、不同背景上识别精度的差异,因此对于所提目标检测算法及其他基于深度学习的目标检测算法,识别精度往往表现出区域上的差异。在目标数量较多的区域提取到的特征更为丰富,对这些背景下的目标识别检测精度较高,但是对于一些复杂背景、内陆城市等样本数量较少的区域,网络学习到的特征往往不够丰富,因此在检测时会出现精度的下降。数据集中目标数量更多的小型车辆、大型车辆等的识别精度略低,这主要是因为高分辨率光学遥感影像中,车辆属于极小目标,所占像素区域少,网络难以提取到更多的有效特征。造成其他一些目标例如田径场、集装箱精度较低的原因一方面是训练集中标注样本数量较少,另一方面是 Ghost 模块及网络剪枝对网络性能所带来的客观影响,对于这些目标可能没有学习到足够的特征,这在一定程度上影响了模型的泛化性能。但考虑到网络的参数量相比于 YOLOv5s 下降了 64.7%,模型大小下降了 62.7%,仅为 5.29 MB,平均单张图片推理时间也降低了 0.2 ms,表明所提算法可提高星上光学遥感影像在轨目标检测的数据处理效率,为深度学习模型在星上部署进行在轨实时的目标检测提供了理论上的基础。

表 8 LW-YOLO 网络对不同目标种类的 AP 与 mAP

Table 8 LW-YOLO network AP and mAP for different types of targets

Type of targets	AP for different types of targets
plane	0.916
baseball_diamond	0.712
bridge	0.473
ground_track_field	0.514
small_vehicle	0.570
large_vehicle	0.787
ship	0.852
tennis_court	0.933
basketball_court	0.639
storage_tank	0.738
soccer_ball_field	0.470
roundabout	0.615
harbor	0.774
swimming_pool	0.681
helicopter	0.587
container_crane	0.110
All classes mAP	0.648

表 9 训练集中各类别目标数量统计

Table 9 Statistics of number of targets in each category in training set

Type of targets	Number of targets
plane	8072
baseball_diamond	412
bridge	2075
ground_track_field	331
small_vehicle	126501
large_vehicle	22218
ship	32973
tennis_court	2425
basketball_court	529
storage_tank	5346
soccer_ball_field	338
roundabout	437
harbor	6016
swimming_pool	2181
helicopter	635
container_crane	142

4 结 论

针对星上存储空间有限、计算资源不足的问题,提出了一种轻量化的光学遥感影像目标检测算法。首先使用 Ghost 模块替换普通卷积,接着提出了一种融合空间与通道的注意力模块 FA 并基于此设计了网络中

的瓶颈层 FABottleneck,为了进一步压缩模型大小对网络进行稀疏参数自适应的通道剪枝。改进后的 LW-YOLO 算法可以以检测效果的小幅度下降为代价,实现模型大小与计算资源的大幅节省,为光学卫星在轨目标检测提供了理论上的支撑。后续可以考虑对剪枝后的算法模型进行进一步的知识蒸馏、参数量化等工作,对深度学习检测模型在星上的部署做出更多的尝试与验证。

参 考 文 献

- [1] 姬晓飞, 秦宁丽. 光学遥感图像多目标检测及识别算法设计与实现[J]. 计算机应用, 2015, 35(11): 3302-3307.
Ji X F, Qin N L. Design and implementation of multi-target detection and recognition algorithm for optical remote sensing image[J]. Journal of Computer Applications, 2015, 35(11): 3302-3307.
- [2] Xu J, Sun X, Zhang D B, et al. Automatic detection of inshore ships in high-resolution remote sensing images using robust invariant generalized Hough transform[J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11(12): 2070-2074.
- [3] 孟樊, 方圣辉. 利用模板匹配和 BSnake 算法自动提取遥感影像面状道路[J]. 武汉大学学报·信息科学版, 2012, 37(1): 39-42.
Meng F, Fang S H. Quasi-automatic extraction of zonal roads from remote sensing images using template matching and BSnake model[J]. Geomatics and Information Science of Wuhan University, 2012, 37(1): 39-42.
- [4] Gu J X, Wang Z H, Kuen J, et al. Recent advances in convolutional neural networks[J]. Pattern Recognition, 2018, 77: 354-377.
- [5] Hoeser T, Kuenzer C. Object detection and image segmentation with deep learning on earth observation data: a review-part I: evolution and recent trends[J]. Remote Sensing, 2020, 12(10): 1667.
- [6] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [7] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.
- [8] Ren S Q, He K M, 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.
- [9] Cai Z W, Vasconcelos N. Cascade R-CNN: delving into high quality object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6154-6162.
- [10] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE

- Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [11] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [12] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2021-05-09]. <https://arxiv.org/abs/1804.02767>.
- [13] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-05-06]. <https://arxiv.org/abs/2004.10934>.
- [14] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Cham: Springer, 2016, 9905: 21-37.
- [15] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2999-3007.
- [16] Wang C Y, Mark Liao H Y, 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), June 14-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1571-1580.
- [17] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [18] 袁秋壮, 魏松杰, 罗娜. 基于深度学习神经网络的 SAR 星上目标识别系统研究[J]. 上海航天, 2017, 34(5): 46-53.
Yuan Q Z, Wei S J, Luo N. Research on SAR satellite target recognition system based on deep learning neural network[J]. Aerospace Shanghai, 2017, 34(5): 46-53.
- [19] 肖化超, 张建华, 王鹏, 等. 基于深度学习的遥感图像在轨目标检测技术研究[C]//第五届高分辨率对地观测学术年会论文集. 北京: 高分辨率对地观测学术联盟, 2018: 10.
Xiao H C, Zhang J H, Wang P, et al. Research on remote sensing image in-orbit target detection technology based on deep learning[C]//Proceedings of the Fifth Annual Conference on High-resolution Earth Observation. Beijing: High-resolution Earth Observation Academic Alliance, 2018: 10.
- [20] 矫腾章, 胡玉新, 吕鹏, 等. 一种在轨海上多运动舰船目标检测和跟踪方法[J]. 中国科学院大学学报, 2020, 37(3): 368-378.
Jiao T Z, Hu Y X, Lü P, et al. A method of multi-ship target detection and tracking by on-orbit satellite[J]. Journal of University of Chinese Academy of Sciences, 2020, 37(3): 368-378.
- [21] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size[EB/OL]. (2016-02-24)[2021-05-04]. <https://arxiv.org/abs/1602.07360>.
- [22] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2021-05-06]. <https://arxiv.org/abs/1704.04861>.
- [23] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.
- [24] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 1314-1324.
- [25] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6848-6856.
- [26] Ma N N, Zhang X Y, Zheng H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design [M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11218: 122-138.
- [27] Han K, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1577-1586.
- [28] 姜春晖. 深度神经网络剪枝方法研究[D]. 合肥: 中国科学技术大学, 2019.
Jiang C H. The study of pruning methods of deep neural network[D]. Hefei: University of Science and Technology of China, 2019.
- [29] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[EB/OL]. (2015-03-09)[2021-05-04]. <https://arxiv.org/abs/1503.02531>.
- [30] Wen W, Xu C, Wu C P, et al. Coordinating filters for faster deep neural networks[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 658-666.
- [31] 张官荣, 陈相, 赵玉, 等. 面向小目标检测的轻量化 YOLOv3 算法[J]. 光学学报, 2022, 59(16): 1610008.
Zhang G R, Chen X, Zhao Y, et al. Lightweight YOLOv3 algorithm for small target detection[J]. Acta Optics, 2022, 59(16): 1610008.
- [32] 蒋镛圻, 叶泽聪, 彭月平, 等. 一种针对弱小无人机目标的轻量级目标检测算法[J]. 激光与光电子学进展, 2022, 59(8): 0810006.
Jiang R Q, Ye Z C, Peng Y P, et al. A lightweight target detection algorithm for small and weak UAV targets[J]. Laser & Optoelectronics Progress, 2022, 59(8): 0810006.
- [33] 程叶群, 王艳, 范裕莹, 等. 基于卷积神经网络的轻量

- 化目标检测网络[J]. 激光与光电子学进展, 2021, 58 (16): 1610023.
- Cheng Y Q, Wang Y, Fan Y Y, et al. Lightweight object detection network based on convolutional neural network[J]. Laser & Optoelectronics Progress, 2021, 58 (16): 1610023.
- [34] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 13708-13717.
- [35] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [EB/OL]. (2015-02-11)[2021-04-06]. <https://arxiv.org/abs/1502.03167>.
- [36] Liu Z, Li J G, Shen Z Q, et al. Learning efficient convolutional networks through network slimming[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2755-2763.
- [37] Xia G S, Bai X, Ding J, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 3974-3983.
- [38] van Etten A. You only look twice: rapid multi-scale object detection in satellite imagery[EB/OL]. (2018-05-24)[2021-06-04]. <https://arxiv.org/abs/1805.09512>.